

N. R. Shetty · L. M. Patnaik ·
H. C. Nagaraj · Prasad N. Hamsavath ·
N. Nalini *Editors*

Emerging Research in Computing, Information, Communication and Applications

ERCICA 2020, Volume 2

Lecture Notes in Electrical Engineering

Volume 790

Series Editors

Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Naples, Italy

Marco Arteaga, Departament de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico

Bijaya Ketan Panigrahi, Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India
Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, Munich, Germany

Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China

Shanben Chen, Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

Rüdiger Dillmann, Humanoids and Intelligent Systems Laboratory, Karlsruhe Institute for Technology, Karlsruhe, Germany

Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China

Gianluigi Ferrari, Università di Parma, Parma, Italy

Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain

Sandra Hirche, Department of Electrical Engineering and Information Science, Technische Universität München, Munich, Germany

Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA

Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

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

Alaa Khamis, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt

Torsten Kroeger, Stanford University, Stanford, CA, USA

Yong Li, Hunan University, Changsha, Hunan, China

Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA

Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore

Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany

Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA

Sebastian Möller, Quality and Usability Laboratory, TU Berlin, Berlin, Germany

Subhas Mukhopadhyay, School of Engineering & Advanced Technology, Massey University, Palmerston North, Manawatu-Wanganui, New Zealand

Cun-Zheng Ning, Electrical Engineering, Arizona State University, Tempe, AZ, USA

Toyoaki Nishida, Graduate School of Informatics, Kyoto University, Kyoto, Japan

Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi "Roma Tre", Rome, Italy

Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Gan Woon Seng, School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Joachim Speidel, Institut für Telecommunications, Universität Stuttgart, Stuttgart, Germany

Germano Veiga, Campus da FEUP, INESC Porto, Porto, Portugal

Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Beijing, China

Walter Zamboni, DIEM - Università degli studi di Salerno, Fisciano, Salerno, Italy

Junjie James Zhang, Charlotte, NC, USA

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

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

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

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

China

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

India, Japan, Rest of Asia

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

Southeast Asia, Australia, New Zealand

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

USA, Canada:

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

All other Countries:

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

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

More information about this series at <https://link.springer.com/bookseries/7818>

N. R. Shetty · L. M. Patnaik · H. C. Nagaraj ·
Prasad N. Hamsavath · N. Nalini
Editors

Emerging Research in Computing, Information, Communication and Applications

ERCICA 2020, Volume 2

 Springer

Editors

N. R. Shetty
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka, India

L. M. Patnaik
National Institute of Advanced Studies
Bengaluru, Karnataka, India

H. C. Nagaraj
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka, India

Prasad N. Hamsavath
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka, India

N. Nalini
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka, India

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-981-16-1341-8

ISBN 978-981-16-1342-5 (eBook)

<https://doi.org/10.1007/978-981-16-1342-5>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022, corrected publication 2022

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

Contents

Enhancement of Low-Light Images Using CNN	1
Avvaru Greeshma Kavya, Uruguti Aparna, and Pallikonda Sarah Suhasini	
Identical Twin Face Recognition Using Gabor Filter, SVM Classifier and SURF Algorithm	11
A. K. Kavya Ashwini, R. Madhumitha, Mary Ann Sandra, S. Supriya, Ullal Akshatha Nayak, and K. Ranjitha	
Modelling and Prediction of Cardiac Dysrhythmia	25
P. Ramesh Naidu, Shourya Thapliyal, Sambhav Bharvesh, and Satyam Kumar	
Lung Cancer Detection and Prediction Using Customized Selective Segmentation Technique with SVM Classifier	37
Ashwini S. S., M. Z. Kurian, and M. Nagaraja	
Design and Implementation of Storage Benchmark Kit	45
Keshava Munegowda and N. V. Sanjay Kumar	
Adaptive Calibration for Camera and Stitching of Images	63
Jharna Majumdar, Shilpa Ankalaki, and Sarala Madolli	
Data Mining Techniques in the Agricultural Sector	87
B. G. Mamatha Bai and N. S. Rashmi	
Design of a Simplified ANN Model for Real Power Prediction Problem	109
V. Parthasarathy, B. Muralidhara, Bhagwan ShreeRam, and M. J. Nagaraj	
Design and Implementation of Power-Efficient and Fast Full Adders Using Hybrid Logics	119
Chilukuri Sai Vamsi, Sanagaram Aravind Kasyap, S. Saiprateeka, and Sonali Agrawal	

Ethereum-Based Approach for Agricultural Marketing	135
B. A. Mohan, K. Sanjay Kumar, H. Sarojadevi, N. Sreenivasa, E. G. Satish, S. Ashok Kumar, and Roshan Fernandes	
Public Risk Deterrence and Security Alert Systems Using Artificial Intelligence	151
N. Saket, K. L. Suchethan, and N. Nalini	
Experimental Analysis of Node Localization Techniques in WSN	163
B. R. Ramyashree, R. Aparna, Promod Rathod, Mahammad Nihaz, and Navneeth Bhat	
Artificial Intelligence for Automated Engineering Design	177
Durgaprasad Janjanam	
Comprehensive Strategy for Power Quality Improvement of Inverter Based Distributed Generation Systems	185
Shreeram V. Kulkarni, Shruti Gatade, N. Samanvita, and D. N. Gaonkar	
Industry 4.0 Researchers Computer Numerical Control Machine Tool to Manufacture Calligraphy Board	197
B. Praveen, S. U. Abhishek, P. B. Shetty, J. Sudheer Reddy, and B. A. Praveena	
ANN-Based Wheat Crop Yield Prediction Technique for Punjab Region	207
Nishu Bali and Anshu Singla	
Trace Element Analysis of Some Vegetables by PIXE Technique	219
S. Abdul Sattar, B. Seetharami Reddy, and K. Ramnarayana	
Fully Automated Waste Management System Using Line Follower Robot	227
V. Geetha, Sanket Salvi, Sampat Kumar Ghosh, Shaikh Sahil Ahmed, and Rohit Sunil Meshram	
Mathematical Model for Application of Natural Language Description in the Creation of an Animation	237
Shradha Kannan and M. K. Vathsala	
Ingenious Bracelet for Women Safety and Security Using IoT	253
K. Hemasagar, R. Manoj Kumar, V. Sai Manojna, and B. K. Priya	
NL2SQL: Natural Language to SQL Query Translator	267
T. J. Revanth, K. Venkat Sai, R. Ramya, Renusree Chava, V. Sushma, and B. S. Ramya	
Cost Optimization for Migration of Data in Cloud Data Centers	279
S. Spoorthi, N. Rakshitha, and K. S. Chandraprabha	

Leveraging X-Ray and CT Scans for COVID-19 Infection Investigation Using Deep Learning Models: Challenges and Research Directions 289
 Vidyadevi G. Biradar, H. C. Nagaraj, and H. A. Sanjay

Predicting Diabetes Using Id3 Algorithm 307
 Pratibha Dandin, D. B. Srinivas, H. Lakshmi, and K. M. Deepika

Expert System: A Fault Diagnosis Expert System for High-Power Industrial Production Platform 317
 S. Jagdeesh Patil

Real-Time Hand Gesture Detection Using Convex Hull and Contour Edge Detection 325
 Smriti Amatya, Ishika, M. V. Manoj Kumar, B. S. Prashanth, H. R. Sneha, Likewin Thomas, and Vishnu Yarlagadda

Mathematical Behavior of a Pandemic Using Models and Growth Curves 337
 Harsh Singh, Ashish Kumar, and N. Nalini

Technological Advances in Airborne Wind Power: A Review 349
 A. Meghana, B. Smitha, and Shefali Jagwani

Four-Layer Authentication with HoneyPot and Cloud Data Encryption 361
 Sanjay Chari, Harshith Umesh, Athul Sandosh, S. Suganthi, and Prasad Honnavalli

Intrusion Resilience Analysis of Smart Meters 377
 C. S. Sandhya, S. Nagasundari, and Prasad B. Honnavalli

UCREDIT—Credit Scoring Using Social Media Platforms 393
 M. Deeptha, G. S. Arpitha, S. D. Anjali, and D. B. Srinivas

Synthesis of Graphene Quantum Dots and Fabrication of Humidity Sensor 407
 Aniruddh Holemadlu, G. Keerthana, Kunal B. Purohit, Habibuddin Shaik, and Veda Sandeep Nagaraja

Optimal Joint Scheduling and Cloud Offloading for Multi-component Applications 417
 S. L. Shiva Darshan, Abdul Mueez, Aakash S. Shetty, B. A. Mohan, S. Ashok Kumar, and Roshan Fernandes

Comparative Performance Analysis of Routing Topology for NoC Architecture 431
 E. G. Satish and A. C. Ramachandra

LoRaWAN-Based Communication Protocol for Wearable Safety Devices in Mining Fields	441
Namratha Karanth, Deepak Choudhary, Jaideep Francis Reddy, and Ubay Athulla	
Software Component Selection in CBSE Considering Cost, Reliability, and Delivery Delay Using PSO-integrated MVO and ALO	455
Anjali Banga and Pradeep Kumar Bhatia	
Analysis of Hybrid Power Generation System on the Highway and Tracking Through IoT	481
Sujata Shivashimpiger, Purshottam Kamat, and S. Anand	
A Software Defect Avoidance Technique for E-commerce Website Development	495
R. Sudarshan, S. K. Srivatsa, K. N. Chaithra, and K. N. Mohan Kumar	
A Novel Methodology to Restructure Legacy Application onto Micro-Service-Based Architecture System	509
T. R. Vinay and Ajeet A. Chikkamannur	
A Survey of Static and Dynamic Metrics Tools for Object Oriented Environment	521
Manju and Pradeep Kumar Bhatia	
Hierarchical Block Chain-Based Authentication Management Scheme for IoT Devices	531
M. Revanesh and V. Sridhar	
Segmentation of Skin Lesion Using Adaptive Contours Color Method	543
Ramya Srikanteswara and A. C. Ramachandra	
Assessing the Cognitive Levels of the Students Based on Their Learning Styles	555
K. S. Mohan and Y. P. Gowramma	
A Comprehensive Review on Deep Learning-Based Community Detection in Networks	567
Naveed Ul Islam and R. Sunitha	
Deep Transfer Learning Model-Based Automated Detection of COVID-19 from X-ray Images and Interpretation of COVID-19 Images Using GLCM Texture Features	581
Shilpa Ankalaki, Kartikeya Shorya, and Jharna Majumdar	
Automatic Symptom Extraction from Unstructured Web Data for Designing Healthcare Systems	599
Priyanka C. Nair, Deepa Gupta, and B. Indira Devi	

Bug Triaging: Right Developer Recommendation for Bug Resolution Using Data Mining Technique 609
 B. H. Chaitra and K. S. Swarnalatha

Analysing Machine Learning Techniques in Python for the Prediction of Diabetes Using the Risk Factors as Parameters 619
 M. S. Akanksha, Kolachana Vinutna, and M. N. Thippeswamy

Session-based Personalized Recommender System for Online Shopping 641
 B. R. Sreenivasa, C. R. Nirmala, and M. V. Manoj Kumar

A Time Series Cryptocurrency Price Prediction Using LSTM 653
 B. Aditya Pai, Lavanya Devareddy, Supriya Hegde, and B. S. Ramya

Smart Bus Stop Reminder with Connected Bus Stops 663
 C. Shashi Kumar, Venkat Charan, K. R. Suhas Gowda, N. Sushmitha Gowda, and M. R. Sowmya

Routing of Flits in Parallel Input Interface Scenario in a Generalized Network-On-Chip Framework Using Wormhole Flow Control Algorithm 679
 R. Uma, H. Sarojadevi, and V. Sanju

Application of Fine Decision Tree Machine Learning Algorithm to Predict the Subclinical Mastitis in Cow Milk Using Prototype E-nose 695
 M. J. Anand, V. Sridhar, and Ramasamy Ravi

Detection of Cyberbullying on Twitter Data Using Machine Learning 703
 A. Sandesh, H. V. Asha, and P. Supriya

Quora Question Pairs Using XG Boost 715
 Abhishek Chunamari, M. Yashas, Aparajitha Basu, D. K. Anirudh, and C. S. Soumya

A New Kaiser-Bessel Constant Modulus Technique for Smart Antenna Beamforming 723
 K. S. Shashidhara, Veerendra Dakulagi, Jasmineeth Kaur, Kim Ho Yeap, Mandeep Singh, and Ratneshwar Kumar Ratnesh

Generic Security Risk Profile of e-Governance Applications—a Case Study 731
 B. S. Kumar, V. Sridhar, and K. R. Sudhindra

Recent Intelligent Optimisation Algorithms for Islanded Microgrids: A Review 743
 H. R. Sridevi, Shefali Jagwani, and H. M. Ravikumar

Design and Analysis of Dual-Band Monopole Antenna Using Star EBG Structures	755
Prasanna G. Paga, H. C. Nagaraj, R. Tejas, and V. R. Sanath	
Database-Cloud Technology Framework Approach for Academia and Industry	769
A. Balachandra, B. R. Kiran, B. S. Ramya, and N. Madhu	
Comparative Study of Directive-based Programming Models on CPUs and GPUs for Scientific Applications	779
C. Navya, H. A. Sanjay, and Sanket Salvi	
Performance Evaluation of HPC Application in Containerized and Virtualized Environment	793
G. Manoj Kumar, Rohit Danti, Odso Amit, R. Guru Raghavendra, B. R. Kiran, and H. A. Sanjay	
Design Improvement of Blade-Less Wind Generator	805
D. Tejas and Shefali Jagwani	
Prediction of Leaf Blotch in Turmeric Plant	815
Manjula R. Chougala and A. C. Ramachandra	
Voice-Based Gender and Emotion Prediction Using Convolutional Neural Network	827
Arshiya Firdos, T. R. Amrutha, Chaithra, N. Tejaswini, and K. M. Deepika	
Architectural Design and Analysis for Economical Operation of Photovoltaic DC Microgrid for Rural Electrification	837
CH. Venkata Ramesh, K. C. Ravi Kiran, T. C. Balachandra, B. L. Namratha, and B. N. Lakshmi Narayana	
Effects of Variable Gravity Field on the Onset of Ferroconvection in an Anisotropic Porous Layer	851
S. Kiran, Y. H. Gangadharaiah, H. Nagarathnamma, and R. Padmavathi	
Graph-Based Keyword Extraction for Twitter Data	863
S. Vijaya Shetty, S. Akshay, B. S. Shritej Reddy, Hemant Rakesh, M. Mihir, and Jyothi Shetty	
DSEORA: Integration of Deep Learning and Metaheuristics for Web Page Recommendation Based on Search Engine Optimization Ranking	873
Ketan Vaish, Gerard Deepak, and A. Santhanavijayan	
Traffic Symbol Detection and Recognition System	885
K. Md. Zakir Hussain, Komal Nagaraj Kattigenahally, S. Nikitha, Pabitra Priyadarshini Jena, and Y. Harshalatha	

Investigation of Gain and SAR in Dual Band Monopole Antenna Using Slotted Electromagnetic band Gap Structures 899
 Prasanna G. Paga, H. C. Nagaraj, V. P. SaiKarthik, and B. R. Bhuvan

Profit Driven Blockchain Based Platform for Land Registry 911
 Md. Amaan Ahmad, Pooja Singh, M. Sushmitha, H. A. Sanjay, and N. Madhu

Bio-Modification—An Emerging Ground Improvement Technique 923
 Divya Viswanath and M. N. Asha

Cropping Pattern Decision for a Canal Distributary for Varied Discharge Using Linear Programming Approach 931
 B. R. Ramesh, S. B. Ganesh Kumar, and H. J. Surendra

Energy-Aware Task Scheduling Approach Using DVFS and Particle Swarm Optimization for Heterogeneous Multicore Processors 943
 K. Siddesha and G. V. Jayaramaiah

Theft Detection and Monitoring System Using Machine Learning 957
 Jatin Arora, Aayush Bangroo, and Shivi Garg

COVID-19 Contact Tracing Using Geolocation 967
 R. Mahanthesh, Surabhi V. Keerthi, and S. Nidhi Gowri

Optimized Two-Dimensional Chaotic Mapping for Enhanced Image Security Using Sea Lion Algorithm 981
 H. R. Latha and A. Ramaprasath

RDRLJ: Integrating Deep Learning Approach with Latent Semantic Analysis for Document Retrieval 999
 Saicharan Gadamsheetti, Gerard Deepak, A. Santhanavijayan, and K. R. Venugopal

Reverse Supply Chain Network for Plastic Waste Management 1009
 Rakshit Shetty, Neha Sharma, and Vishal A. Bhosale

Real-Time Implementation and Analysis of Different Adaptive Enhancement Algorithms Using Embedded Hardware Boards 1027
 Jharna Majumdar, Manish Verma, Prajwal Shah, Gagan Karthik, Srinath Ramachandhran, and Thribhuvan Gupta

Proxy Re-Encryption Using Vector Decomposition 1041
 A. U. Kaviya and I. Praveen

Correction to: A New Kaiser-Bessel Constant Modulus Technique for Smart Antenna Beamforming C1
 K. S. Shashidhara, Veerendra Dakulagi, Jasmineeth Kaur, Kim Ho Yeap, Mandeep Singh, and Ratneshwar Kumar Ratnesh

About the Editors

Prof. N. R. Shetty is Chancellor of Central University of Karnataka, Kalaburagi, and Chairman of the Review Commission for the State Private University Karnataka. He is currently serving as an advisor to the Nitte Meenakshi Institute of Technology (NMIT), Bengaluru. He is also Founder Vice-President of the International Federation of Engineering Education Societies (IFEES), Washington DC, USA. He served as Vice Chancellor of Bangalore University for two terms and President of the ISTE, New Delhi three terms. He was also Member of the Executive Committee of the AICTE and Chairman of its South West Region Committee.

Prof. L. M. Patnaik obtained his Ph.D. in 1978 in the area of Real-Time Systems and D.Sc. in 1989 in the areas of Computer Systems and Architectures, both from the Indian Institute of Science (IISc), Bengaluru. During March 2008 to August 2011, he was Vice Chancellor, Defence Institute of Advanced Technology, Deemed University, Pune. Currently, he is Honorary Professor with the Department of Electronic Systems Engineering, Indian Institute of Science, Bengaluru, and INSA Senior Scientist and Adjunct Professor with the National Institute of Advanced Studies, Bengaluru. During the last 50 years of his long service, his teaching, research, and development interests have been in the areas of parallel and distributed computing, computer architecture, CAD of VLSI systems, high-performance computing, mobile computing, theoretical computer science, real-time systems, soft computing and computational neuroscience including machine cognition. In these areas, he has 1286 publications in refereed international journals and refereed international conference proceedings including 30 technical reports, 43 books and 26 chapters in books.

Dr. H. C. Nagaraj was born on 4th April, 1959 in Bengaluru. He obtained his B.E. Degree in Electronics and Communication Engineering securing first class with distinction from the University of Mysore in 1981. He obtained his M.E. in Communication Systems from P. S. G. College of Technology, Coimbatore, Bharathiyar University in 1984 and secured first rank. He got his doctoral degree in the area of Biomedical Signal Processing and Instrumentation from Indian Institute of Technology Madras, Chennai in 2000. He assumed charge as Principal, Nitte Meenakshi

Institute of Technology, Bengaluru in September 2003 and till date he is serving as Principal. He has about 38 years of experience in teaching, research and administration. During his term as Principal in NMIT, Bengaluru, the institution has been conferred with Autonomous Status by the Visvesvaraya Technological University, Government of Karnataka and UGC, New Delhi; accredited by the National Assessment and Accreditation Council, National Board of Accreditation, AICTE, New Delhi for UG programmes, and a pico satellite weighing less than 1 kg was designed and fabricated and handed over to ISRO and was launched on 12th July 2010 from Sriharikota, Andhra Pradesh, India. Dr. H. C. Nagaraj has published a number of papers in National and International Journals, Conferences and delivered invited talks in the field of Biomedical Signal Processing, Image Processing, Mobile Communication etc. He got the Best Paper Award in the National Conference of Biomechanics held during February 22–24, 1996 at IIT Madras. He was awarded the Best Paper Award at State Level Seminar on “Introduction of Flexible System in Technical Education” held by the Government of Karnataka during February 1999 in Bengaluru. He has guided several students for their Ph.D. programmes and written a book titled *VLSI Circuits* published in 2006 for B.E. Students. He is a member of expert teams visiting various colleges for accreditation by NAAC and affiliation of UG, PG and Ph.D. programmes under Electronics and Communication, Biomedical Engineering Streams, on behalf Institution of Engineers, Visvesvaraya Technological University, Belgaum and Kuvempu University, Shimoga. He is a Fellow of the Institution of Electronics and Telecommunication Engineers (IETE), Life member of ISTE, Life Member of Biomedical Engineering Society of India and Member of Global Engineering Deans Council India Chapter (GEDCIC). Dr. Nagaraj has visited countries like USA, Canada, UK, Singapore, Malaysia, South Korea, Hungary etc. in connection with collaboration for academic activities. Presently, he is the Member of the Academic Senate at Visvesvaraya Technological University (VTU), Belagavi. He is also the Member of Court at Pondicherry University and Member of Expert Team for the evaluation of R&D projects funded by VTU.

Dr. Prasad N. Hamsavath is currently working as Professor and Head of the Department of Master of Computer Applications at Nitte Meenakshi Institute of Technology, Bengaluru. He completed his Ph.D. at Jawaharlal Nehru University, New Delhi, India. Dr. Prasad has more than 12 years of experience in different roles in both public and private sector enterprises, including the Ministry of Human Resource and Development, New Delhi, Government of India. He has received the prestigious “Dr. Abdul Kalam Life Time Achievement Award” and also received a “Young Faculty” award at the 2nd Academic Brilliance Awards.

Dr. N. Nalini is currently working as Professor at the Department of Computer Science and Engineering at Nitte Meenakshi Institute of Technology, Bengaluru. She received her M.S. from BITS, Pilani, in 1999, and her Ph.D. from Visvesvaraya Technological University in 2007. She has more than 17 years of teaching and 10 years of research experience. She has several international publications to her credit, and has received the “Bharath Jyothi Award” by Dr. Bhishma Narain Singh, Former

Governor of Tamil Nadu and Assam, given out by the India International Friendship Society. She is Lifetime Member of the ISTE, CSI, ACEEE and IIFS.

Enhancement of Low-Light Images Using CNN



Avvaru Greeshma Kavya, Uruguti Aparna, and Pallikonda Sarah Suhasini

1 Introduction

High-quality image plays a critical role in object detection and classification. Image enhancement has important contributions in object detection, classification, segmentation, recognition and speech detection and been used in many real applications like remote sensing, video processing, medical field, etc. To address the enhancement issues, there are many techniques of image enhancement, e.g., histogram equalization, log transformation, gamma correction, bilateral, CLAHE, etc. But, in recent research, it is found that deep learning contributes more than any other techniques of image enhancement. In this paper, the enhancement of low-light images using a deep learning platform, convolutional neural network (CNN), is implemented. With the popularity of smartphones and digital cameras, images play an important role in our daily life. Noise is present in any imaging system, but it makes imaging particularly challenging in low light. Since it is a low-light image, the photon number is low to build the actual image. The quality of an image is analyzed in subjective type, by eye vision, and objective type, by considering various parameters of noise like mean square error (MSE), structural similarity index measurement (SSIM), peak signal-to-noise ratio (PSNR), etc.

Low-light imaging is challenging due to low photon count [1] and low SNR. A variety of processes like denoising, deblurring in which, with the support of pipelining, image is enhanced by convolutional network. The limitation is that this is not applied to humans and dynamic objects. So, we take the low-light images with only object-level orientation [2], in which a simple illumination invariance method is proposed. But in real time, we need to take images in RGB illumination [3] in which an imposing structure can be illuminated and an illumination map is created. Furthermore, we need to take images as a set to propose a machine learning method

A. G. Kavya (✉) · U. Aparna · P. S. Suhasini
Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India

on it. So, we take this into another level, by taking some simple unpaired training images [4], which are further made into a proper training set as MSR-net [5], and the trained set should go through many layers of the network we designed and the network layers with required functions to predict the performance [6]. Apart from all this, noise is the major problem in image enhancement, so a model with the joint framework of denoising and image enhancement [7] is proposed. A multi exposure framework [8], in which the input image and the synthetic image are fused according to their weights.

Apart from all this, in real time an image does not only contain a single noise, but come with a combination of noises. So, in this paper we analyzed with different combinations of noise and performance is evaluated with different parameters [9–11]. Different techniques are used to enhance the image added noises and are compared results [12].

2 Trained Model

In this paper, a model is defined using CNN, in which a dataset of images is trained. Noises like salt and pepper, Gauss, Poisson and speckle are defined in this model. An image captured contains more than a single noise in it. As the four different noises are defined in this model, a single noise and different combinations of noises are applied on each image in the dataset.

Figure 1 shows the model that is trained with the dataset such that the images go through different convolutional layers and add up so on, resulting in the enhanced image.

Since CNN consists of numerous layers, like convolutional layer, activation layer, pooling layer and fully connected layer, and several convolutional layers, i.e., a neural network, each convolutional layer in CNN performs convolution operation with the kernel. The size of the kernel is set through the term steps per epoch, and epoch is the term used to train the network number of times to understand the model defined. The Inception-ResNet-V2 architecture is used. Activation layer is the rectified linear unit, and its function is

$$F(x) = 0, \quad \text{for } x < 0 \\ = x, \quad \text{for } x = 0, x > 0$$

Pooling layer is a function that reduces the spatial size and performs max pooling; i.e., it subdivides the image pixels and places the maximum value of that subdivided image. The raw image will be with perfect illumination, and this image is gone through certain operations to result in a low-light image. This low-light image is used to enhance through the model enhancer.

Firstly, we applied a single noise to the dataset of 100 images resulting in the enhanced image as a low-light image. Due to the result that we have obtained, a

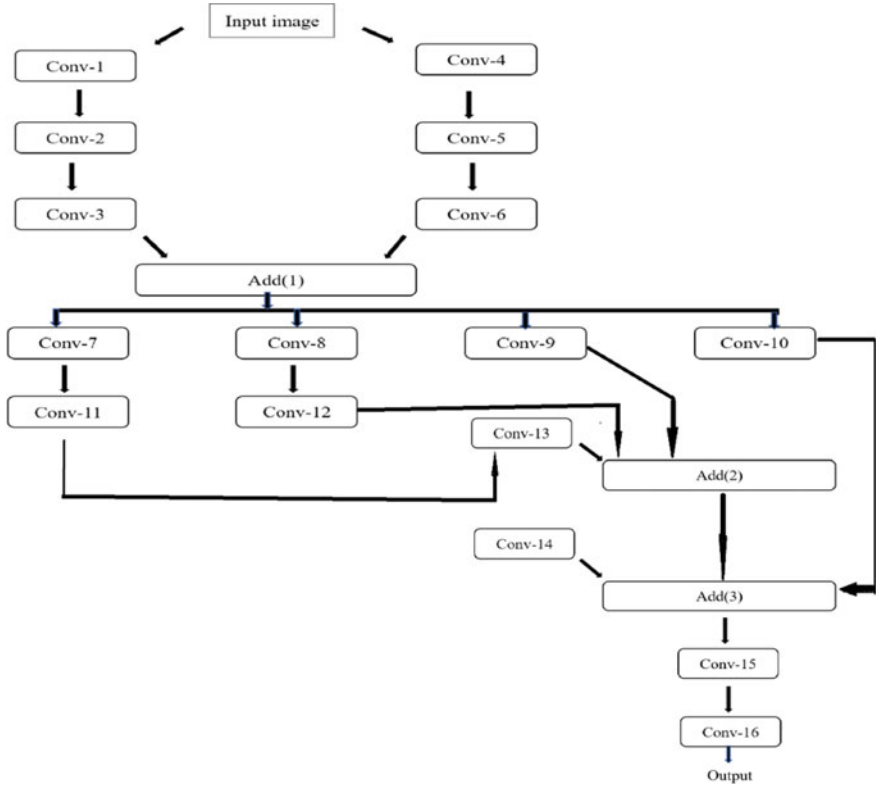


Fig. 1 Trained model

dataset of images of around 2000 (~2117) is trained with different noises with a number of epochs and steps per epochs. The enhanced image size is not similar to the original image size to compare. So, the enhanced image is reshaped to original image size. The results are evaluated, and comparison is made using the parameters like mean square error (MSE), peak signal-to-noise ratio (PSNR), and structure similarity index measurement (SSIM).

2.1 Evaluation Metrics

$$MSE = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N [I(x, y) - I'(x, y)]^2 \tag{1}$$

where $I(x, y)$ is the original image, $I'(x, y)$ is the enhanced image and $M \times N$ is the size of the image.

Peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation.

$$\text{PSNR} = 20 \log \frac{\text{MAX}}{\sqrt{\text{MSE}}} \quad (2)$$

2.2 Algorithm

Figure 1 shows the architecture of trained model.

Algorithmic steps are as follows.

Step 1: Installing required software like TensorFlow and OpenCV.

Step 2: Uploading the images to form a dataset.

Step 3: Introducing virtual noise to every image in the dataset.

Step 4: Converting the RGB image to HSV image.

Step 5: Defining the dataset into an array with 2 elements X and Y . Element X is for input noisy images, and element Y is for predicted output images.

Step 6: Defining the convolution model.

Step 7: For the input sample defining the model enhancer, i.e., optimizer and loss function.

Step 8: Generating the X and Y elements by running the model for different time intervals, i.e., different combinations of epochs and steps.

Step 9: Predicting the output image for the given input image.

Step 10: Get the output for the given input.

Step 11: Reshaping of enhanced image to the size of original image.

Step 12: Comparing original image with the reshaped image to evaluate metrics.

2.3 Experimental Results and Observations

Single noises like salt and pepper, Poisson, Gauss and speckle are applied to the trained images as shown in Fig. 2. The enhanced outputs with CNN are compared, and it is found that, for Poisson noise, the results are better with less error as shown in Table 1.

The following experiments illustrate the performance of CNN when the images are added with different combinations of noise types. Further, the results with different numbers of epochs are also investigated.

With Single Noise:

Salt and pepper noise is applied with an increasing number of epochs, and steps per epoch is 3. The number of epochs used is 10, 35, 100, 500 with 3 steps per epoch and

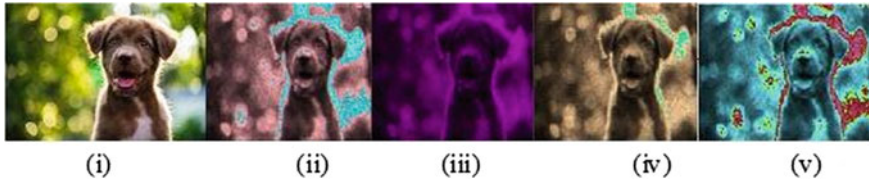


Fig. 2 Original image with single noise [original (i), salt and pepper (ii), Gauss (iii), Poisson (iv), speckle (v)]

Table 1 Comparison of different parameters with single noise (S—salt and pepper, P—Poisson, G—Gauss, SP—speckle)

Noise	Epochs	Steps	Technique	MSE	PSNR	SSIM
Original image	–	–	–	0.0	100.00	1.00
S	10	2	CNN	2559.17	14.05	0.42
P	10	2	CNN	826.97	18.96	0.76
G	10	2	CNN	7560.74	9.35	0.31
SP	10	2	CNN	3402.86	12.81	0.43

other combinations as 53 number of epochs and 39 steps per epoch. As the number of epochs and steps per epoch are increasing, the time taken for training the model is also increasing. The enhanced images for corresponding number of epochs and steps per epochs are shown in Fig. 3. And the parameters are given in Table 2.

This analysis proves that while increasing the number of steps and epochs, the enhancement will be better.

With Double Noise:



Fig. 3 Enhanced images for different numbers of epochs

Table 2 Comparison of parameters for different numbers of epochs

Noise	Epochs	Steps	Technique	MSE	PSNR	SSIM
S	10	3	CNN	907.67	18.55	0.70
S	35	3	CNN	706.11	19.64	0.80
S	53	3	CNN	779.54	19.21	0.73
S	100	3	CNN	803.40	19.08	0.80
S	500	3	CNN	443.45	21.66	0.83

Two different types of noises are applied to images. Figure 4 shows the low-light image, and the input is given with double noise as salt and pepper along with Poisson and trained 469 images with 10 epochs and 2 steps. Table 3 is related to Fig. 5.

With Three Different Noises:

Here three noises as salt and pepper with gauss and poisson are added, trained 1071 images with 10 epochs and 2 steps per epoch, 35 epochs and 3 steps per epoch respectively. The parameters are given in Table 4.

Figure 6 shows the enhanced images after applying three noises, viz., salt and pepper with Gauss and speckle, trained 1071 images with 10 and 35 epochs and 3 steps per epoch, respectively. The parameters are given in Table 5.

With Four Different Noises:

The model is investigated with a combination of four different types of noises. The parameters with the number of trained images are observed. Finally, the results of CNN when trained with 929 images, epochs of 5 and 2 steps per epoch are

Fig. 4 Low-light image with two types of noises and enhanced image

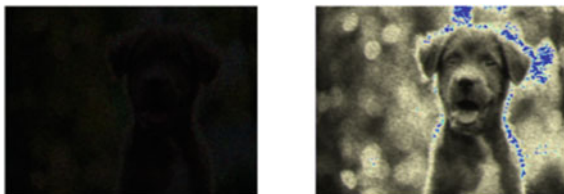


Table 3 Comparison of parameters

Noise	No. of images trained	Epochs	Steps	Technique	MSE	PSNR	SSIM
S&G	1071	10	2	CNN	2978.39	13.39	0.44
S&G	1071	35	2	CNN	1962.412	15.20	0.49

Fig. 5 Enhanced images for S&P and Gaussian noises

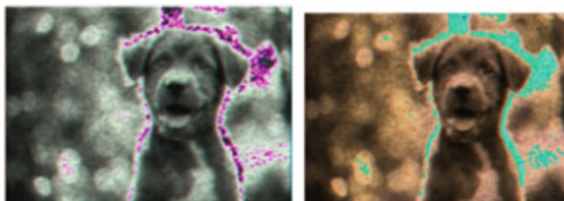


Table 4 Comparison of parameters

Noise	Epochs	Steps	Technique	MSE	PSNR	SSIM
S,G&P	10	2	CNN	2559.17	14.05	0.42
S,G&P	35	3	CNN	1983.01	15.16	0.48

Fig. 6 Enhanced images for three types of noises (S&P, Gauss and speckle)

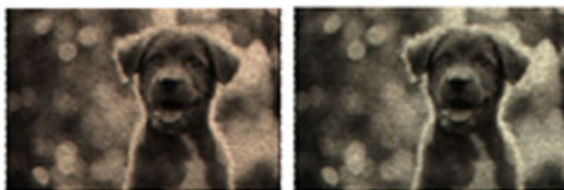


Table 5 Parameter comparison

Noise	Epochs	Steps	Technique	MSE	PSNR	SSIM
S,G&SP	10	3	CNN	3084.69	13.24	0.47
S,G&SP	35	3	CNN	2605.52	13.97	0.49

compared with other existing techniques like CLAHE, total variation, bilateral, log transformation, gamma correction and histogram equalization. Figure 7 shows the output images with different enhancement techniques, and Table 6 gives performance comparison.

The enhancement using CNN yields better results compared to other techniques.

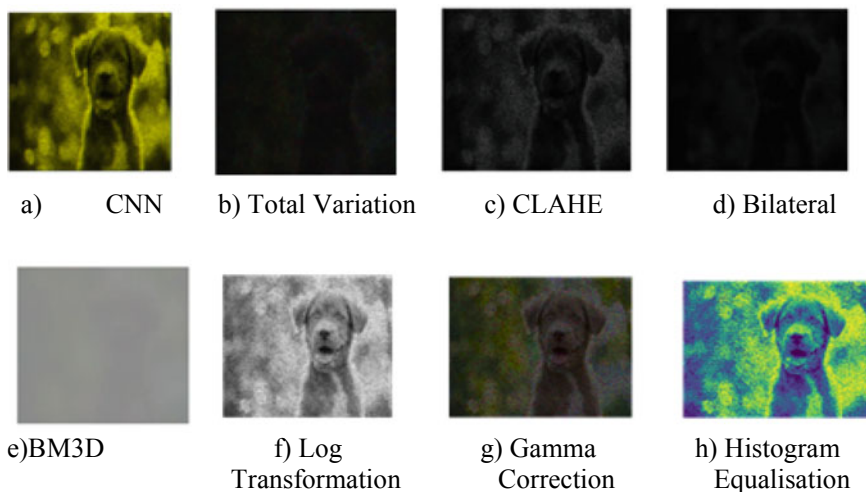


Fig. 7 Comparison of CNN with other techniques

Table 6 Comparison with other techniques

Technique	MSE	PSNR	SSIM
CNN	706.11	19.64	0.80
Total variation	11,371.71	7.57	0.14
CLAHE	11,321.80	7.59	0.15
Bilateral	14,195.65	6.61	0.05
Log transformation	5133.08	11.03	0.09
Gamma correction	7977.57	9.11	0.14
Histogram equalization	4652.49	11.45	0.11

3 Conclusion

Through parameters, it is observed that, by increasing the number of images in the dataset, by increasing the number of epochs and by increasing the steps per epoch, the performance of the enhanced image is improving correspondingly. This comparison with other techniques revealed that CNN technique is better than the earlier techniques of image enhancement. Google Colaboratory is used to execute the program; as we increase the images' number, epochs and steps, it takes much time to execute a particular cell of the program and the memory will be drained out. So, less number in images, epochs and steps are taken in these experiments. Further using GPUs, with large datasets the performance of CNN may be explored.

References

1. Chen C, Chen Q, Xu J, Kolthun V (2018) Learning to see in the Dark. In: IEEE/CVF conference on computer vision and pattern recognition, pp 3291–3300. <https://doi.org/10.1109/CVPR.2018.00347>
2. Loh YP, Chan CS (2018) Getting to know low-light images with the exclusively dark dataset. In: Computer vision and image understanding, vol 178, pp 30–42. <https://doi.org/10.1016/j.cviu.2018.10.010>
3. Guo X, Li Y, Ling H (2017) LIME: low-light image enhancement via illumination map estimation. *IEEE Trans Image Process* 26(2):982–993. <https://doi.org/10.1109/TIP.2016.2639450>
4. Jiang Y, Gong X, Ding Liu Y, Cheng CF, Shen X, Yang J, Zhou P, Wang Z (2019) EnlightenGAN: deep light enhancement without paired supervision. *Comput Vis Image Underst* 1:1–11
5. Shen L, Yue Z, Feng F, Chen Q, Liu S, Ma J (2017) MSR-net: low-light image enhancement using deep convolutional network. In: Computer vision and image understanding, pp 1–9
6. Jones Y (2017) Deep learning for image enhancement & visibility improvement, Article, pp 1–132
7. Tao L, Zhu C, Song J, Lu T, Jia H, Xie X (2017) Low-Light enhancement using CNN and Bright channel prior. In: International conference on image processing, pp 3215–3219. <https://doi.org/10.1109/ICIP.2017.8296876>
8. Ying Z, Li G, Gao W (2015) A bio inspired multi-exposure fusion framework for low-light image enhancement. *J Latex Class Files* 14(8):1–10

9. Arora S (2015) Performance evaluation of image enhancement techniques. Article 8(8):251–262. <https://doi.org/10.14257/ijcip.2015.8.8.27>
10. Jaya VL, Gopikakumari R (2013) IEM: a new image enhancement metric for contrast and sharpness measurements. Int J Comput Appl 79(9):19. <https://doi.org/10.5120/13766-1620>
11. Hanumantharaju MC, Ravishankar M, Rameshbabu DR, Manjunath Aradhya VN (2011) An efficient metric for evaluating the quality of color image enhancement. In: Proceedings of the 5th Indian international conference on artificial intelligence, pp 1016–1026
12. Lv F, Li Y, Lu F (2019) Attention-guided low-light image enhancement, pp 1–12

Identical Twin Face Recognition Using Gabor Filter, SVM Classifier and SURF Algorithm



A. K. Kavya Ashwini, R. Madhumitha, Mary Ann Sandra, S. Supriya, Ullal Akshatha Nayak, and K. Ranjitha

1 Introduction

Previously, manual experiments were done to identify the difference, and many more systems existed to show differences between twins. In the existing system, many processes are used for twins' identification like finger print, voice and iris recognition. A person can be identified uniquely using finger print identification process. The method proposed takes a scan image from the person and compares it with the database for identification. The iris recognition is also similar to finger prints identification. Every person has a unique iris, which helps us to identify the person through iris recognition method. The process of voice recognition works based on the voice to identify the twins. Finger print recognition has the drawbacks of slow processing and is easy to hack. Iris recognition also has drawbacks such as mismatching due to identification. In general, it is time consuming for identification, and if we have some faults in the eye, then iris recognition is not suitable for identification.

Similarly voice recognition also has demerits of easily being misused by another person. Hence, due to the above-stated drawbacks of fingerprint, iris and voice recognition, we use a more advanced system such as an identical twin face recognition system which extracts different features from the dataset and compares the two images precisely based on the similarities and dissimilarities using different algorithms such as multiSVM, Gabor features, Gabor filter bank and open surf. Upon completion of the above task, the two images are then combined to form a warp image from which we can identify whether the person is a twin, not a twin or is a different person by viewing the distortion level of the warped image.

A. K. Kavya Ashwini · R. Madhumitha · M. A. Sandra · S. Supriya · U. Akshatha Nayak (✉) · K. Ranjitha
Nitte Meenakshi Institute of Technology, Bangalore 560047, India
e-mail: akshatha.n@nmit.ac.in

2 Proposed System

There has been a huge progress in facial recognition field of study and research, but there are still many challenges when it comes to differentiating between identical twins. Low accuracy has been the main drawback of face recognition, and this would be the one important objective to be accomplished. The main objective of this project is to create a face recognition system that can recognise similar faces and be able to differentiate between the identical twins. The purpose of the project is to develop an efficient algorithm in terms of low complexity, with maximum possible number of face detection factors, to reduce false identification of identical twins which in turn will provide a solution for differentiating between them. The sample images of identical twins are collected and are used in training the system in .pgm, .jpg and .bmp formats. Image processing is done using Gabor filter and speeded up robust features (SURF) algorithm. Machine learning dataset training and classification is done using multiSVM. Twins face recognition system will detect, extract and recognise frontal faces from the acquired images. The system should work and satisfy all the conditions of the proposed system and recognize the images of few twins at least.

This system can be designed using various other coding techniques, but MATLAB can do the same with ease using its image processing applications, its functions and tool box (Fig. 1).

3 System Design and Implementation

System architecture has six stages. The first stage includes the input of the two images from the dataset. In the second stage of the system architecture, the system will detect the facial features of both the images. The third stage is the feature extraction; using algorithms such as Gabor filter and open surf, features of the input image are extracted. In the fourth stage, the identification of the face for the first image is done. In the fifth stage, features of the second input image are extracted using the same algorithms as specified above. The last stage is the classification, in which the images are classified using SVM (Fig. 2).

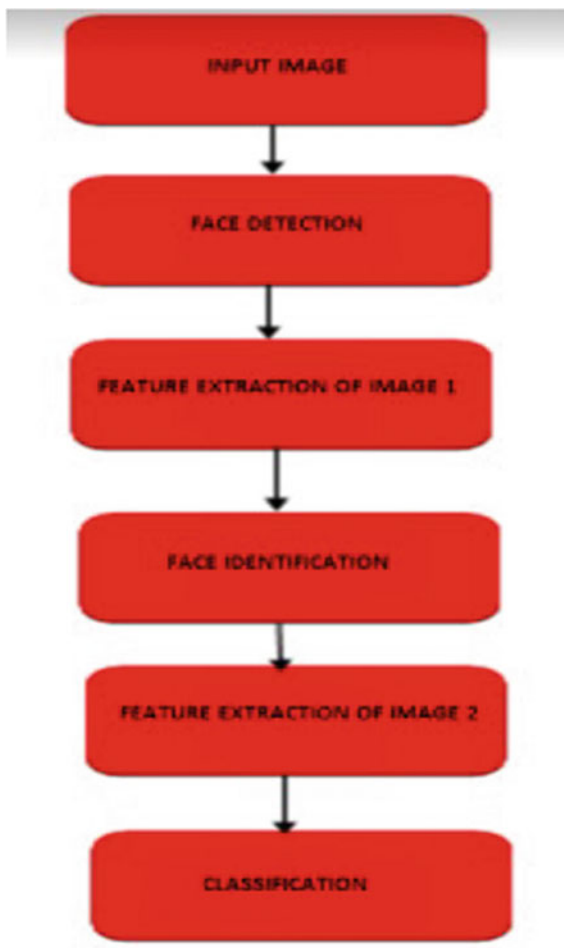
The above figure represents the system architecture and the overview of the system design. The steps involved are.

1. Training—in this process, the collected data is loaded, and it is used for training and feature extraction.
2. Classification—in this process, using the SVM, the given image is classified; this is done after training.

The system design mainly consists of.

- Image collection
- Image pre-processing
- Feature extraction

Fig. 1 Proposed system architecture



- Training
- Classification using multiclass SVM.

3.1 Image Collection

For training purpose, images of different identical twins are collected. The images will be stored in a standard format; in this system, we use the portable grey map (.pgm) images; while storing them, the sample images are taken from the internet or Kaggle dataset.

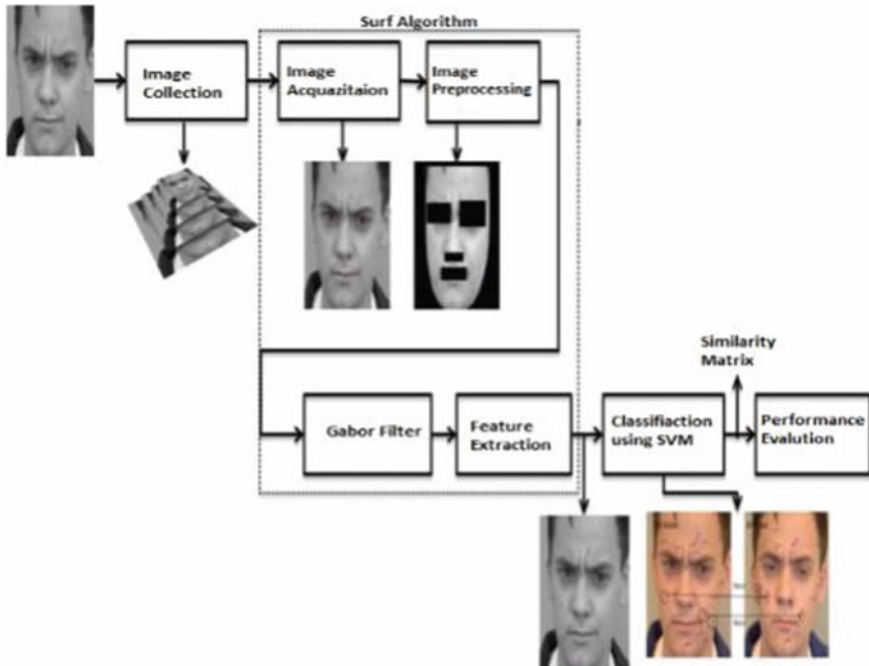


Fig. 2 System design

3.2 Image Pre-processing

For data that is uneven and noisy, image pre-processing plays a major role. In this process, the image is transformed to another image which has improved quality that is better for analysing. This transformation basically removes unwanted components of the image and enhances the image features. This step is important because it helps the image to be effective for subsequent task like feature extraction, data mining techniques that depend on image quality. The images and its statistical properties are observed in greyscale format as it makes boundaries and edges easy to analyse when in black and white. Also, the image is seen well in RGB when it comes to information related to colour formats. Here, the intensity image is converted into binary image by resizing the given image to 256×256 , and as a result, the threshold is calculated. This also helps in converting the RGB image format to greyscale image. A histogram is used to compute the mean using this, it is scaled between 0 and 1 which is a normalized value.

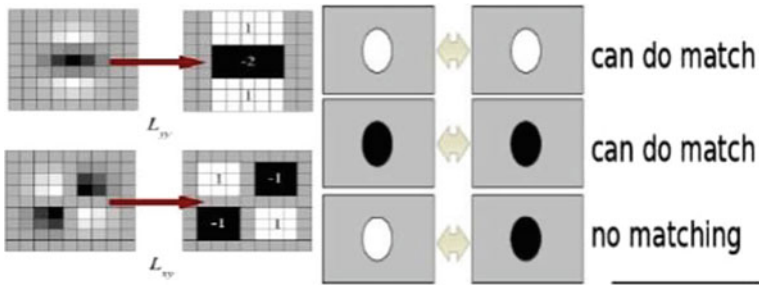


Fig. 3 SURF (Speeded up robust feature)

3.3 Feature Extraction

Feature extraction is done by using two different algorithms, speeded up robust features (SURF) and Gabor filter. These two methods will enable us to extract the features from the twin’s images.

3.4 Feature Extraction in SURF

Key point detection can be done by feature extraction using SURF which in turn uses a very basic Hessian matrix approximation for storage of values. Hessian matrix exploits each column of the matrix which can be approximated by taking the difference between two instances which are evaluated at the two nearby points. The image is used in a quick and effective way for finding the sum of values of the pixels for the given image or for the grid of a given image. Faster matching occurs with very minimum information, and it also does not reduce the performance of the feature description of SURF algorithm (Fig. 3).

3.5 Feature Description in SURF

The descriptor is designed to be scale invariant and rotationally invariant and is partly inspired from scale invariant feature transform (SIFT) descriptors. These descriptors are used to locate and recognise objects, faces to reconstruct 3D scenes, to track object and to extract points of interest, and it uses Hessian matrix to find interest.

3.6 *Gabor Filter*

Gabor filter is used in image processing as a linear filter used for detection of edges. We use this filter to extract features at different angles. Optimization-based feature selection algorithm is used to find the features for the optimal feature subset. At each stage, an improvement is made. In image processing, this filter can also be used for texture analysis which means that it checks if there is frequency specifically for the image in particular direction of that region around the point.

3.7 *Training*

Training is done to assign the desired output of -0.9 to non-face feature vector and 0.9 to face feature vector; we will be able to obtain the network desired output based on these features. In this training process, first the datasets are trained, and then image acquisition is done, where in the image is resized and is pre-processed during which features are added.

3.8 *Classification Using Multiclass SVM*

The red and blue colours are two types of data in the below given image. Calculation of the distance to all the training samples is done, and the one with the minimum distance is taken for a test data in KNN. It is time and space consuming to measure all the distances to store all samples. But referring the below given image, much memory is not needed. Another idea is to find a line, which divides both the data into two regions. We get a new test data, and if it belongs to a blue group, we substitute it. This line is called as the decision boundary. This method is memory efficient and very simple. Such data which can be divided into two with a straight line is called linear separable (Fig. 4).

The above image shows that plenty of such lines are possible. The incoming data can be a noise because of which it is better if the lines pass as far as possible from all the points. The classification accuracy should not be affected by the noise. So therefore, the line with the greatest distance is chosen as it has lesser effect of the noise. So what SVM does is it finds hyperplane with largest minimum distance to the training samples. The bold line in the below image passing through the centre is the hyperplane (Fig. 5).

So to find the decision boundary, we need the proper training data. The data which are close to the opposite group are sufficient. In the above image, the blue-filled circle and two red-filled squares represent the support vectors and the lines passing through them are called support planes; these are used for finding our decision boundary. We may ignore other data as they are not needed for this purpose.

Fig. 4 Multiclass SVM graph

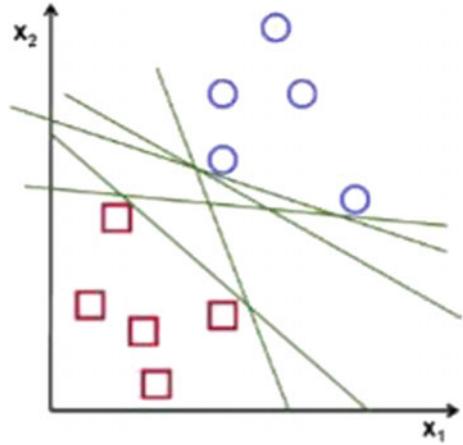
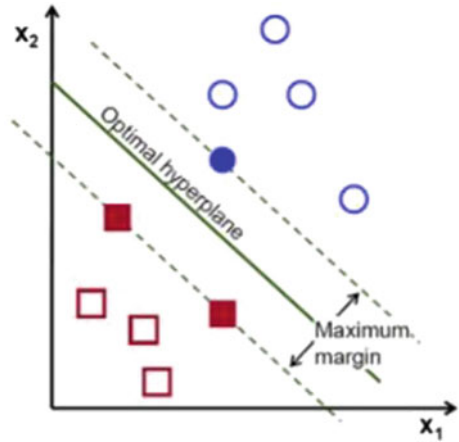


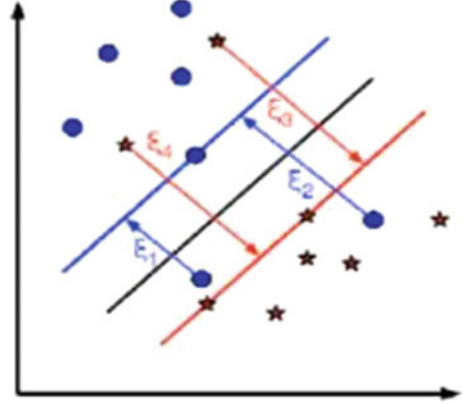
Fig. 5 Optimal hyperplane graph



3.9 Support Vector Machine (SVM)

SVM is a machine learning algorithm which helps with classification. SVM uses a technique to transform the data, and then based on the transformation, it searches an optimal boundary between the possible outputs. The given examples are marked into the categories; SVM training algorithm does the work of building a model that will assign new examples to the categories which makes it a non-probability binary linear classifier.

Fig. 6 Misclassified feature graph



3.10 Misclassification

It is not enough if we just find a decision boundary with maximum margin because we need to consider the problems of misclassification also. In some cases, it may be possible to find a decision boundary with fewer margins but with reduced misclassification. Therefore, here we need to modify our model such that it should find decision boundary with maximum margin but with less misclassification. The modification is done as below.

$$\min \|w\|^2 + C (\text{distance of misclassified samples to their correct regions})$$

The below image represents this concept. A new parameter is defined for each sample of training data. The distance from its corresponding training sample to their correct decision region.

The ones that are not misclassified fall on the corresponding support planes; therefore, their distance is zero (Fig. 6).

3.11 Test Case and Design Testing Results—Unit Testing

The dataset is to be loaded in the initial stage which will process and train the images and also replaces the previous dataset with the current new dataset with the input being the images taken from the internet. This test case is passed if the expected output matches with the actual output (Figs. 7 and 8).

Images of the twins are to be selected one at a time as an input to the project. The image is resized and algorithms such as SURF, Gabor filter and SVM are applied on it. The resulting values are compared and the output is displayed as twins found and shows the matching point and non-matching point values, and thereby passes this particular test case if the expected output matches with the actual output. If images of same person are to be selected, i.e. one at a time as an input to the project, the image is

# Test Case:	Test Case-1
Name of Test:	Loading the dataset
Items being tested:	Dataset
Sample input:	Images
Expected output:	Depending on the inputs, it must load the images
Actual output:	Successful Loading
Remarks:	Pass

Fig. 7 Test case 1

resized and then goes through the algorithms such as SURF, Gabor Filter and SVM. The resulting values are compared, and the output is displayed as same person found and shows the matching point and non-matching point values and thereby passing this particular test case if the expected output matches with the actual output. Images of same persons are to be selected, i.e. one at a time as an input to the project. The image is resized and then goes through the algorithms such as SURF, Gabor Filter and SVM. The resulting values are compared, and the output is displayed as same person found and shows the matching point and non-matching point values and thereby passing this particular test case as the expected output matches with the actual output (Figs. 9 and 10).

Images of two different persons are to be selected, i.e. one at a time as an input to the project. The image is resized and then goes through the algorithms such as SURF, Gabor filter and SVM. The resulting values are compared, and the output is displayed as twins not found or different persons found and shows the matching point and non-matching point values and thereby passing this particular test case if the expected output matches with the actual output (Figs. 11 and 12).

Images of a pair of twins are to be newly inserted in the dataset, and we select that particular newly added ones, i.e. one at a time as an input to the project. The image is resized and then goes through the algorithms such as SURF, Gabor filter and SVM. The resulting values are compared, and the output is displayed as different persons found and shows the matching point and non-matching point values and thereby failing this particular test case if the expected output did not match with the actual output (Fig. 13).

4 Results and Discussion

The performance of the proposed system is discussed in this section. Face recognition system among biometric systems has ease of use, reliability, cost effectiveness, etc.



Testcase 3

# Test Case:	Test Case-3
Name of Test:	Same Person Found
Items being tested:	Dataset
Sample input:	Two images of a same person
Expected output:	Display as same person found
Actual output:	Same Person found
Remarks:	Pass

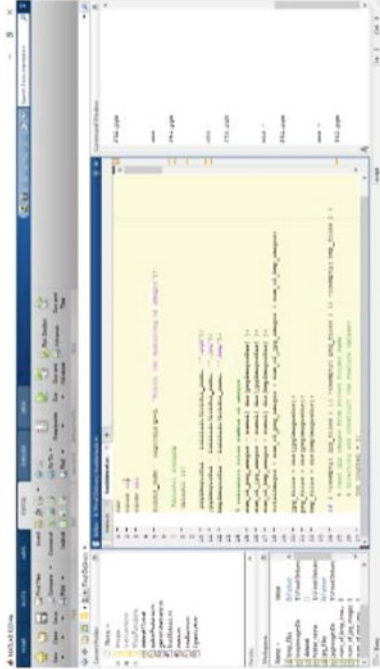


Fig. 8 Output of test case 1



Fig. 9 Output of test case 3

# Test Case:	Test Case-4
Name of Test:	Different Person Found
Items being tested:	Dataset
Sample input:	Images of two different person i.e. non twins
Expected output:	Display the error as different person found i.e. not twins
Actual output:	Different persons found
Remarks:	Pass

Fig. 10 Test case 4

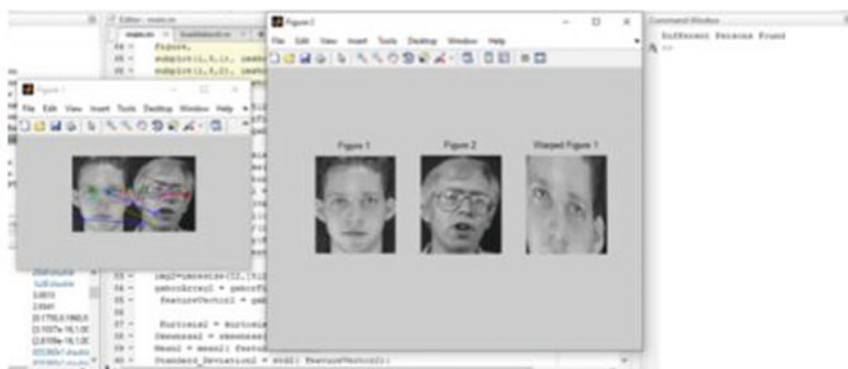


Fig. 11 Output of test case 4

# Test Case:	Test Case-5
Name of Test:	Adding new image in dataset and testing
Items being tested:	Dataset
Sample input:	Two images i.e. of identical twins
Expected output:	Display as twins and show matching and non-matching points
Actual output:	Different persons found
Remarks:	Failed

Fig. 12 Test case 5

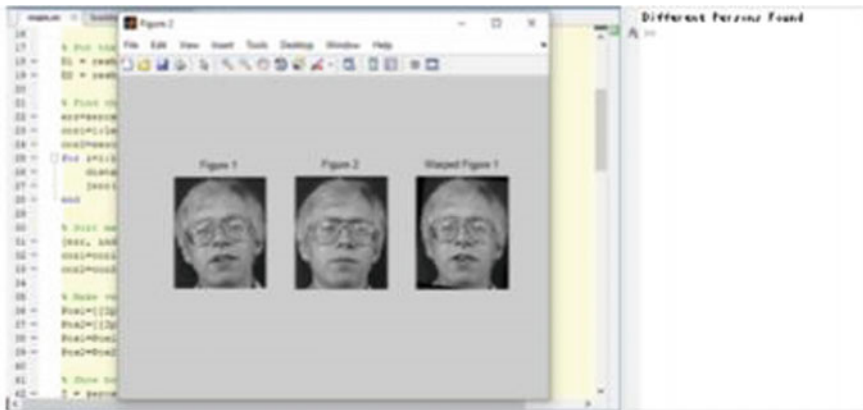


Fig. 13 Output of test case 5

Many public institutions, researches and businesses. It is the need of the hour for face recognition systems to work quickly and precisely by matching the images and differentiating between the identical twins as the main priority. The performance of the proposed system is 100% when the dataset is containing images of 10 pairs of twins, i.e. for 20 images in total. The performance goes on to deteriorate as the dataset gets bigger. The overall success rate of this system is 85–90% when the dataset is extended for few more images. The smaller the dataset, the better this project works.

5 False Acceptance Rate (FAR) and False Rejection Rate (FRR)

The false acceptance rate is a way to measure the number of instances where an unauthorised person is accepted incorrectly. FRR is the process by which a system rejects an access attempt by an unauthorized user. The system's FRR is typically

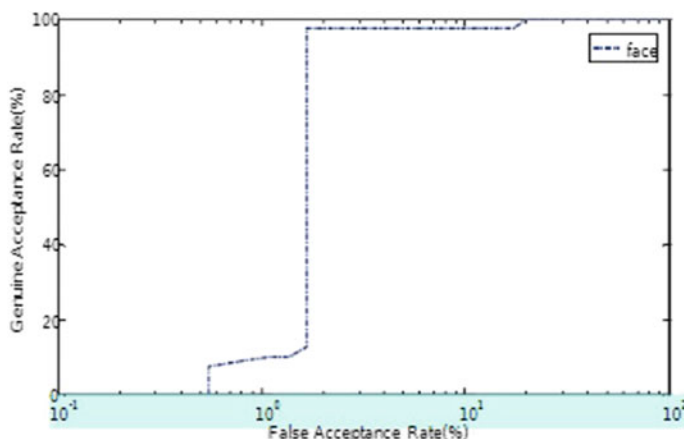


Fig. 14 Curve for face recognition performance

stated as the ratio of the number of false recognitions to the number of identification attempts (Fig. 14).

At the false acceptance rate (FAR) of 2%, the genuine acceptance rate (GAR) performance is 97.5%. The system performance based on equal error rate (EER) achieves the value of 2.5%

6 Conclusion

The current work mainly concentrates on identifying the twins, similar faces and differentiating between the identical twins on the basis of the features extracted. The proposed project uses MATLAB for developing the project. This uses various numbers of methods or algorithms such as SURF algorithm, Gabor filter, SVM, multiple SVM, and Hessian matrix. The SURF method is an efficient algorithm for local similarity difference representation and to compare images. Gabor filter is used for edge detection and for capturing different angles. SVM is a machine learning algorithm which can be used for classification; it uses technique to transform the data, and then based on the transformations, it finds an output. The practical performance of this project is analysed, and it shows if the project works perfectly for the limited dataset. This project recognises the identical twins and shows the matching points and gives the differences between them.

References

1. Zafaruddin GM (2014) Face recognition: a holistic approach. Department of MCA Millennium Institute of Management Aurangabad, IEEE
2. Patil KD, Kopargaon MS, Patil SA, Sanjivani KBP, Kopargaon MS (2016) Identification of identical twins using face recognition with results. Information Technology Department SRES College of Engineering, IJRITCC 2016
3. Jonathon Phillips P, Flynn PJ, Bowyer KW, Vorder Bruegge RW, Grother PJ, Quinn GW, Pruitt M (2011) Distinguishing identical twins by face recognition, IEEE 2011
4. Subash Chandra Bose S, Christopher T (2019) Aggregate linear discriminate analyzed feature extraction and ensemble of bootstrap with KNN classifier for malicious tumor detection, IJRTE 2019
5. Patil K, Sachin Bojewar PG (2015) Identification of identical twins using face recognition. ARMIET, Mumbai University, MS, India 2VIT, Mumbai University, MS, India, IJETAE 2015

Modelling and Prediction of Cardiac Dysrhythmia



P. Ramesh Naidu, Shourya Thapliyal, Sambhav Bharvesh,
and Satyam Kumar

1 Introduction

Cardiac dysrhythmia is a classification of infirmities and disorders where the heartbeats are irregular that is excessively fast or excessively slow. Numerous types of dysrhythmias exist of which many have no symptoms. However, when symptoms are present, they may include palpitations or feeling of discontinuous heartbeats and rests in between. In some extreme cases, there could be chest pain, faint attacks, breath shortage, giddiness and dizziness [1]. While most types of dysrhythmia are not extreme, some of them can expose a person to extreme conditions such as stroke and heart failure, while many others may even result in cardiac arrest.

Dysrhythmia affects millions of people all over the world. Sudden cardiac death is the major cause of about nearly half of deaths due to cardiac disease which comprises about 15% of all deaths worldwide [2–4]. About 80% of sudden cardiac death is the result of ventricular dysrhythmias. Dysrhythmias may occur at any age but are more common among older people.

Any disruption to the impulses is root cause for the heart to contract fast resulting in dysrhythmia [5, 6]. The healthy person is having a resting heart rate of 60–100. The fit individual is having lower resting heart rate.

Many factors can affect the heart to work inaccurately, they include:

- Liquor exploitation
- drug abuse
- diabetes
- heart disease
- excessive coffee consumption

P. Ramesh Naidu (✉) · S. Thapliyal · S. Bharvesh · S. Kumar
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore 560064, India
e-mail: ramesh.naidu@nmit.ac.in

Fit being do not undergo from dysrhythmia, if they do not have any bad habits. It may be caused due to some internal situations which may source electrical impulses not to travel through heart properly, thus snowballing the chances of dysrhythmia. When your heartbeat is excessively fast, as well excessively slow [7–10], otherwise beats in an uneven rhythm, it is identified as a cardiac dysrhythmia (abnormal heart rhythm), which is very common and dangerous heart conditions. Maximum individuals have occasional cardiac dysrhythmias.

1.1 Research Motivation

The nature of the cardiac dysrhythmia and its resemblance to cardiac arrest has led to confusions and discrepancies even among the professionals in the medical field. Due to the subtle nature of the cardiac activity, its perils and hazards can hide in plain sight and become unrecognizable. This has led us to make use of the ECG waves which can be mapped and measured using an electro-cardiograph machine and later quantified into numerical values to form a dataset [11–14]. The entire graph on the Cartesian plane can be converted to float numerical values and stored as a dataset. By making use of this dataset, we can make attempts to create models and learn to predict the outcomes under the influence of various parameters. Successful design and models can help improve the detection of dysrhythmias and allow early detection of the same resulting in better diagnosis.

1.2 Problem Statement

To design and develop algorithms which are perfect for detection of cardiac dysrhythmia, to develop a system that correctly models, detects and classifies cardiac arrhythmia and to develop models with higher accuracy than the ones which act as an inspiration.

2 Literature Survey

A literature review survey, simply put, is an extensive analysis of books, papers, documents, or any relevant materials which have already been published pertaining to the topic at hand. Literature surveys are done as to get an idea of the work that has been already done in the concerned field, to determine the extent and coverage of the project.

The key components of a literature review can be:

- Give a new elucidation of topic or combine the new and old meanings,

- Keep a track of all major developments in the field,
- Advise the reader on most apt research by judging the sources, or
- Find disparity in methodology of the work done on the topic.

Literature review serves the following objectives:

- Rank each work on basis of information and knowledge provided.
- Express a connection or similarities between the school of thought of related works.
- Discover new ways to understand prior works.
- Reveal any gaps that exist in the literature.
- Fathom discrepancy among seemingly contradictory previous studied
- Distinguish key points of related work to prevent duplication
- Explicitly point the need for new research to be done

Predictive and classification of cardiac arrhythmia, Vasu Gupta, Sharan Srinivasan, Sneha S Kudli et al: The paper aims at using a variety of machine learning algorithms like SVM, Naive Bayes, random forest for predicting and classifying heart blocks into different classes [15, 16]. Experimentation was done with a couple of filter feature selection techniques. Using limited number of records helps in overcoming overfitting and helps in pin pointing the important features which will have the maximum effect on prediction and analysis out of 257 possible attributes. In the first approach, approximation of continuous values was done using discrete quantities and then quantify the mutual information $I(Y, X)$ between each feature and the output label vector. In the following approach, Mrmr which is a MATLAB feature selection package was used, thus selecting the features that have both minimum correlation among themselves and maximum correlation with the output labels (Fig. 1).

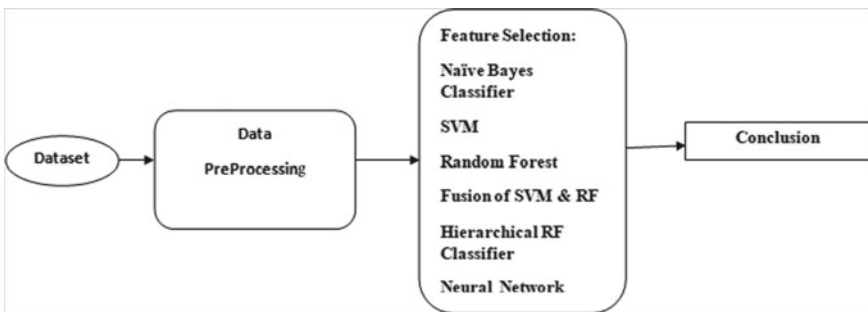
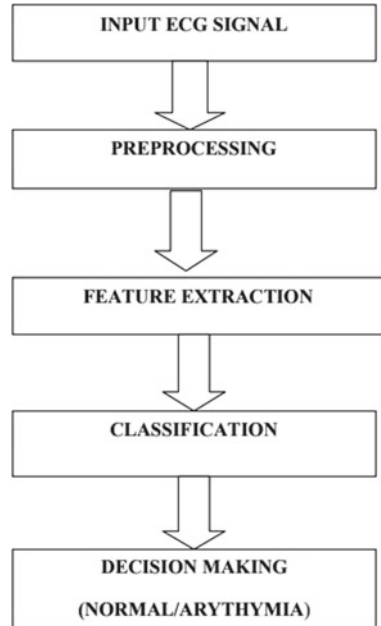


Fig. 1 Architecture of the system designed by the author

Fig. 2 Flow chart



2.1 Classification of Arrhythmia Using Machine Learning Techniques, Thara Soman and Patrick O. Bobbie

The set of following algorithms was used:

- **oneR**: stands for “one rule” as the name suggests it a very simple algorithm which formulates a single rule for predictor, then the predictor is tested against a testing set and the predictor with the least error frequency is chosen [17, 18]. Though it is a very simple algorithm, its results and accuracy are comparable to complex algorithms.
- **J48**: it is an implementation of C4.5 algorithm which generates decision trees which can then be used for classification. The decision trees are made by using the concept of information entropy. At each node, the algorithm performs a greedy technique to select an attribute which splits the set into one class or another.
Limitation: only applicable for Naïve Bayes (Figs. 2 and 3).

2.2 Prediction of Cardiac Arrhythmia, Varun Kathuria, Prakhar Thapliyal

Feature Selection: Data cleaning is performed on the training dataset. In the process, any attributes which were giving value of either 0 or 1 90% of times were gleaned out.

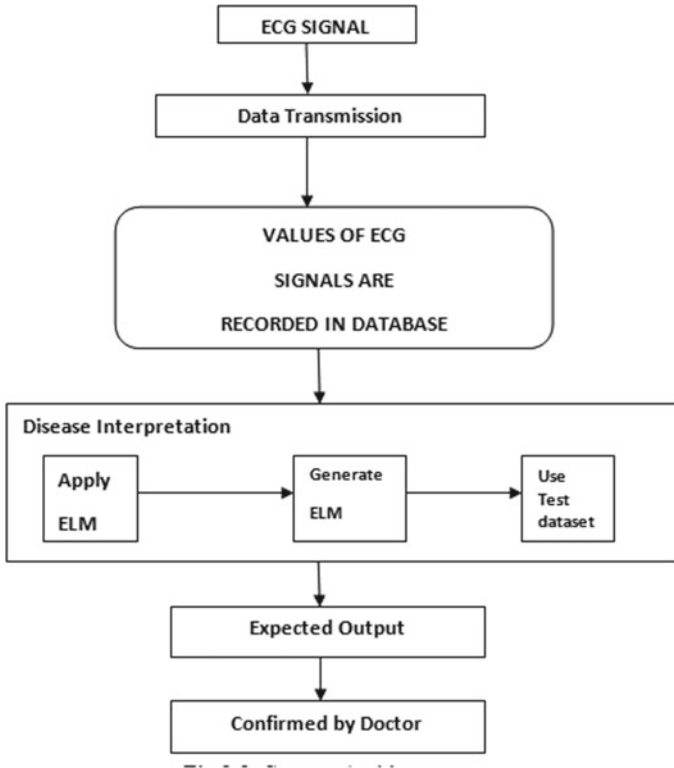


Fig. 3 Architecture of the system designed by the author

For missing values such as null, nil and NA, we replace them for better computation [19, 20]. The mean and standard deviation for such attributes found and the null data is replaced by mean \pm standard deviation. If for an attribute majority of values are missing, it is removed from the dataset.

Random Forest and Decision Tree: Classification is done using the random forest algorithm. The algorithm formulates random decision trees on every iteration through the dataset and then tries to test the data by fitting the test data a randomly generated tree. On every node, it decides by analyzing the values of the attributes that had been chosen at random while forming the tree. Several such iterations are done on randomly generated trees. The tree with the highest accuracy on test data is selected as the classification model for prediction.

Limitation: The accuracy is only 73%.

Summary

Through extensive study of the research papers mentioned above, it is clear that most of them consider feature selection and extraction an important step in the pre-processing phase. The technique of random forest is considered by two of the three above-mentioned papers. There is a stress of the study of support vector

Machines and Naïve Bayes where the former provides an accuracy of 77%, and the latter is applied only in specific cases, and its result with respect to other cases is left unobserved. Decision trees provide poor results leaving room for better alternatives. The process or data flow is very standard in case of machine learning algorithms and is common to all. This remarkable effort by such esteemed researchers is surely going to pave a way for the trails and experimentation of different models to further their efforts.

3 Implementation

Agile methodology is an application that encourages and assists us in several iterations of software development as well as testing it throughout the development lifecycle of the software. Agile methods mostly endorse very disciplined and strict project management processes which promote a greater number of inspections and adaptations, a philosophy that encourages the sense of teamwork within a team, their capability to self-organize and ultimately, encourages. Any development process that is associated with the theories of the agile manifesto falls under agile development.

3.1 Description of Process

The first testing was done in the first module, i.e., data pre-processing which is to ensure that the dataset does not contain any missing value or unknown value. The original CSV file is taken as input, and data cleansing is performed successfully.

The dataset is highly complex and multivariate and consists of multiple types of data comprising of categorical, integer, and real values. It contains 279 features and many of them have missing values which we have overcome by assigning it the mean of the corresponding attribute (Formula 3.1).

Further testing, basically, second and third testing is done in second module, i.e., feature extraction to reduce the dimensionality of dataset. The pre-processed csv file is taken and PCA, and random forest is successfully applied separately to get the reduced feature dataset.

The pre-processing stage comprises of applying two techniques for the extraction of features, namely principal component analysis and random forests. These two

$$A = \frac{1}{n} * \sum_{i=1}^n x_i$$

Formula 3.1 Arithmetic mean where A stands for arithmetic mean; n stands for the number of items whose average is being calculated; x_i stands for the value of each item in the list of numbers whose average is being calculated

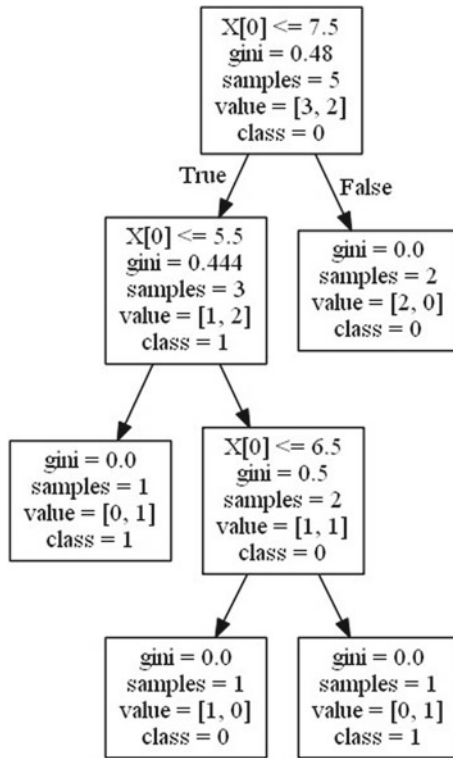
Formula 3.2 Covariance
used in PCA

$$cov(X,Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y})$$

Formula 3.3 Gini impurity
used in random forest

$$I_G(n) = 1 - \sum_{i=1}^J (p_i)^2$$

techniques allow us to know only the features with the positive contributions to the outcome. This allows us to remove any feature that is negatively impacting the outcome of the predictability of our model. Also it was observed that certain features provide high weightage to the outcome only in groups with the other features (Formulas 3.2 and 3.3).



The last four tastings are performed for one classifier each, i.e., KNN, weighted KNN, logistic regression, SVM and Naive Bayes to predict and classify the class of cardiac arrhythmia. The csv file with reduced features is taken as input, and the accuracy and classification are done.

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$

Formula 3.4 K-nearest neighbors formula

$$y' = \underset{v}{\operatorname{argmax}} \sum_{(\mathbf{x}_i, y_i) \in D_z} w_i \times I(v = y_i)$$

Formula 3.5 Weighted K-nearest neighbors formula

K-Nearest Neighbor and Weighted KNN: It is a type of instance-based learning where k-closest training examples is considered part of feature space. Weights were also given to each instance which improved overall results (Formulas 3.4 and 3.5).

Logistic Regression: This method is used to model probability of certain event or class existing as “pass/fail”. A sigmoid function represents change or growth in values from 0 to 1. During classification, each instance is either classified as either 0 or 1, and this is done to each outcome category (Formulas 3.6 and 3.7).

Naïve Bayes: Naïve Bayes is based on Bayes’s Theorem which is a probabilistic classifier that assumes a great degree of independence between the features. This

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)}}$$

Formula 3.6 Logistic regression sigmoid curve formula

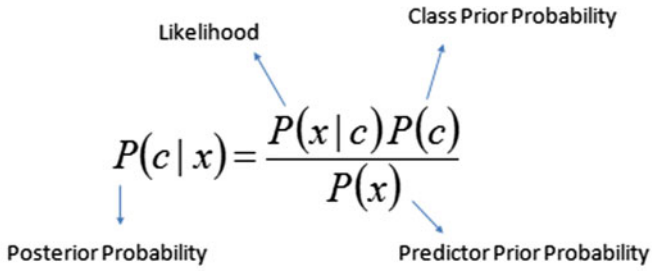
$$\beta^1 = \beta^0 + [X^T W X]^{-1}$$

β is a vector of the logistic regression coefficients.

W is a square matrix of order N with elements $n_i \pi_i (1 - \pi_i)$

μ is a vector of length N with elements $\mu_i = n_i \pi_i$.

Formula 3.7 Logistic regression matrix formula



$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Formula 3.8 Naïve Bayes formula

model of classification is purely mathematical in nature and uses continuous real values (Formula 3.8).

4 Results

See Figs. 12, 13, 14 and 15.

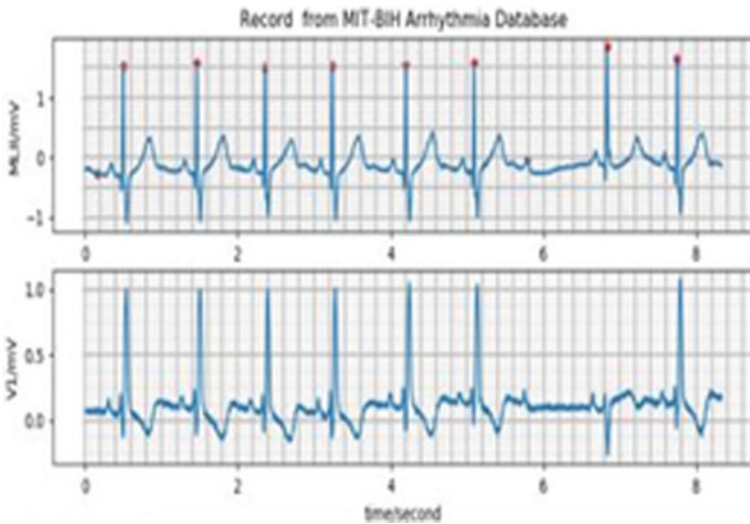


Fig. 12 ECG using wfdb python module

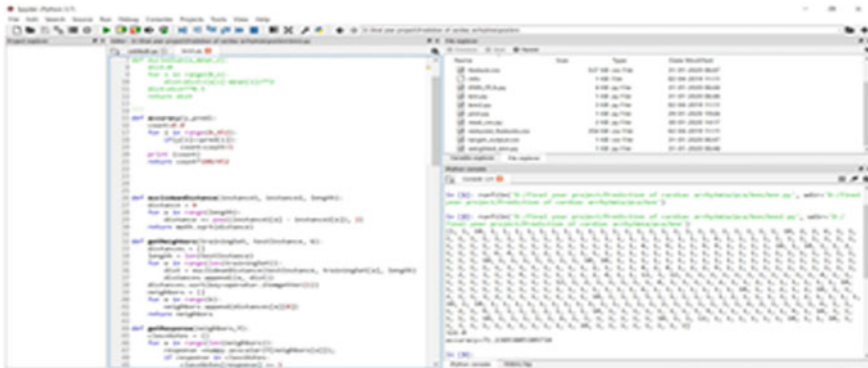


Fig. 13 KNN algorithm accuracy: 71%

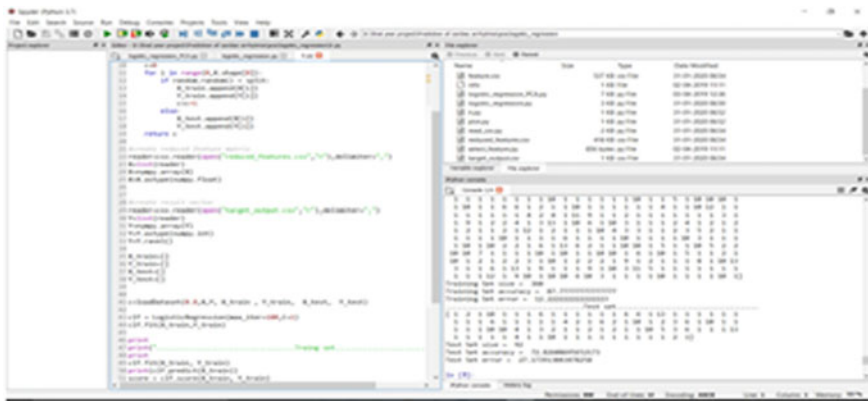


Fig. 14 Logistic regression accuracy: 72%

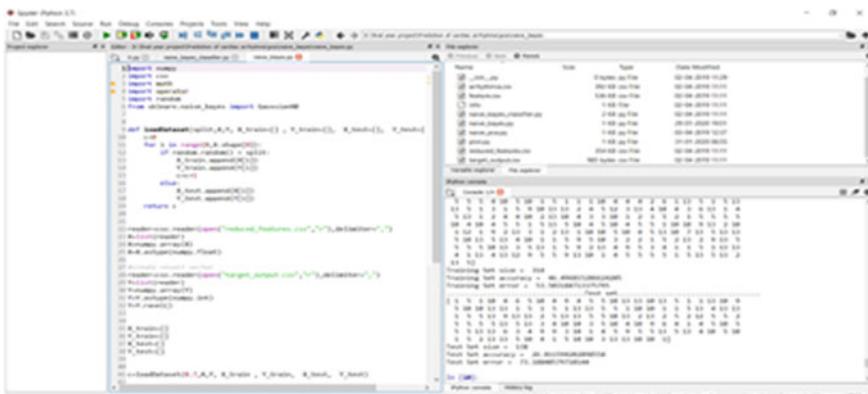


Fig. 15 Naïve Bayes accuracy: 26%

5 Conclusion and Future Work

The results of our project suggests that machine learning can be a great aide in diagnosing cardiac arrhythmias. This project helps in the prediction of cardiac arrhythmias in its earliest stage. The earliest prediction of cardiac arrhythmias would help to take precautions in earlier stages. Prediction of Atrial Fibrillation: A chronic condition that is progressive in nature and which affects the pumping capacity of the heart muscles is known as congestive heart failure. Early detection and diagnosis can prolong the human life. Prediction of congestive heart failure: Atrial fibrillation occurs when heart chambers do not work together as they should because of faulty electrical signaling. This too is extremely fatal and early detection and diagnosis can lead to improvement in overall health. There exist various other types of dysrhythmias that may not be as fatal but can lead to complications in old age, whereas there are some that show signs of fatality but might be simply a reason for poor health. Successfully categorizing into one of the 14 categories of dysrhythmias help us know the exact severity of the heart.

The project can be deployed in hospitals so that it can be continuously checked and verified with the new datasets of the patients. The project can be coupled with Internet of Things devices to provide continuous monitoring and reporting of cardiac cycles. The project can become more user friendly by adding the functionalities that the doctor may require in the near future.

References

1. Padmavathi K, Ramakrishna KS (2015) Classification of ECG signal during Atrial Fibrillation using Autoregressive modeling. In: *Procedia Computer Science*, vol 46, Issue ICICT 2014, pp 53–59
2. Sharma A, Bhardwaj K (2015) Identification of normal and abnormal ECG using neural network. *Int J Information Res Rev* 2(05):695–700
3. Subbiah S, Patro RK, Subbuthai P (2015) Feature extraction and classification for ECG signal processing based on artificial neural network and machine learning approach. In: *International Conference on Inter Disciplinary Research in Engineering and Technology [ICIDRET]*, vol 1, Issue, ICIDRET007, pp 50–57
4. Huang G, Huang GB, Song S, You K (2015) (2015), -Trends in extreme learning machines: a review. *Neural Netw* 61:32–48
5. Dankan V, Gowda D, Kishore V, Shivashankar A, Ramachandra C, Pandurangappa C (2017) Optimization of motorcycle pitch with non linear control. In: *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016—Proceedings*, pp 1656–1660. <https://doi.org/10.1109/RTEICT.2016.7808114>
6. Dankan V, Kishore D, Gowda V, Shivashankar, Mehta S (2016) MANET topology for disaster management using wireless sensor network. In: *International Conference on Communication and Signal Processing, ICCSP 2016*, pp 0736–0740. <https://doi.org/10.1109/ICCSP.2016.7754242>
7. Dankan Gowda V, Ramachandra AC, Thippeswamy MN, Pandurangappa C, Ramesh Naidu P (2019) Modelling and performance evaluation of anti-lock braking system. *J Eng Sci Technol* 14(5):3028–3045

8. Penna M, Gowda DV, Jijesh JJ, Shivashankar (2017) Design and implementation of automatic medicine dispensing machine. In: RTEICT 2017—2nd IEEE international conference on recent trends in electronics, information and communication technology, Proceedings, vol 2018, pp 1962–1966. <https://doi.org/10.1109/RTEICT.2017.8256941>
9. Gowda DV, Ramachandra AC, Thippeswamy MN, Pandurangappa C, Ramesh Naidu P (2018) Synthesis and modeling of antilock braking system using sliding mode controller. *J Adv Res Dyn Control Syst* 10(12):208–221
10. Gowda DV, Varun CA, Shivashankar, Sahana M, Varun RS, Rajesh T (2018) Implementation of swarm intelligence in obstacle avoidance. In: RTEICT 2017—2nd IEEE international conference on recent trends in electronics, information and communication technology, proceedings, vol 2018, pp 525–528. <https://doi.org/10.1109/RTEICT.2017.8256652>
11. Gowda DV, Sridhara SB, Naveen KB, Ramesha M, Pai GN (2020) Internet of things: Internet revolution, impact, technology road map and features. *Adv Math Sci J* 9(7):4405–4414. <https://doi.org/10.37418/amsj.9.7.11>
12. Ramesh Naidu P, Guruprasad N, Gowda DV (2020) Design and implementation of cryptcloud system for securing files in cloud. *Adv Math Sci J* 9(7):4485–4493. <https://doi.org/10.37418/amsj.9.7.17>
13. Ramesha M, Gowda DV, Sridhara SB, Naveena Pai G (2020) FPGA Implementation of low power high speed BTED algorithm for 8 bit error correction in cryptography system. *Int J Emerg Trends Eng Res* 8(7):3893–3897. <https://doi.org/10.30534/ijeter/2020/158872020>
14. Gowda DV, Ramachandra AC (2018) Importance of non-linear controller in implementing anti-lock braking system—a technical review. *Int J Adv Res Comput Sci* 9(2):193–199. <https://doi.org/10.26483/ijarcs.v9i2.5676>
15. Gowda DV, Ramachandra AC (2017) Slip ratio control of anti-lock braking system with bang-bang controller. *Int J Comput Tech* 4(1):97–104 (online). Available: <http://www.ijctjournal.org>
16. Gowda DV, Penna M, Palle S (2017) Digital intercommunication system in advanced light helicopter (ALH). *Adv Comput Sci Technol* 10(5):699–708
17. Gowda DV, Chakrasali S (2014) Comparative analysis of passive and semi-active suspension system for quarter car model using PID controller. In: Proc. of Int. Conf. on Recent Trends in Signal Processing, Image Processing and VLSI, ICrtSIV, no. May, pp 510–517, doi: 03.AETS.2014.5.131
18. Müller AC, Guido S (2016) Introduction to machine learning with python: a guide for data scientists, O'Reilly Media
19. Kathuria V, Thapliyal P (2017) Prediction of cardiac arrhythmia. MIT Publications. Jitendra Kumar, Gajendra E (2015) A novel approach of ecg classification for diagnosis of heart diseases. *IJARECT*, vol 4, issue 11
20. Soman T, Bobbie PO (2015) Classification of arrhythmia using machine learning techniques. School of Computing and Software Engineering Southern Polytechnic State University (SPSU) Publications

Lung Cancer Detection and Prediction Using Customized Selective Segmentation Technique with SVM Classifier



Ashwini S. S., M. Z. Kurian, and M. Nagaraja

1 Introduction

Malignant cancer is a rapid growth of cells in a certain area of the body. The transformation and multiplication of cells are two important aims behind this rapid development. Due to this multiplicative cell growth, an abnormal organ growth will result in tumor formation. This can happen in any part of the body. It is supposed to be a lung malignant growth during the multiplication process of cell development in lungs, and there are two types of lung malignancies: non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). In the early stages, identification of lung disease is not possible. Therefore, it can be helpful to identify the initial period of lung malignancy to control the rate of death. The accuracy of the lung malignant growth acquired by radiologists is much lower while dealing with enormous volumes [1–3].

CT scanned images are better compared to other imaging modalities for recognition of small lung nodules. But there are numerous challenges in CT, for example the constraint of the human visual prediction and experience, and factors like weariness or interruption that leads high paces of miss predictions of disease. Here, an alternate method is normally required to diminish wrong predictions. This can be normally performed by utilizing computer-aided design (CAD) frameworks that are helpful for radiologists [4–7].

The processing of images can help in the early stages to identify the malignant growths that promote early treatment. The processing of images comprises four key phases, preprocessing, segmentation, characteristic extraction, and classification. This paper provides machine learning techniques by using the CT image filtering

Ashwini S. S. (✉) · M. Z. Kurian

Department of ECE, SSIT, Sri Siddhartha Academy of Higher Education, Tumakuru 572105, India

M. Nagaraja

Department of Physics, SSIT, Sri Siddhartha Academy of Higher Education, Tumakuru 572105, India

and auto-contrasting and finally distinguishes lung malignancy by the segmentation with the SVM classification. We are aiming to enhance the accuracy of lung cancer identification with various segmentation techniques.

2 Related Work

Although there are several image processing techniques to identify early lung disease, but new strategies are currently required to improve in terms of accuracy, impacts on the ability, characteristics, and the search for new techniques.

Tanushree et al. [8] carried out a research by focusing on arrangement of images of the lung, whether or not it is influenced by malignant ailments. Fuzzy interface mechanism and neural are used for classification of the image. The accuracy of the developed approach is 94.12%.

Ritika Agarwal et al. [9] considered the idea of clinical substance dependent picture recovery. As the author suggests, various techniques are available to determine whether nodules are present in the lungs at its initial stage, but none of the approach is successful. They have developed CAD system based on retrieval of content of medical CT image of lung which helps to identify nodule in it. Process contains acquisition of image from database, segmentation, CBIR feature extraction, SVM as training classifier and followed it by testing.

Mukherjee et al. [10] presented a programmed CAD framework dependent on CT-examine pictures. The framework introduced uses the traditional histogram-dependent thresholding strategy in an iterative manner. The nodules inside are recognized through a rule-based separate approach after extraction of the lungs area.

To classify the type of lung cancer, Emre Dandil et al. [11] developed a computer-aided diagnosis (CAD) system. In this system, image was enhanced by using median filter and histogram equalization was performed to increase the contrast by altering the intensity of the image. Extraction of Lung lobes was done by applying morphological operation, and remaining pieces on the edges were removed using thresholding approach. Self-Organization Map (SOM) was employed for segmentation of image. For feature extraction GCLM was used and output of this given to principle component analysis (PCA) for feature reduction. Finally, artificial neural network (ANN) is applied for classification, and at the end, this is implemented by CAD system which has given the accuracy of 90.63%.

Anum et al. [12] have been designed and implemented a CAD support system for lung nodule detection and classifying the stages of cancer. In this work, CT images are collected from various databases, and from data recording using wearable sensor devices (IoT), these images are enhanced using Gabor filter, and lung region of interest is extracted using thresholding approach, classification is done using deep fully convolution network (DFCNet). Obtained accuracy of implemented system is 84.58%. Thus, the performance measure of designed system effectively helps radiologist for classification of lung cancer stages.

Generally, various methods have been utilized for the image processing to encourage early location of lung malignancy. The fact is, CAD system can identify early lung malignancy. But, there remain issues in terms of the needs of precision. The SVM classifier can be used to limit the issue with the combination and custom selection of various segmentation techniques to give better outcomes to the CAD framework to improve accuracy of detection.

3 Segmentation

3.1 Threshold Segmentation

Segmentation refers to the division of a picture into sections (or contours) which corresponds to images. In general, by defining basic characteristics, we seek to segment regions. Similarly, by recognizing variations among regions (contours), we define contours. Intensity is the simplest thing pixels can share in a region. Thus, the threshold, the separation between light and dark areas is a normal way to segment such areas. Thresholding produces gray-level binary pictures by converting all the pixels under a threshold to zero and all the pixels above the threshold to one.

If $g(x, y)$ represent threshold version of $f(x, y)$, at certain global threshold T .
 $g(x, y) = 1$ if $f(x, y) \geq T$, otherwise '0'.

3.2 K-Means Segmentation

The segmentation of K-means is focused on

- Partitioned clustering approach.
- The centroid is affiliated with every cluster.
- A cluster with a nearby centroid is assigned to every point.
- The centroid (center point) is affiliated with each cluster.
- A cluster with a nearby centroid is assigned to each point.
- K-number of clusters is specified.

Fundamental **k-means** algorithm is as follows:

```
{
Choose k-points as the first center
Repeat
Create k-clusters in the nearest centroid distributing all the points.
Calculate each region's centroid.
Until—the centroids do not alter
}
```

3.3 Marker-Controlled Watershed Segmentation

Marker-controlled watershed segmentation follows the below steps:

- Step-1: Gradient magnitude is used to compute a segmentation function in which gradient is high at borders of object and low inside object.
- Step-2: Foreground objects are marked with the help of morphological operations: opening–closing reconstruction. Followed by it, computation of its regional maxima operation creates flat maxima inside object.
- Step-3: Background objects are marked using thresholding operation.
- Step-4: Watershed transform of segmentation function is computed, so that it has regional minima only at certain desired areas.
- Step-5: Resultant image is obtained by superimposing foreground markers, background markers and segmented object boundaries on original image.

4 Proposed Method

This research work proposes a machine learning dependent lung malignant detection and prediction with improvement in the accuracy of prediction. The procedure followed to implement the identification of lung malignant as mentioned below and Fig. 1 shows the flow of the proposed method.

- Acquisition of CT lung image from database.
- Median filter is used to remove noises as it does not blurs image much together it preserves the edges of images which is necessary.
- Then lung is segmented using thresholding, K-means, and marker-controlled watershed segmentation algorithm as shown in flow diagram.
- GLCM (texture features) of the segmented images are extracted and trained to SVM classifier.
- SVM classifier gives different classification results based on specific segmented image features.
- Customized decision is performed to select the accurate results from the classified results of the SVM classifiers.

5 Experimental Results

Effectiveness of proposed method is showed by acquiring nearly 150 images from cancer imaging archive database. Experiment has been carried out using MATLAB environment. Figure 2 shows the entire simulation process of our proposed method using GUI feature.

CT lung image (normal/affected) is loaded as input image as shown in GUI (Fig. 2). Salt and pepper noise is added to the input image and filtered using median filter.

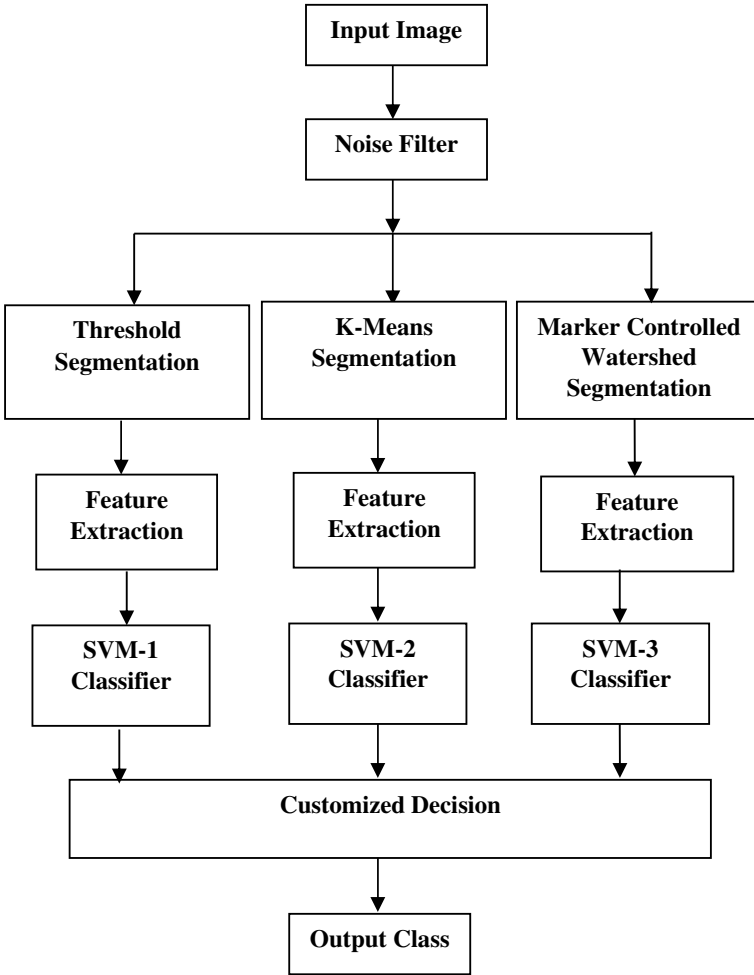


Fig. 1 Proposed flow diagram

Further image is enhanced by using auto contrast enhancement. Figure 3 illustrates noisy and filtered image.

Different segmentation methods such as thresholding, K-means clustering, and marker-controlled watershed transform are applied to the preprocessed images. Figure 4 presents output images of respective segmentation methods. From the visualization of output images of different segmentation methods, it can be identified that marker-controlled watershed transform segmentation gives the better output result.

The customized decision process has been carried out for various cases (for normal/affected images) by using three SVM classifiers. Customized decision is based on the higher accuracy of predicted class. Accuracy of each case and for different SVM is evaluated. The obtained accuracy results are shown in Table 1.

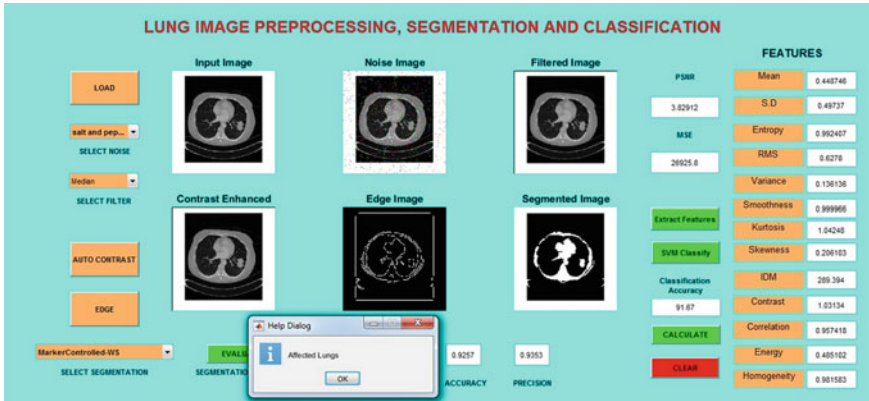


Fig. 2 Simulation of proposed work

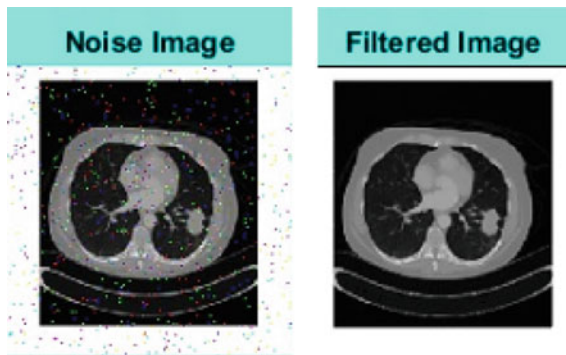


Fig. 3 Noise added and filtered image of lung

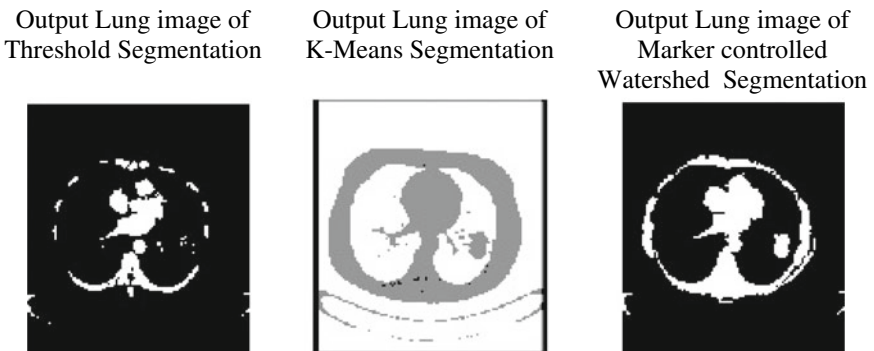


Fig. 4 Output Images of different segmentation methods

Table 1 Accuracy and customized decision of lung cancer images

Cases	Class	Normal/Affected	Accuracy (%)	Predicted class	Customized decision
CASE-I	SVM-1	NORMAL	83.25	NORMAL	NORMAL
	SVM-2		85.26	NORMAL	
	SVM-3		92.78	NORMAL	
CASE-II	SVM-1	AFFECTED	84.23	NORMAL	AFFECTED
	SVM-2		86.65	AFFECTED	
	SVM-3		92.67	AFFECTED	
CASE-III	SVM-1	NORMAL	81.65	NORMAL	NORMAL
	SVM-2		87.45	AFFECTED	
	SVM-3		91.83	NORMAL	
CASE-IV	SVM-1	AFFECTED	81.33	AFFECTED	AFFECTED
	SVM-2		86.24	AFFECTED	
	SVM-3		92.35	AFFECTED	
CASE-V	SVM-1	AFFECTED	82.76	AFFECTED	AFFECTED
	SVM-2		89.25	AFFECTED	
	SVM-3		92.72	AFFECTED	
CASE-VI	SVM-1	NORMAL	83.78	AFFECTED	NORMAL
	SVM-2		85.76	NORMAL	
	SVM-3		91.54	NORMAL	

Table 2 Average accuracy of SVM and their segmentation

SVMs	Segmentation method	Average accuracy (%)
SVM-1	Threshold	82.81
SVM-2	K-means clustering	86.76
SVM-3	Marker-controlled watershed transform	92.31

As calculated and indicated in Table 2, the average accuracy of SVM-1 classifier with threshold segmentation is 82.81%, SVM-2 classifier with K-means clustering method has given the average accuracy of 86.76%, whereas SVM-3 classifier with marker-controlled watershed transform segmentation achieved the average accuracy of 92.31%.

6 Conclusions

In this paper, we proposed a technique for lung cancer identification and customized selective segmentation techniques with SVM classifiers and showed the improvement

in accuracy. Experimental outcomes signify that our proposed method has given the improved average accuracy of 92.31%. The proposed method not only improved the accurate estimation of cancer, it also explores the combined advantages of various segmentation techniques in machine learning for biomedical image processing in future. This proposed work can be further extended for different stages of lung cancer, and accuracy can be increased by training the model with more number of data sets.

References

1. Ammi RP, Giri Babu K, Rao VK, Ramesh Babu I (2016) Automated lung segmentation from HRCT scans with diffuse parenchymal lung diseases. *J Digit Imaging* 29:507–519. <https://doi.org/10.1007/s10278-016-9875-z>
2. Dignam JJ, Huang L, Ries L, Reichman M, Mariotto A, Feuer E (2009) Estimating breast cancer-specific and other-cause mortality in clinical trial and population-based cancer registry cohorts. *Cancer* 115:5272–5283. <https://doi.org/10.1002/cncr.24617>
3. Shruthi I, Robin J (2015) Computer aided lung cancer detection system. In: *L 2015 Global Conference On Communication Technologies (GCCT)*, pp 555–558. <https://doi.org/10.1109/GCCT.2015.7342723>
4. Armato SG, Sensakovic WF (2004) Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis. *Acad Radiol* 11:1011–1021. <https://doi.org/10.1016/j.acra.2004.06.005>
5. Sluimer I, Prokop M, Ginneken BV (2005) Towards automated segmentation of the pathological lung in CT. *IEEE Trans Med Imag* 24:1025–1038. <https://doi.org/10.1109/TMI.2005.851757>
6. Xu Y, Sonka M, McLennan G, Guo J, Hoffman EA (2006) MDCT-based 3-D texture classification of emphysema and early smoking related lung anthologies. *IEEE Trans Med Imag* 25:464–475. <https://doi.org/10.1109/TMI.2006.870889>
7. Mithuna BN, Pushpa R, Arpitha CN (2018) A quantitative approach for determining lung cancer using CT scan images. In: *Proceedings of the 2nd International conference on Electronics Communication and Aerospace Technology (ICECA 2018)*, pp 1786–1790. <https://doi.org/10.1109/ICECA.2018.8474670>
8. Tanushree SR, Neeraj S, Arti P (2015) Classification of lung image and nodule detection using fuzzy inference system. In: *International Conference on Computing, Communication and Automation (ICCCA2015)*, pp 1204–1207. <https://doi.org/10.1109/CCAA.2015.7148560>
9. Ritika A, Ankit S, Raj Kumar S (2015) Detection of lung cancer using content based medical image retrieval. In: *5th international conference on advanced computing and communication technologies*, pp 48–52. <https://doi.org/10.1109/ACCT.2015.33>
10. Mukherjee M, Biswal PK (2018) Segmentation of lungs nodules by iterative thresholding method and classification with reduced features. In: *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp 450–455. <https://doi.org/10.1109/ICICCT.2018.8473287>
11. Emre D, Murat C, Ziya E, Murat O, Ozlem KK, Arzu C (2014) Artificial neural network-based classification system for lung nodules on computed tomography scans. In: *6th international conference of soft computing and pattern recognition*, pp 382–386. <https://doi.org/10.1109/SOCPAR.2014.7008037>
12. Anum M, Bin S, Ping L, Xuhong H, Xiaoe W, Jing Q, Dagan F (2018) Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *J Biom Info* 79:117–128. <https://doi.org/10.1016/j.jbi.2018.01.005>

Design and Implementation of Storage Benchmark Kit



Keshava Munegowda and N. V. Sanjay Kumar

1 Introduction

The Storage Benchmark Kit (SBK) [1–3] is an open source, vendor neutral, high-performance storage benchmarking software framework. The SBK is containerized using dockers [3], and it is cloud deployable too. It is designed to support any storage device/client with any data type as a payload. The SBK supports multiple writers and readers/callback (push) readers performance benchmarking. The design of SBK is inspired by the Pravega benchmark tool [4, 5]. The Pravega benchmark tool [5] is specific to Pravega [6] and Kafka [7, 8] performance benchmarking, whereas the SBK supports a variety of storage systems such as Apache Bookkeeper [9, 10], Hadoop Distributed File System (HDFS) [11, 12], RabbitMQ [13, 14], RocketMQ [15], NATS [16], NATS Streaming [17], ActiveMQ Artemis [18], NSQ [19], Apache Pulsar [20] along with the existing Pravega and Kafka streaming storage systems.

The SBK also supports the performance benchmarking of database systems such as Apache Derby [21], MySQL [22], PostgreSQL [23], Microsoft SQL [24] and SQLite [25] through Java DataBase Connectivity (JDBC) [26].

The SBK also supports the performance benchmarking of a persistent key-value store such as RocksDB [27] and the distributed key-value store such as FoundationDB [28]. It also supports the performance benchmarking of Protocol buffer-based [29] Record layer [30] of FoundationDB and document-based databases such as FoundationDB Document layer [31] and MongoDB [32]. The performance benchmarking of object storage systems such as MinIO [33] is also supported in SBK. The SBK implements the periodic logging of benchmarking results to the Grafana [34] analytics platform through Prometheus monitoring systems [35].

K. Munegowda (✉)
DellEMC, Bangalore, India

N. V. Sanjay Kumar
KIT, Tiptur, India

In this paper, the design and implementation of SBK are detailed. To demonstrate the performance benchmarking capabilities of SBK, the XFS file system, HDFS, and Kafka performance benchmarking are conducted. Since Red Hat Enterprise Linux (RHEL) version 7.4 supports XFS as the default file system, we selected XFS for file system benchmarking. In our experiments, it is confirmed that HDFS is stable and high performing in the distributed file systems category and Kafka is stable and performs well in the distributed streaming storage systems category.

2 Design of SBK

The following internal components of SBK are shown in Fig. 1.

2.1 *SBK Benchmark*

The SBK benchmark parses and processes the application/user supplied or command line arguments, configures the multiple writers, readers, and the component “SBK performance processor.” For some of the storage systems/distributed messaging platforms like RabbitMQ [13, 14] and RocketMQ [15], the SBK initiates the callback (push) asynchronous readers too.

2.2 *Writers and Readers/Callback (Push) Asynchronous Readers*

These components initiate the performance benchmarking of write and read operations. These components implement Burst Mode/Max Throughput Mode [4] to measure the maximum write/read throughput of a storage system, Throughput Mode [4] and Rate Limiter Mode [4] to analyze the latency variations under controlled throughput or rate of events/records and End to End Latency Mode [4] to determine the total time duration between time to write a data record and reading the same data record.

2.3 *Data Type Handler*

The Data Type handler defines the type of data and methods/operations to operate on the data. Example data type handlers are Byte Array [36], Java NIO Byte Buffer [36], Java String [36], and Protocol buffers [29].

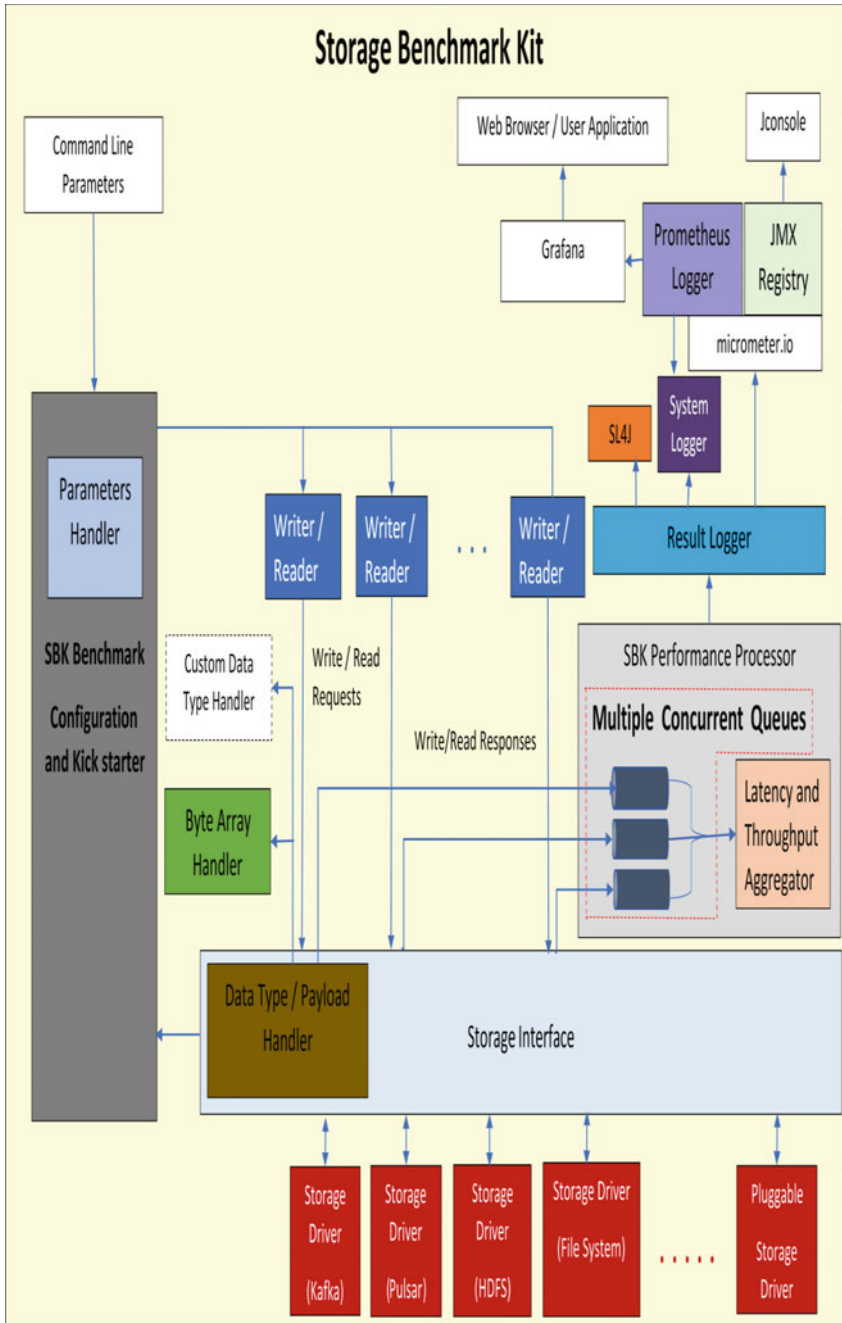


Fig. 1 Design of SBK

2.4 *Storage Interface and Driver*

The SBK defines and implements default methods for the storage interface which are extended and used to implement a custom and pluggable storage driver for any storage device/client. This pluggable storage driver component defines the write and read operations of the storage device/client. A single storage driver component implements a single or multiple instances storage device/client. The pluggable storage driver either chooses one of the available data type handlers or defines a new custom data type handler.

2.5 *SBK Performance Processor*

The SBK Performance Processor solves the synchronization issues between the multiple writers, readers and the response threads which are created upon completion of asynchronous write or read operations. This component uses multiple concurrent queues [37–39] to store an information record enclosing performance values such as start time, end time, number of records, and number of bytes in the total number of records for a single or multiple write/read operations. The java concurrent linked queue [38, 39] provides a thread-safe and wait-free/non-blocking [37] application programming interface (APIs) for enqueue and dequeue operations. Multiple response threads of the write/read completion or writers and readers enqueue the performance values to these multiple concurrent linked queues, but a single dedicated thread named “Latency and Throughput Aggregator” dequeues the stored performance values from these multiple concurrent linked queues to calculate the latency and throughput values. The SBK treats the response time of write/read as the latency value. The latency counts are stored in an array in which latency value is used as an index. These latency counts are extracted for calculation of the latency percentiles [4]. The latency percentiles calculation method used in SBK is inspired by counting-sort algorithm with time complexity of Big O (Maximum Latency).

2.6 *Result Logger*

This component receives the benchmark results such as throughput values, average, and maximum latency values and latency percentiles for every predetermined time interval from the SBK performance processor component. These benchmark results are logged to a local output device. The SBK uses the micrometer [40] software interface to log the results to the Prometheus [35] monitoring system and SL4J (Simple Logging façade For Java) [41] logging system. The Grafana [34] analytics platform receives these benchmark results from Prometheus [35] monitoring system.

3 Implementation Details

The SBK is implemented in Java 8 and is open sourced in the Git repository hub [1]. The SBK Docker images are available at the Docker Hub [3] too. The SBK Release Version 0.8 [2] is used for performance benchmarking of the file system, HDFS and Kafka presented in this paper. The SBK GitHub [1] details the guidelines for open source developers to improve the SBK source code and to add a new driver for performance benchmarking of any other storage system.

4 Results and Discussion

4.1 File System Performance Benchmarking

The Flexible IO (FIO) [42] and IOzone [43, 44] are the commonly used file system benchmarking tools. These tools are implemented in the programming language C. The SBK uses the Java file channel [36] APIs and ByteBuffer [36] as data record for the file write and read operations. The hardware and software configurations used for performance benchmarking are listed in Table 1.

Single writer File System Performance Benchmarking: Fig. 2 shows the Grafana snapshot of the single writer file system performance in terms of MB/s (Mega Bytes/second). In our benchmarking experiments, we set the record size as 1,000,000 bytes (approximately 1MB) and the total file size to write is 1 TB (Tera

Table 1 Hardware and software configuration of the test setup

Components	Remarks
Number of compute nodes	4 Nodes 1 for HDFS client/Kafka client 1 XFS File system node/HDFS Name Node 3 for Kafka brokers/HDFS Data nodes
CPUs (Central Processing Unit) per compute node	32 CPUs. Each of CPU is 64 Bit, 2.6 GHz (Giga Hertz)
RAM (Random Access Memory) per node	350 GB (Giga Bytes)
Hard Disk per node	HDD (Hard Disk Drive) of size 3 TB (Tera Bytes)
Ethernet per node	10 Gbps (Giga Bits/second) Network
Operating System	RHEL (RedHat Enterprise Linux) Release version 7.4

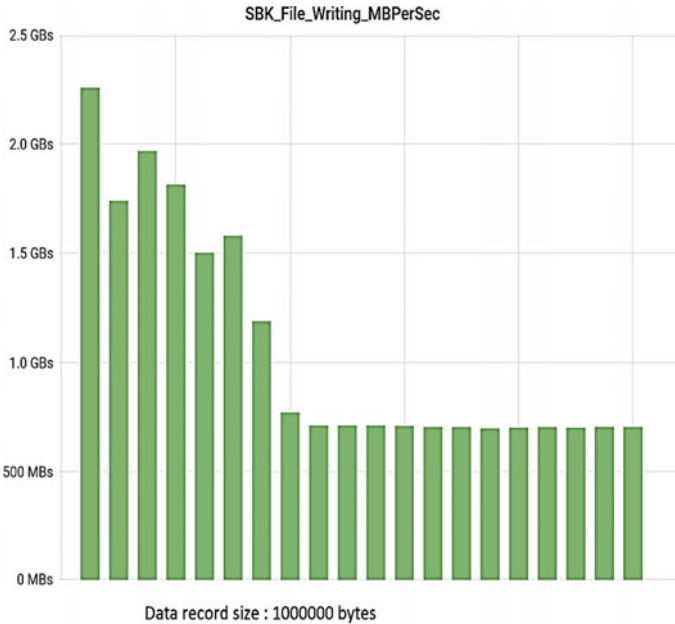


Fig. 2 Single File Writer throughput performance in MB/s

Bytes). Maximum throughput of 2.2 GB/s and average write throughput range of 750–780 MB/s are observed in our test setup.

Single and Multiple Readers File System Performance Benchmarking: Fig. 3 shows the read throughput of the single reader with a record size of 1,000,000 bytes (approximately 1MB) to read a 1 TB size file. The peak throughput is 4–5 GB/s and an average read throughput is 650–680 MB/s.

The SBK scales high with multiple readers. Figure 4 shows the read throughput of 10 file readers, the peak performance of 40–50 GB/s (Giga Bytes/second), and an average performance of 6–6.4 GB/s are observed and it indicates that the SBK scales high with multiple readers.

4.2 Hadoop File System (HDFS) Performance Benchmarking

TestDFSIO (Distributed File System Input-Output) [12, 45] is the de-facto performance benchmark tool for HDFS. This tool uses the Map-Reduce framework/programming model and hence achieves more parallelism for HDFS write and read operations. But, SBK uses the HDFS Java stream APIs rather than Map-Reduce framework. Thus, the SBK does the performance benchmarking of the raw

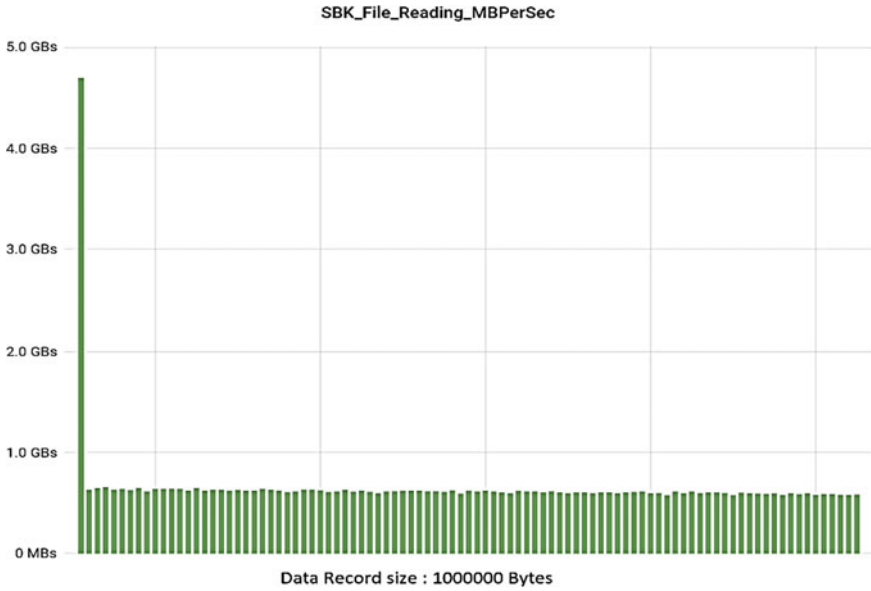


Fig. 3 Single File Reader throughput performance in MB/s

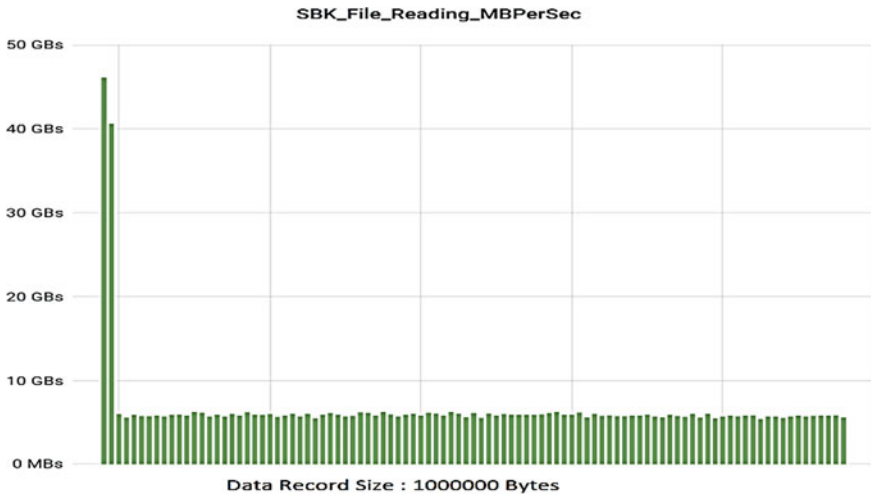


Fig. 4 10 File Readers throughput performance in MB/s

distributed file system with Byte Array as data type. The Hadoop version 3.2.0 [45] is used for performance benchmarking.

Single writer HDFS performance benchmarking: Like any other file systems, the HDFS supports write, append and overwrite of a file by a single writer only.

Figure 5 shows the Grafana graph snapshot of the write throughput of the single writer of one file. Note that, the max throughput is in the range of 500–550 MB/s (Mega Bytes per Second), and the average throughput range is 200–280 MB/s. In our benchmarking experiments, we set the block size as 1,000,000 bytes (approximately 1 MB) and the total file size to write is 1 TB (Tera Bytes).

Single and Multiple HDFS Readers performance benchmarking: The SBK scales high with multiple readers. Figure 6 shows the read throughput of the single reader and 10 readers of each reader of block size 1,000,000 bytes (approximately 1MB) to read 1 TB size file. Note that, with 10 readers SBK records the peak throughput of more than 2 GB/s (Giga Bytes per second).

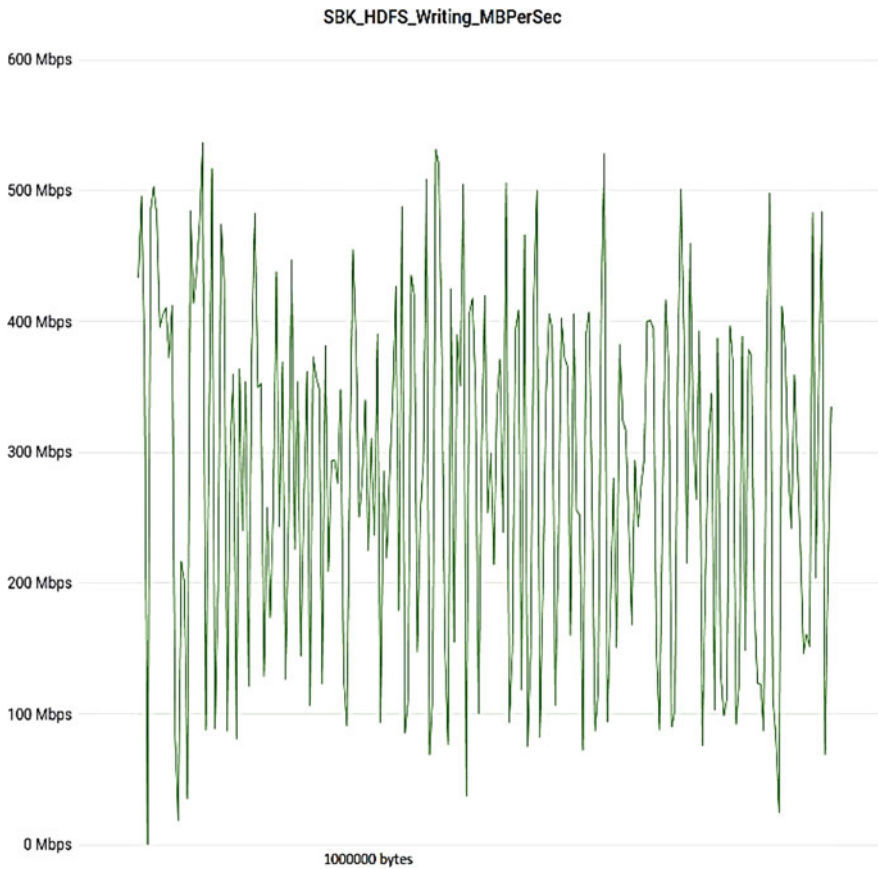


Fig. 5 Single HDFS Writer throughput performance in MB/s

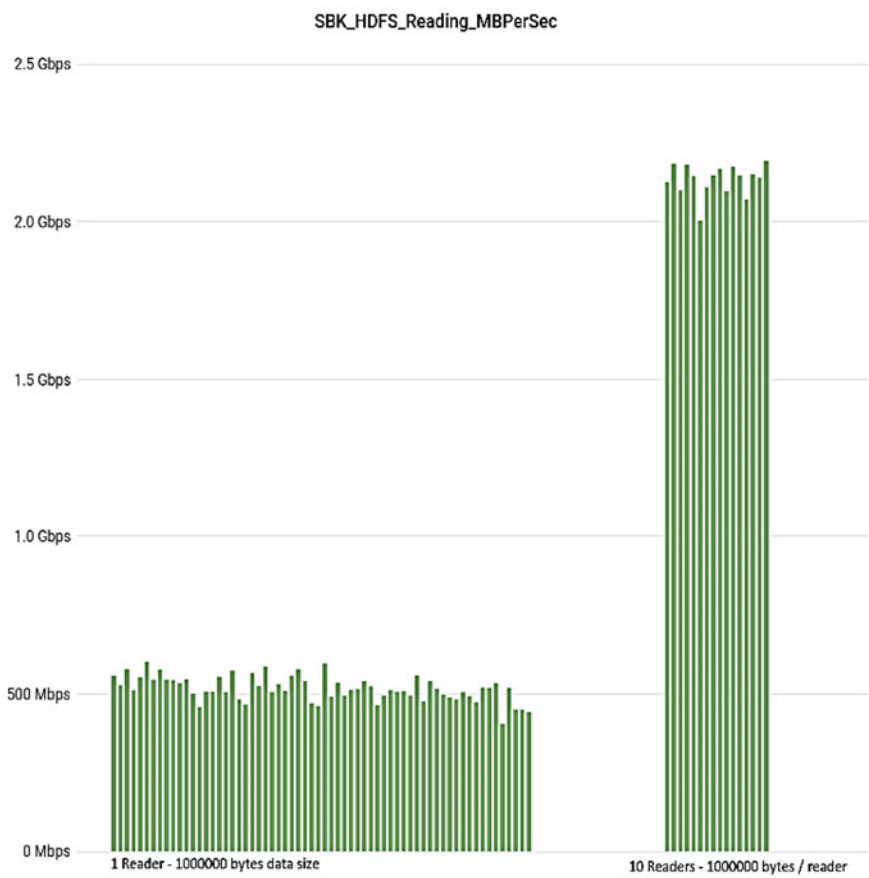


Fig. 6 Single and 10 HDFS Readers throughput performance in MB/s

4.3 Kafka Performance Benchmarking

The Kafka release version 2.4.1 [46] is used for performance benchmarking. A Kafka topic with 15 partitions, 3 replicas in which 2 are in-sync replicas, with the hardware configuration shown in Table 1 is used for performance benchmarking. The max poll (maximum poll records for consumer) configuration is set to 32-bit integer max value to read the maximum available records from the Kafka broker. Assigning the higher value for max.poll configuration value improves the Kafka read performance. The Byte Array [36] is used for data serialization and deserialization.

An existing Kafka (version 2.4.1) producer benchmark tool [4, 8, 46] for write performance benchmarking and consumer benchmark tool [4, 8, 46] for read performance benchmarking, are limited only to a single producer/writer and

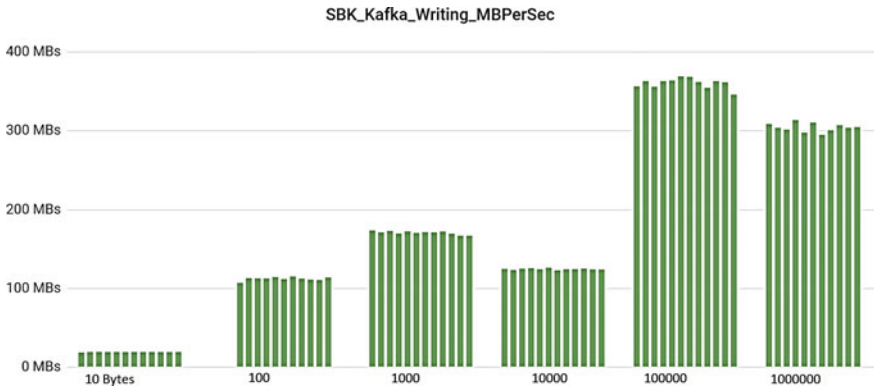


Fig. 7 Kafka Single producer throughput performance in terms of MB/s

consumer/reader, respectively, whereas SBK supports benchmarking of Kafka write and read with multiple writers and readers.

The open messaging benchmark [47, 48] software supports the Kafka performance benchmarking. As of today, the open messaging benchmarking is developed with the concept of benchmarking distributed messaging and streaming platforms, whereas SBK adheres to the concept of benchmarking any storage system with any type of data payload. The SBK supports the benchmarking of distributed messaging and streaming platforms along with any generic persistent or non-persistent storage platform.

Single Kafka Producer Benchmarking: The Kafka performance benchmarking is conducted with varying data sizes from 10 bytes to 100,000 bytes. The Grafana results snapshot of throughput variations for these different data sizes for a single producer is shown in Fig. 7.

Note that, for a single producer, the SBK records the peak performance in the range of 280–370 MB/s for approximate data sizes of 100 K and 1 MB. The same throughput results can be depicted in terms of records/second or writes/second (wps) as shown in Fig. 8. For smaller record/event sizes, the wps values are high and for larger data size the wps values are very low. The peak throughput is recorded in the range of 1.5–1.8 million wps. Due to a large gap between 1.8 million wps to 1000 wps, the lower wps values for data size of 1MB are not clearly visible in Fig. 8.

In our experiments, we observed that, reducing the value for the configuration parameter `log.flush.interval.messages` improves the durability of writing data by flushing data to the storage disk/device more frequently. But, it reduces performance. Hence, in our test setup, we have set this value to 64-bit long integer maximum value as default. Since the Kafka write operations are asynchronous, Fig. 9 shows the latency values of the write operations of the single producer. In our experiments, we observed that the median write percentile of 5 ms (milliseconds) for 10 bytes of data size and 60 ms for 1MB data size with the maximum possible throughput.

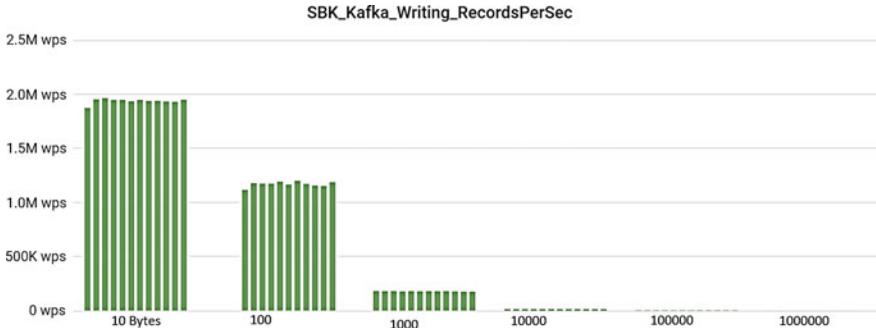


Fig. 8 Kafka single producer throughput performance in terms of Records/Second or Writes/Second (wps)

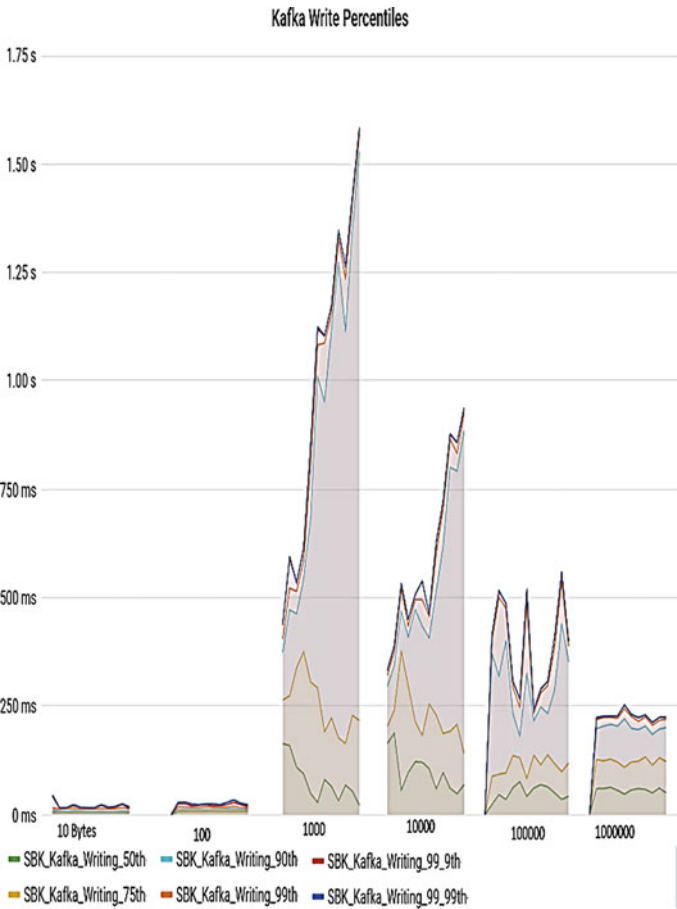


Fig. 9 Single Kafka Producer latency Percentiles

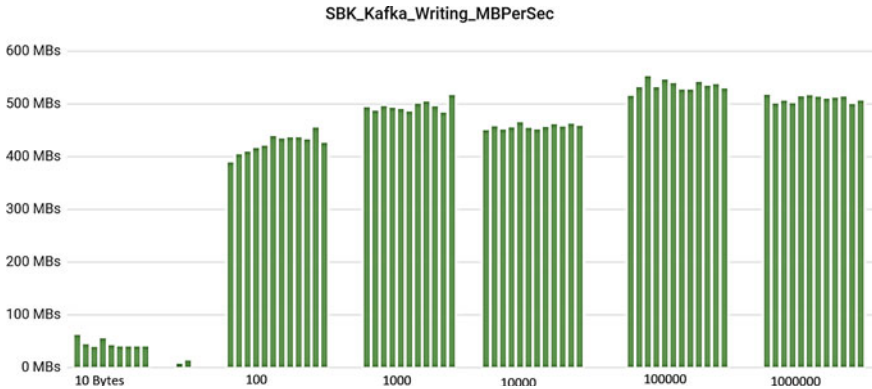


Fig. 10 Kafka 10 producers throughput performance in terms of MB/s

Multiple Kafka Producers Benchmarking: The SBK scales high with multiple writers/producers. Figure 10 shows the performance benchmarking of 10 Kafka producers. Note that, the SBK achieves the maximum throughput range of 500–550 MB/s for record (data) sizes of 100 K to 1 MB. The same throughput results can be depicted in terms of records/second or writes/second as shown in Fig. 11. Note that, for smaller data sizes less than 100 bytes, the SBK records the peak throughput of more than 4 million wps.

Single Kafka Consumer benchmarking: Fig. 12 shows a single Kafka consumer performance for the data sizes of 10, 100, 1000, 10000, 100000, and 1000000 bytes. The SBK records the maximum throughput range of 35–45 MB/s for 10 bytes data record size. The throughput range of 400–450 MB/s is observed for data sizes of 10 K to 1 MB. Figure 13 shows the single Kafka consumer throughput performance in terms of Records/Second or Reads/Second (rps). In our experiments, we have observed that by decreasing the Kafka reader configuration parameter max.poll drastically reduces the read performance. Hence, in our test setup, the max poll is set to 32 Bit Max integer value.

Multiple Consumers benchmarking: Fig. 14 shows the 10 consumers performance benchmarking for the data sizes of 10, 100, 1000, 10000, 100000, and 1000000 bytes. It shows that the SBK records the peak performance range of 1.5–1.8 GB/s (Giga Bytes/second) for the data sizes of 100 K to 1 MB. Figure 15 shows the same performance benchmarking in terms of Records/Second or Reads/Second.

End to End Latency Benchmarking: Fig. 16 shows the End to End latency of the Kafka single writer and reader for 10, 100, 1000, 10000, 100000, and 1000000 bytes of data sizes. In our experiments of End to End latency mode, the SBK is set to flush every record and hence it attempts to record the least latency at a low throughput rate. For the data size of 100K or less, the median (50th percentile) latency is in the range of 1–3 ms. For the data size of 1MB, the median latency is in the range of 10–15 ms.

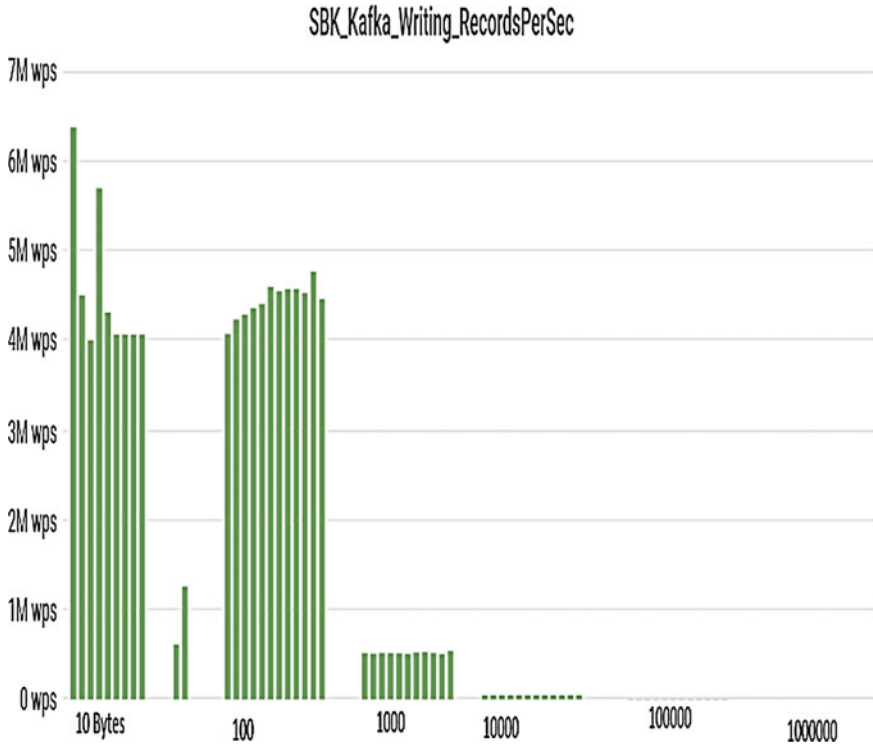


Fig. 11 Kafka 10 producers throughput performance in terms of Records/Second or Writes/Second (wps)

5 Conclusion

The design and implementation of the SBK framework, presented in this paper, scales high with multiple readers and writers. This design solves the synchronization issues between multiple writers, readers/callback (push) readers, and it also describes the best suited data structures such as multiple concurrent queues to measure the maximum throughput and low latency for any storage device/client/cluster. The design of SBK exports the standard storage interface APIs which can be extended to include a storage driver to conduct the performance benchmarking of any custom storage device/client. The SBK currently supports benchmarking of a wide category of storage systems such as local mounted file systems, distributed file systems, distributed messaging, streaming storage platforms, key-value storage systems, database systems, and object storage systems. SBK can be used as a common framework to conduct performance benchmarking among similar category storage systems. For example, different database systems such as MySQL, Apache Derby, PostgreSQL, FoundationDB document layer, and MongoDB can be compared with respect to performance using SBK. Different streaming storages such as Kafka and

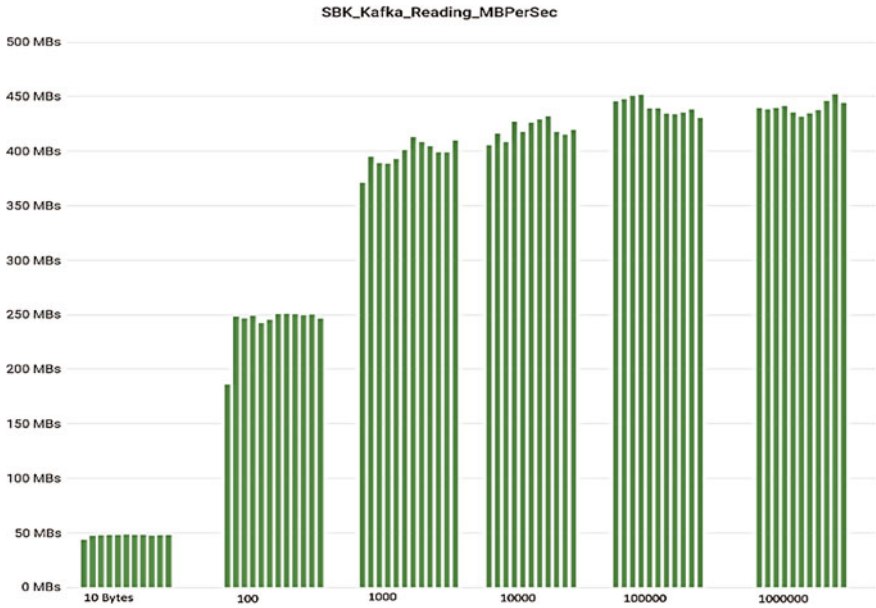


Fig. 12 Single Kafka consumer throughput performance in terms of MB/s

Pulsar can be compared using SBK. The SBK can be used to compare the performance of different file systems too. The design of SBK is flexible to enclose the performance benchmarking of non-persistent in-memory message queues also.

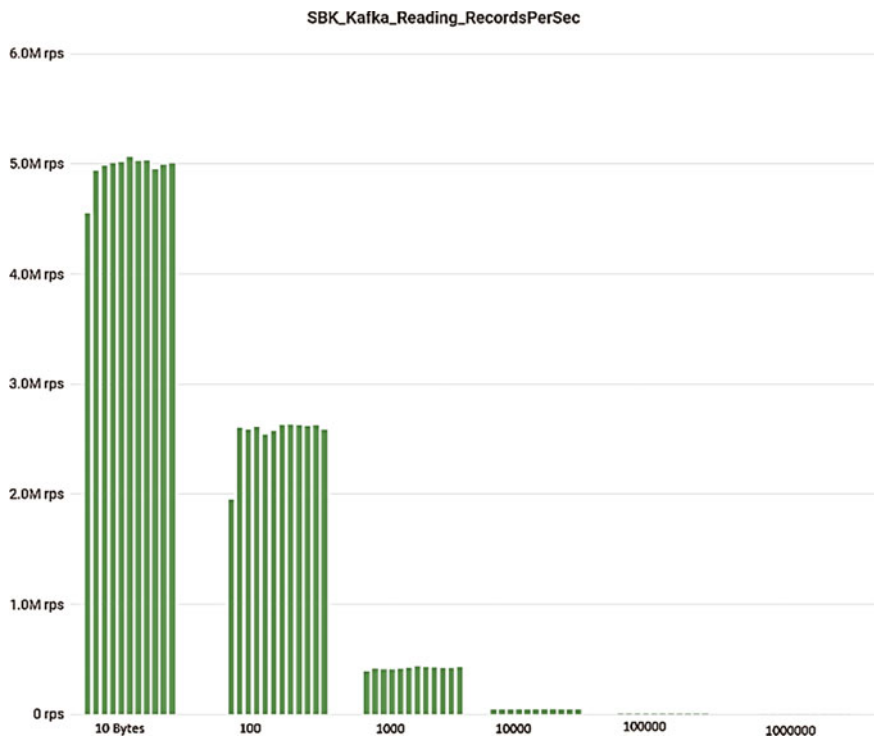


Fig. 13 Single Kafka consumer throughput performance in terms of Records/Second or Reads/Second (rps)

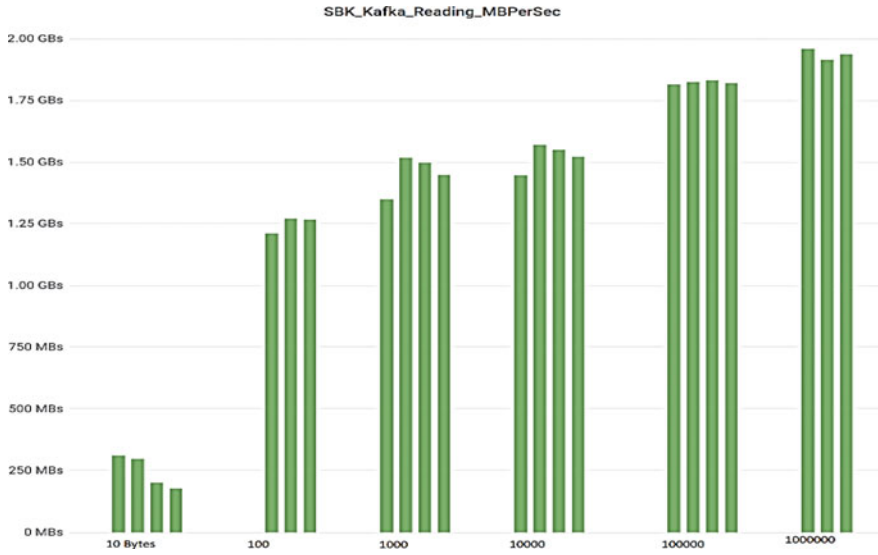


Fig. 14 Kafka 10 Consumers throughput performance in terms of MB/S

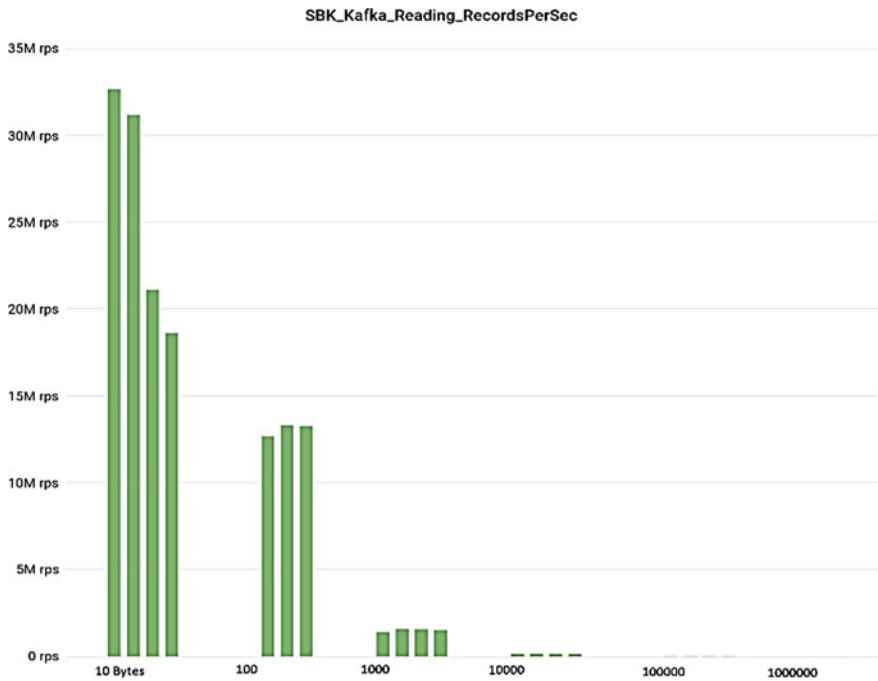


Fig. 15 Kafka 10 Consumers throughput performance in terms of Records/Second or Reads/Second (rps)

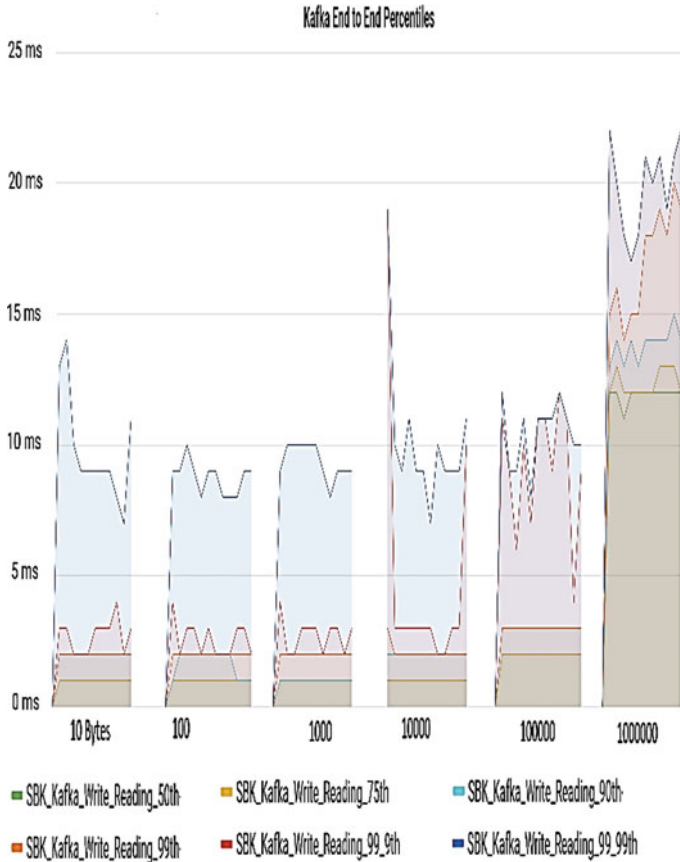


Fig. 16 Kafka End to End Latency percentiles

References

1. Storage Benchmark Kit (SBK). <https://github.com/kmgowda/SBK> (2020)
2. SBK Releases. <https://github.com/kmgowda/SBK/releases> (2020)
3. SBK docker images. <https://hub.docker.com/repository/docker/kmgowda/sbk> (2020)
4. Sanjay Kumar NV, Munegowda K (2019) Distributed streaming storage performance benchmarking: Kafka and Pravega. Int J Innov Technol Exploring Eng (IJITEE) 9(2S). ISSN: 2278-3075
5. Pravega Benchmark Tool, Releases. <https://github.com/kmgowda/pravega-benchmark/releases> (2020)
6. Pravega website. <http://pravega.io> (2020)
7. Narkhede N, Shapira G, Palino T (2017) Kafka, The Definitive Guide, O'reilly series, 1st edn
8. Apache Kafka website. <https://kafka.apache.org/> (2020)
9. Junqueira F, Kelly I, Reed B (2013) Durability with Bookkeeper. In: ACM SIGOPS operating systems review, vol 47, issue 1, January 2013
10. Apache Bookkeeper website. <https://bookkeeper.apache.org/> (2020)
11. White T (2015) Hadoop definitive Guide, O'reilly series, 4th edn

12. Apache Hadoop website. <https://hadoop.apache.org/> (2020)
13. Roy G (2018) Rabbit MQ in Depth. Manning Publications
14. RabbitMQ website. <https://www.rabbitmq.com/> (2020)
15. RocketMQ website. <https://rocketmq.apache.org/> (2020)
16. NATS 2.0 website. <https://nats.io/> (2020)
17. NATS Streaming. <https://nats.io/download/nats-io/nats-streaming-server/> (2020)
18. Apache ActiveMQ Artemis. <https://activemq.apache.org/components/artemis/> (2020)
19. NSQ website. <https://nsq.io/> (2020)
20. Apache Pulsar website. <http://pulsar.apache.org/> (2020)
21. Apache Derby website. <https://db.apache.org/derby/> (2020)
22. MySQL website. <https://www.mysql.com/> (2020)
23. PostgreSQL website. <https://www.postgresql.org/> (2020)
24. Microsoft SQL website. <https://www.microsoft.com/en-in/sql-server/sql-server-2019> (2020)
25. SQLite website. <https://www.sqlite.org/index.html> (2020)
26. Deitel and Deitel (2015) Java How to Program, 10th edn. Pearson Publications
27. RocksDB. website <https://rocksdb.org/> (2020)
28. FoundationDB, Website. <https://www.foundationdb.org/> (2020)
29. Google Protocol Buffers. Website <https://developers.google.com/protocol-buffers> (2020)
30. Chrysafis C, Collins B, Dugas S, Dunkelberger J, Ehsan M, Gray S, Grieser A, Herrstadt O, Lev-Ari K, Lin T, McMahon M, Schiefer N, Shraer A (2019) FoundationDB record layer: a multi-tenant structured datastore. In: SIGMOD '19: Proceedings of the 2019 international conference on management of data, pp 1787–1802
31. FoundationDB Document layer. website: <https://foundationdb.github.io/fdb-document-layer/> (2020)
32. MogoDB. website: <https://www.mongodb.com/> (2020)
33. MinIO. website: <https://min.io/> (2020)
34. Grafana Website. <https://grafana.com/> (2020)
35. Prometheus Website. <https://prometheus.io/> (2020)
36. Schildt H (2014) Java: the complete reference, 9th edn. Oracle Press
37. Michael MM, Scot ML (1996) Simple, fast, and practical non-blocking and blocking concurrent queue algorithms In: Proceedings of the fifteenth annual ACM symposium on Principles of distributed computing
38. Goetz B (2010) Java concurrency in practice. Addison-Wesley publications, 9th print
39. Java 7/8, Concurrent Linked queue. <https://docs.oracle.com/javase/8/docs/api/java/util/concurrent/ConcurrentLinkedQueue.html> (2020)
40. Micrometer IO website <https://micrometer.io/> (2020)
41. SL4J website: <http://www.slf4j.org/> (2020)
42. FIO, website: <https://github.com/axboe/fio> (2020)
43. Munegowda K, Raju GT, Poddar S (2012) FFSB and IOzone: file system benchmarking tools, features and internals. In: Proceedings of Embedded Linux Conference (ELC), Barcelona, Spain, Europe
44. IOzone website: <http://www.iozone.org/> (2020)
45. Hadoop version 3.2.0 release: <https://hadoop.apache.org/release/3.2.0.html> (2019)
46. Kafka version 2.4.1 release: <https://kafka.apache.org/downloads> (2020)
47. Open messaging framework. website: <http://openmessaging.cloud/docs/benchmarks/> (2020)
48. Open messaging source code. <https://github.com/openmessaging/openmessaging-benchmark> (2020)

Adaptive Calibration for Camera and Stitching of Images



Jharna Majumdar, Shilpa Ankalaki, and Sarala Madolli

1 Introduction

Camera calibration is the procedure of obtaining intrinsic and extrinsic parameters. The internal features of the camera like focal length, skew, distortion, and center of image are covered in intrinsic parameters. Extrinsic parameters portray their world orientation and location. The first step for 3D computer vision is knowledge of intrinsic parameters. Geometric camera calibration is also known as camera resection, calculates the lens parameters and video camera imaging sensor. These parameters may also be used to correct the distortion of the lens, to evaluate the size in world units of an object, or to establish the camera prospect. These errands are utilized to identify and determine objects in machine vision applications as well as in for navigation systems, robotics, and 3D scene reconstruction. The original purpose of surround view was to offer the driver a novel and previously unachievable view of his car, usually by using 4 or 6 cameras and. In various situations, the driver benefits from verifying blind spot while changing lanes on the motorway, or drawing a parallel garage without grinding the wheels. Optimal purpose of camera calibration is to obtain camera parameters, for which it is required to have 3D world points and their analogous 2D points of image.

These correspondences could be got by using multiple images of a calibration pattern, for example, a checkerboard. Using the correspondences, camera parameters can be solved.

Stitching of image is utilized to incorporate information from various images with overlapping areas of perspective to create an all-inclusive panoramic perspective of

J. Majumdar (✉) · S. Ankalaki · S. Madolli
Department of CSE, Nitte Meenkashi Institute of Technology, Bengaluru 560064, India
e-mail: jharna.majumdar@nmit.ac.in

S. Ankalaki
e-mail: shilpa.a@nmit.ac.in

one frame. The literature demonstrates that the stitching of pictures is still a challenge for individual panoramic images. Many of the algorithms have been commonly proposed in past years to deal with the issue of image stitching. For image processing professionals, the use of image stitching in real-time apps is a difficult field. It is widely used in the fields of video conferencing, video matting, video stabilization 3D imaging, video resuming, satellite imagery, video compression, and a variety of medical applications. The capacity to summarize and compress images taken using panning camera is an exciting application of image stitching.

The contributions of this paper are as follows:

- Camera calibration has performed by determining the calibration patches and corner points.
- Employed SIFT, SURF for feature extraction and matching
- Employed FAST algorithm for extracting key points for image stitching
- RANSAC algorithm is used for matching the key points and image stitching.

2 Related Work

Choi et al. [1] proposed an approach in which four cameras are automatically calibrated for around view monitor framework in a usual driving condition. This technique approximate orientation angles of four cameras that frame system as well as assuming that their positions. Intrinsic parameters are well identified earlier. This method uses path marks as they emerge across adjoining camera images and are present in nearly all on-road circumstances. The process begins by identifying path marks by the captured images of four cameras in a lucrative way. Wrong path marks are discarded with analyzing perceived path mark arithmetical properties. This approach initially calibrates the front and back cameras when path marks are adequately drawn together, afterward calibrates the left and right cameras by using previous results of calibration.

Satouri et al. [2] devised an innovative and vigorous technique for self-calibration of camera that has variable intrinsic parameters by an image series of unknown 3D object. This approach is depended on the construction of a nonlinear cost task by identifying an association among 2 scene points with comparative against to abscise axis and projections in image planes. The imperative of this system occurs in the application of a solitary image couple that gives less calculation, facilitates the numerical complications then reduces the performance time, and the usage of information for initial image. The trial outcomes on an artificial and physical information evidences.

Satar et al. [3] performed a research of camera calibration technique, and the camera calibration impact was assessed using distinct effects. As number of images and feature points on camera parameters are changed, a noise-level impact was assessed. Lens distortion models are proposed to estimate camera parameters, reducing the overall minimum errors. This approach produced the optimum evaluation of error for set of images when compared to the rigid model.

Komorowski et al. [4] introduced a technique for calibration of extrinsic camera from a picture sequence. The camera intrinsic matrix is presumed to be known, and distortion coefficients are fixed during the whole progression. A number of stereo sample datasets assess the performance of the proposed technique. The algorithm presented may be used as a first phase in the densely reconstructed stereo system.

Tan et al. [5] declares that in 3D computer vision assignments, camera calibration plays a critical role. The most frequently used technique of calibration uses a flat checkerboard and can be carried out automatically. But during the capture phase, the user must either move the checkerboard or the camera. This manual procedure takes time and makes the outcomes of calibration volatile. In order to solve the above problems caused by physical operation, this method presents a full-automatic camera calibration method using a virtual pattern instead of a physical one. The virtual pattern is actively transformed and displayed on a screen so that the control points of the pattern can be uniformly observed in the camera view. The proposed method estimates the camera parameters from point correspondences between 2D image points and the virtual pattern. The technique suggested here estimates the camera parameters from point match between 2D picture points and the virtual model. The camera and the display are fixed throughout the entire process, so no physical activities are required for the suggested technique.

Huang et al. [6] discussed the significance of calibration of multi-camera in many areas. In this document, a new technique of flat-refractive geometry multi-camera calibration is presented. Every camera can obtain Transparent Glass Calibration Board (TGCB) calibration of pictures all together. The implementation of TGCB contributes to an occurrence of refraction that may lead to calibration errors. To eliminate the fault, the theory of flat-refractive geometry is applied. The novel approach can fix the TGCB refractive phenomenon. In order to minimize re-projection errors and get optimized calibration outcomes, the technique for bundle adjustment is also used.

Yan et al. [7] suggested a versatile and high-precision technique for calibrating a camera. Initially, the radial distortion center was calculated because ideal findings are essential. The intrinsic parameters and camera distortion coefficients were then solved separately for the radial distortion of the divisional representation. Ultimately, through Levenberg–Marquardt method, the camera intrinsic parameters were improved. The suggested technique effectively decouples intrinsic parameters and coefficients of distortion and enhances the precision of calibration through optimization consequently. In addition, the methodology used in this case yields excellent outcomes, whether it is comparatively tiny image distortion or bigger image distortion. The robustness and precision of the suggested technique are established by both simulation and actual information experiment. Findings of experiments indicate that the technique suggested can be more accurate than conventional techniques.

In the context of correspondence of image and 0 skew, Tang et al. [8] provided a flexible and effective camera automatic calibration explanation from more than or equal to three viewpoints. In this proposed work, the knowledge of camera movement, 3D data, the scene or inner restrictions is not necessary. The methodology proposed is based on basic matrices, with three primary virtues. The first point is that the focal

length and aspect ratio can be estimated independently of a significant principal point change in the center-aligned matrix coordinates. Thirdly, for quick final refining of four internal parameters, the Levenberg–Marquardt algorithm is implemented to achieve better efficiency.

Gasbarro et al. [9] discussed the aim to determine a series of camera settings in camera calibration which describe mapping between 3D and 2D image reference coordinates. For precise data from vision systems, correction distortion in cameras is a significant subject. The various methods of camera calibration observed in literature include an examination of calibration techniques for visual devices. These methods are helpful for contemporary industrial competition in order to know the efficient distortion of the lens and the chance to measure the vision system uncertainty. The novel camera calibration technique named two phase method was suggested by Tsai. The research begins with Brown's theoretical views of the same camera model, but the method is designed to characterize an actual camera calibration operation.

In this work, a mathematical model that defines its unknown generic geometry to describe the geometric calibration of a particular multi-projective camera (MPC) is presented by Khoramshahi et al. [10]. Industrial level 360° non-metric cameras are calibrated with the use of an altered bundle block adjustment. A calibration space that has been precisely determined by close distance photogrammetric can retrieve the composition of any MPC as a collection of relative, interior orientation. Following the adjustment, residual pixel images and acceptable object-space mistakes were noted.

Anjana et al. [11] stated that image mosaicing is the way to join the two pictures in the panorama. The accuracy of the picture is based on precise identification and correspondence of features. Extraction of feature is one of the major steps in the processing of images. The detection of feature is a processing of low-level images. Different methods of extraction for picture mosaicing may be used according to the suggested technique. Two robust techniques scale invariant feature transform and speeded up robust features were compared here. It also allows distinguishing invariant characteristics to be drawn from pictures which may be used to match distinct views of an object or scene.

Jayanthi et al. [12] have been provided an insight on how a range of image detection and tracking techniques work on a variety of datasets. There have been a broad variety of datasets, from hand gestures to items and objects, to text of manuscript. The analyzed performance of algorithms consists of the detection method of Blob, templates, and the algorithm of surf. These image matching algorithms have been compared using different measures such as accurateness, speed of processing, flexibility to use for different information sets, rotational invariances, illumination, and scales.

The various methods such as Sift, Surf, Fast, and Orb for identifying and corresponding of features were discussed by Mistry et al. [13]. Because Sift and Surf are invariant in translation, scale, rotation, blur, and illumination, they are most helpful methods for detecting and match characteristics. The comparison of Sift and Surf methods is discussed here. Invariant, blurred and warped, Surf is better than Sift. In various scale pictures, Sift is better than Surf. Because integral image and box filter

are used, Surf is three times quicker than Sift. In the illumination change pictures, Sift and Surf perform well.

Arya [14] researched on different approaches of stitching of images. In order to provide a better outcome, most image stitching methods need to overlap exactly between images and their exposures. Image stitching method is commonly used for the recovery of initial ripped information. Imagery stitches are used for the reconstruction of torn paper, which is a large issue in forensic and investigative science. In picture mapping, stitching is performed to map the entire area. The image stitching algorithms are utilized in digital maps and satellite photographs form high-resolution image mosaics. It is a broad field of computer vision and photogrammetric study.

Jenitta et al. [15] devised an enhanced technique to overcome the distortion. Incorporating modified choosing manner for the indication image and projecting an approach which could calculate the matrix of transformation for any of the image of the series to arrange in a line by means of the indication image in the similar space of coordinate. The enhanced stitching technique also chooses the subsequent input picture dynamically according to the amount of Sift matching points. The enhanced technique improves the amount of matching feature points and decreases the reference images Sift feature detection range as compared to the conventional stitching process. The experimental findings indicate that the enhanced technique not only speeds up the effectiveness of stitching, but also reduces panoramic distortion problems and ultimately provides a pleasant panoramic outcome.

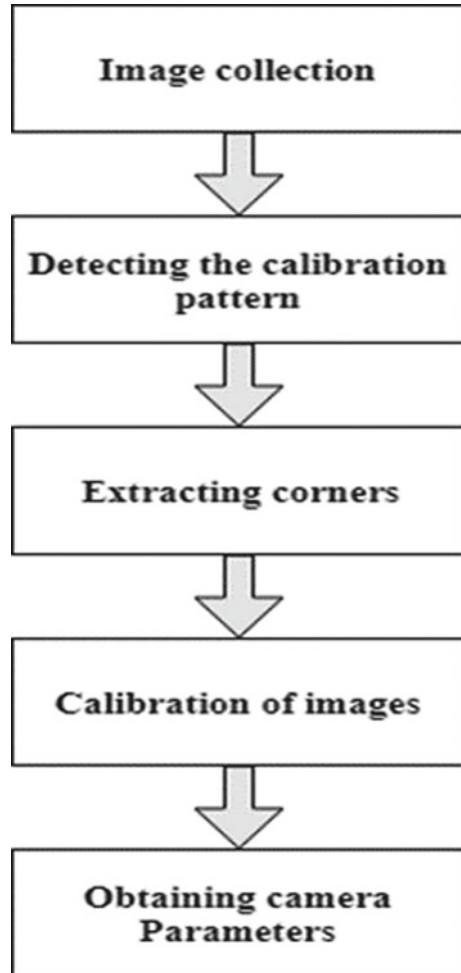
Sruthi et al. [16] presented an effectual approach for formation of panoramic image. A Sift technique has been applied here to detect and match the features mechanically. It is a highly robust technique for detecting and describing local picture characteristics. The overlapping region of the two pictures can then be found. A minimal cost path in the overlapping area of two images is found by dynamic programming method. Active programming methods detect a minimum price route in the overlapping region of two pictures. The minimum cost route is used for stitching images. Dynamic programming is quicker and utilizes little memory than other seam finding techniques. The pictures overlap and fuse together along the seam.

3 Proposed Methodology

This section gives a brief description of assumptions, methodologies and detailed architecture of the proposed system. It helps to achieve the objectives of the proposed system. The proposed work is to calibrate a camera. Figure 1 shows the work flow of camera calibration.

Initially, the images are captured from the webcam of resolution 1920*1080. The images are stored, and the checkerboard pattern in the images is detected. The size of checkerboard is 20*20, and it consists of black and white squares. The corner points are extorted from pattern. After detecting corner point's calibration is carried out which results in the fundamental matrix that leads to estimate the camera intrinsic and extrinsic parameters.

Fig. 1 Flow diagram for camera calibration



A. Intrinsic Parameters

The intrinsic parameters matrix comprises of the following values:

- Focal length: The distance between the image sensor and lens is referred to as focal length.
- Principle point: The point on the image plane on which center is projected, as well as the point from which the focal length is calculated is termed to be principle point.
- Skew: The skew coefficient in each direction on the sensor is the number of pixels per unit length.

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where f, f_y are focal lengths and u_0, v_0 are principle points.

B. Extrinsic Parameters

Extrinsic parameters comprise of:

- Rotation: Rotation is circular movement of camera around a center.
- Translation: Translation is horizontal movement of camera.

$$[R|T] = \begin{bmatrix} r_{11} & r_{12} & r_{13} | t_1 \\ r_{21} & r_{22} & r_{23} | t_2 \\ r_{31} & r_{32} & r_{33} | t_3 \end{bmatrix}$$

3.1 Methods Comprising Calibration

SIFT: It is a technique used for detecting feature points in an image. Primarily, key points in Sift are stored in a database. In order to filter out excellent matches, key points subsets are recognized from the whole set of matches which consent about the item and its place, extent and direction in the novel picture. Reliable clusters are quickly determined using an effective application of hash table of the comprehensive Hough transformation. A further thorough model test is then carried out for every cluster containing more than three characteristics that have the same opinion on an item and its position, and then outliers are eliminated.

SURF: It is also a method used for identifying features of an image which works slower when compared to sift. In order to detect significant points, surf utilizes a Hessian blob detector determinant integer approach which can be calculated using a pre-computed integral image with three integer operations. Its descriptor of the function is based on the amount of the reaction of the Hair wavelet. The integrated picture can also be used to calculate these. The image is converted to coordinates by using the pyramidal method to replicate the initial picture in the pyramid of Gaussian or Laplace pyramidal form to achieve an image with the same volume but with decreased bandwidth.

Fundamental Matrix: This matrix is a relation between two pictures in the same scene, which limits the screening of points from both pictures. In view of the screen point projection into one of the pictures, the respective point in the other picture is restricted to a row, helping the search to detect false correspondences. The relationship between the respective picture points represented by the basic matrix is known as epipolar limitation, corresponding restriction, discrete constraints, or incidence relationship.

$$\begin{aligned}
 &x^T Fx = 0 \\
 &(u \ v \ 1) \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \\
 &F_{11}uu + F_{12}vu + F_{13}u + F_{21}uv + F_{22}vv + F_{23}v + F_{31}u + F_{32}v + F_{33} = 0 \\
 &\begin{bmatrix} u_1u_1 & v_1 & u_1 & u_1 & u_1v_1 & v_1v_1 & v_1 & u_1 & v_1 & 1 \\ u_2u_2 & v_2 & u_2 & u_2 & u_2v_2 & v_2v_2 & v_2 & u_2 & v_2 & 1 \\ u_3u_3 & v_3 & u_3 & u_3 & u_3v_3 & v_3v_3 & v_3 & u_3 & v_3 & 1 \\ u_4u_4 & v_4 & u_4 & u_4 & u_4v_4 & v_4v_4 & v_4 & u_4 & v_4 & 1 \\ u_5u_5 & v_5 & u_5 & u_5 & u_5v_5 & v_5v_5 & v_5 & u_5 & v_5 & 1 \\ u_6u_6 & v_6 & u_6 & u_6 & u_6v_6 & v_6v_6 & v_6 & u_6 & v_6 & 1 \\ u_7u_7 & v_7 & u_7 & u_7 & u_7v_7 & v_7v_7 & v_7 & u_7 & v_7 & 1 \\ u_8u_8 & v_8 & u_8 & u_8 & u_8v_8 & v_8v_8 & v_8 & u_8 & v_8 & 1 \end{bmatrix} \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0
 \end{aligned}$$

Fig. 2 Basics of fundamental matrix

After pattern detection followed by corner detection, the calibration is done for obtaining camera parameters. These parameters are also obtained manually by computing fundamental matrix which is shown in Fig. 2.

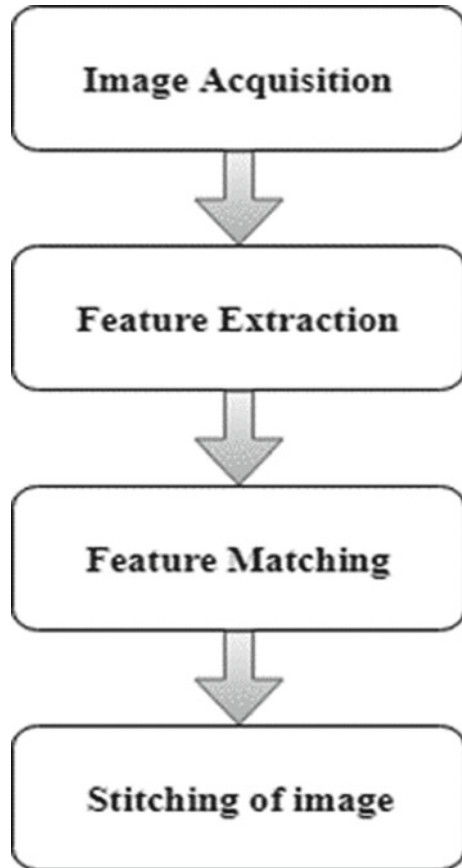
3.2 *Stitching of Images*

The workflow for image stitching is depicted in Fig. 3. The foremost step here is to collect images with their size and angle. The images are then used for feature detection and matching. After obtaining the homogeneity among images, they are blended.

FAST (Features from Accelerated Segment Test) Detector: It is an approach for detecting corners that is utilized to extort feature points and then used for tracking and mapping things in a lot of computer vision errands. The most promising advantage of the FAST corner detector is its computational efficiency. It is actually quicker than many other renowned techniques of extortion by reference to its name.

Pixel p is selected in image and let the intensity of pixel p be I_p . Threshold t is fixed. 16 pixels around pixel p within radius 3 are considered. For the algorithm to be faster, in these selected 16 pixels first minimum 4 pixels say 1, 9, 5, 13 must have intensity above or below $I_p + t$. If this executes, then n continuous pixels are checked for the same condition. Here n is taken as 12. If the pixel satisfied, the condition it is

Fig. 3 Flowchart of image stitching



selected as corner point. All the pixels are treated by same procedure. The flowchart describing each step is shown in Fig. 4.

RANSAC (Random sample consensus): It is an iterative way of estimating mathematical model parameters from a number of observed information containing outliers where external values are to be left uninfluenced. It is therefore nondeterministic method in the logic that a sensible outcome can only be achieved with certain likelihood, with the likelihood rising as more iterations are permitted. This is also an outlier detective method. The aim was to identify the points in space on a picture to a number of points with knowledgeable places.

Data point are selected randomly and a model such as line is defined. Threshold t is chosen manually. The data points whose distance to the defined line is less than t are selected, and others are treated as outliers. The basic outline as explained here is shown in Fig. 5.

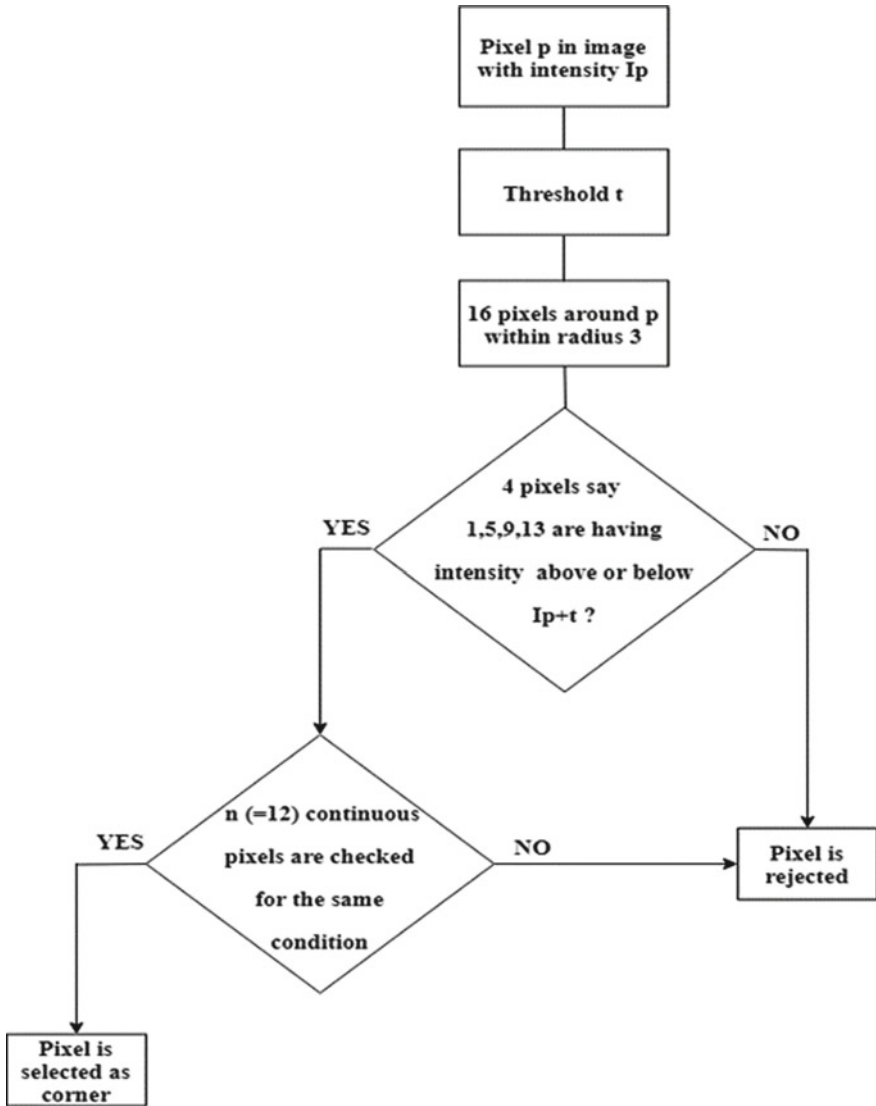


Fig. 4 Flowchart of FAST detector

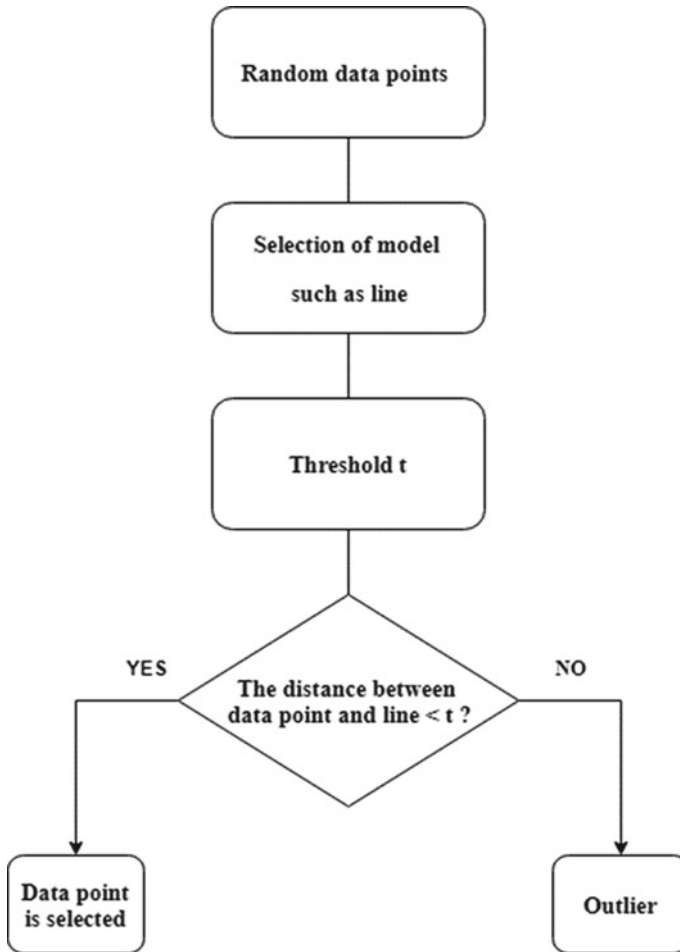


Fig. 5 Flowchart for RANSAC

4 Results

4.1 Calibration Results

For the collected images first, the calibration is carried out in MATLAB toolbox. The images captured are loaded. 3 among 17 images are rejected because of improper orientation. And 17 images are considered for calibration. The corner points are extracted, and re-projection errors are given in Fig. 6. Re-projection error is defined as geometric error related to the distance of image. Camera centric and pattern centric views are shown in Figs. 7 and 8. Camera centric view shows the different

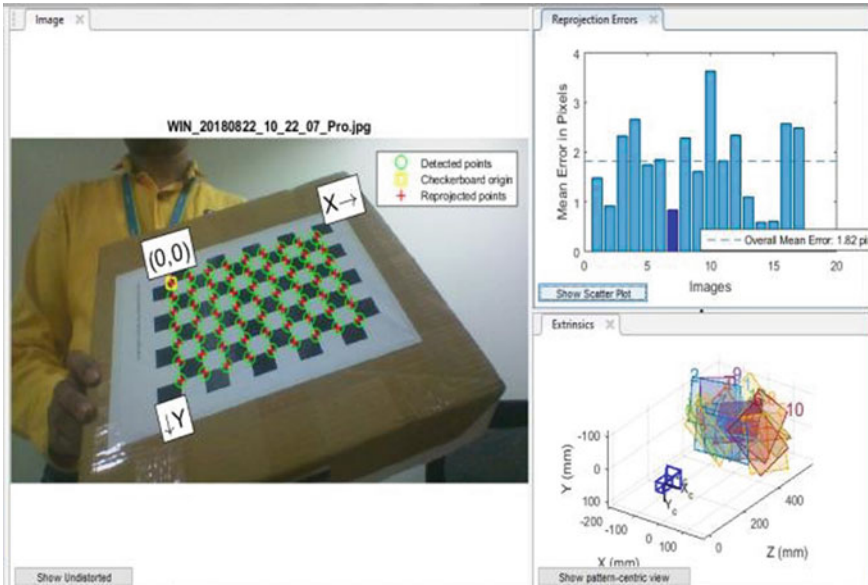


Fig. 6 Re-projection errors

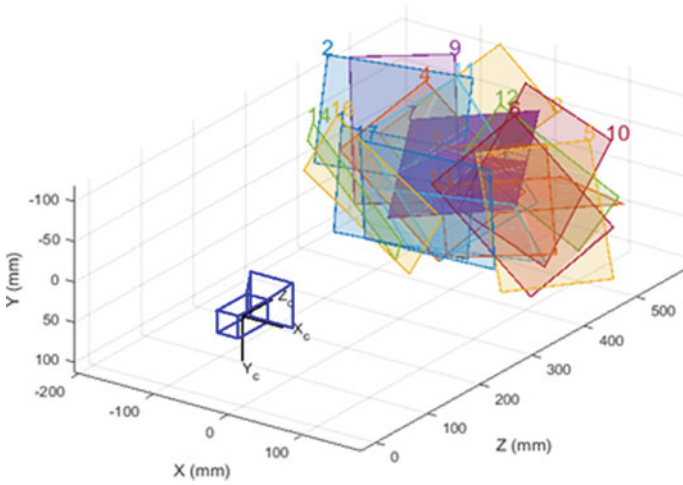


Fig. 7 Camera centric view

image orientations in the view of camera. Pattern centric view depicts the camera orientations when pattern is fixed.

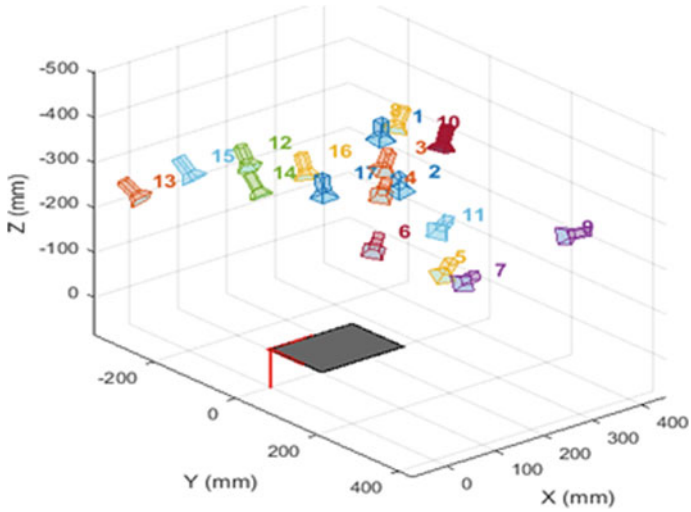


Fig. 8 Pattern centric view

Initially, images with squares in checkerboard are captured. The input images are shown in Figs. 9 and 10. The images are captured from the webcam which consists of the checkerboard pattern to detect the feature points. The camera parameters are

Fig. 9 Image p1

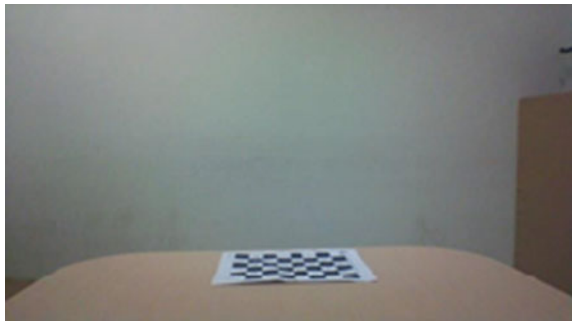
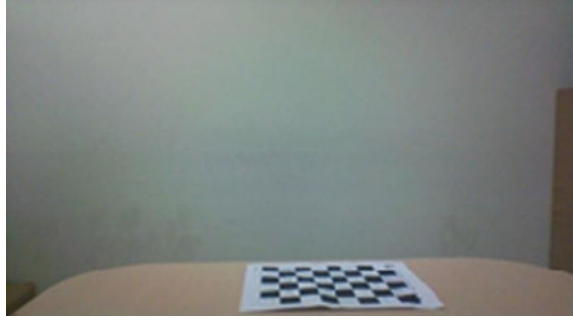
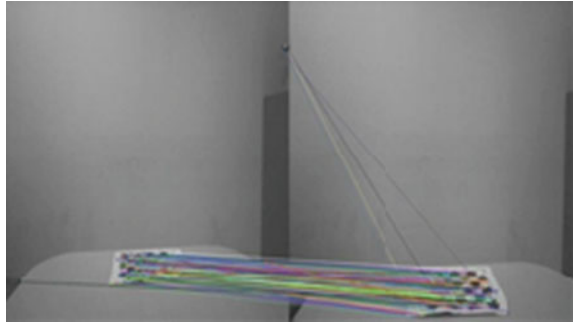


Fig. 10 Image p2**Fig. 11** Matching between image p1 and image p2

obtained for the calibration. From kruppas equation, fundamental matrix and focal length are obtained.

$$R1 = \text{inv}(k) * H_1 \quad (1)$$

$$R2 = \text{inv}(k) * H_2 \quad (2)$$

$$R3 = \text{cross}(R) \quad (3)$$

The matching between above two images is shown in Fig. 11. Here matching is performed using sift. The feature points are detected in the checkerboard and the detected points are matched. From image p1 and p2 fundamental matrix and focal length:

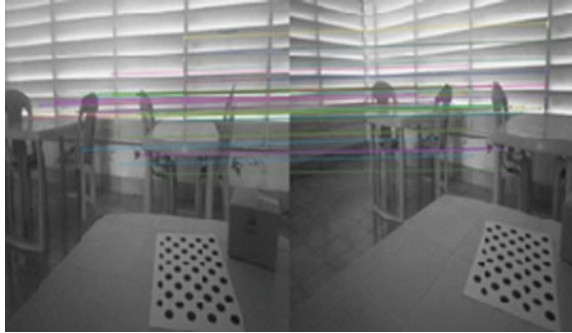
$$F = \begin{bmatrix} 9.848007e - 07 & 6.315377e - 05 & -0.036952 \\ -5.402217e - 05 & -1.608876e - 05 & 0.050256 \\ 0.032642670 & -0.044993497 & 1 \end{bmatrix}$$

<p>Focal length from Eq. 1:</p> <p>-97.017347736433329317779540870884 97.017347736433329317779540870884 -183.55932794213615305629608252565i 183.55932794213615305629608252565i</p>	<p>Focal length from Eq. 2:</p> <p>-91.817043813871127501533236930484 91.817043813871127501533236930484 -186.64672660395639767145576375646i 186.64672660395639767145576375646i</p>	<p>Focal length from Eq. 3:</p> <p>-94.398256171289465271319728945101 94.398256171289465271319728945101 -185.04216842311840060509689053905i 185.04216842311840060509689053905i</p>
---	---	---

Fig. 12 Image q1**Fig. 13** Image q2

Later images with circles in checkerboard are captured. Those input images are shown in Figs. 12 and 13. The images are captured in outdoor placing checkerboard on table.

Fig. 14 Matching between image q1 and image q2



The matching between these two images is carried out using sift which is shown in Fig. 14. The feature points are detected in the outdoor scene, and the detected points are matched.

From image q1 and q2 fundamental matrix and focal length:

$$F = \begin{bmatrix} 2.166096e - 07 & 8.669577e - 06 & -0.002381 \\ -8.854528e - 06 & 1.300576e - 06 & 0.006138 \\ 0.001947908 & -0.009453051 & 1 \end{bmatrix}$$

Focal length from Eq. 1: -733.28932243040054694576174466219 733.28.932.243.040.054.694.576.174.466.219 -104.95995184957076352416522364965i 104.95995184957076352416522364965i	Focal length from Eq. 2: -788.34789530299628895463073788703 788.34789530299628895463073788703 -107.5295234070847562020223409614i 107.5295234070847562020223409614i	Focal length from Eq. 3: -846.94157861145036323771731099333 846.94157861145036323771731099333 -111.21155439210435078496108345893i 111.21155439210435078496108345893i
---	--	--

Fig. 15 Image a

4.2 Image Stitching Results

The dataset is collected from the below link: <http://www.cvl.isy.liu.se/research/datasets/passta/>. The images are resized for the resolution of 480*720 and angle of 15° difference between each image. The input images are shown in Fig. 15 and 16.

The key points are detected by using FAST algorithm in both the images as depicted in Fig. 17a, b. The key points are indicated by colored circles in both the figures. Matched points between image a and image b using RANSAC algorithm are indicated by black circles in Fig. 18. The stitching is done by using the homogeneity of two images depicted by matched points. Stitched result for two images is displayed in Fig. 19.

And in the similar way, the stitching result for multiple images of the scene is displayed in Fig. 20.

Fig. 16 Image b

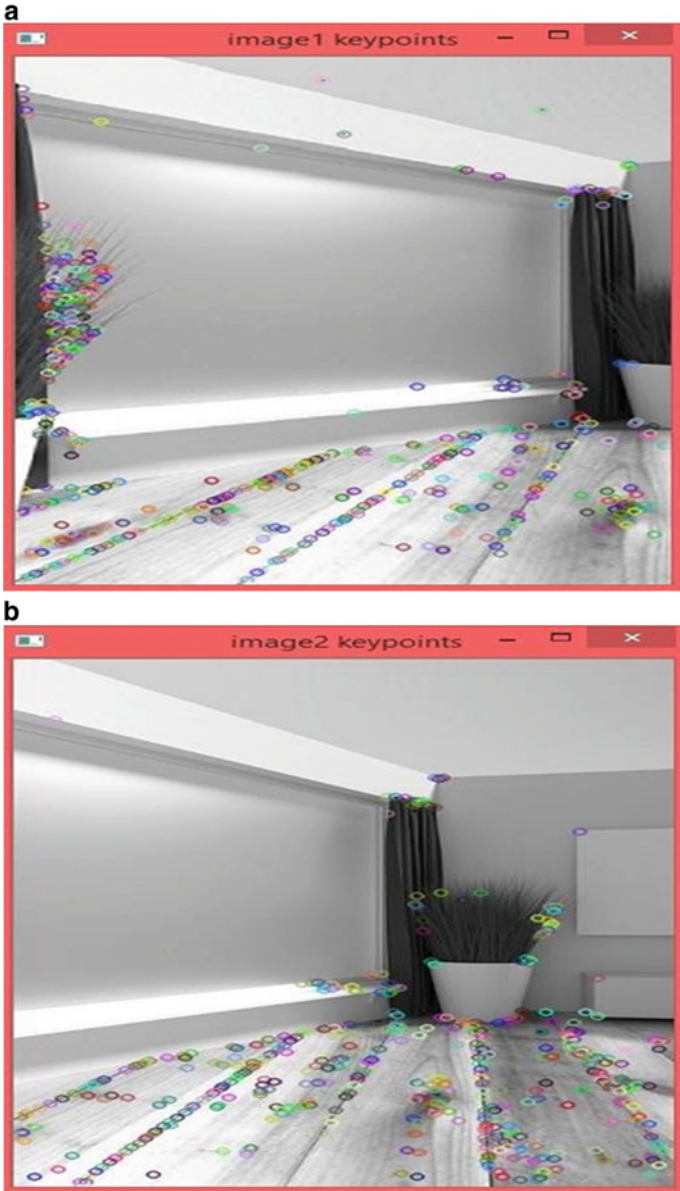


Fig. 17 a Key points identified in image a, b Key points identified in image b

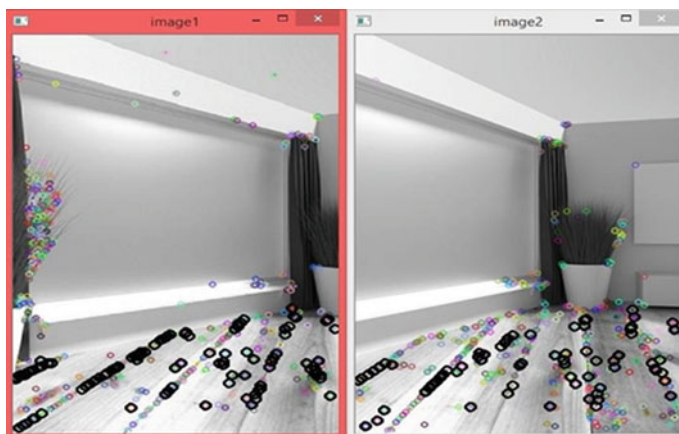


Fig. 18 Matching between image a and image b

Fig. 19 Stitching of image a and image b



Fig. 20 Stitching of multiple images



5 Conclusion

To compute the camera model intrinsic and extrinsic parameters, conventional calibration approaches have used the constrictor link among the paradigm reference and its images. Different approaches are performed for camera calibration such as from MATLAB and using homography matrix and then using Krupp's equation and fundamental matrix. For getting precise estimate of focal length, Krupp's equations can be solved. A Krupp's equation for focal lengths is solved, but obtained imaginary values for the same. So there is a need of optimization procedures and recalculation of the equations as a future work.

Stitching of images is helpful for a range of computer graphics and vision activities. It has a huge number of distinct approaches for the process of features point identification and matching. The selection of approach is based on the kind of problem. Here are provided very basic and fundamental steps for the image stitching. The stitching of images proposed here is for more than two images. Due to high speed results and quicker than many of the other familiar extraction techniques, like Sift or Surf, the technique used in the proposed system for feature recognition is very appropriate for real-time video processing applications.

References

1. Choi K, Jung HG, Suhr JK (2018) Automatic calibration of an around view monitor system exploiting lane markings. *Int J Comput Appl* 33:2–26
2. Satouri B, Tairi H, El abderahmani A, Satori K (2016) Camera self-calibration with varying intrinsic parameters by an unknown three-dimensional scene. *Int J Adv Comput Sci Appl* 7:77–87
3. Satar SMA, Hussein WY, Saud BM (2018) Accurate camera calibration algorithm for laser applications. In: *International conference on laser applications and advanced materials*, vol 15, pp 1–20
4. Komorowski J, Rokita P (2018) Extrinsic camera calibration method and its performance evaluation. In: *International conference on computer vision and graphics*, vol 7594, pp 129–138
5. Tan L, Wang y, Yu H, Zhu J (2017) Automatic camera calibration using active displays of a virtual pattern. *Int J Comput Appl* 10:1–13
6. Huang S, Feng MC, Zheng TX, Li F, Wang JQ, Xiao LF (2018) A novel multi-camera calibration method based on flat refractive geometry. In: *International conference on robotics and mechatronics*, vol 320, pp 14–26
7. Yan K, Tian H, Liu E, Zhao R, Hong Y, Zuo D (2016) A decoupled calibration method for camera intrinsic parameters and distortion coefficients. *Int J Math Problems Eng* 16:1–12
8. Tang R, Deng B, Li j, Yan Y (2018) An efficient and flexible solution for camera auto calibration from $N \geq 3$ views. *Int J Math Problems Eng* 18:1–10
9. Emilia GD, Di Gasbarro D (2017) Review of techniques for 2D camera calibration suitable for industrial vision systems. *J Phys: Conf Ser* 841:41–47
10. Khoramshahi E, Honkavaara E (2018) Modelling and automated calibration of a general multi-projective camera. *Int J Photogrammetric Rec* 33:86–112
11. Anjana MV, Sandhya L (2017) Implementation and comparison of feature detection methods in image mosaicking. *IOSR J Electronics Commun Eng* 25:07–11

12. Jayanthi, Indu S (2018) Comparison of image matching techniques. *Int J Latest Trends Eng Technol* 7:396–401
13. Mistry D, Banerjee A (2017) Comparison of feature detection and matching approaches: SIFT and SURF. *GRD J Glob Res Dev J Eng* 4:7–13
14. Arya S (2017) A review on image stitching and its different methods. *Int J Adv Res Comput Sci Softw Eng* 5:299–303
15. Jenitta A, Abinandhini G, Geerthanadevi M, Latha M (2017) A fast panorama stitching method of image sequence. In: *International conference on electrical, information and communication technologies*, vol 29, pp. 58–63
16. Sruthi P, Dinesh S (2017) Panoramic image creation. In: *National conference on discrete mathematics and computing*, vol 35, pp. 12–24

Data Mining Techniques in the Agricultural Sector



B. G. Mamatha Bai and N. S. Rashmi

1 Introduction

Data Mining denotes discovering the useful information from large volume of data. It is also referred as KDD, i.e., Knowledge Discovery in Databases. Collection of data and various storage techniques has made possible for administrations to gather enormous amount of data at best price. Manipulating this stored data with the aim of extracting useful insights is the main objective of Data Mining. Choosing a proper dataset or concentrating on a subset of factors is one of the objectives to start with. After the collection of dataset, pre-process the data in order to remove the noise data and missing values. Introduce different strategies for handling the missing values. Select the appropriate attributes as important features to show the data based on the aim of the task [1].

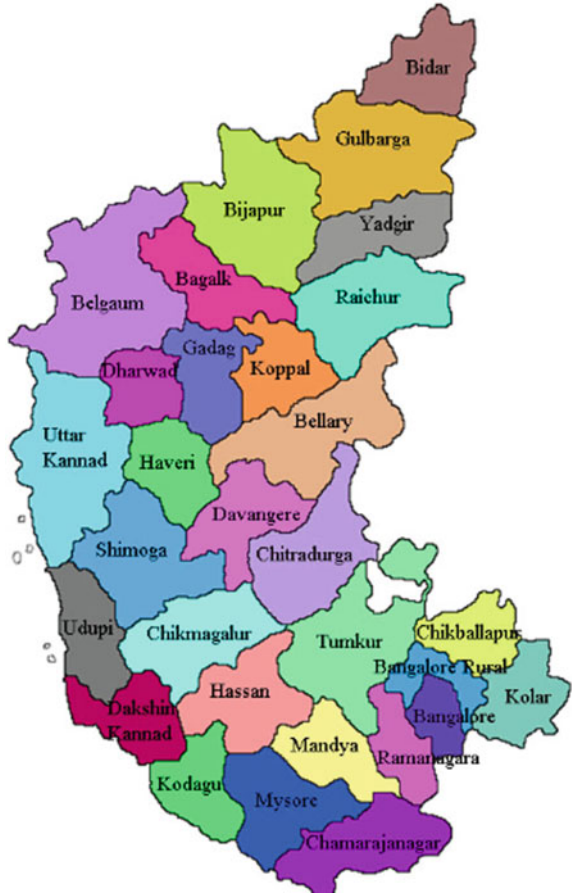
Agriculture is the backbone of our Country. The Economy is greatly affected by the Agricultural sector in a nation like India. Agricultural sector's success or failure depends on the Rainfall and the Climate of all seasons. Although there is a lot of technological use in the field of Agriculture, there is still minimal use of Data Mining Techniques in India's Agricultural sector.

1.1 Agriculture in Karnataka

Karnataka is located in India's South-West region. It is one of India's most prosperous states. In the areas of Education, Business, Agriculture, Literature, and Tourism, Karnataka has produced tremendous gains. With a total of 27,481 villages, the state has 30 Districts and 176 Taluks. For a majority of the rural population, Agriculture

B. G. Mamatha Bai (✉) · N. S. Rashmi
Department of CSE, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India

Fig. 1 Districts of Karnataka



is the main Occupation. There are five million hectares under cultivation supplied with rain. In agriculture sector, Karnataka stands second highest after Rajasthan. Karnataka’s agricultural sector is defined by a big amount of drought-prone areas and dispersed irrigated land patches. Agriculture in Karnataka is defined by broad crop diversification and remains extremely dependent on the southwestern monsoon [2]. Figure 1 shows region under study.

1.2 Background

Farmer still follows traditional methods of growing crops in the agricultural sector; this reports less production in the field of agriculture. Because of the critical problems faced in the agricultural sector, farmers feel more compelled to embrace intensive

farming methods and sustainable farming methods to maximize Financial and Environmental expenses. Farmers randomly forecast the yield of grains and other crops which affects farmer's economy. The reasons articulated for the significance behind farmer's suicides were sudden environment changes like less rainfall and change in temperature, low production, low prices, stress and family duties, bad irrigation, and higher cultivation costs. The primary reason is the low product prices and the higher cultivation costs.

2 Literature Survey

Vishal et al. [3] summarized the notion of data mining and demonstrated its importance. Data mining is studied in detail on the basis of the Neural Network and Genetic Algorithm. According to him, Genetic Algorithm attempts to integrate natural assessment thoughts. The main technology and methods of achieving data mining on the neural network and genetic algorithm are also investigated. A significant quantity of business and media attention has been attracted by data mining and knowledge discovery in databases. Smyth et al. [4] offer an overview of the evolving field, explaining how data mining and knowledge discovery in databases are linked to one another as well as associated fields such as machine learning, statistics, and databases. It cites specific real-world apps, specific data mining methods, difficulties involving the discovery of information in real-world, present and future directions in the field of study.

In terms of pattern recognition duties, many real-world systems can be researched so that adequate use of machine learning techniques becomes crucial in practical applications. Although many techniques of classification have been suggested, there is no agreement on which techniques are more appropriate for a specified dataset. As a result, comprehensive comparison of techniques in many possible situations is essential. Rodriguez et al. [5] conducted a systematic comparison of nine well-known clustering techniques in the R language, assuming normal distribution of information. The TanujWala's focus [6] is on how clustering techniques help in the agricultural field. He thinks that data mining methods such as Agglomerative Clustering, DBSCAN, EM algorithms, and K-Means will deliver progress that has long waiting in the agricultural sector. Developing models to assist and deliver the correct seed quality would go a long way in shortening the process time between growing in areas to the retail store racks. Such models will also greatly assist to reduce the expenses of warehousing, in particular the expenses of storage, losses due to rotting.

Kaur et al. [7] say "The general objective of the data mining method is to obtain and convert information from a big data set into a comprehensible form for further use." In data analysis and data mining applications, clustering is essential. Aroquiaraj et al. [8] explain multiple farm yield forecast regression methods. The suggested research in this article focuses primarily on using regression methods to obtain predictor model. Predictor formula is most helpful in agricultural crop production in tons crop

forecast. Sugarcane and tapioca are the highest yield rate in the Tamil Nadu state, especially in the northwest area. Banana, maize, ragi, turmeric, coconut, cotton and jowar have the next largest output relative to other plants, depending on the predictor model outcomes.

The crop yield prediction can be made with different models, but the Adaptive Neuro Fuzzy Inference System (ANFIS) model handles the non-linear connection between the information as opposed to the other crop yield prediction models when the input characteristics are improved by incorporating all the vital parameters necessary for crop development says Menaka et al. [9]. This method includes improving the vital inputs and adding them to the ANFIS model by reducing the input consumption and feeding into the model all the vital variables. Sandeep et al. [10], estimated the connection for chosen plants cultivated in Karnataka between crop yields and climatic factors. The three primary climatic factors chosen were complete seasonal rainfall, complete heat exposure in increasing days, and exposure in extreme days to above-critical temperatures. They have studied the option of a linear connection at the taluk's stage between crop yield variability and the three climatic variables. Taluks have been regarded the spatial unit for assessment since they are likely to be more homogeneous in terms of temperature and rainfall pattern than the district.

Shahane et al. [11] states the advanced scheme can thus provide the foundation for fertilizer suggestions for a specified crop and assess the soil's fertility status and schedule a nutrient management program. Venkatkumar et al. suggest in [12] "In today's globe, where we produce big amounts of data, we can harness the advantages of concealed information, i.e., patterns or correlations in these data." This information can only be used in different constructive areas if we are prepared to manage big data effectively. One such method used to obtain and manage the confidential information is Data Mining. There are different methods in Data Mining, namely grouping objects into clusters, predicting the future, classifying objects, association, etc.

Suman et al. [13] mainly concentrate on grid-based clustering. Grid-based clustering technique creates hierarchical structure from information and effectively responds to multiple queries. STING runs from the database to study and evaluate the cells into statistical parameters, and $O(n)$ is the time complexity, where n is the complete amount of objects. $O(p)$ is the processing time of query after generating the hierarchical structure, where p is the grid cells count at the very low level, usually much lower than n . He concludes that compared to scalability of STING, CLIQUE is better. Bellundagi et al. [14] examines the development and instability of region, manufacturing and productivity, as well as the forecast of region and output of Finger Millet crop in the states of India. Growth rates were calculated using compound growth rates in the region, manufacturing and productivity. In both phases, the trend lines showed a rise in both manufacturing and productivity in Karnataka's main ragi-growing districts, although the region under the plant is declining.

Shreya et al. [15] demonstrates altered k-Means of crop conjecture clustering by enhancing quality and precision. The modified clustering algorithm k-Means is assessed by comparing k-Means and k-Means++ algorithms, achieving the highest amount of high-quality clusters, correct crop forecast and highest precision count.

Information Mining plays a significant role in improving harvest forecast in the agricultural sector.

3 Proposed Methodology

Agriculture is the main source for so many people in India. Approximately, half of India's population relies on Agriculture for their livelihood. The objective and motivation of this work is to help Farmers in appropriate crop planning and reduce their loss. Farmers are still following the traditional approach to grow crops today. No scheme is available to educate farmers on what crops to grow, what are the required parameters for the better yield. So, here we present a DMTA Model based on analyzing trends followed in previous years to forecast Optimal Parameters that are required to obtain the better yield for Farmer's land using different methods of Data Mining. The proposed DMTA Model is shown in Fig. 2.

3.1 Clustering

Clustering method is one among the most significant methods under Data Mining. Clustering means establishing object groups based on their characteristics so that objects belonging to the same groups are equivalent and objects belonging to different groups are distinct. There are various types of clustering techniques as shown in Fig. 3.

3.2 Algorithms Used in DMTA Model

In order to propose DMTA Model, the following Data Mining techniques are implemented and are analyzed with different distance functions for distinct crops.

3.2.1 Bisecting K-Means

It is the most commonly used algorithm under Partitioning based. Initially, it divides a collection of objects according to their resemblance into groups of objects or clusters. The Bisecting K-Means algorithm is an updated version of the standard K-Means algorithm. We often "refine" the clusters with the fundamental K-means algorithms for various clusters, but we do not refine the nesting clusters. Detailed algorithm is given in Appendix 1.

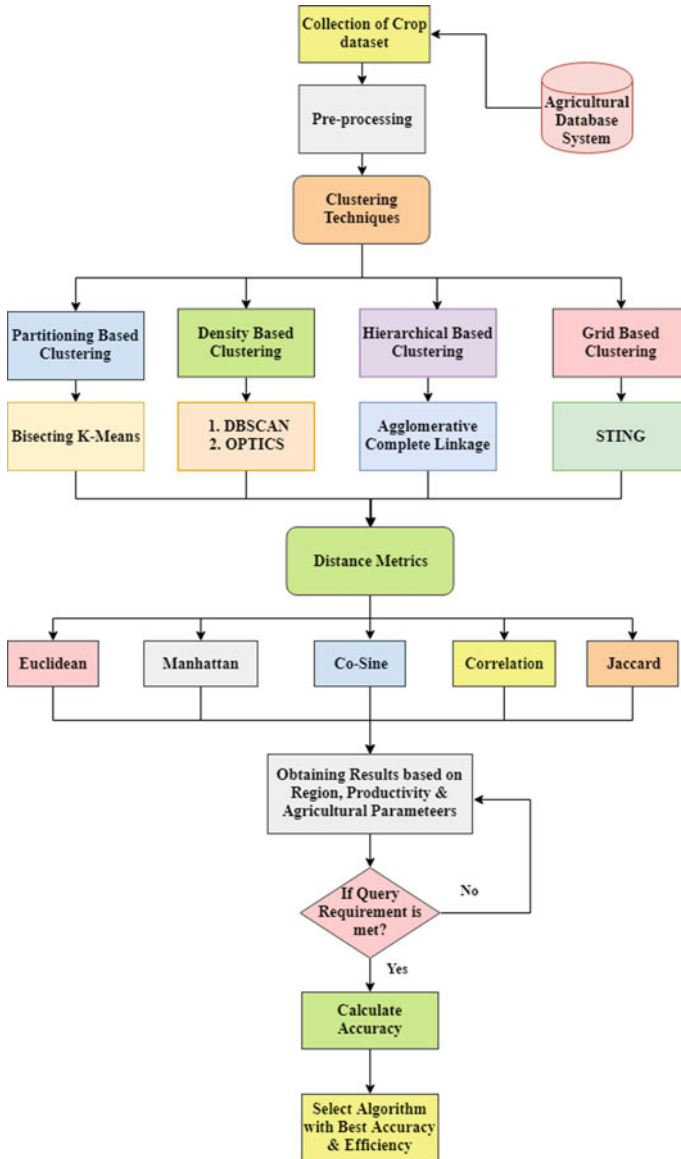


Fig. 2 Proposed DMTA model

3.2.2 DBSCAN

It is a clustering method used to separate high-density clusters from low-density clusters as part of machine learning. It finishes an amazing activity in finding areas in the data that has strong observation density versus data areas that are not highly

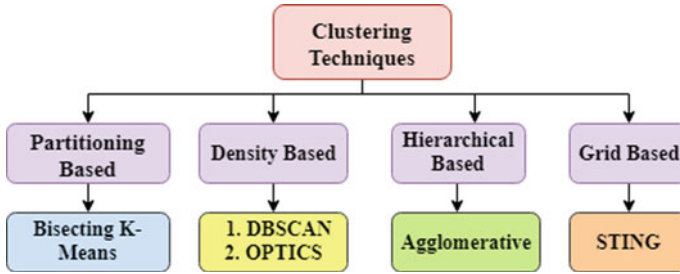


Fig. 3 Clustering techniques

dense with observations. DBSCAN can also sort data into clusters of fluctuating forms, another strong preferred point of view. Two parameters have to be defined for this, i.e., epsilon and min-point.

3.2.3 OPTICS

To identify density in spatial data, OPTICS is the most commonly used algorithm. Its fundamental concept is comparable to DBSCAN, but it overcomes one of the main disadvantages or weaknesses of DBSCAN: the issue of identifying significant clusters in variable density information. Two parameters are required for OPTICS as of DBSCAN:

- Epsilon (ϵ)—This denotes the maximum distance to be considered.
- Minimum point (MinPts)—This denotes Number of points required to be in cluster.

Detailed algorithm is given in [16].

3.2.4 Hierarchical Clustering

It is used on the basis of their resemblance to group objects in clusters. It is also addressed as the Agglomerative Nesting (AGNES). Initially, each object is treated as a cluster of singletons. Next, pairs of clusters are combined successively until all clusters are merged into one large cluster with all the objects. Detailed algorithmic steps are given in Appendix 2.

3.2.5 STING

STING and CLIQUE are part of the Grid-based clustering. The spatial region is split into rectangular cells in the STING algorithm. There are several distinct levels corresponding to distinct resolution of such rectangular cells, and these cells form a

hierarchical framework. Each high-level cell is divided into a number of low-level cells. Each cell's statistical data is calculated and is stored in advance which later responds to queries. Detailed Algorithm is given in Appendix 3.

3.3 Distance Measures

Classifying observations into groups needs certain techniques to calculate the distance or (dis)similarity between each pair of observations. There are various methods to calculate distance function they are as follows:

3.3.1 Euclidean Distance

The Euclidean distance in either plane or 3-dimensional space between two points measures the length of a section that connects the two points. It is the most evident way to show the distance between two points.

$$d_{\text{euc}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where x and y are the variables of length n .

3.3.2 Manhattan Distance

Two points distance can be measured using Manhattan distance in the correct angles along the axes.

$$d_{\text{man}}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

where x and y denote the variables of length n .

3.3.3 Correlation Distance

The function of correlation distance calculates the distance between two double vectors and returns a value in $[0, 1]$. An outcome of 0 implies a favorable linear connection exists. An outcome of 1 implies a linear negative connection exists. An outcome of 0.5 implies no linear connection.

$$\text{Pearson}(x, y) = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{N}\right)\left(\sum y^2 - \frac{(\sum y)^2}{N}\right)}} \quad (3)$$

where x and y are the data objects and N refers to the total number of attributes.

3.3.4 Jaccard Distance

Sometimes it is called as Jaccard relevance coefficient, a statistics used to compare the likeness and variety of sample sets. The Jaccard coefficient measures likeliness between sample sets and is described as the size of the boundary separated by the size of the sample sets union.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4)$$

When A and B are empty, we define $J(A, B) = 1$.

3.3.5 Cosine Distance

It refers to a metric of resemblance between two n -dimensional vectors by discovering the perspectives between them. Given two attribute vectors, A and B , cosine similarity is calculated using a magnitude and dot Product as:

$$\theta = \arccos \frac{X \cdot Y}{|X||Y|} \quad (5)$$

where x and y are the variables.

4 Dataset

Agriculture is regarded as one of the Karnataka residents "main occupations." In Karnataka, the majority of individuals are engaged in growing crops, particularly in rural regions. Many attributes are considered for the analysis and is shown in Table 1. Past 10 years data has been considered for analyzing and obtaining better results. Distinct agricultural crops of Karnataka are considered for analysis. Yield of Primary crops like Ragi, Paddy, and oil seed like Groundnut are analyzed based on different districts of Karnataka. All crops dataset is taken from government and raithamithra website [17]. Weather and soil parameters play a major role in cultivation and are listed below:

Table 1 Attributes of dataset

Sl. No	Attributes	Description	Sl. No	Attributes	Description
1.	Districts	29 Districts of Karnataka	8.	Average annual humidity (%)	Average humidity for each district is considered
2.	Type of Soil	Major Six types of Soil are considered	9.	Area under cultivation	Total area considered in a particular district for cultivation
3.	Crop year	10 years of data considered	10.	pH value	Soil pH is a metric of soil acidity and alkalinity
4.	Crop name	Ragi, Groundnut and Maize	11.	P ₂ O ₅ (kg/ha)	Phosphorus content in various districts of Karnataka
5.	Average annual rainfall (mm)	Average rainfall for each district is considered	12.	Manganese (kg/ha)	Manganese content in various districts of Karnataka
6.	Season	Kharif, Rabi, Summer	13.	Iron (kg/ha)	Iron content in various districts of Karnataka
7.	Average annual temperature (Celsius)	Average temperature for each district is considered	14.	Production	Total production of the crop (ton/ha)

- **Seasons:** State agricultural output is spread over three seasons:
 - a. Kharif—Starts from July through October.
 - b. Rabi—Starts from October through March.
 - c. Summer—Starts from April through June.
- **Types of soil:** Depending on the soil’s agricultural capacity, the soil types are categorized as Red soil, Black soil, Brown soil, Laterite soil, Loamy soil, etc.

5 Experimental Results

Data Mining Techniques are applied using DMTA Model for three distinct crops: Ragi, Groundnut and Maize. Clustering techniques used are Bisecting K-Means, DBSCAN, OPTICS, Hierarchical Complete Linkage and STING. Figures 4, 5, and 6 represent the clusters formed for the instances considered for the crops Ragi, Groundnut, and Maize respectively. Each bar plot shows the Cluster Formation.

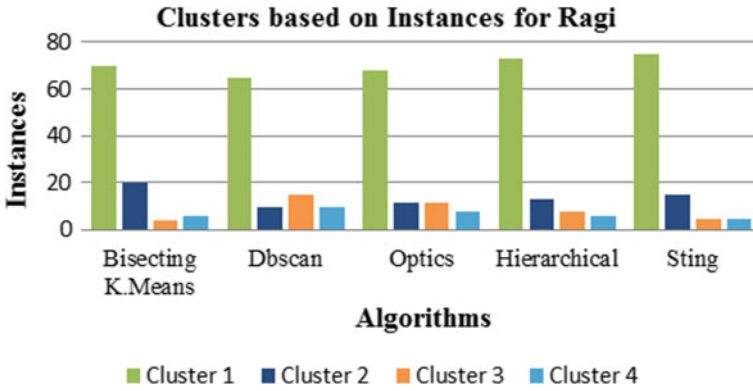


Fig. 4 Clusters based on instances for the Ragi crop

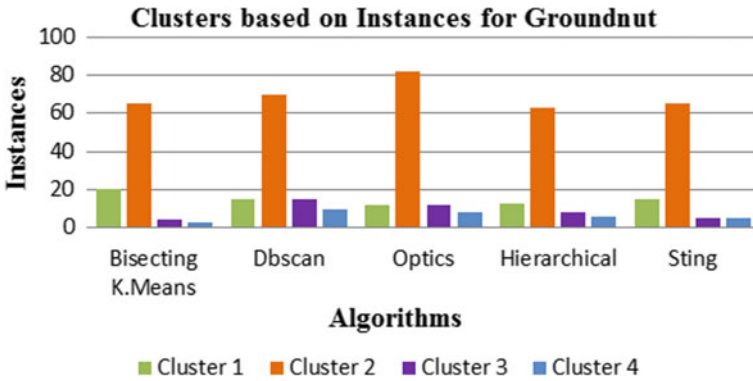


Fig. 5 Clusters based on instances for the Groundnut crop

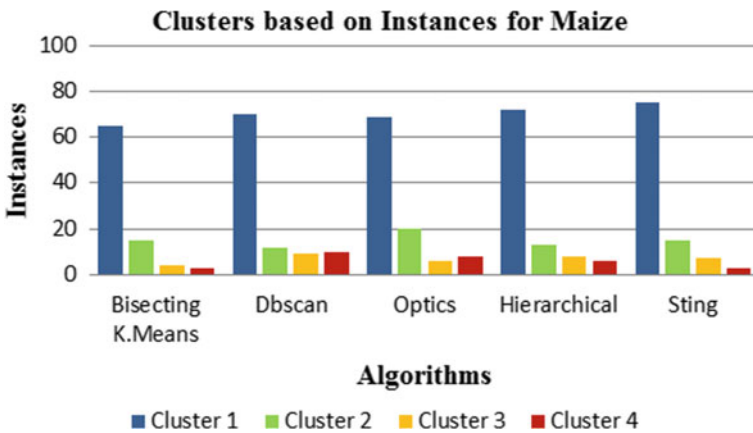


Fig. 6 Clusters based on instances for the Maize crop

Performance measures for all three crops are calculated based on time and memory consumed by each algorithm and are shown in Fig. 7. For all five algorithms, performance measure is evaluated. Least time and memory consumed shows good performance. In Fig. 8, error rate is calculated for all five distance measures for the different algorithms implemented. Hierarchical and STING algorithms for Jaccard distance and correlation distance gives best result compared to other measures. Less error rate indicates good performance.

Performance measures for Bisecting K-Means are shown in Table 2. Five Distance Measures are used for analysis. According to the observation, Jaccard and Correlation give less error rate, execution time, and memory consumed compared to other distances. Figure 9 shows a graphical representation of Bisecting K-means algorithm’s performance as explained above.

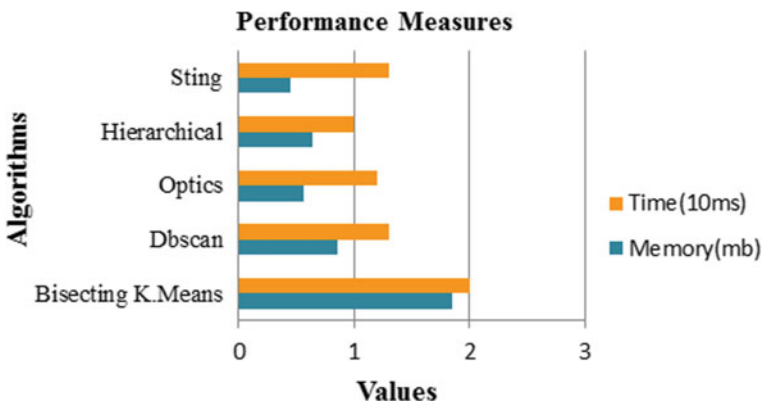


Fig. 7 Overall performance measures

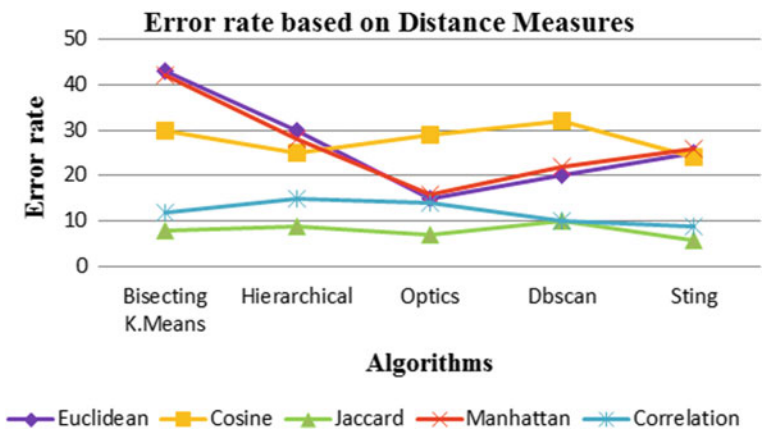


Fig. 8 Error rate based on distance measures

Table 2 Performance of bisecting K-means algorithm

Algorithms	Method	Dataset	Distance measure	Error rate (SSE)	Execution time (ms)	Memory consumed (mb)
Bisecting K-means	Partitioning method	Ragi Groundnut Maize	Euclidean	43	20	6.256
			Cosine	30	18	6.86
			Jaccard	15	15	5.56
			Manhattan	42	19	7.643
			Correlation	15	14	5.45

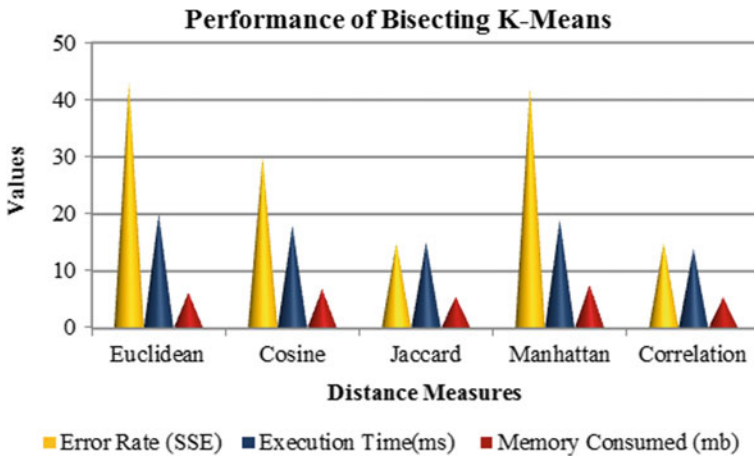


Fig. 9 Performance of Bisecting K-means algorithm

Performance measures for DBSCAN are shown in Table 3 and Fig. 10. Five distance measures are used for analysis. According to the observation, Jaccard and Correlation show less error rate, execution time, and memory consumed compared to other distances.

Table 3 Performance of DBSCAN algorithm

Algorithms	Method	Dataset	Distance measure	Error rate (SSE)	Execution time (ms)	Memory consumed (mb)
DBSCAN	Density-based method	Ragi Groundnut Maize	Euclidean	20	19	3.9
			Cosine	32	17	5.86
			Jaccard	10	14	5.6
			Manhattan	22	18	7
			Correlation	17	15	4.8

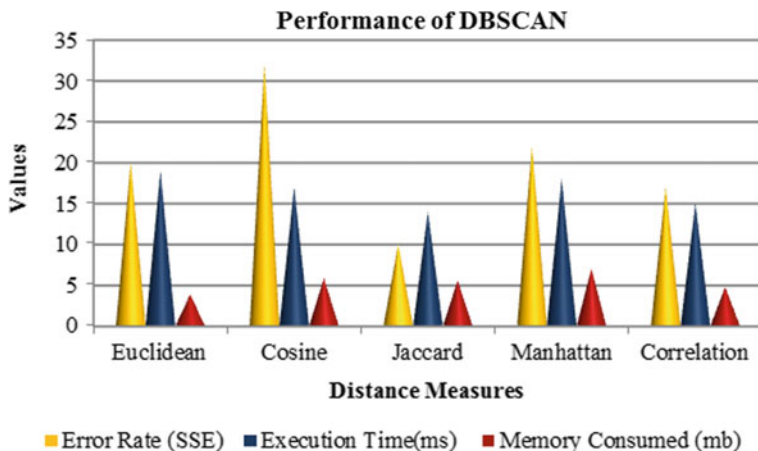


Fig. 10 Performance of DBSCAN algorithm

Table 4 Performance of OPTICS algorithm

Algorithms	Method	Dataset	Distance measure	Error rate (SSE)	Execution time (ms)	Memory consumed (mb)
OPTICS	Denisty-based method	Ragi Groundnut Maize	Euclidean	15	20	5.9
			Cosine	29	17	5.86
			Jaccard	15	15	4.2
			Manhattan	16	19	7
			Correlation	14	15	4.8

Performance measures for OPTICS are shown in Table 4 and Fig. 11. Five distance measures are used for analysis. According to the observation, Cosine shows more error rate which leads in missing data or outliers; Outliers are nothing but the data which does not belong to any Clusters. Jaccard, Euclidean, Manhattan, and Correlation give less error rate, execution time, and memory consumed compared to other distances.

Performance measures for Agglomerative Complete Linkage are shown in Table 5 and Fig. 12. According to the observation, Jaccard and Correlation show less error rate, execution time, and memory consumed compared to other distances.

Performance measures for STING are shown in Table 6 and Fig. 13. Five distance measures are used for analysis. According to the observation, Jaccard and Correlation show less error rate, execution time, and memory consumed compared to other distances. Based on the result analysis, we can conclude that STING algorithm and Hierarchical algorithm with Jaccard or Correlation distance give less error rate, good time, and memory consumption.

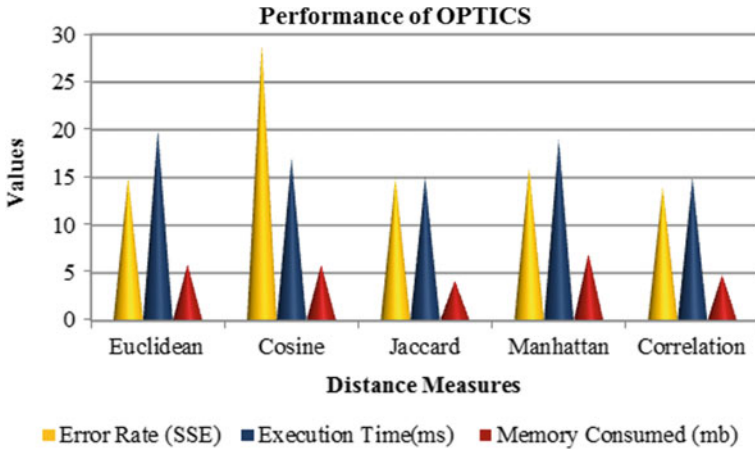


Fig. 11 Performance of OPTICS algorithm

Table 5 Performance of agglomerative complete linkage algorithm

Algorithms	Method	Dataset	Distance measure	Error rate (SSE)	Execution time (ms)	Memory consumed (mb)
Agglomerative complete linkage	Hierarchical method	Ragi Groundnut Maize	Euclidean	30	19	3.9
			Cosine	25	17	5.86
			Jaccard	7	14	4.6
			Manhattan	28	18	7
			Correlation	10	15	4.8

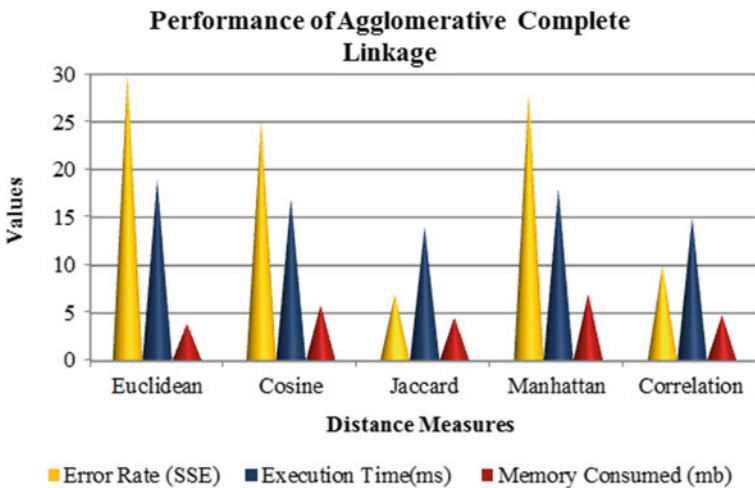


Fig. 12 Performance of agglomerative complete linkage algorithm

Table 6 Performance of STING algorithm

Algorithms	Method	Dataset	Distance measure	Error rate (SSE)	Execution time (ms)	Memory consumed (mb)
STING	Grid-based method	Ragi Groundnut Maize	Euclidean	25	18	4.9
			Cosine	24	17	5.86
			Jaccard	6	13	4.2
			Manhattan	26	18	7
			Correlation	9	15	4.8

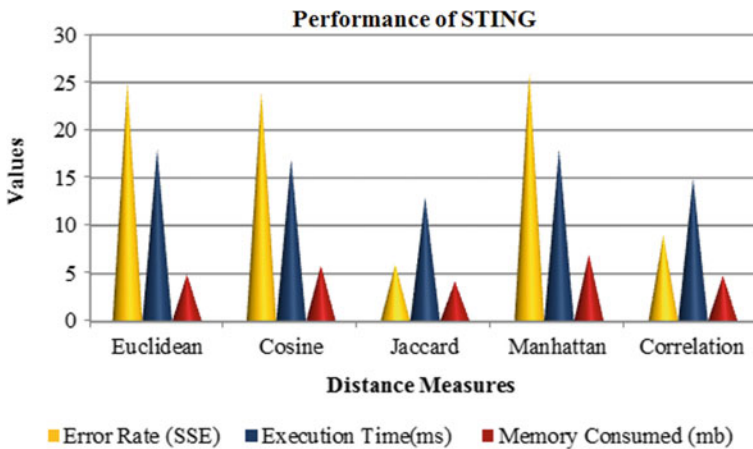


Fig. 13 Performance of STING algorithm

This work is analyzed by considering 10 years of agricultural data w.r.t. Highest Ragi Production in various Districts of Karnataka. The data is given as input to the STING algorithm and results are obtained. According to the analysis, districts such as Bangalore Rural, Tumakuru, Chikaballapura and Kolar are the major Ragi production regions in Karnataka State as shown in Fig. 14. There are various attributes which affects the cultivation of crops. Some of them are average temperature, average humidity, average rainfall, pH value, p2o5, and K2O. Figure 15 shows the weather and soil parameters which are necessary to get highest production for the Ragi crop.

Figure 16 shows the analysis of highest Groundnut production Districts of Karnataka. Chitradurga, Koppal, Raichur, Gadag, and Chikballapur are the major groundnut production districts. Optimal parameters necessary for groundnut cultivation are presented in Fig. 17. Figure 18 represents the analysis of highest Maize production districts of Karnataka. As per the analysis, Bagalkote, Bijapura, Gulbarga, and Bidar are the major Maize production districts. Figure 19 depicts the optimal parameters considered for Maize production.

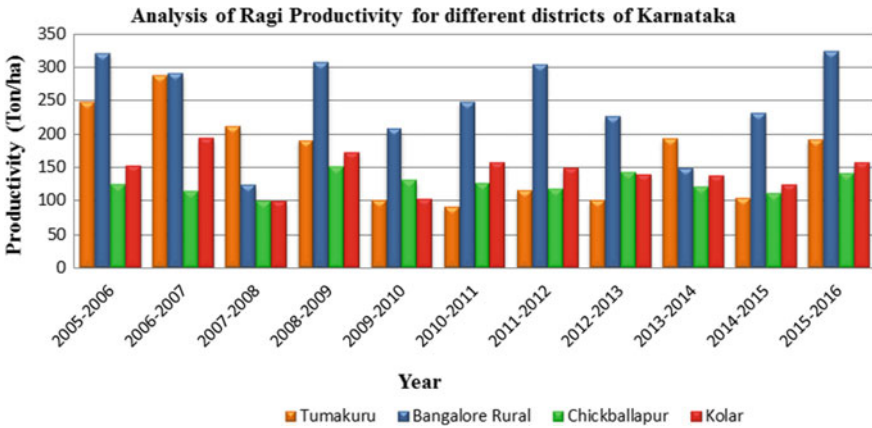


Fig. 14 Analysis of highest Ragi production districts of Karnataka

Optimal Parameters for Ragi Cultivation

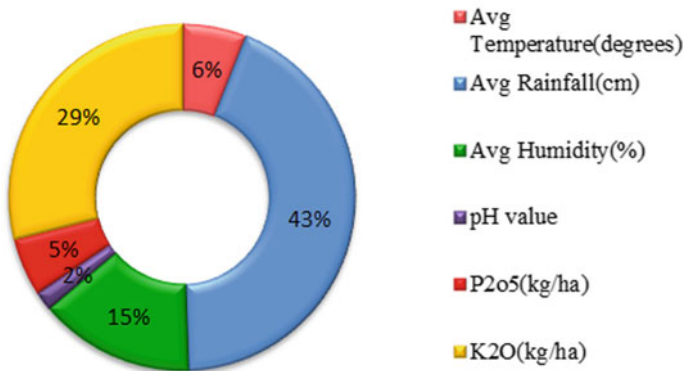


Fig. 15 Optimal parameters for Ragi cultivation

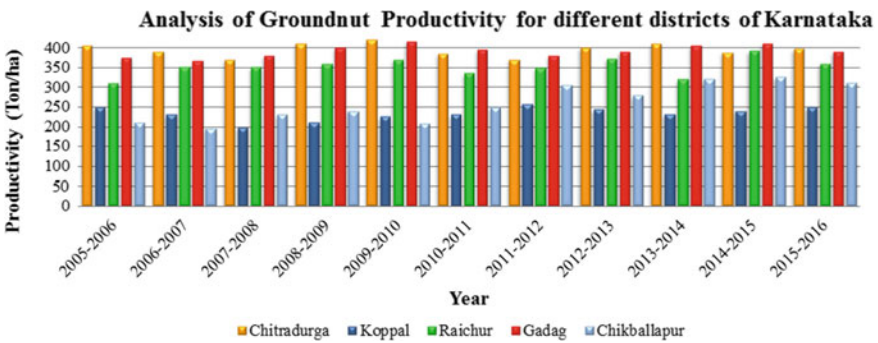


Fig. 16 Analysis of highest groundnut production districts of Karnataka

Optimal Parameters for Groundnut Cultivation

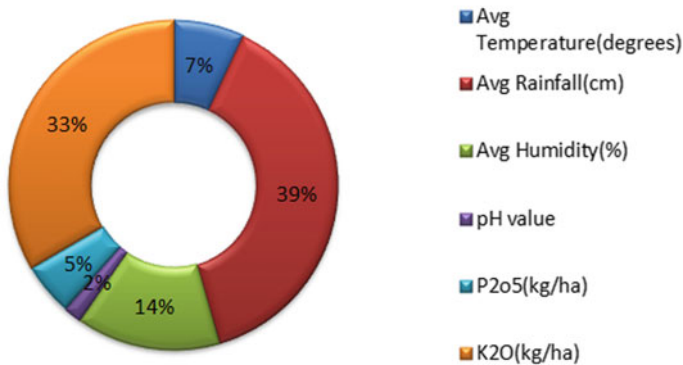


Fig. 17 Optimal parameters for Groundnut cultivation

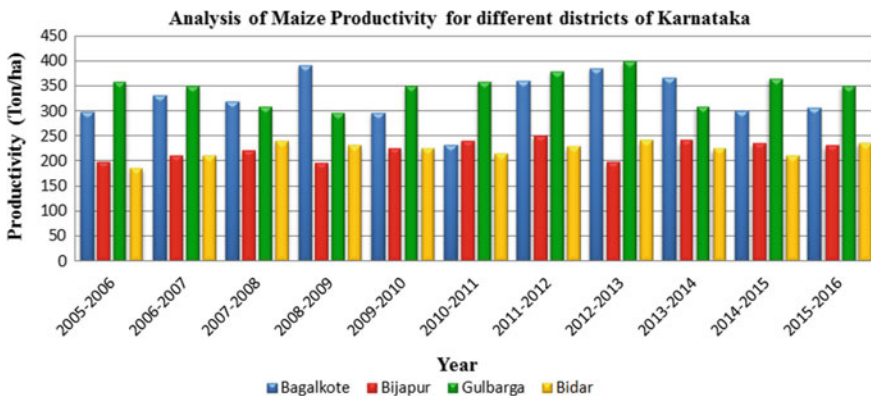


Fig. 18 Analysis of highest Maize production districts of Karnataka

6 Conclusion

This research work is based on Data Mining in Agricultural Sector. Comparative analysis is done for distinct crops using proposed DMTA Model (Data Mining Techniques in Agriculture sector). In this, five algorithms from four different types of clustering techniques are studied, namely Bisecting K-Means from Partitioning clustering; DBSCAN and OPTICS from Density-based clustering; Agglomerative Complete Linkage from Hierarchical clustering and STING from Grid-based clustering.

It is observed that among five algorithms STING and Hierarchical clustering has good performance measure, and when it comes to distance measures, Jaccard distance and Correlation distance have less error rate. Using the analysis of highest production, optimal parameters for ragi, groundnut, and maize are obtained. This

Optimal Parameters for Maize Cultivation

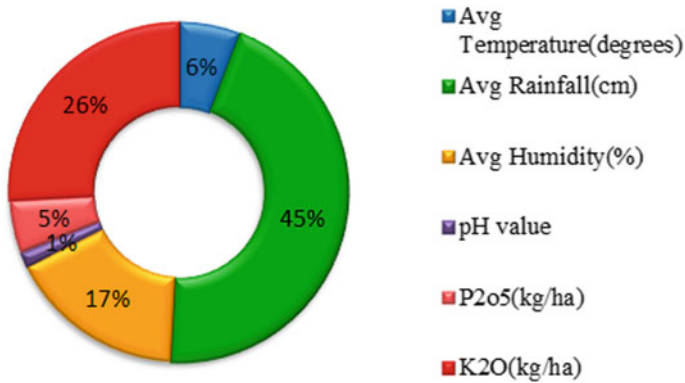


Fig. 19 Optimal parameters for Maize cultivation

helps the farmer in choosing the agricultural parameters that result in better yield and production. This helps farmers to adopt smart farming and gain considerable profit.

This work can be further experimented by considering other crops as well as different data mining techniques.

Acknowledgements The authors express their sincere gratitude to Prof. N. R Shetty, Advisor, Dr. H C Nagaraj, Principal, Dr. Jharna Majumdar, Dean R&D, Dr. Ramachandra A C, HoD, ECE, Nitte Meenakshi Institute of Technology for giving constant encouragement and support to carry out research at NMIT. The authors extend their thanks to Vision Group on Science and Technology (VGST), Government of Karnataka, to acknowledge our research and providing financial support to setup the infrastructure required to carry out the research.

Appendix

Appendix 1

Algorithm: Bisecting K-means.

Input: Distinct crops dataset, K value, Distance function.

Output: Grouping into Clusters.

Step 1: Choose a cluster to divide

Step 2: Use the fundamental K-means algorithm to find 2 sub-clusters. (Step bisecting)

Step 3: Repeat step 2 and take the split with the greatest general resemblance producing the clustering.

Step 4: Repeat first 3 steps until you reach the required amount of clusters.

Appendix 2

Algorithm: Agglomerative Complete Linkage

Input: Dataset of Distinct crops

Output: Clusters

Step 1: Preparation of data computing information (dis)similarity between each pair of objects in the data set.

Step 2: Using the linkage feature to group objects in a hierarchical cluster tree based on the distance data produced at step 1.

Step 3: Using the linkage feature, objects / clusters in close proximity are connected together.

Step 4: Determine where the hierarchical tree should be cut into clusters. This generates a data partition.

Appendix 3

Algorithm: STING

Input: Dataset of Distinct crops

Output: Clusters

Step 1: To start with, determine a layer.

Step 2: Calculate the trust interval (or estimated range) of probability to the cell which is applicable to the request for each cell in the layer.

Step 3: Label the cell as appropriate or not appropriate from the interval calculated above.

Step 4: If the bottom layer is this layer, go to Step 6; if not, go to Step 5.

Step 5: We're going one level down the hierarchy system. For those cells that form the appropriate higher-level layer cells, go to Step 2.

Step 6: Go to Step 8 if the request requirement is met; else go to Step 7.

Step 7: Retrieve and process these information into the appropriate cells. Give the outcomes that fulfill the query's requirement. If requirement is met then stop

Step 8: Find the appropriate cell areas. Return those areas that fulfill the query's requirement. Go to the 9th step.

Step 9: Stop.

References

1. Bharati M, Ramageri (2015) Data mining techniques and applications. *Indian J Comput Sci Eng* 1(4):301–305. ISSN: 0976–5166
2. Gadge Y, Sandhya (2017) A study on various data mining techniques in agriculture. In: International conference on electrical, electronics, communication computer and optimization techniques, vol 3, pp 420–423
3. Jain N, Srivastava V (2013) A study on data mining techniques. *Int J Res Eng Technol* 2:116–119. eISSN: 2319-1163 | pISSN: 2321-7308
4. Zhu X, Davidson I (2014) Knowledge discovery and data mining: challenges and realities. *Int J Adv Res Comput Eng Technol* 301–304. ISBN 9781-59904-252, Hershey, New York
5. Rodriguez MZ, Comin CH (2016) Clustering algorithms: a comparative approach. *Int J Plus One Res* 4:921–930
6. Wala T (2015) Various data mining techniques in agriculture. *Int J Adv Res Comput Eng Technol* 3:235–240
7. Mann AK, Kaur N (2013) Survey paper on clustering techniques. *Int J Sci Eng Technol Res* 2(4):803–806
8. Surya P, Laurence Aroquiaraj I (2018) Crop yield prediction in agriculture using data mining predictive analytic techniques. *Int J Res Anal Rev* 5(4)
9. Menaka K, Yuvaraj N (2016) A survey on crop yield prediction models. *Indian J Innov Dev* 5(12):783–786
10. Murari K, Sandeep (2019) Extreme crop yields and temperature in Karnataka. *Research Gate publications*, vol 3, pp 112–116
11. Silas NM, Nderu L (2017) Prediction of tea production in Kenya using clustering and association rule mining techniques. *Am J Comput Sci Information Technol* 5(2):1–8. ISSN 2349-3917
12. Venkatkumar IA, Jayantibhai S, Shardaben K (2016) Comparative study of data mining clustering algorithms. In: IEEE international conference on data science and engineering, vol 3, issue 3, pp 111–116. ISSN (P): 2349-3968, ISSN (O): 2349-3976
13. Suman, Pinkirani (2017) A survey on STING and CLIQUE grid based clustering methods. *Int J Adv Res Comput Sci* 8(5):245–250. ISSN No. 0976–5697, May–June 2017
14. Bellundagi V, Umesh KB, Ravi SC (2016) Growth dynamics and forecasting of finger millet (Ragi) production in Karnataka. *Econ Affairs* 2:195–201
15. Bhanose SS, Bogawar KA (2016) Crop and yield prediction model. *Int J Adv Sci Res Eng Trends* 1(1):23–28. ISSN: 2456-0774
16. Mamatha Bai BG, Nalini BM, Majumdar J (2018) Analysis and detection of diabetes using data mining techniques—a big data application in healthcare. *ERCICA 2018*, vol 1, ISSN 2194-5357 ISSN 2194-5365 (electronic), *Advances in intelligent systems and computing*, ISBN 978-981-13-5952-1. ISBN 978-981-13-5953-8 (eBook). <https://doi.org/10.1007/978-981-13-5953-8>
17. Agricultural crops dataset including weather parameters and soil parameters are taken from <http://raitamitra.kar.nic.in/statistics.html>. <https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics>

Design of a Simplified ANN Model for Real Power Prediction Problem



V. Parthasarathy, B. Muralidhara, Bhagwan ShreeRam, and M. J. Nagaraj

1 Introduction

Various models for Real Power Prediction (RPP) have been proposed in the last few decades and sufficiently analyzed by researchers[1–3]. Particularly, when the actual issue is to develop a prediction algorithm in hardware, the power Engineers opted for conventional techniques instead of the artificial intelligence mechanisms. For example, in the unit commitment problem, the number of generators that are to be engaged for the given loading condition can be decided by the curve fitting technique where the available system data have been plotted as a graph with reference to a common axis. This common axis may be any of the dependent or independent variables of the system control. And these variables are to be selected such that they are not producing any sort of nonlinearity at any given operating condition.

The recent innovations in the operation and control domain are the development of intelligent controllers. Instead of developing a computer-based algorithm using the system variables, one can develop customized hardwares. These hardwares not only producing the crucial decisions related to the system operation but also implementing the same. In the RPP problem, prediction, engagement and control (in short the PEC) strategy can be effectively modeled by a dedicated hardware wherein case the role of a computer-based solution is restricted to prediction part alone. But the limitation of such a process is the unavailability of “exact” equivalent hardware element for a software syntax. An approximated identification of hardware element may be ending

V. Parthasarathy (✉) · M. J. Nagaraj
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: parthasarathy.v@nmit.ac.in

B. Muralidhara
Higher College of Technology, Al-Khuwair, Muscat, Oman

B. ShreeRam
Lovely Professional University, Phagwara, Punjab, India

with the erratic outcome for the selected control problem [4]. In this paper, the first stage of a PEC system is proposed where only the prediction is discussed. It is ensured in the coding part that the syntaxes used can be easily transformed into their equivalent hardware using and description coding.

The ultimate limitation of the conventional ANN software design is the enormous time required for the training of the database [5]. It is to be noted that the learning duration will be exponentially increasing with the size of the network. If the weight of every segment has been fixed at the beginning itself and remains constant throughout the process, a mere software design may provide some solution by adopting any programming language in the computer with acceptable error [6]. Where in case the weights are to be dynamically adjusted depending on the algorithm, an adoptive approach is required. A dedicated hardware with variable model can be more effectively implement the same. Still lot of researches are going on better adoption with reduced memory requirement.

The unique feature of ANNs, the inherent parallelism can be achieved in a better manner in a hardware implementation [7]. This can be justified by the higher performance of the generalized ANN systems designed so far with reasonable cost. Also for a real-time computing process, only hardware part is not sufficient and the basic software is also required. But for the custom neuro computing, developing the software is comparatively tougher task due to its limited user base. Since the FPGAs are widely used this is an easier task [8]. Before going for selecting an ANN model, the factors which are affecting the daily load profiles must be identified [9, 10].

Liu et al. [11] have discussed the physical and mathematical relations between the load forecasting and the unit commitment problems. They have suggested the genetic algorithm-based solution for unit commitment. Jahromi et al. [12] suggested that the optimum allocation of generating units can be incorporated with the ANN-LF models and a single integrated approach can be better than separation of these problems.

Sharif et al. [13] extended the conventional FFNNs with recurrent systems and they have proved particularly for the smart grid scenario, where the nonlinearity is higher than the conventional grids. They did a comparative study in this regard. Zhao et al. [14] analyzed the impact of load forecasting models in the Distributed Generation (DG) scenario. In the DG, the primary constraint will be the system security rather than the cost wherein the normal LF problems concentrating on the cost factors only. The authors have been tried to balance between the system stability and the optimum allocation by a new sets of differential equations.

Hammad et al. [15] reviewed almost 45 papers and observed that despite the relative simplicity of all reviewed models, the regression analysis is still widely used and efficient for long-term forecasting. As for short-term predictions, machine learning or artificial intelligence-based models such as Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Fuzzy logic are favored.

2 Preliminary Considerations

For the beginning level analysis, randomly selected structures with different number of neurons have been considered. Every model is trained for same set of data and tested with a common testing value. The model that produces lesser error has been picked for further analysis. The details are shown in the table in Fig. 1.

Let “x” be the date for which the prediction is required. Every day of a week is assigned with a number varying from 1 to 7 and every month of a calendar year is assigned with 1 to 12. The objective is to train the ANN using the x-n’ data and make it to predict about the requirement on “x”. It is to be noted here that the real-time variations on the customer end may be slightly varying from the understanding of the network. Since the power requirement is directly impacted by the temperature of a particular day that is also considered as a variable. Here, T_1 and T_2 are the minimum and maximum temperatures of any given day. There are two difficulties related to this parameter. The first one is the unavailability of accurate temperature data. And the other issue is some cases, even though the temperature is same, the power consumption got differed. It is depending on other operating conditions including pollution level, effective spacing and line specifications.

MODEL	PARAMETER-1	PARAMETER-2	PARAMETER-3	PARAMETER-4	TOTAL
1	Day for ‘x-1’ (1-7)	Month for ‘x-1’ (1-12)	Real power consumption during every period on ‘x-1’ & ‘x-2’	T_1 and T_2 for x-1 & T_1 and T_2 of x-2	54
2	Day for ‘x-1’ (1-7)	Month for ‘x-1’ (1-12)	Real power consumption during every period on ‘x-1’ (24)	Real power consumption during every period on ‘x-2’ (24)	50
3	Day for ‘x-1’ (1-7)	Month for ‘x-1’ (1-12)	Real power consumption during every period on ‘x-1’ (24)	-	26
4	Day for ‘x-1’ (1-7)	Month for ‘x-1’ (1-12)	Real power consumption during every period on ‘x-1’ (24)	T_1 and T_2 for x-1 & T_1 and T_2 of x-2	30
5	Day for ‘x-1’ (1-7)	Month for ‘x-1’ (1-12)	Real power consumption during every period on ‘x-7’ (24)	-	26

Fig. 1 ANN Models for initial analysis

3 Neural Network Training

The training database first filtered, pre-processed and the training process were conducted using the standard protocols. As mentioned earlier “one year training and one month testing” has been adopted. Instead of selecting continuous training patterns, the total raw data have been rigorously scrutinized and the patterns were identified. The weights for various stages are initialized with 0.3 and during it gets modified. Whenever the patterns were changed for updating, the corresponding weights were adjusted by using the legrangian approximation.

Since the original database has almost 12,000 numerical values, to pick and tabulate the identified values, a separate program has been developed. From excel format, the values were changed to data format and they were placed in the same working directory. The MATLAB program with testing and training loops is also placed in the same folder. The data which has been given by government authorities is consist of time, power and date so among these the time and date have been saved separately and power database has been saved separately such that during the training and testing both will execute separately and gives predicted output as power. After separation of data, the data is converted into normalized form in between 0 to 1 using normal standard formulas for date, month, year and power. This is because the neural network that can work efficiently if data range should be in between 0 to 1.

All the five ANN designs have been trained for 365 days (8760 samples) from May 2017 to April 2018 and eventually tested for 30 days during May 2018, and their Mean Average Percentage Error (MAPE) has been evaluated. It is an indication of how effectively the ANNs absorb the nonlinearity of the input data. The results show that the models 1 and 2 are producing more MAPE (>15%). Hence, the two models were not considered for prediction process. The remaining three models were further tested with different set of data. The conclusion here is that when the number of input variables is more, the ANN is unable to be trained properly. The models 1 and 2 are each having roughly 50 inputs and the mapping of data become very cumbersome. Also, the temperature data seems to be inaccurate and not enough to get the desired accuracy.

Also when the number of training patterns is changed, the MAPE has been drastically changed. When the samples are very less (say 720 or 1440), the error produced by the network was very high. Also among the three models, model with 26 inputs has been provided error within the limits and the same have been selected for final implementation (Fig. 2).

4 Final Analysis

The final ANN model is fixed with a structure of 26-5-24. From the hardware realization perspective, the number of multipliers per neuron must be equal to the number of connections to this neuron. The number of full adders equals to the number of

Fig. 2 Error comparison of 3 models

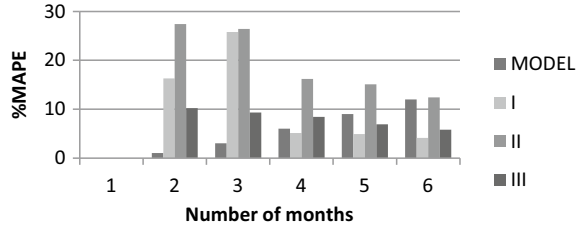
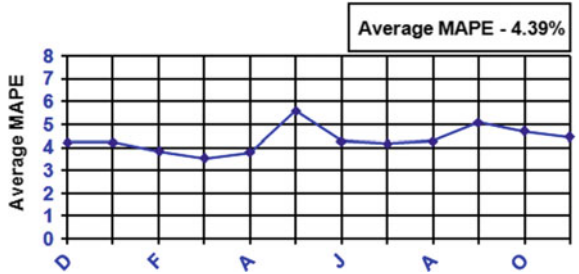


Fig. 3 Month wise MAPE produced by selected model



connections to the previous layer minus one. Hence, every output neuron must have 5 multipliers and four adders and every hidden neuron should have 26 multipliers and 25 adders. Both these elements since readily available with the DSP boards, the hardware realization for this work is feasible. As soon as the input details including day, month and power consumption of previous day have been given as input, the present power requirement has been predicted by the system.

The graph shown below gives the MAPE produced by the final ANN model monthly wise for the year 2018.

From the graph in Fig. 3, it is clear that when the weather is playing the crucial role, the model is unable to predict the accurate power requirement details. In the month of May, the climate is unpredictable and sudden showers reduce the power consumption drastic reduction of number of units engaged in the system. Apart from this spike, the predictions for other months are within the acceptable limits.

The performance of ANN has been analyzed under three different scenarios. A normal working day, a holiday and a semi-holiday have been considered and every day the prediction curves have been plotted.

(a) **A working day:**

With respect to the graph shown, the working day prediction was approximately following the actual load requirement. The average MAPE in this case is 5%. But during the peak load demand and the late night consumptions have been predicted not so accurately. The training data set is to be changed, and there is a possibility of improvement of performance (Fig. 4).

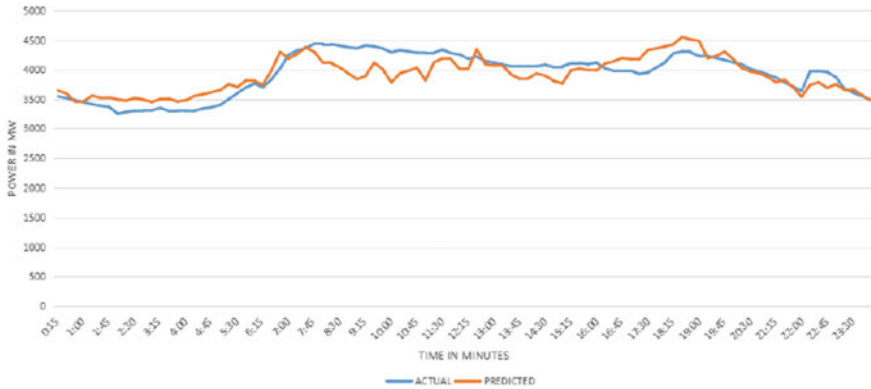


Fig. 4 ANN prediction for a working day

(b) **A Holiday:**

As shown in Fig. 5, the prediction was more or less mapped with the actual power consumption except 7.00 AM in the morning. Sudden drop is predicted and the reason was analyzed and the conclusion is that during the 7.00 AM load data is experienced the sudden drop twice in the subsequent weeks. The average MAPE stands around 8% for all the holiday predictions.

(c) **A Restricted Holiday:**

Since the number of restricted holidays are less, the prediction was made with acceptable accuracy. During the restricted days, certain loads will be functioning normally and the remaining will be shut down. The trend can be easily studied and the theory developed based on this. The average MAPE is 6% (Fig. 6).

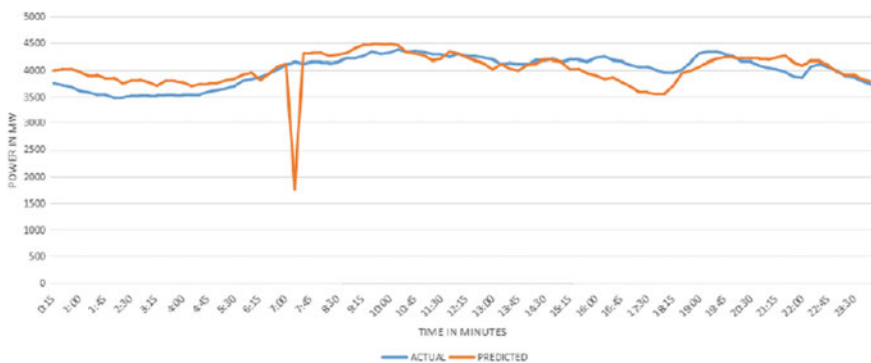


Fig. 5 ANN prediction for a holiday

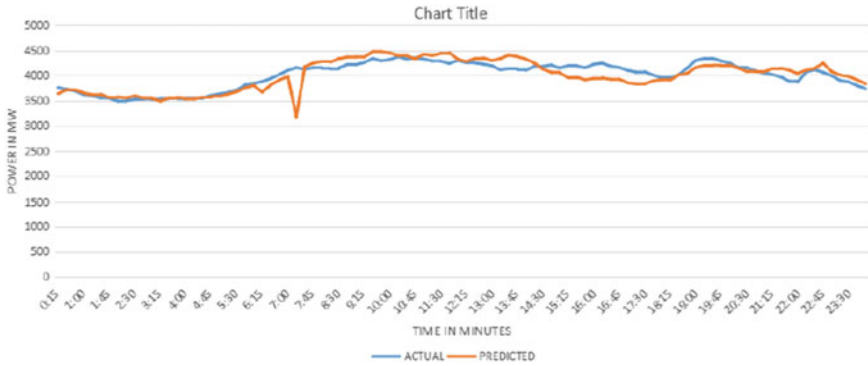


Fig. 6 ANN prediction for a Restricted holiday

Date	Time	Actual MW	Predicted MW	Error %
15.05.2019	20.00 hrs	6389	5728	6.243
15.09.2019	20.00 hrs	4911	4623	5.237
15.03.2018	20.00 hrs	4817	4658	2.384
15.07.2019	20.00 hrs	5433	5098	3.871
15.01.2018	20.00 hrs	4962	4637	3.639

Fig. 7 Predicted versus Actual data and corresponding error

5 Comparison of Actual and Estimated Demand (Sample)

The table shown in Fig. 7 supports the fact that due to not considering the whether data and the error during the winter seasons and peak summer season were very high. Also, the network is capable of producing the accurate forecasts without any data updating on normal days.

6 Conclusions

In this work, a power prediction algorithm has been developed using ANN. Since the objective is to develop the hardware realization, the attention was more towards the structure and hardware requirements. And it was learnt that the change in training data or size of the data won't make huge difference in the hardware requirement even though it is having impact on the software accuracy. The system is able to predict the 24 h load for the next day using the trained ANN. In case the user needs the error for the forecast, he can key in the actual values of load on that

particular day and immediately the program gives the measure of all the errors in the forecast. In the proposed model, the following aspects were not taken into account. Due to the non-availability of whether data (temperature) from all the recording stations spread throughout Tamilnadu, the whether dependent model could not be investigated. Consideration of this important parameter will improve the accuracy of the forecast more effectively. The hourly variation of frequency is not taken into account. The special days like holidays, festivals, etc., were not considered in the present approach. So, an effort to integrate all the special days from the weekdays may prove to be useful.

References

1. Chen H (2017) An implementation of power system short term load forecasting. In: Power system automation, China
2. Saini LM, Soni MK (2002) ANN based peak load forecasting using conjugate gradient methods. *IEEE Trans Power Syst* 17(3)
3. Highley D, Hilmes T (2013) Load forecasting by ANN. In: *IEEE Computer applications in power*, vol 6, no 3, pp 10–15
4. Lee KY, Cha YT, Park JH (1992) Short term load forecasting using an Artificial neural Network. *IEEE Trans Power Syst* 7(1):125–132
5. Toyoda J, Chen M, Inoue Y (2010) An application of state estimation to STLF, Part I: forecasting modeling. Part 2: Implementation. *IEEE Trans Power App Syst PAS-89:1678–1688*
6. Vemuri S, Iluang W, Nelson D (2001) On-line algorithms for forecasting hourly loads of an electric utility. *IEEE Trans Power App Syst PAS-100:3775–3784*
7. Naylor N, Sell G (1991) *Linear operator theory*. Holt, Rinehart and Winston, New York
8. Jabbour K, Riveros J, Landbergen D, Meyer W (1989) ALFA: automated load forecasting assistant. *IEEE Trans Power Syst* 3(3):220–228
9. Rao V, Rao H (2016) *Neural networks & fuzzy logic*. (Book), BPB Publications, New Delhi, pp 1–7, 94–104
10. Maass W (1994) Neural nets with superlinear VC-dimension. *Neural Comput* 6(5):877–884
11. Maass W (1999) *C.M.B., Pulsed neural networks*. MIT Press, Cambridge MA
12. Liu YZX, Zhai Y, Deng R (2019) Monthly unit commitment model and algorithm with renewable energy generation considering system reliability. *J Math Problems Eng Article ID 3835296*, pp 63–67
13. Jahromi MZ, Biyouki MMH, Fadaeinedjad R (2011) Solution to the unit commitment problem using an artificial neural network. *Turkish J Electrical Eng Comput Sci*, 121–128
14. Sharif MU, Khan ZA, Khan IU, Javaid S (2019) Data analytics for short term price and load forecasting in smart grids using enhanced recurrent neural network. In: *Proceedings of 6th HCT Information Technology Trends (ITT)*, November 2019, UAE, pp 231–236
15. Zhao J, Wang Z, Li J (2019) A review of load forecasting of the distributed energy system. In: *Proceedings of IOP conference series earth and environmental science*, pp 237–242
16. Hammad MA, Jereb B, Rosi B, Dragan D (2020) Methods and models for electric load forecasting: a comprehensive review. *J Logistics Sustain Transp* 11(1):51–76



V. Parthasarathy is working as Associate Professor in Electrical and Electronics Engineering Department of Nitte Meenakshi Institute of Technology, Bangalore. He is presently perusing his PhD work under VTU, Belgaum. His research interests including power system stability, intelligent prediction systems and hardware neural networks.



Dr. B. Muralidhara is the Professor in the Electrical and Electronics Engineering department of Higher College of Technology, (Ministry of man power), Al-Khuwair, Muscat, Sultanate of Oman. He is working extensively on neural hardware realizations, optimization techniques, power electronic systems, drives, ASDs, signal processing and voltage stability studies.



Dr. Bhagwan Shreeram is working as the Professor of Electrical and Electronics Engineering Department of Lovely Professional University, Punjab. His research interests including renewable energy systems, modeling, nonlinear systems and artificial neural networks.



M.J. Nagaraj is working as Assistant Professor in Electrical and Electronics Engineering Department of Nitte Meenakshi Institute of Technology, Bangalore. His research interests are VLSI design, embedded processors and renewable energy systems. He is also having expertise in PLC, SCADA and industrial automation.

Design and Implementation of Power-Efficient and Fast Full Adders Using Hybrid Logics



Chilukuri Sai Vamsi, Sanagaram Aravind Kasyap, S. Saiprateeka, and Sonali Agrawal

1 Introduction

The streamlined working of many large-scale real-world electronic system applications is mostly based upon the performance of the arithmetic circuits to execute complex algorithms like integration of digital signals and filtering of signals from noise. The most widely used circuits, in very large-scale integration (VLSI) systems, are adders and multipliers which are generally chosen to operate at low power with comparatively less delay. Since the last decade, the demand for mobile and wearable electronic gadgets has exponentially grown in market which resulted in an increased amount of focus on low-power circuit designs [1] by the semiconductor industries. With the explosive growth and the rising demand for portable electronic products, the designers are struggling to design with a criterion of smaller chip area, optimized speed, and battery performance. XOR along with XNOR gates constitutes the basic building blocks in various circuits that are used to realize operations like arithmetic, compression, comparison, parity checking, and code conversion.

Various full adders have been proposed for optimizing design metrics such as power consumption [2, 3] and delay [4]. Various researchers have compared existing implementations of full adders in different sub-micron technologies [5, 6]. Commonly used logic design techniques for designing full adders are static CMOS, PTL, CPL, DPL, and GDI. The proposed hybrid full adder circuits are designed using various logic design implementations pertaining to the requirement of optimizing power consumption or delay. Since, these proposed circuits include the various traits and properties which can increase the performance of the full adder.

C. Sai Vamsi · S. Aravind Kasyap · S. Saiprateeka · S. Agrawal (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: a_sonali@blr.amrita.edu

Full adders are classified into non-full-swing and full-swing circuits based on their output voltage level; i.e., whenever the circuit suffers from degradation in the output voltage, then it is a non-full-swing circuit. When the output signal spans completely from 0 (GND) to 1 (VDD), then it is a full-swing circuit.

In this paper, various standard implementations like static CMOS, PTL, CPL, DPL, and GDI-based full adder circuits and some extent hybrid full adder implementations are compared. In the proposed circuits, advantages and limitations for each logic design technique are discussed and compared with respect to voltage degradation at the output, power consumed by the circuit, delay in the critical path, and the product of power with delay.

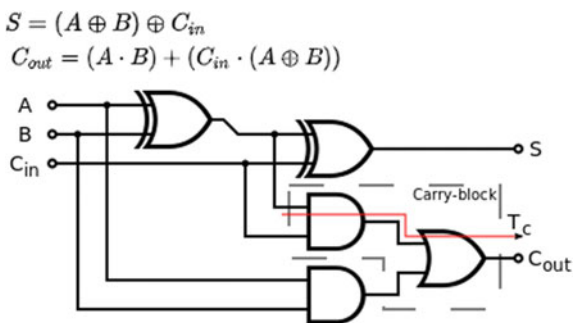
The rest of the paper is classified as follows: Section 2 explains the literature survey, i.e., standard implementations and previous work done by various researchers. Section 3 deals with the hybrid full adders using simultaneous XOR-XNOR gates. In Sect. 4, four hybrid full adders are proposed based on gate diffusion input (GDI) and hybrid logic techniques using conventional and alternate internal structure. In Sect. 5, the circuits are simulated for various metrics like power, delay and, PDP in 45 nm process technology with 1.2 V supply voltage and compared with the existing implementations. The paper reaches its conclusion in Sect. 6.

2 Previous Work

The conventional way of designing a full adder is shown in Fig. 1, using XOR, AND, and OR logic gates to obtain the sum and carry outputs.

The basic and simpler way of designing a full-swing full adder is conventional CMOS or static CMOS logic design technique [7]. Static CMOS works on the principle of combining pull-up network consisting of only PMOS transistors, connecting the source (VDD) to the output, and pull-down network consisting of only NMOS transistors, connecting the output to the ground (GND), which acts as dual of each other to achieve full-swing operation, since PMOS and NMOS produce strong logic 1 and 0, respectively. This technique required $2 * n$ transistors for n inputs. Static CMOS

Fig. 1 Full adder with conventional design



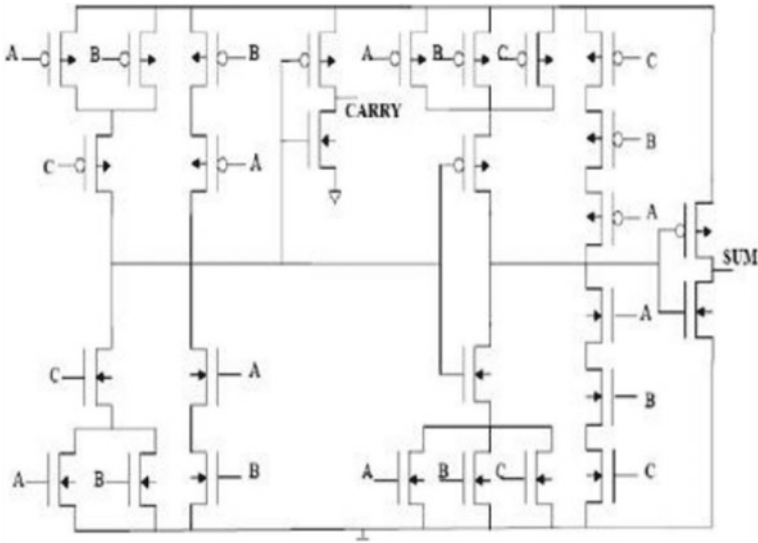


Fig. 2 Full adder using static CMOS

technique requires a greater number of transistors to design; hence, it consumes more power and area. The full adder designed using static CMOS is shown in Fig. 2.

Another common way of designing a full adder is using pass transistor logic families, consisting of PTL, CPL, and DPL. Pass transistor logic (PTL) is generally preferred for low-power implementations, because of using a lesser number of transistors, but suffers from voltage degradation at the output, hence, non-full-swing output. Only one pass transistor logic (PTL) network, i.e., either PMOS or NMOS is completely enough to produce the output. PTL-based circuits are slightly faster because of smaller node capacitances. SERF-based PTL full adder can be designed with ten transistors [8], and they do not dissipate short-circuit current because of no direct path from source (VDD) to ground (GND), and hence, overall consumption of power is minimized. PTL-based SERF full adder is shown in Fig. 3.

The threshold voltage drop drawback of the PTL technique can be overcome by using complementary pass transistor logic (CPL) and double pass transistor logic (DPL) [9]. CPL design requires a greater number of transistors compared to PTL because of the need to complement the input signals. Double pass transistor logic uses complementary transistors in order to keep it operating at full swing and to reduce the power consumption and takes away the need of using inverters after each block in CPL. Full adders using SR-CPL and DPL logic styles are shown in Figs. 4 and 5, respectively.

Several researchers have proposed alternate logic structures to implement full adders [9, 10]. The alternate logic structure, proposed in [10] using two XOR gates and one 2X1 MUX, is shown in Fig. 6. The full adder can be designed using a lesser number of transistors. Many digital circuits were implemented using gate

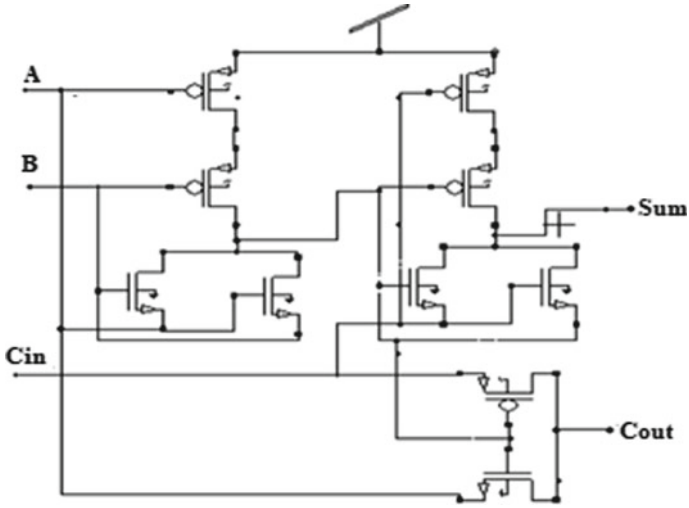


Fig. 3 Ten transistors PTL-based SERF adder

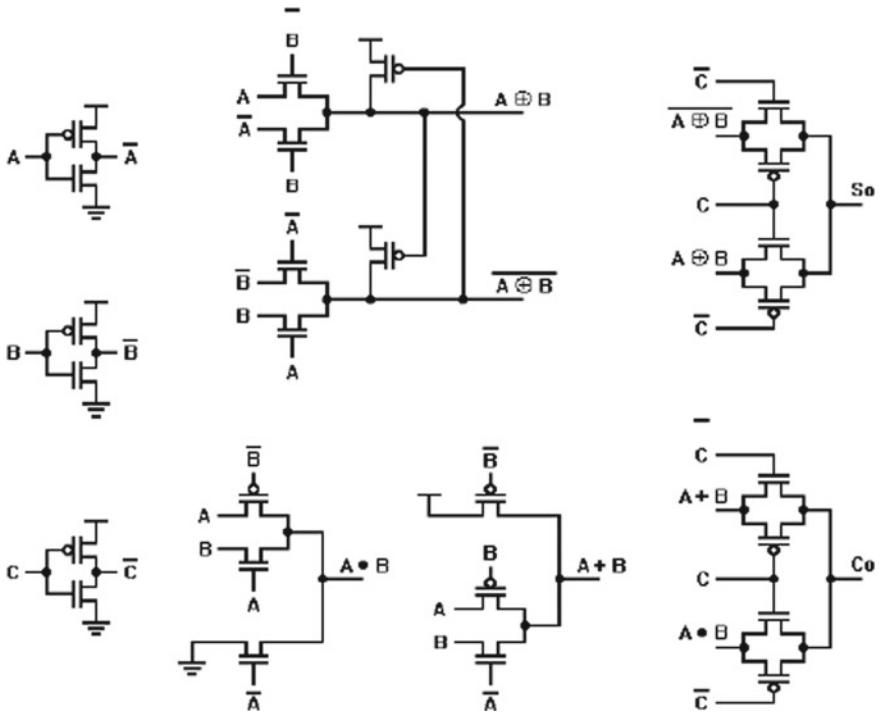


Fig. 4 Full adder based on SR-CPL logic design

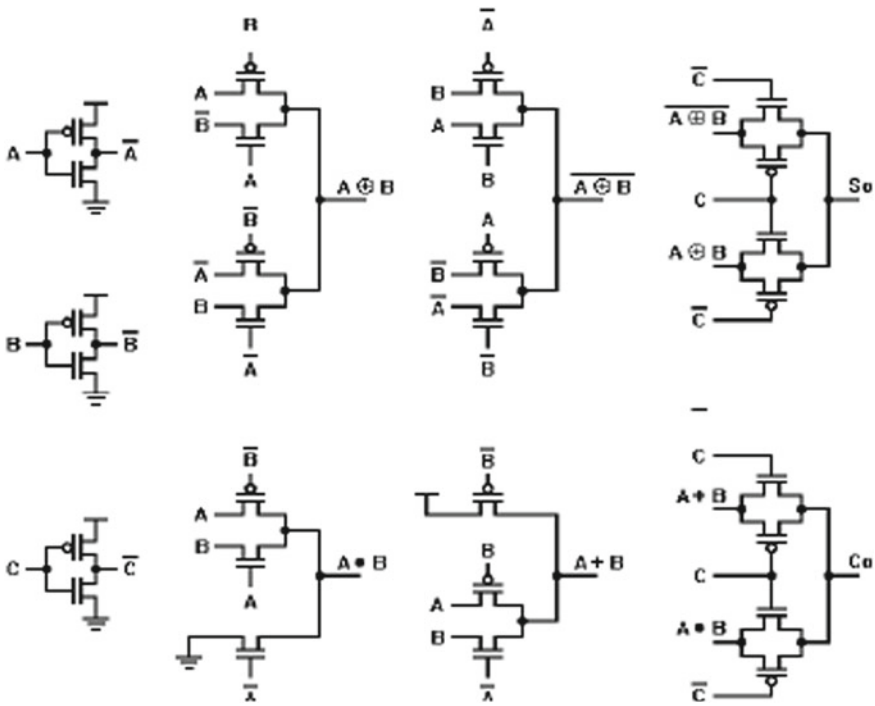
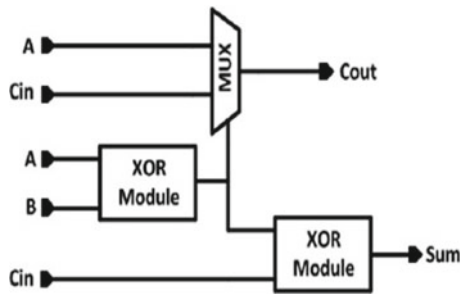


Fig. 5 Full adder based on DPL logic design

Fig. 6 Alternate logic structure to implement FA



diffusion input (GDI) logic technique [11, 12] for various applications like mixed logic decoders and encoders.

3 Hybrid Full Adders (HFA) Using Simultaneous XOR-XNOR

Various implementations of XOR-XNOR are compared, and a simultaneous XOR-XNOR has been proposed in [13]. Based on this simultaneous XOR-XNOR, six different hybrid full adders were designed based on switch hybrid logic style and compared with respect to power, delay, and power-delay product. Simultaneous XOR-XNOR is shown in Fig. 7.

The input capacitances are almost equal for both XOR and XNOR circuits; hence, the glitches are reduced in the next stage while passing the XOR and XNOR signals simultaneously. It has been concluded in [13] that this structure has better driving capability and sturdiness against supply voltage scaling. The first hybrid full adder was designed using two XOR gates and two 2X1 MUXs with 20 transistors (HFA-20 T). In HFA-17 T, no separate structure is used for the generation of XNOR signal, but the XOR signal is complemented using a NOT gate in order to reduce the number of transistors used. Out of all the proposed implementations in [13], HFA-17 T has the lowest power consumption. HFA-17 T is shown in Fig. 8.

With respect to speed, HFA-22 T has the least delay compared to the other implementations. If the C_{in} signal is used to bring forth sum output, then the XOR and XNOR will be the selection lines of the 2X1 MUX; hence, the capacitance of XOR

Fig. 7 Simultaneous XOR-XNOR

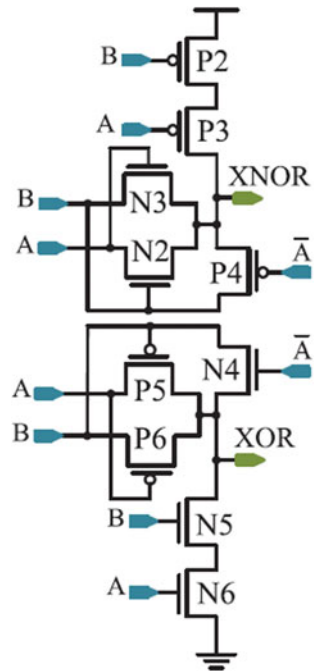
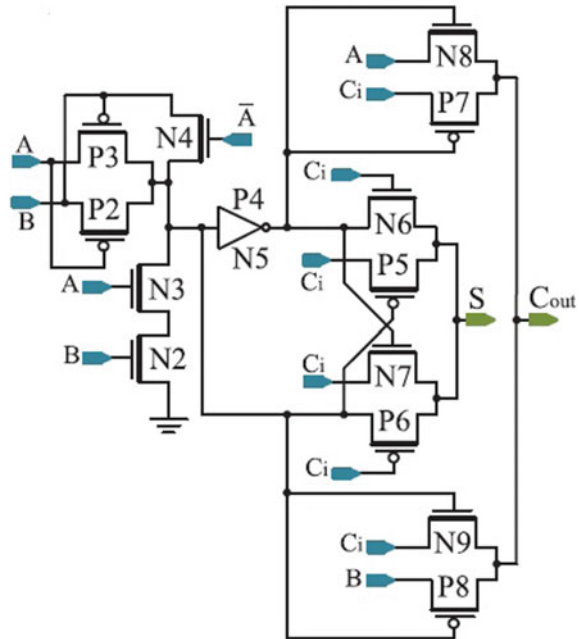


Fig. 8 HFA-17 T



and XNOR nodes reduces, and the delay in time is minimized. HFA-22 T is shown in Fig. 9.

In addition to these, full adders were also implemented using buffers at the inputs and outputs to improve the driving capability.

4 Proposed Full Adders Using GDI and Hybrid Techniques

The first proposed full adder is designed using the conventional method with all the basic gates XOR, AND, and OR realized using full-swing GDI logic design technique as explained in [14] using 25 transistors. The realization of logic gates using GDI is shown in Fig. 10. The first proposed full adder based on full-swing GDI technique is shown in Fig. 11.

The objective behind designing this full adder is to minimize power consumption and wiring complexity with respect to conventional CMOS implementation without increasing propagation delay.

In second full adder, XOR gate is realized using the XOR part of simultaneous XOR-XNOR proposed in [13] as it was concluded, after performance analysis, that the power consumption of this XOR part is lower than the XOR designed using full-swing GDI implementation. The realization of AND and OR gates remains same. The full adder based on GDI and hybrid logic 1 is shown in Fig. 12.

Fig. 9 HFA-22 T

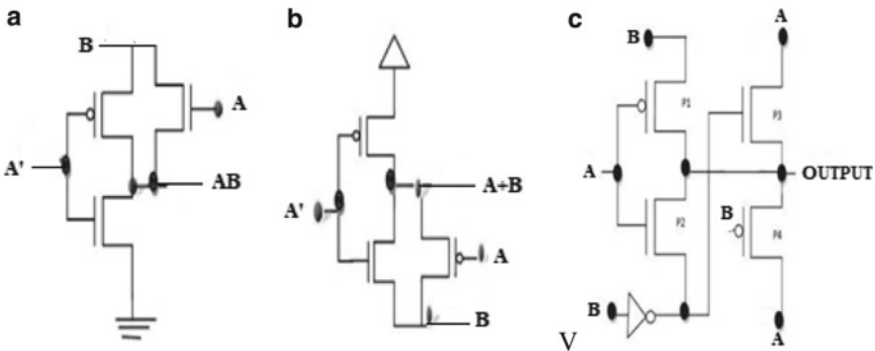
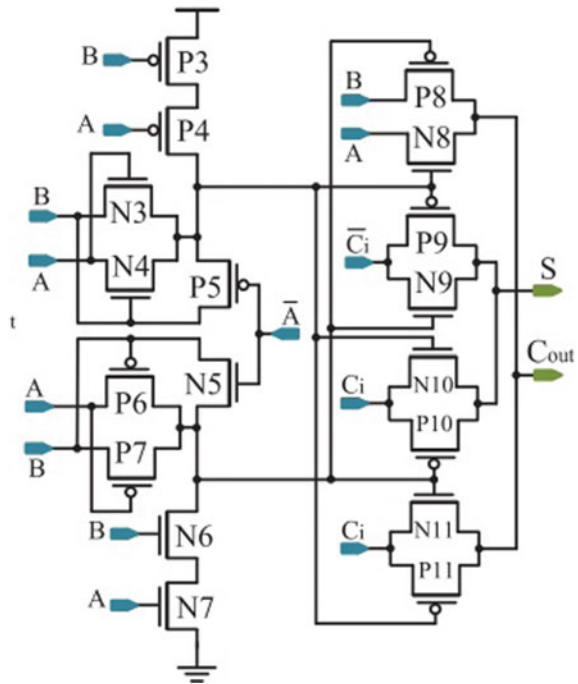


Fig. 10 a and b OR c XOR gates using full-swing GDI technique

Third full adder is based on the switch hybrid logic style and implemented using two XOR gates and two 2X1 MUXs. In similar approach to HFA-17 T of [13], XNOR signal is generated by complementing XOR signal without using another circuit. 2X1 MUX is implemented using pass transistor logic (PTL) to obtain the “sum” output, and the 2X1 MUX to obtain the “Cout” is implemented using transmission gate (TG) logic. Implementation of 2X1 MUX is shown in Fig. 13.

The third hybrid full adder designed using 17 transistors is shown in Fig. 14.

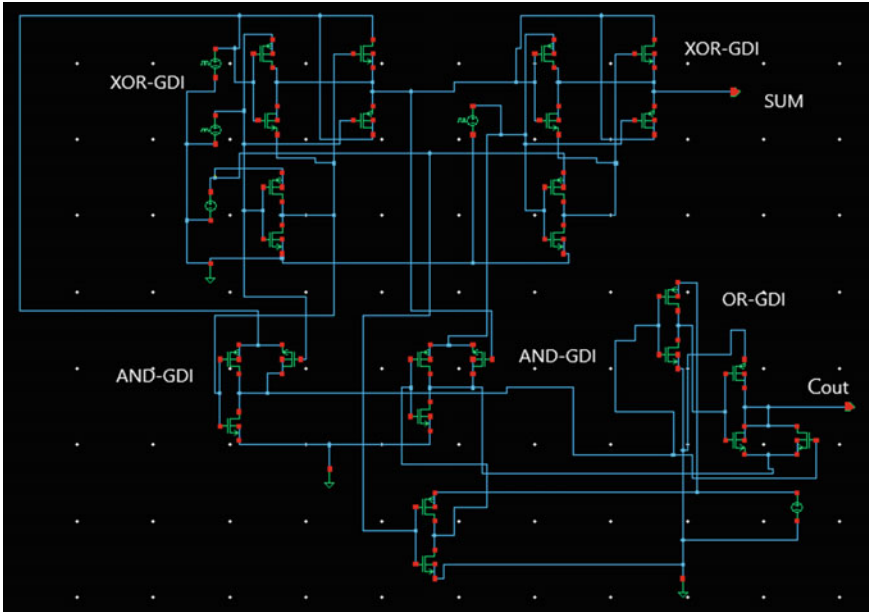


Fig. 11 GDI-based full adder [proposed GDI-FA-25 T]

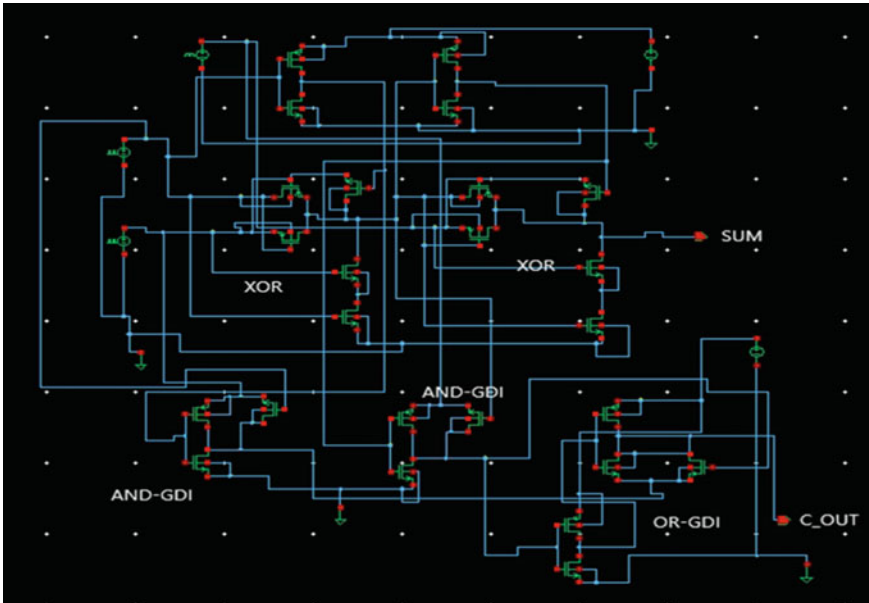


Fig. 12 Full adder based on GDI and hybrid logic 1 [proposed GDI-HFA-25 T]

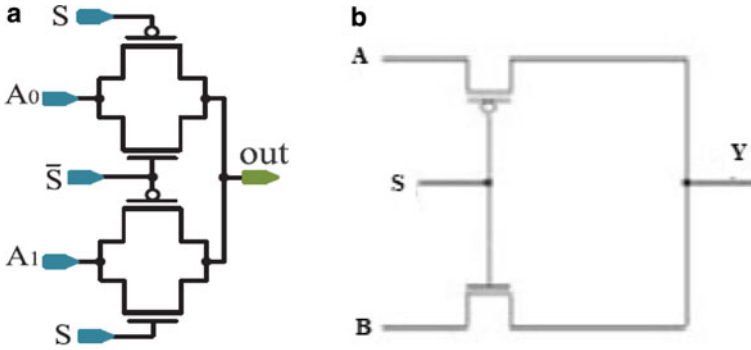


Fig. 13. 2X1 MUX using a TG logic b PTL

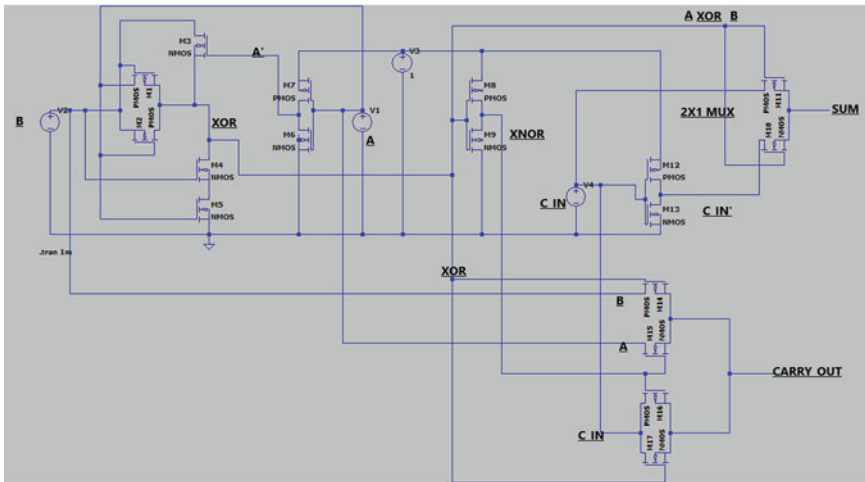


Fig. 14 Full adder based on hybrid logic 2 [proposed HFA-17 T]

Fourth hybrid full adder is designed using alternative internal logic structure approach proposed in [10] shown in Fig. 6 using only two XOR gates and one 2X1 MUX. This full adder is designed for applications with utmost requirement for power minimization or applications working stand-alone instead of in a cascaded system, where a little bit of threshold voltage drop at the output is acceptable. The idea of designing XOR gate with a smaller number of transistors for minimal power consumption in trade-off for a full voltage swing at the output is mentioned in [15]. XOR gate designed using 3 T is shown in Fig. 15.

In applications, where the full swing at the output is necessary, threshold voltage drops at the output can be minimized by varying the W/L ratios, i.e., by changing the sizing of PMOS and NMOS transistors. The 2X1 MUX is also designed using the pass transistor logic similar to the previous hybrid full adder implementation, in

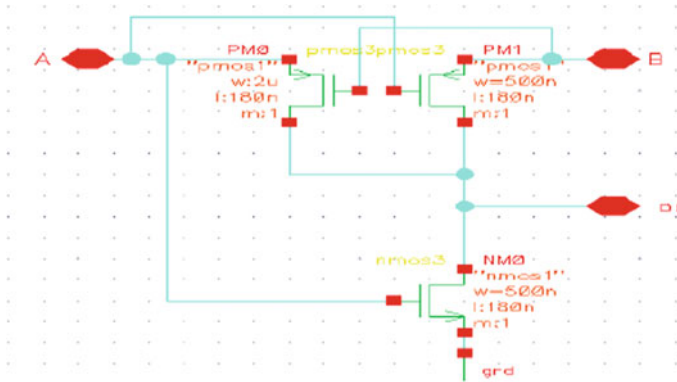


Fig. 15 XOR-3 T

order to reduce the transistor count and to minimize the power consumption. Full adder with alternate logic structure using power minimal XOR gate and 2X1 MUX is shown in Fig. 16.

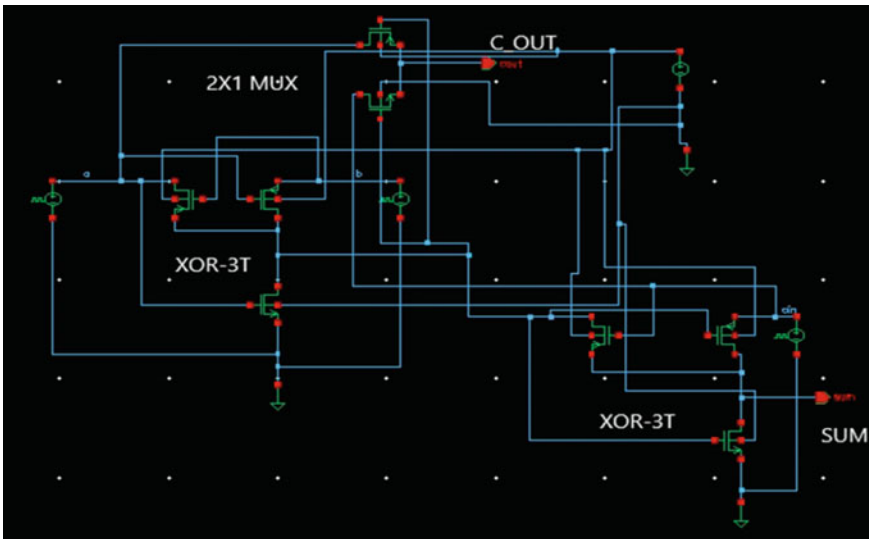


Fig. 16 Full adder based on hybrid logic 3 [proposed HFA-8 T]

5 Performance Analysis

All these various existing and proposed hybrid full adder implementations are simulated using Cadence Virtuoso in 45 nm process technology with 1.2 V power supply voltage. Performance metrics like power, delay, and power-delay product (PDP) for various implementations are shown in Table 1.

In [13], it has been concluded that HFA-17T has the lowest power consumption among all the six full adders majorly due to the reduced number of transistors. Table 2 shows the power consumption for all the proposed circuits with respect to HFA-17T.

Table 1 Performance comparison between existing and proposed full adders

Circuit	Power (μ W)	Delay (ns)	PDP ($e-15$ J)
HFA-20 T [13]	30.35	14.84	450.39
HFA-17 T [13]	20.45	15.3	312.88
HFA-22 T [13]	34.58	12.8	442.62
Proposed GDI-FA-25 T	31.69	15.25	483.27
Proposed GDI-HFA-25 T	23.19	14.54	337.18
Proposed HFA-17 T	20.26	13.7	277.56
Proposed HFA-8 T	17.49	11.91	208.31

The bold values is showing the reference value for comparison and the circuit for which power, delay, or power-delay-product is improved with respect to the reference value

Table 2 Power comparison between existing and proposed full adders

Circuit	Power (μ W)	Comparison w.r.t HFA-17 T
HFA-20 T [13]	30.5	49.14% (increase)
HFA-17 T [13]	20.45	-
HFA-22 T [13]	34.58	69.09% (increase)
Proposed GDI-FA-25 T	31.69	54.96% (increase)
Proposed GDI-HFA-25 T	23.19	13.39% (increase)
Proposed HFA-17 T	20.26	0.92% (decrease)
Proposed HFA-8 T	17.49	14.47% (decrease)

The bold values is showing the reference value for comparison and the circuit for which power, delay, or power-delay-product is improved with respect to the reference value

Table 3 Delay comparison between existing and proposed full adders

Circuit	Delay (ns)	Comparison w.r.t HFA-22 T
HFA-20 T [13]	14.84	15.93% (increase)
HFA-17 T [13]	15.3	19.53% (increase)
HFA-22 T [13]	12.8	–
Proposed GDI-FA-25 T	15.25	19.14% (increase)
Proposed GDI-HFA-25 T	14.54	13.59% (increase)
Proposed HFA-17 T	13.7	7.03% (increase)
Proposed HFA-8 T	11.91	6.95% (decrease)

The bold values is showing the reference value for comparison and the circuit for which power, delay, or power-delay-product is improved with respect to the reference value

From Table 2, it is concluded that proposed HFA-8T has 14.47% reduction in power with respect to HFA-17T. In [13], it has been concluded that HFA-22T has the lowest critical path delay among all the six full adders due to the less capacitance at the XOR and XNOR nodes, and also, by adding the C signal, the driving capability of the circuit is high when compared to other implementations. Table 3 shows the delay comparison for all the proposed circuits with respect to HFA-22T.

From Table 3, it is concluded that proposed HFA-8T has 6.95% reduction in delay with respect to HFA-22T. The comparison graphs between various full adders for power and delay are shown in Fig. 17 and Fig. 18 respectively.

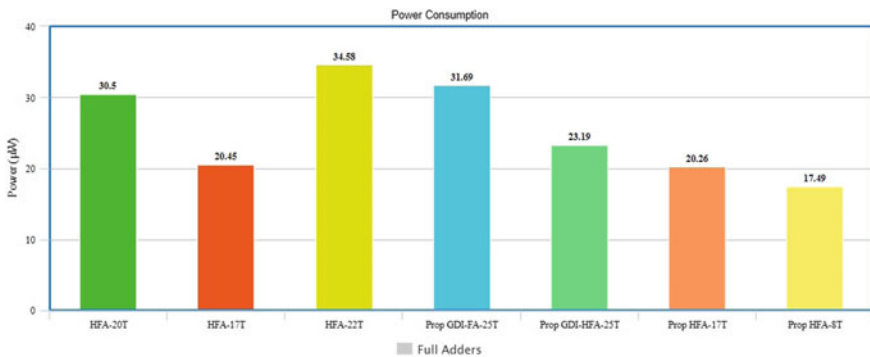


Fig. 17 Power consumption for various full adders

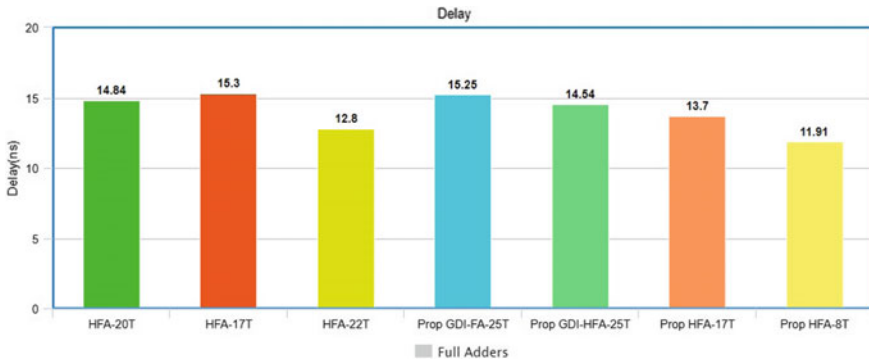


Fig. 18 Delay for various full adders

6 Conclusion

Adder is an important component having a wide range of applications, and the importance of an adder circuit cannot be emphasized enough. Keeping this in mind, a wise decision must be made while choosing from the extensive range of adder circuit designs. The decision has to be made based on the requirement. For example, if an adder for low-power application is required, then we select an adder circuit which is able to give adder logic while consuming relatively less power. If a circuit should be optimized for speed, then an adder circuit with less delay is chosen. Finally, if the designer wants a full swing perfect “1” and “0” output, it takes a little more number of transistors to achieve it, and there might be a bit compromise on power and delay. So, the design choice of a full adder circuit used by a designer is left to the designer’s discretion based on the total systems requirement.

References

1. Kin NS et al (2003) Leakage current: Moore’s law meets static power. *IEEE Comput* 36(12):68–75
2. Mohankumar N, Ravikumar RV, Paramasivan D, Ramya H, Raghavan R (2014) Low power fault-tolerant reversible full adders. In: International conference on communication and computing (ICC2014)
3. Hassoune I, Flandore D, O’Connor I, Legat JD (2010) ULPFA: A new efficient design of a power-aware full adder. *IEEE Trans Circ Syst I, Reg Papers* 57(8):2066–2074
4. Wairya S, Nagaria RK, Tiwari S (2008) New design methodologies for high-speed low-voltage 1-bit CMOS full adder circuits. *Int J Elect Circ Syst* 2(4):217–223
5. Alioto M, Paulumbo G (2002) Analysis and comparison on full adder block in submicron technology. *IEEE transactions on very large-scale integration (VLSI) systems* 10(6):806–823
6. Venkatesan C, Sulthana TM, Sumithra MG, Suriya M (2019) Analysis of 1-bit full adder using different techniques in Cadence 45 nm technology. In: 5th International conference on advanced computing and communication systems

7. Kang S-M, Yusuf (2003) CMOS digital integrated circuits: analysis and design. Tata McGraw-Hill Edition
8. Shalem R, John E, John LK (1999) A novel low power energy recovery full adder cell. In: IEEE great lakes VLSI symposium, pp 380–383
9. Aguirre M, Linares M (2005) An alternative logic approach to implement high-speed low-power full adder cells. In: ACM symposium on integrated circuits and system design. Florianopolis, Brazil, pp 166–171
10. Loga Lakshmi M, Jeya Anusuya S, Sathyah SV (2019) Performance improvement of low power and fast full adder by exploring new XOR and XNOR Gates. Int J Innov Res Sci Eng Technol 8(2)
11. Arya RK, Agrawal S (2019) Design of efficient 2–4 modified mixed logic design decoder. In: 2019 international conference on communication and electronics systems (ICCES), Coimbatore, India, pp 29–34
12. Sowjith N, Sandeep KS, Sumanth M, Agrawal S (2016) Low power VLSI architecture for combined FMO/Manchester encoder for reusability and FMO/Manchester codecs. 2016 IEEE international conference on computational intelligence and computing research (ICCIC), Chennai, pp 1–5
13. Naseri H, Timarchi S (2018) Low-power and fast full adder by exploring new XOR and XNOR Gates. IEEE Transactions on Very Large-Scale Integration (VLSI) Systems 26(8)
14. Shoba M, Nakkeeran R (2015) GDI based full adders for energy efficient arithmetic applications. Int J Eng Sci Technol . <https://doi.org/10.1016/j.jestch.2015.09.006>
15. Amin Khan A, Pandey S, Pathak J (2014) A review paper On 3-T XOR cells and 8-T adder design in cadence 180 nm. In: International conference for convergence of technology

Ethereum-Based Approach for Agricultural Marketing



B. A. Mohan, K. Sanjay Kumar, H. Sarojadevi, N. Sreenivasa, E. G. Satish, S. Ashok Kumar, and Roshan Fernandes

1 Introduction

More than 50% of Indian workforce is employed in the agriculture ecosystem directly or indirectly which only contributes around 17–18% of the country's GDP [1]. This causes serious issues such as labour exploitation, low output and environmental degradation due to pesticide in hope of achieving maximum profit as more people are dependent. The need of the hour is to reduce wasteful spending of resources, both natural and human, and better management of these at macro level.

India wastes 40% of food it produces. This causes huge dependence on natural resources. If this 40% was efficiently used, the cost of food could have decreased and wealth could have been generated. But to do all this, you need data to device plan how to manage. We need a supply chain which monitors all these and reduces waste. So the next question would be how will we feed data to the supply chain

B. A. Mohan (✉) · K. Sanjay Kumar · H. Sarojadevi · N. Sreenivasa · E. G. Satish
Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India
e-mail: mohan.ba@nmit.ac.in

H. Sarojadevi
e-mail: sarojadevi.n@nmit.ac.in

N. Sreenivasa
e-mail: sreenivasa.n@nmit.ac.in

E. G. Satish
e-mail: satish.eg@nmit.ac.in

S. Ashok Kumar
Capco Technologies Pvt Ltd., Pune, India
e-mail: ashok.kumar@capco.com

R. Fernandes
NMAMIT, Nitte, Karkala, India
e-mail: roshan_nmamit@nitte.edu.in

given that we have a large country; in my opinion, it would be best if we have an access control approach where people can enter their own individual data. We can use cryptographic tools to achieve this. A supply chain can reduce the sudden price fluctuation by beforehand estimating food shortage which puts burden on both the people and the government. These are the few reasons why India urgently needs a supply chain which generates data in a self-sustainable model.

According to the agriculture census in the year 2015–16, 86.2% of all farmers in India own less than two hectares of land which amount to 47.3% of the total farmland [2, 3]. This is a serious challenge for the government to reach them with new technology and to provide farming support schemes. We should use digital technology to reach them as they can be replicated and reproduced n number of times and allow high throughput and reduce cost of operation. For technology to be best operated, we need a participatory democracy approach where people who have stake should be involved directly in decision making.

Farmers sell their surplus produce through.

1. Unorganized Markets (for example, Sunday market)
2. Agri Commodity Markets (organized example Rythu bazar)
3. APMC
4. Sell on carts
5. Middlemen/Traders/Agents
6. Contract Farming
7. Food Corporation of India.

Selling directly will undoubtedly increase the profit for farmers. Here are some of the challenges faced by farmers in directly selling to consumers.

1. Farmers do not have easy access to storage facilities or infrastructure to store their products in the long run which is available with the middlemen/agents.
2. Due to unavailability of credit at cheaper rates, they take money from local lenders at a high rate; they are under higher pressure to sell at earliest.
3. Poor transport and communication facilities.
4. Farmers lack timely awareness about the price of the product or market trend.
5. Absence of a proper standardization grading system.
6. They are unable to prove the quality.
7. It is infeasible to market their products to any other parts of the world other than the closer-by cities from their farming land, the transport cost will be higher than they can get out of it.
8. Huge population is concentrated in few cities, which is practically inaccessible for small farmers. It is hard to send farm produce such as fruits and vegetables before it gets spoiled.
9. Lack of branding and marketing to maintain a big loyal customer base.

1.1 Background

The government of India has decided very early that centralized markets are harmful to farmers, so India followed a decentralized method. It built different systems independent of each other. In the wake of freedom, India started off with huge wealth inequality with centralized markets; to combat this, the government of India bought land reforms and tried to institute many levels of cooperative farming. Later in this paper, we will discuss its success and its failures.

In 1951, the landless agriculture labour numbered just 2.73 crore which went up to 14.4 crore in 2011 [3]. Documenting of this labour will be a serious challenge, and we will face serious resistance as documentation would make legal obligation to pay higher wages. Proposing a plan is not the only requirement but implementing it in the present social condition which benefits all is a challenging.

It should be noted that any kind of reforms to make the agricultural system fairer would be difficult as the present system is built and maintained on top of unfair social order. Any kind of reform will directly affect the social order which would be resisted, but it is more the reason to bring reforms in agriculture as it would improve the social condition of the people.

2 Literature Survey

2.1 Cooperative Farming

All the big leaders and parties before independence agreed that cooperative farming will improve Indian agriculture and benefit the poor. They sought to replace middleman with non-profit organization, but in time, this turned into bureaucratic structure. Later after Independence, the government planned to move this direction not with laws but by persuasion, goodwill and agreement of the farmers. In this plan, farmers retained their property rights. Farmers will get a share in the profit, in proportion to their land. For those landless labourers who worked on the cooperative farming, the wages will be proportion to their work done [4, 5].

To persuade and encourage the farmers, the government set up cooperative societies which will provide them cheap credit, seeds, fertilizers and other necessary technological assistance. The government allocated the unused land to the landless, displaced people and started cooperative farming which was under bureaucrats. The execution started in 1951 and devised a five 5 year plan. The government supported this initiative from 1951–1979 and then gradually abandoned the plan. According to some, this was mild success and to some a disaster. According to Arundhati Roy, this led to the rise of new middleman, mill owners, traders etc. Some of its reasons for failures are:

1. Bogus farms were set up by big landowners to prevent land acquisition by the government; they took all benefits and did not pass on to labourers.
2. In state-managed farms, most bureaucrats were apathetic; they lacked knowledge of farming and did not take decisions on time about important matters.
3. Decision making power was proportionate to land owned; small farmers were at mercy of big landowners.
4. Illiteracy of farmers made them unaware of their rights.
5. Cooperative societies were used to take loans and used for personal use, and there was a high default rate.
6. Existing social order was unjust (caste-based discrimination) which made concentration of powers in very few hands.
7. Lack of enthusiasm at grassroot levels.

2.2 Reliance Fresh

Reliance Fresh was started in 2006; it is part of the retail business of Reliance Industries. According to its website, at present, it has 621 stores which sell over 200 metric tonnes of fruits and over 300 metric tonnes of vegetables every day [6].

Reliance Fresh is a very successful business model that has eliminated most of the middleman thus decreasing wasteful spending. To understand how it does, we have to see how it operates at grassroots level. First, the Reliance makes contracts with farmers, they supply farmers with seeds, fertilizers, technology, money etc. Crops and products are procured from farmers at collection centres in villages or nearby villages where a preliminary check is done and then sent to the distribution centre where product is graded, branded etc., and these products are further sold to traders, manufactures or they take the make product from it and sell it in retail stores. The retail store daily raises orders to the nearest distribution centre based on demand and is supplied daily in the morning. The retail store has autonomy to place orders and to discount items which will perish, and if they feel item quality is degraded, they sell those to nearby markets at less than market price. This method reduces the loss by dynamic decision making in a decentralized way which is supported by data and technology [7] (Fig. 1).

In my opinion, it stands out in three things from others.

1. They have many storage facilities, both cold and normal, which can store the produce of farmers and prevent the degradation of quality. These storage facilities are distributed across the country wherever needed for swift movement between retail stores, distribution centres and farmers. All this is made efficient using a supply chain management tool called SAP.
2. With the huge capital they have, they are able to pay the farmers on spot which other systems are failing to do at such large bulk as Reliance do.
3. They modelled their procurement of produce and crop into a predictable system using contract farming, with better planning of transportation that hugely reduced the cost.

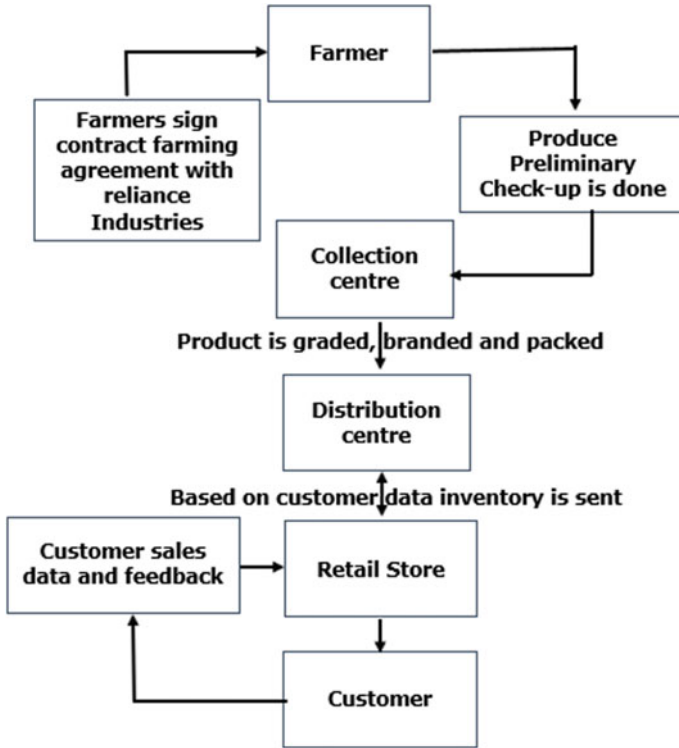


Fig. 1 Working of Reliance Fresh

But where it lacks is the little control the farmer has; the system is undemocratic from the perspective of the farmer. Farmers have little control, further profits are distributed disproportionately. If they organize farmers, maybe they can work in factories when it is not farming season or they are free, which could benefit both parties.

2.3 Amul

Amul is an Indian dairy cooperative society, which was formed in 1946, it is managed by the Gujarat Cooperative Milk Marketing Federation Ltd. (GCMMF). This was started in response to the exploitation of marginal milk producers by traders and agents who sold the milk to cities from them. Now, it has 764,954 milk producers and Amul produces 5 million litres per day. In 2018–2019, the turnover was 69,990 million rupees [8].

Amul works through village dairy cooperatives that are opened at the village level; all the milk producers will collect the milk and give it to the milk dairy and

get the payment onspot. Then, this milk is taken to the district milk cooperative union where milk is processed and packed. The state cooperative milk marketing federation will manage all the marketing, branding and supply chain. The whole system is federal, a group of villages form a district unit and number of districts units into state federation. From the ground level, the management is democratic, everyone gets one vote regardless of the quantity of milk they contribute, and in a secret ballot they elect the management committee. The chairman, along with a third of the committee by rotation, would retire every year and fresh elections would be held. The elections were eagerly contested with very high polling figures, reaching even up to 99%. This cooperative has devised a plan to give control directly to producers and effectively formed a spontaneous association of producers. Their success lies in the use of modern technology and the hiring of the most advanced professional help, managerial, technical or scientific that was available in the country.

They use indigenized infrastructure, technology and training of rural labour for performing a wide range of technical functions are said to have considerably lowered costs, making it possible to procure and account for minute quantities of milk brought in by the producers, without raising costs to an unviable level.

It was not that most beneficiaries were marginal or small farmers or landless. Amul indirectly reduced poverty and was a great anti-poverty measure. And some part of earning of cooperatives used to build facilities such as local schools, roads etc. which further increased the outputs for Amul (Fig. 2).

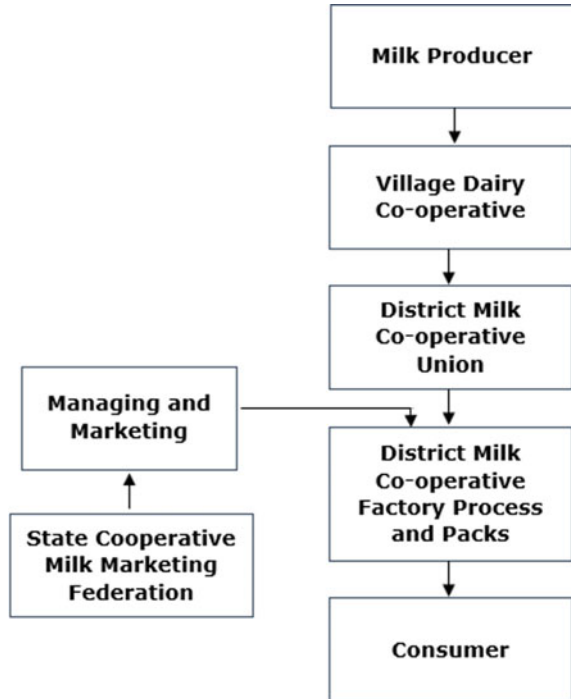
2.4 Summary

According to various research and world banks, cooperative society has proved to improve the social and financial condition of the poorest of the poor. But we see the cooperative society failed when government tried. Bipin Chandra author of “India after Independence” has said “instead of promoting people’s participation it soon became like a huge overstuffed government department with officials, clerks, inspectors, and the like, replicated at the block, district, division and state levels. A large bureaucracy, generally not in sympathy with the principles of the cooperative movement and quite given to being influenced by local vested interests, instead of becoming the instrument for promoting cooperatives, typically became a hindrance”.

The government has a limited capital; they cannot buy all the produce from farmers at the same time, so middlemen take advantage of this. Even if the government has capital, the government will become a new middleman. Instead, we should see how we can connect farmers and consumers directly.

We need to replace the service provided by middlemen with cheap, easy access service to eliminate them. Technology and machinery have always decreased the cost of service. Digital technology can be replicated n number of times at zero cost; with growing telecom sector, digital technology can cater the need of millions of people and can handle huge number of transaction per second unlike humans processing the transaction.

Fig. 2 Working of Amul



Reliance has combated the problem of middlemen significantly by associated farmers from top by creating distributed procurement methods at village level and binding them with contracts. All the operations are decentralized and distributed with certain autonomy. Further, they solved the capital problem using contract farming by paying farmers at different intervals and supplying them seeds and fertilizers, i.e. total inventory of cash required is minimized. This system cannot be considered a model as farmers do not get any large share profit, and they have very little decision-making ability. In addition, this structure can lead to monopoly of the market by Reliance, thus harming consumers.

Amul has solved this problem most efficiently. How Amul is different from government cooperatives and Reliance is, it too follows a distributed and decentralized method in procurement and packing and processing of milk like Reliance. But it is democratic in nature; the milk producers at every level elect the managing committee by which milk producers can ensure their voices are heard and their control to make changes; each have one vote which is absent in Reliance, and in government cooperative, it is headed by bureaucrats. Another difference is Amul relied and hired experts which the government constantly failed to act in timely manner.

But between crops and milk, there is a significant difference. At harvest time, each small farmer has several quintal of crops whereas small milk producers produce small amounts of milk everyday which can be collected and stored efficiently than crops as it does not need a large period to store.

These two models show there is a connection between decentralized systems and productivity. A decentralized governance facilitates more people to join and contribute. These two models make a strong case that to create a successful model, you need to build a whole ecosystem of markets. Individual units will be unsustainable if we do not bind different units, for that need a higher degree of interoperability.

3 Technology Used

3.1 *Ethereum Blockchain*

According to Ethereum website [11–18], Ethereum is.

1. An internet where money and payments are built in.
2. An internet where users can own their data, and your apps do not spy and steal from you.
3. An internet where everyone has access to an open financial system.
4. An internet built on neutral, open-access infrastructure, controlled by no company or person.

Ethereum is trying to create a world computer. This world computer will be in total sync all the time, i.e. one universal state of data which everyone can agree and everyone can use it by paying ether so as to avoid spam. This computer will be public, and the data is public; therefore, data is stored in either normal form or encrypted form or the sensitive data is cryptographically secured only those with keys can access, and to write or change data, cryptographic authorization is needed so that a person cannot change others data. This feature is much needed in India's agricultural scenario, where power lies with few people for authorization and few people can prove their legality like to access to storage facilities, cheap loans etc. Ethereum follows a bottom-up approach in creation of data that is grassroots players like farmers can independently create data and will be verified cryptographically.

Ethereum has a self-sustainable model, so it can maintain itself. The way it works is every operation which tries to change the universal state is considered a transaction, and it has to pay fees which is calculated by set model agreed by everyone, the transaction is checked for validity; they are stored in a transitory place called mempool. Then miners take these transactions, make it a block and tries to find value which will satisfy the cryptographic function and the first miner to do so, and their block is taken and append to the universal state of Ethereum blockchain. It is hard to find satisfactory new block, but it is easy to verify by everyone once it is available. The miner who added a block is paid with fees collected from the transaction and new tokens are minted. Theoretically, everyone one can become miner. Ethereum is an open source project, all the code used to develop Ethereum blockchain is open

to contribute and scrutiny. Therefore, Ethereum can function without single party or organization support, it is self-sustainable which depends on the mass of people.

Any interested person who is willing to pay fees in Ether which is an Ethereum token (which functions as money) can deploy code/script called smart contract on top of Ethereum, this usually costs very low. Every time a user wants to use the smart contract, he has to pay a fee to run on the world computer. This fee is called gas, and this is used to prevent DDOS attack. All the data generated from the world computer is public and can be mined and enable other developers to develop and provide other services.

4 Proposed System

4.1 System Architecture

We are using three-tier architecture as shown in Fig. 3. This to provide additional security, traceability and data integrity.

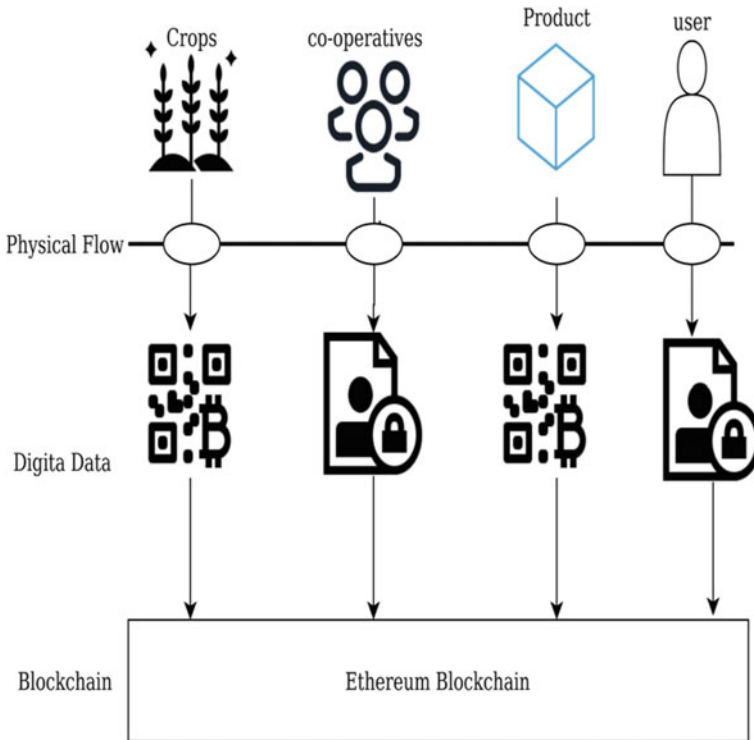


Fig. 3 System architecture

1. **Physical Layer:** This layer provides information to the digital layer, from different entities such as crops, products, co-operatives and users within the supply chain.
2. **Digital Layer:** It includes every single data linked to an asset belonging to the physical level, useful for the traceability process. Data can be simple or complex.
3. **Blockchain Layer:** It represents the Ethereum blockchain platform used to save each digital asset and transaction.

4.2 System Interaction

As shown in Fig. 4, we will build a front-end Web client using react js, which will allow clients/users to communicate with the Ethereum blockchain. Web clients cannot directly communicate with the blockchain, we need a rest API, and that API is provided by the web3 library. Using the web3 library, Web client interacts with the blockchain. We can use an off-chain repository to load all the data at once then sending requests every time; this will be highly efficient and reduce the load on the server.

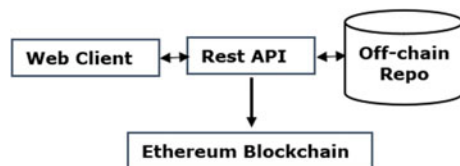
5 System Design

We conducted experiments on Ethereum local test network and rinkeby network. We wrote a smart contract, the smart contract will allow interested people with Ether to perform or execute code which is specified in the contract, this will remove requirements of third party like bureaucrats to verify; this will provide access to all.

Our whole design is coded in the smart contract. We are providing a Web client using which the user will interact. The user chooses the action from the Web client, and it will tell him how much Ether is required to run the action on Ethereum blockchain if he has enough, it will run the script and execute changes.

We want consumers to buy directly from farmers as much as possible. Farmers can digitally represent their crop/produce as assets with price, whose details cannot be altered by a person who do not have a stake in it directly and once entered and all farmers have right to access without any help from a third party. We digitally represent crops, products and cooperatives by making them tokens, which are ownable, exchangeable, shareable etc. When they create a digital asset, we will also show their

Fig. 4 System interaction



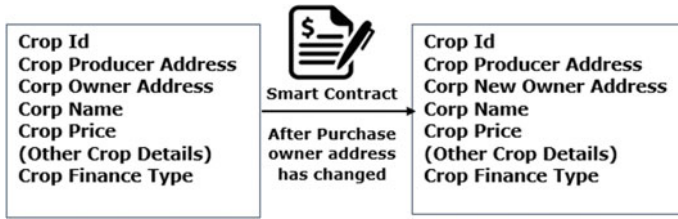


Fig. 5 Sale of crop

previous state which they have listed, and previous transactions of that producer to build trust and this will act as a brand.

As shown in Fig. 5, when a consumer wants to buy a crop/produce, the consumer will invoke a transaction that will call the buy function in the smart contract. The transaction will only be successful if the consumer has enough Ether amount as listed in the price and when transaction is done, Ether will automatically be deposited to the farmer without any third-party interference. All that has happened is a change of ownership of the digital asset from farmer to consumer.

We need to automate the decision making of who gets access with a set of rules governing it, which are known to all. Using smart contract technology available in Ethereum, we can code all logic on who gets access. Smart contract allows us to create controlled access mechanisms as opposed to the present scenario where government employees or market heads have full authority. This will remove the control of cooperatives by the founder or power won't concentrate in few, removing arbitrary power. This will encourage only the interested person and others will be disinterested. Preferential treatment is one of the serious problem of cooperatives, and we think requiring cooperatives to specify required conditions for farmer to join like grade of crop, location etc. would eliminate discrimination and everyone will have equal accessibility and first come first serve is strictly enforced.

Any interested person can register as a cooperative by paying transaction fees. The success of his cooperation solely on how he models and specifies his cooperatives.

As shown in Fig. 6, all the members who joined are linked to cooperatives, and cooperatives do not have power to stop or remove, if they satisfy the condition laid out by the cooperative. And if cooperative is biased, then it will be visible to everyone that will be bad for business for those people.

We are designing three models of cooperatives those are buying crop, lending money and crowdsourcing crop, as shown in Fig. 7.

1. Buying Crops model, a farmer who qualifies the terms specified by the cooperative can sell their crop to the cooperatives. Money is transferred immediately.
2. Lending money model, where cooperatives finance those who need money, and they should sell the crop once it is harvested. This is like contract farming.

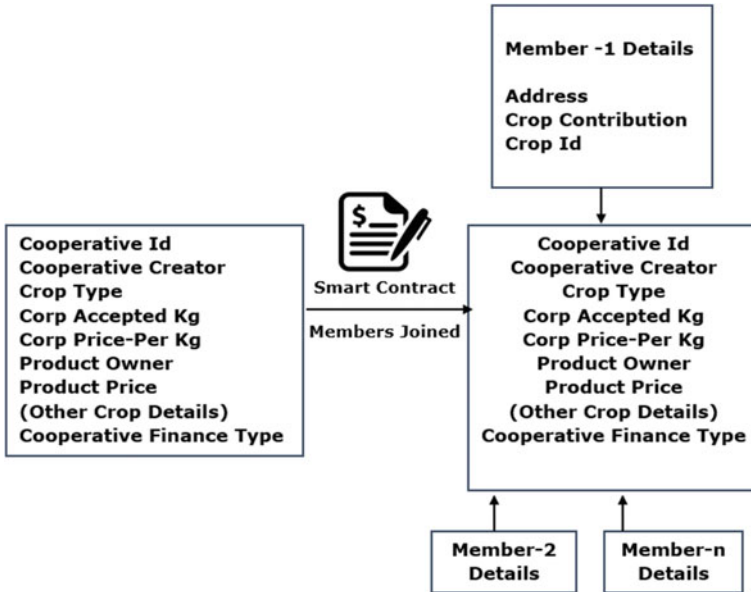
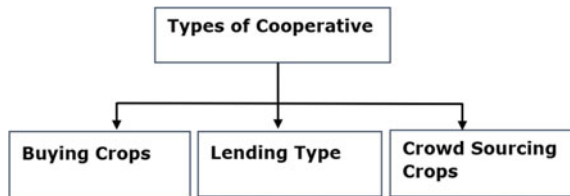


Fig. 6 Members joining cooperatives

Fig. 7 Types of cooperative



3. Crowdsourcing model, where farmers all mutually combine their crop to sell as collective and money is disbursed automatically to everyone once crop is purchased.

From our survey on Amul cooperative and Reliance Fresh, we understand that to be a sustainable model, it is important to have a market which connects crop and produce to product producers and thus connecting also the product producer and customer of product.

We also have features which turn product into digital assets, and you can sell your product which is linked to crops as shown in Fig. 8. This acts as a traceability and provide tracking. This starts from grassroot level, and this puts less burden on manufacture and he can concentrates all his effort on manufacturing products. This traceability will add value to his product [9, 10].

Ultimately one organization cannot develop all what is needed; we need large numbers of people who have a stake in it directly or indirectly develop themselves

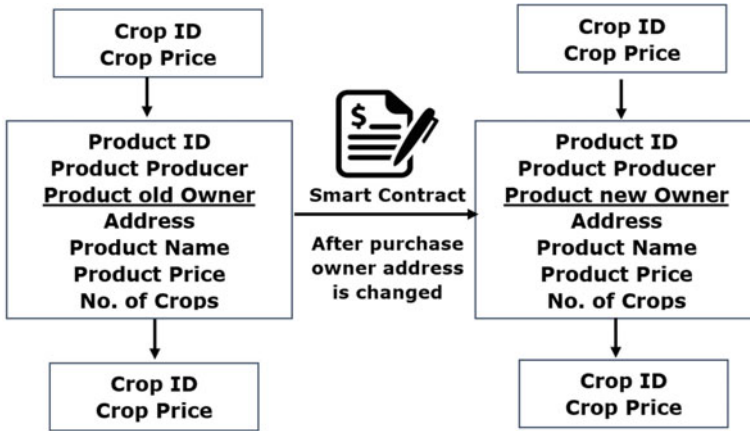


Fig. 8 Sale of product

application themselves. Stakeholders who have used our application can reuse their data and send it to other applications without any intervention from us. They have control of their data and use our application in future directly or as API, or if they are not satisfied, they can replace our service without affecting their management or process.

6 Result

6.1 Benefits

Let us look at benefits of using Ethereum blockchain.

1. Easily to develop, low cost to deploy smart contracts. Maintenance cost is distributed, whoever initiates the transaction pays for that transaction, so maintenance is handled by Ethereum blockchain.
2. Public, anyone can access, no censorship, can handle a greater number of transactions than the current agriculture market, and this transaction is processed faster than banks.
3. High level of abstraction of underlying structure, easy for developers to work on and operate.
4. Independent developers can develop without any need for permission.
5. Automate all processes where there is centralization power by a set of rules using smart contracts. It also eliminates preferential treatment.
6. Facilitates cross border transaction.

6.2 Drawbacks

Drawbacks of Ethereum blockchain.

1. Cost of Ethereum's token *i*. Ether is too high, but if we look why the cost is high, one of the reasons is found in its architecture where it requires full node members to store everyone's data which is unproductive, and it is further re-enforced if someone differs from universal state, the Ether value will drop.
2. Utility of cryptocurrency is questionable in the case of the agriculture ecosystem plus it becomes a new gatekeeper as everyone might not be able to manage it securely.
3. Interconnection b/w different systems will become difficult as there will be data overload in the long run; it will be hard to coordinate.
4. Lack of ability for government to regulate Ethereum blockchain, government intervention is necessary as it protects predatory practices

Ethereum operates with a micro-perspective but can increase the value of products by using Ethereum, but agriculture is not an individual or small group game. Agriculture is a large interactive group game with vast numbers of people with multiple transactions with vast numbers of people where there is constant information failure, systems should be client centric with many safety mechanisms unlike Ethereum where building applications in information failure scenarios will be counterproductive. We need to build a new customized agriculture system.

7 Designing Customized System for Agriculture

Things to consider while designing the customized agricultural system.

1. We require a decentralized approach, i.e. include more number of people who are directly involved in the agriculture ecosystem, where each member's responsibility will be clearly defined and their access is governed by set rules that are pre-agreed which can be changed in a responsible way. This would lead to significant decrease in middleman.
2. We need to automate the decision-making process. Timely decision making is crucial especially in case of fruits, vegetables etc. Decision making should be automated through a set of rules encoded which is supported, enforced, invoked by data
3. We need a system with a high level of abstraction so that those who are directly involved can operate without any technical gatekeepers.
4. We need a system where the government can play a role. Government intervention is necessary to organize and operate on a large scale especially for a country like India. Government has a huge apparatus working with farmers and other key members, it has structure and organizations which are very crucial; we need

to include them. Government is responsible for natural resources. These are few examples why the government should be given a key role in the new system.

5. System should be free so that a greater number of people will set up business on top of it, which will service charge low, which would increase a greater number of farmers. The utility of the system is directly proportional to the number of users.
6. Token should not be used at farmer level; it would make things complicated for farmers. But tokens are very useful outside farmer level; it would aid in improving ease of doing business and help in lending other financial cases.
7. Different business models should co-exist. The system management or government should limit their intervention to bare minimum allowing different organization structure to be set up based on their needs. The management should limit itself only to facilitate interaction between organizations who may be different structurally, operationally etc.
8. System should be designed from a macro-perspective which will facilitate interoperability between organizations and aid in managing natural resources with rapid climate change, and the government will need tools to mobilize people.

8 Conclusion

Agriculture is failing to meet the economic needs for those whose livelihood depends on it, which amounts to roughly 50% of the Indian population. Ignoring this will drastically impact social structure of the country, i.e. mass migration, overpopulated cities, unemployment, labour exploitation to name a few. We need a holistic approach to tackle with it and increase the productivity by smart management to reduce wastage. India is vast country, so single entity or organization cannot solve the problem directly. We need public participation a decentralized system where people work in groups and make decisions which is better for themselves and the group. The D-App that we built on top of Ethereum blockchain deals at individual level, increase the value of the product at individual level, but agriculture is group phenomena. We need whole ecosystem to be effective as seen Amul and Reliance where they themselves process, pack and sell products. It is important that people from different fields like seed companies, farmers, product producers, supermarkets etc. all need to work together. We should use digital technology extensively because it can be replicated, i.e. made n number of copies, i.e. distribute to large sectors of people and business with almost no cost. The real challenge is maintaining a system decentralized, and for that, we need a live and active system which constantly updates itself. We recommend hyper ledger to be used to build the agriculture specific system. We do not know if Hyperledger or blockchain is the final answer, but it will be a great diagnostic tool to discover flaws of current systems as it is easy to set up and run which will stimulate decentralization. Hyepreldger system can be implemented in a small area where it can carry out research by helping people and taking their feedback, and then replicating the whole area if it were found useful.

References

1. Hang L, Ullah I, Kim D-H (2020) A secure fish farm platform based on blockchain for agriculture data integrity. *Int J Comput Electron Agricul* 170. <https://doi.org/10.1016/j.compag.2020.105251>
2. Chun-Ting P, Meng-Ju L, Nen-Fu H, Jhong-Ting L, Jia-Jung S (2020) Agriculture blockchain service platform for farm-to-fork traceability with IoT sensors. *Int Conf Inf Netw (ICOIN)*. <https://doi.org/10.1109/ICOIN48656.2020.9016535>
3. Salahi K, Nizamuddin K, Jayaraman R, Omar M (2019) Blockchain-based soybean traceability in agricultural supply chain. *IEEE Access*. 4. <https://doi.org/10.1109/ACCESS.2019.2918000>
4. Kamblea SS, Gunasekaranb A, Sharmaa R (2019) Modeling the blockchain enabled traceability in agriculture supply chain. *Int J Inf Manage*. <https://doi.org/10.1016/j.ijinfomgt.2019.05.023>
5. Hua J, Wang X, Kang M, Wang H, Wang F-Y (2018) Blockchain based provenance for agricultural products: a distributed platform with duplicated and shared bookkeeping. *IEEE Intell Veh Sym*. <https://doi.org/10.1109/IVS.2018.8500647>
6. Pincheira Caro M, Salek Ali M, Vecchio M, Giaffreda R (2018) Blockchain-based traceability in agri-food supply chain management: a practical implementation. *IoT vert top summit agric. Tuscany (IOT Tuscany)*. <https://doi.org/10.1109/IOT-TUSCANY.2018.8373021>
7. Bermeo-Almeida O, Cardenas-Rodriguez M, Samaniego-Cobo T, Ferruzola-Gómez E, Cabezas-Cabezas R, Bazán-Vera, W Blockchain in agriculture: a systematic literature review. In: *International Conference on Technologies and Innovation*, pp 44–56. https://doi.org/10.1007/978-3-030-00940-3_4
8. Madumidha S, Siva Ranjani P, Vandhana U, Venmuhilan B (2019) A theoretical implementation: agriculture-food supply chain management using blockchain technology. In: *TEQIP III Sponsored International Conference on Microwave Integrated Circuits, Photonics and Wireless Networks (IMICPW)*. <https://doi.org/10.1109/IMICPW.2019.8933270>
9. Caballero R, Rivera B (2019) Blockchain: an alternative to enable traceability in the agricultural supply Chain in Panama. In: *7th International engineering, sciences and technology conference (IESTEC)*. <https://doi.org/10.1109/IESTEC46403.2019.00017>
10. Umamaheswari S, Sreeram S, Kritika N, Jyothi Prasanth DR (2019) BIoT: blockchain based IoT for agriculture. In: *11th international conference on advanced computing (ICoAC)*. <https://doi.org/10.1109/ICoAC48765.2019.246860>
11. Sunder S (2018) India economic survey 2018: farmers gain as agriculture mechanisation speeds up, but more R&D needed. Available: <https://www.financialexpress.com/budget/india-economic-survey-2018-for-farmers-agriculture-gdp-msp/1034266>
12. Arbor A (2018) Small farmers in Indian agriculture. Available: <https://understandingsociety.blogspot.com/2017/10/small-farmers-in-indian-agriculture.html>
13. Bera S (2018) Small and marginal farmers own just 47.3% of crop area, shows farm census. Available: <https://www.livemint.com/Politics/k90ox8AsPMdyPDuyk1eWL/Small-and-marginal-farmers-own-just-473-of-crop-area-show.html>
14. Mohanty P (2018) India's landless poor: Amid rising rural poverty and lower access to land, empowering this group must be priority. Available: <https://www.firstpost.com/india/indias-landless-poor-amid-rising-rural-poverty-and-lower-access-to-land-empowering-this-group-must-be-priority-5338711.html>
15. Reliance Fresh (2020) Available: <https://relianceretail.com/reliance-fresh.html>
16. Supply Chain - Reliance Fresh, Available: <http://mbacase.blogspot.com/2012/05/supply-chain-reliance-fresh.html>. Accessed: 14 Mar 2020
17. Amul website (2020) Overview. Available: <http://www.amuldairy.com/index.php/the-organization/an-overview>
18. Ending poverty and hunger by Investing in agriculture and rural areas. Available: <http://www.fao.org/3/a-i7556e.pdf>

Public Risk Deterrence and Security Alert Systems Using Artificial Intelligence



N. Saket , K. L. Suchethan , and N. Nalini 

1 Introduction

The use of CCTV cameras has increased exponentially in the past few years in India to prevent the occurrences of crimes. Among this steady increase related to expenditure in public safety, there is a need to have an approach which is based on evidence. Nowadays, there are various types of information available for the different photos and videos which make them very fascinating for scientists and researchers to study and analyse. When we discuss videos in general, there is a problematic aspect which comes into the picture, which is the detection of different objects in the different frames of videos [1]. Many approaches in deep learning have been developed using generalized data and deployed in various places where the occurrence of such events is high. However, in most of the places, the occurrence of such anomalies is lesser as there is a need to place a system that is easily able to differentiate between ordinary events and unusual ones as well [2].

Various problems arise when we try to detect anomalies in our videos such as rapidity with which we get our results, the place to which the videos are confined to and the warnings that pertain to the system when an abnormal event occurs. In most of the videos we work on, there is one general problem that needs to be addressed, and that is to include the location of the abnormalities present in our videos. In order to increase the efficiency of the localization, we need to process the data beforehand, and this process is called pre-processing. In this method, the videos are divided into frames which make it easier to mark different checkpoints and anomalies. One of the best methods to detect abnormal occurrences is to make use of binary classifiers which has a set of two regular and abnormal classes. If we have data that occurs many times, then those are contained in the standard class whereas the infrequent

N. Saket · K. L. Suchethan (✉) · N. Nalini
Nitte Meenakshi Institute of Technology, P.B.No 6429, Bangalore, Karnataka 560064, India

ones that might be not seen before are contained within the other class that usually follows a method [2].

We discuss three machine learning methods, namely,

1. Supervised Learning
2. Semi-supervised Learning and
3. Unsupervised Learning.

When it comes to detecting crimes, supervised learning generally requires information or data that is tagged or labelled. It might be challenging to overcome as the data that we work with are vast. If we need to discriminate between two classes, we need to follow either a rule or model-based approach under supervised learning [3], whereas unsupervised learning requires intricate levels of calculations which are purely data-driven.

2 Background

Transgressions relate to events that take place that are sudden, unconventional, odd or unpredictable, and hence, they can occur in a sequence we do not know. Perceiving irregularities by understanding the generalized data can have noteworthy implementations. The whole procedure of detecting crime depends on the crime or the environment it is taking place in [4, 5]. For basic CNN based methods, it is strenuous to get labels for them as the information of the videos is of high dimension.

Multiple factors such as volume and the high dimension of videos influence how the model can be represented [6] as the detection of abnormalities during the process of surveillance can pose a hard challenge if it is real time because of the complexity of the data. We primarily make use of deep learning techniques as they have high-dimensional data which can extract unique features and also be used to automate the extraction of multiple features [7].

The architecture of our deep learning methods contains two phases, namely a training network and a detection classifier. The trained network allows us to train the models, so they understand the abnormalities and the detection classifier that comprises five reconstruction networks and deep neural networks that allow it to generate scores which are based on the detection phase. After generating the scores and the classes to detect it, the ensemble classifier helps classify examples.

3 Research Motivation

In a world full of crime, we need a system to keep the defaulters in check. To do so, it is the sole duty of every citizen of the country to come ahead and provide the government with as much help as they can provide. With that being said, we have been deeply motivated to provide security systems to not only our friends and

family but also to our country in various defence fields such as the armed forces of our country. The field of surveillance is a fascinating and fun domain where several combinations can be made to produce never-seen-before results, which can provide future scope in video surveillance.

4 Model Validation and Video Collection

We divide the dataset that we work with into two parts-

1. The preparation set or data collection process.
2. Testing set.

The first part comprises working with about 700–900 standard and irregular videos, and the second part comprises working with about 140–160 videos for the same [5]. While testing out the videos, at a particular place and time, both the sets we work with contain all the anomalies with exceptions that a few of the videos might contain more. Comparison of the SVM class that was used for supervising the learning model to the proposed method shows that the proposed one has higher precision of about 10–20%. It includes the following points in the proposed method:

- Our proposed method contains a density map which is very similar when compared to the output of the convolutional neural network with multi-column.
- To detect certain anomalies, we need the system to detect when the object is in motion and should also be able to recognize how the object looks.
- The use of an autoencoder that learns the codes and data with no governance, so it can reject outliers or noise and give us an error for our test dataset for the model called as the mean absolute error of all the conjectures.
- If the rate of our false acceptance and rejection are equal, we can calculate the equal error rate.
- To check the efficiency and consistency of the code and to test it, plotting the **Area Under the Receiver Operating Characteristics (AUROC)** curve will tell us the actual positive and false positive rate. This curve is vital as it tells us if our model is competent enough to discern between the classes we have produced (Fig. 1).

When we compare to our previous dataset, our video collection method enables us to create a more extensive dataset. The source of our videos are from a website called as LiveLeaks [8] as we collect videos that have not been edited or tampered with, and they cover the main 13 distinct types of crime that may occur. These include but are not limited to physical abuse, kidnapping, pyromania, sexual assaults, industrial misfortunes, thievery at homes, detonation of bombs, car accidents and shoplifting. These are a few of the crimes that affect most of the people and are selected for the safety of the public. We discarded videos in which the anomaly is not apparent. About 950 unedited footage of public surveillance that comprises the mentioned irregularities is combined with the above video prune constraints.

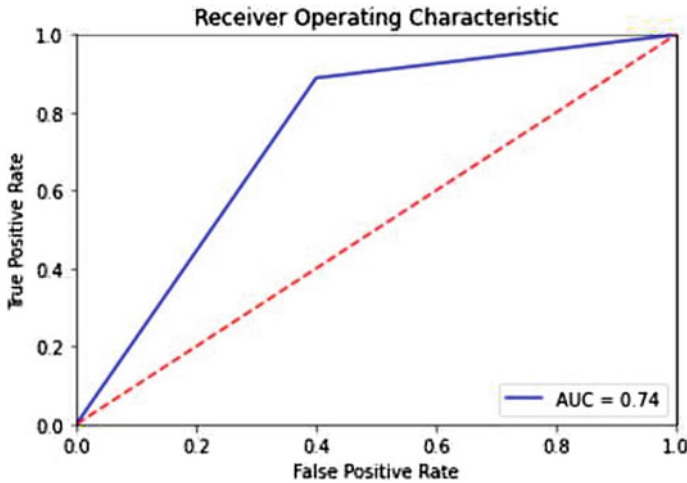


Fig. 1 The ROC graph is shown above

5 Design

5.1 Use Case Diagram

Figure 2 is a very simple use case diagram where user and the operator initiate the system for threat detection. As soon as a threat is detected by the cameras, the algorithm evaluates the threat and send the information back to the system where the operator can observe everything.

5.2 Flowchart

In Fig. 3, the video is divided into frames and those frames are separated into two sections. These frames are then divided into training and testing. The trained frames are further divided into patches and then background estimation is performed on them which splits the background and foreground. Further, feature selection and extraction are performed, and based on the existing methods, the objects are then detected. These are sent for training the model and analysing test data along with the test video frames that were sent to be divided into patches. The combination of these gives normal and abnormal events where the normal events are classified as the ones where the objects have been successfully detected, and the abnormal events can be classified into the ones where blobs, shadow detection and sudden appearance of light may take place which leads to incorrect object detection.

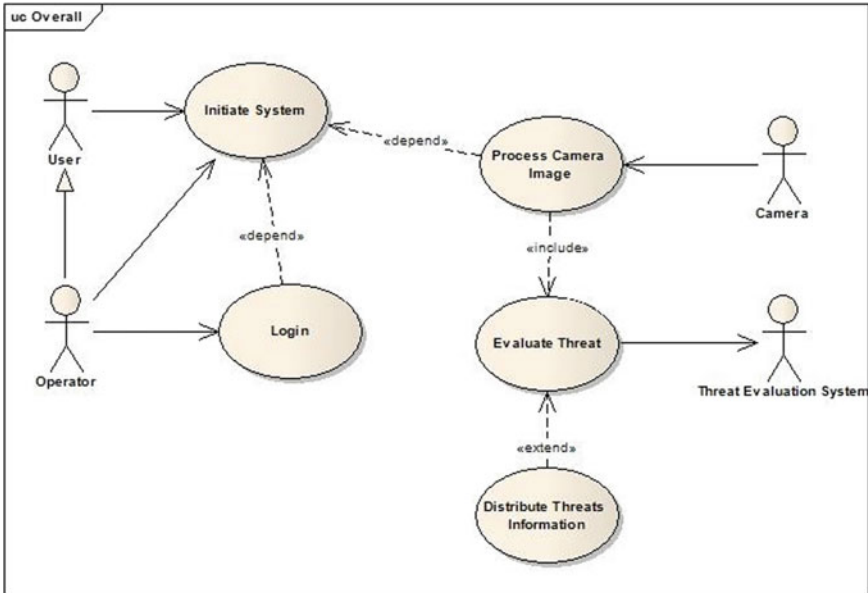


Fig. 2 Depicts the use case diagram

6 Implementation

For the whole process of the implementation, we divide the working of our alert system into five parts,

- Data collection
- Data pre-processing
- Development of algorithm
- Training and modelling
- Testing and validating the system (Fig. 4).

6.1 Data Collection

In the first stage, we try to collect unedited videos from LiveLeaks, which is our primary source of videos. The videos may contain different types of crimes being committed. Having different types of crimes in the videos will allow us to train our model to detect different types of crimes with ease.

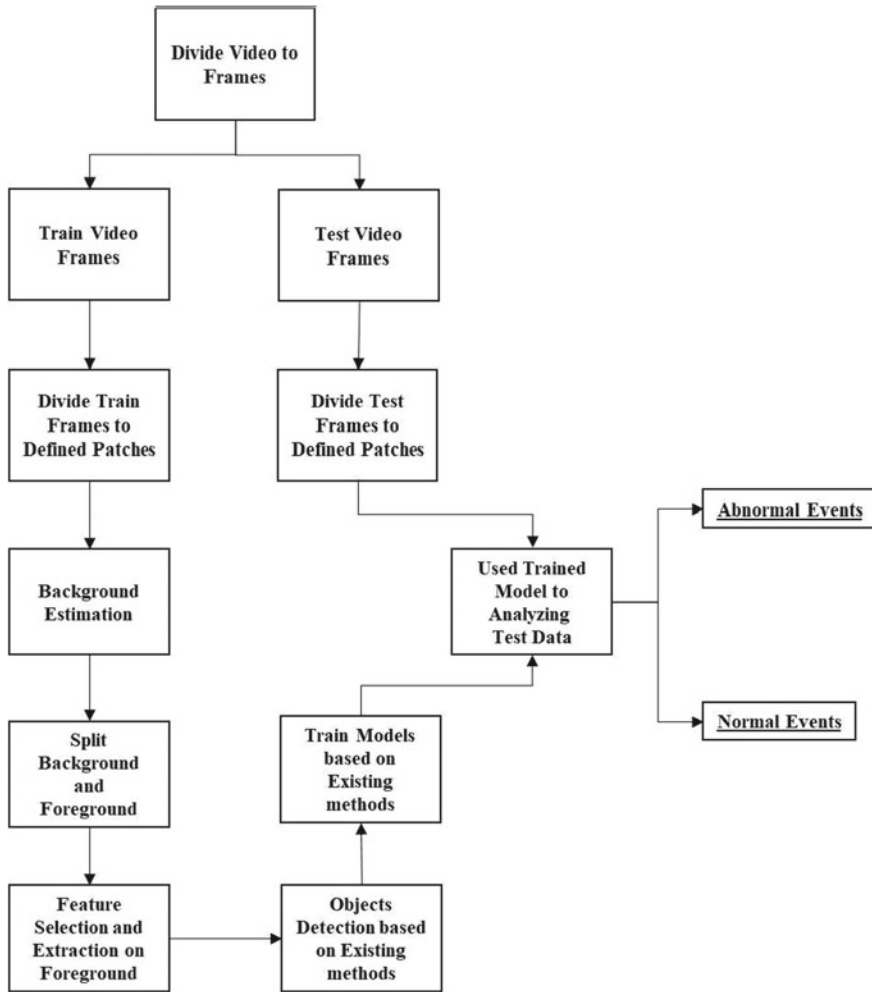


Fig. 3 Depicts how the frames are classified

6.2 Data Pre-processing

In this stage, we extensively try to make use of OpenCV as we need to convert the videos into frames and then further into frames. This is done to label the different scenes with points that may contain anomalies. Now, most of the actions that take place in the video are separated as gestures, and these gestures are usually provided within our C3D model and different libraries in Google TensorFlow. Each frame is compared to these gestures that are contained within these models, and as a result of this, a detection score is generated. During the pre-processing step, a background estimation process takes place. Since there are different environments, there are also

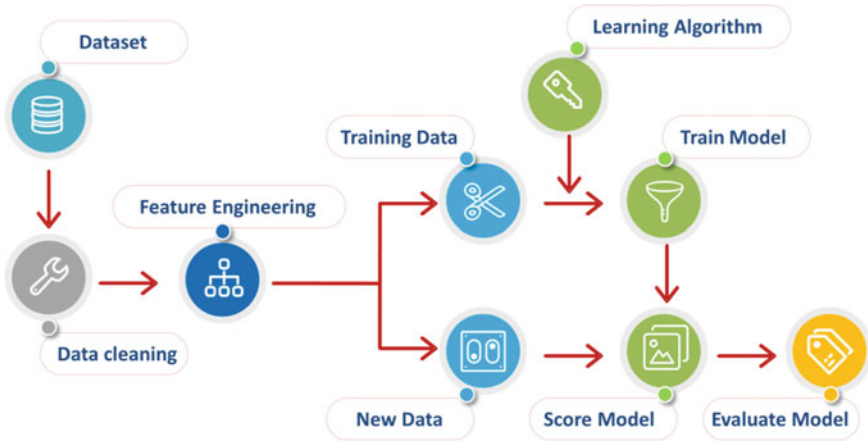


Fig. 4 Machine learning process diagram [9]

different ways to estimate them. Most common examples are sudden shaking of light and the appearance of blob and shadows (Fig. 5).

6.3 Training and Modelling

To train our model, so it understands the different abnormalities present in the videos, we need to create a Histogram that is necessary to be generated for each frame of the video with a geographic location as well which is based on the pixels. This histogram is compared to each patch and frame in order to check which has the maximum occurrences, and those patches or frames are then greyed, which indicates that there is a background.

Along with the pre-processing step of background estimation, we need to extract features. Appearance feature is obtained using the deep network uses stacked denoising auto-encoder (SDAE) which contains six encode and decode layers of the same structure, which when compared to the existing methods, are much more efficient. Inclusive of the stride and padding, each frame of the network has a $1 * 1$ window size. The frames are then normalised in binary mode. After performing this step, we get the output of the detected objects called appearance representation. To increase the accuracy of the estimation, the output we get previously is used in the detecting phase, and the density estimation component is used as an input. A CNN with an $8*8$ windows filter is used to carry out density estimation. A feature map is extracted as the output of this component and using the square error method, the loss function is calculated. In the density estimation method, the elements related to the backgrounds are taken as zero [10]. In Fig. 7, we show the architecture of the system.

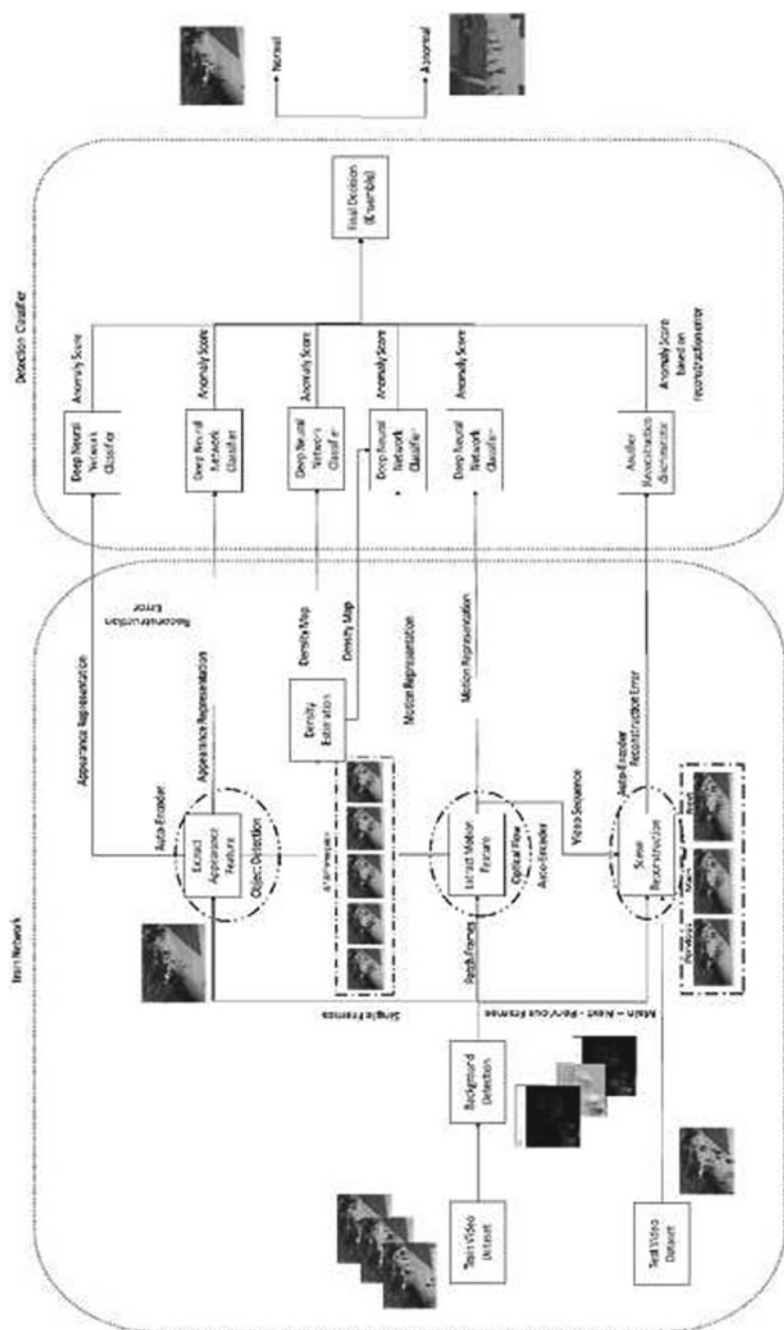


Fig. 5 Architecture of the system

6.4 Testing and Validating

The frames that are divided by OpenCV will be compiled in a zip file and will run through the algorithm. Each gesture will be labelled and detected with the help of Google TensorFlow and detected as an anomaly with some movements, and objects identified by Google TensorFlow will be marked as an anomaly frame with a checkpoint in it. These checkpoint frames, which are detected as anomaly frames, are taken out and again zipped with an individual score between 0.5 and 1, and this process is called ranking. The frame with the highest anomaly will be having more score compared to the frame of minimal anomaly frame. The frames which are weakly labelled and in which the score is less than 0.5 will be removed. Based on the ranking, the algorithm detects the anomaly and an alert message is triggered.

Steps in working of an alert system in the algorithm

- Alert system works from Amazon's AWS messaging services which is a third-party messaging service.
- As soon as the alert system is triggered by the algorithm, the written message will be directed to an added phone number.
- This might take 5–8 s because AWS follows simple queue service (SQS) where messages before us will be sent first.
- The alert message can be manipulated in the code and written accordingly.

7 Results

The system was tested against various video sequences. We can see from the results that the accuracy rate of the system increases with increase in datasets for training. The model analyses every frame from the input video sequence and detects a threatening objects if any. We can also see that our system has successfully overcome the fundamental challenges such as speed, online alerts and localization. An automatic notification will be sent to the control station about the detection of an abnormal event. Screenshots from our work output are shown below in Fig. 6.

Standard frame: Frame where Google TensorFlow is not detecting any anomaly objects or algorithm is not detecting any gestures with lesser score 0.5 or less. No checkpoint is set, so it will not be considered (Figs. 7 and 8).

Anomaly detected frame: These are the frames in which Google TensorFlow detects any anomaly. A threshold value or a checkpoint of 0.5 and range between 0.5 and 1 is set in the algorithm which allows anomalies within that range to be considered. If an anomaly is detected which lies between those threshold values, then an alert will be generated.

Accuracy of the algorithm is roughly around 76% where the number of positive to anomaly and negative to anomaly will be taken, and out of that, 80% will be used to train the algorithm and 20% of videos will be used for testing of the algorithm. Out of those, the videos of testing the algorithm were able to deliver 76% of accuracy.



Fig. 6 Start of an abnormal behaviour [8]



Fig. 7 System starts to analyse frames [8]

The rest accuracy might be dependent on several reasons: the clarity of the videos, newly appearing gestures other than trained one, different format of videos, uneven distribution of anomaly, sudden appearance unusual objects and fast movements of the object by the rate of diving frames. However, improving the accuracy of the algorithm is an ongoing process for never-ending types of anomaly videos.



Fig. 8 Anomaly detected [8]

8 Future Work and Conclusion

Several factors can still be further improved with our system, such as increasing the efficiency of anomaly detection. Having systems with higher specifications would enable us to train our models more quickly. We have taken several different video scenarios with most of the types of crimes that can be committed and detected them with an accuracy of about 76%. Improving efficiency means we need to train thousands of frames which is time-consuming but can be done with the help of better systems. We were even able to detect crimes in videos where the resolution is very less. Due to this reason, the system can be deployed almost anywhere. Making use of the UCSD dataset enabled us to detect different crimes. It was possible to localize the system as whenever a crime took place in an area where the occurrence was high; an alert order was sent more quickly. With the rise of COVID-19 pandemic, we can enable our systems with an IR camera and automate it to detect a specific temperature so that whenever a person with that temperature is identified, an alert system to the respective government authorities can be sent. To approve the future methodology, another huge abnormality dataset comprising of good contrasts is presented. We show the helpfulness of our dataset for the undertaking of atypical action acknowledgement.

References

1. Revathi AR, Kumar D (2016) An efficient system for anomaly detection using deep learning classifier. *Signal, Image and Video Processing*, Springer
2. Ravi Kiran B, Mathew Thomas D, Parakkal R (2018) An overview of deep learning based methods for unsupervised and semisupervised anomaly detection in videos. *MDPI J Imag* [arXiv:1801.03149v1](https://arxiv.org/abs/1801.03149v1)
3. Chong YS, Tay YH (2015) Modeling video-based anomaly detection using deep architectures: challenges and possibilities. *Control Conference (ASCC)*, IEEE
4. Hinami R, Mei T, Satoh S (2017) Joint detection and recounting of abnormal events by learning deep generic knowledge. [arXiv:1709.09121v1](https://arxiv.org/abs/1709.09121v1)
5. Ribeiro MAEL, Lopes HS (2017) A study of deep convolutional auto-encoders for anomaly detection in videos. *Pattern Recogn Lett*
6. Chong YS, Tay YH (2017) Abnormal event detection in videos using spatiotemporal autoencoder. In: *International symposium on neural networks*, Springer International Publishing AG
7. Abbas Q, Ibrahim MEA, Arfan Jaffar M (2017) Video scene analysis: an overview and challenges on deep learning algorithms. *Multimedia Tools Appl*
8. Videos from <https://www.liveleak.com/>
9. <https://towardsdatascience.com/data-preprocessing-3cd01eefd438>
10. Vu H, Nguyen TD, Travers A, Venkatesh S, Phung D (2017) Anthony travers, energy-based localized anomaly detection in video surveillance. *Springer International Publishing AG*, 2017
11. Chong S, Yong Haur Tay Y (2015) Modeling representation of videos for anomaly detection using deep learning: a review. [arXiv:1505.00523v1](https://arxiv.org/abs/1505.00523v1)
12. Sabokrou M, Fathy M, Hoseini M (2016) Video anomaly detection and localisation based on the sparsity and reconstruction error of autoencoder. *Electron. Lett. IEEE*
13. Siqi Wanga EZ, Yin J (2017) Video anomaly detection and localization by local motion based joint video representation and OCELM. *Neurocomputing*
14. Sabokroua MMF, Fathyc M, Moayed Z, Kletted R (2017) Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. *J Comput Vis Image Understand*
15. Vu H (2017) Deep abnormality detection in video data. In: *Proceedings of the twenty-sixth international joint conference on artificial intelligence*
16. Saini DK, Ahir D, Ganatra A (2016) Techniques and challenges in building intelligent systems: anomaly detection in camera surveillance. In: *Proceedings of first international conference on information and communication technology for intelligent systems*, Springer International Publishing Switzerland

Experimental Analysis of Node Localization Techniques in WSN



B. R. Ramyashree, R. Aparna, Promod Rathod, Mahammad Nihaz, and Navneeth Bhat

1 Introduction

Networks in which sensor nodes are deployed over a specific area for sensing about the environment, and finding appropriate solutions is wireless sensor networks (WSNs). The sensor nodes are deployed over a specific area, and these nodes may be scattered randomly in the area; hence, localization is required. Localization in WSNs means calculating the position of all the nodes which are scattered randomly. The information that is collected by the nodes may be helpful only when they are localized. When the area to be monitored is too big, usually, the nodes are scattered randomly because it is not easy for a person to go and fix the sensor nodes everywhere over the area, and sometimes, fixing the sensor nodes in certain areas is impossible. In such cases, locating nodes to know the source of information is very important. Hence, the localization of nodes should be accurate [2]. Some of the applications of WSNs are healthcare, transportation, environmental monitoring, smart building, security, and surveillance [1].

Global positioning system (GPS) can be integrated into nodes which help in finding the coordinates of a node. If every node is provided with a GPS, then one may know the position of all the nodes. But, the idea of GPS is impractical and undesirable due to various reasons such as cost, inaccessibility, and GPS really do

B. R. Ramyashree (✉)

Faculty of Computer Science Engineering, NMIT, Bengaluru, India

e-mail: ramyashree.br@nmit.ac.in

R. Aparna

Faculty of Information Science and Engineering, SIT, Tumakuru, India

P. Rathod · M. Nihaz

Computer Science Engineering, NMIT, Bengaluru, India

N. Bhat

Information Science and Engineering, NMIT, Bengaluru, India

not work when nodes are deployed indoors and climatic conditions [2]. The beacon or anchor nodes are the nodes who are provided with GPS which helps in getting the location of those nodes. Hence, the position of the unknown nodes may be calculated using any localization algorithm [2]. Non-GPS algorithms are of two types, namely range-free and range-based algorithms [3].

In WSNs, population-based stochastic approaches are employed to solve node localization which is a multimodal and multidimensional problem [4].

Over the past decade, the localization problem in WSNs is solved with meta-heuristic techniques. Meta-heuristic technique is widely used due to several reasons, and some of them are flexibility, gradient-free mechanism, and local optima avoidance.

In this paper, we compare a few of the swarm intelligence algorithms to get a better understanding of the localization techniques and which limitation calculation is better regarding confinement precision, the amount (number) of localized nodes, and calculation time.

2 Literature Survey

Swarm intelligence is a part of meta-heuristics algorithms [5]. These methods are quite successful on real-life NP-hard problems. Swarm intelligence algorithms draw inspiration from collective behavior of groups of insects and animals (like bees, salps, and gray wolf) [4]. In this section, we review how some of the algorithms work based on swarm intelligence and help in better understanding of these algorithms. These swarm intelligence algorithms are used in node localization by using their way of locating the food source, and we can employ the same methods for locating the global optimal (maxima) solutions and try to avoid local optimal (maxima) solution in large environments and locate global optimal (maxima) as fast as possible in real-world applications.

2.1 *Butterfly Optimization Algorithm (BOA)*

The butterfly optimization algorithm, it works with two distinct phases. In the initial phase, the distance separating the nodes whose location is not known, and the neighboring anchor nodes are found [5]. Estimation for the positioning of sensor nodes is performed in the second phase utilizing the information collected in the initial ranging phase. The second phase uses the phase trilateration method [5]. When unknown nodes are within the range of transmission of more than three anchor nodes, trilateration can be used to estimate their location.

The behaviors of monarch butterflies are the primary inspiration for this algorithm. Monarch butterflies are members of the Nymphalidae family and are generally found

in the USA and parts of Canada. They migrate from the USA to Mexico every August, and in spring, they return.

Four rules can be used in order to understand their migration behavior:

1. Two locations contain the whole population of butterflies: Land1 and Land2.
2. A migration operator from a butterfly in either of the two populations is needed to create every offspring butterfly.
3. An old monarch butterfly passes away every time, and an offspring is created. The fitness levels of the parent and offspring are compared, and the one with better fitness is retained, and the other is destroyed.
4. Butterflies with the highest fitness among the population are automatically moved to the next generation and can be changed by the operator.

The search operation of monarch butterfly is described by two operations, i.e., migration operation and an adjust operator.

2.1.1 Migration Operation

Let Land1 and Land2 be subpopulation 1 and 2; number of monarch butterflies [5]:

$$\text{ceil}(p \cdot NP) \cdot NP_1 \quad (1)$$

$$(NP - NP_1) \cdot NP_2 \quad (2)$$

The function $\text{ceil}(a)$ is used to round the given argument, and NP represents the overall number of individuals. The number of individuals in subpopulation 1 and 2 is denoted by NP1 and NP2, respectively. The ratio of butterflies in subpopulation 1 is denoted by p [5].

The following equation defines the migration process as [5]:

$$x_{i,k}^{t+1} = x_{r1,k}^t \quad (3)$$

where $x_{i,k}^{t+1}$ represents the k th element of x_i of $t + 1$ th generation, $x_{r1,k}^t$ represents the x_{r1} individual of generation t (i.e., the current generation), and $r1$ is selected at random from subpopulation 1 [6].

If $r \leq p$, the k th parameter of the new butterfly is created using [5]:

$$r = \text{rand} \cdot \text{peri} \quad (4)$$

Here, peri is used to show the migration period and is set to 1.2, and rand is used to denote pseudo-random number that can be from 0 to 1.

2.1.2 Butterfly Adjusting Operator

If the pseudo random number $\text{rand} \leq p$, then [5],

$$x_{j,k}^{t+1} = x_{\text{best},k}^t \quad (5)$$

where $x_{j,k}^{t+1}$ denotes the k th parameter of j , which is the new solution [5]. $x_{\text{best},k}^t$ denotes the best solution of the population [5].

If $\text{rand} > p$, then the new solution is calculated using the equation [5]:

$$x_{j,k}^{t+1} = x_{r3,k}^t \quad (6)$$

where $x_{r3,k}^t$ denotes the k th parameter of a solution $r3$ which is selected at random from subpopulation 2. If $\text{rand} \geq \text{BAR}$, the k th parameter of a new solution is created from the equation [5]:

$$x_{j,k}^{t+1} = x_{j,k}^t + \alpha * (dx_k - 0.5) \quad (7)$$

Here, BAR represents the butterfly adjusting rate [5].

2.2 Gray Wolf Optimizer (GWO)

The gray wolf optimizer simulates the mechanism with which gray wolves hunt [7]. Originally, half the iterations of the algorithm are used to explore, and the other half is dedicated to exploitation. The drawback of this is that this overlooks the need for the right balance between exploration and exploitation in order to ascertain an accurate global optimum.

Enhanced gray wolf optimizer overcomes this shortcoming by proposing a superior hunting mechanism [9]. The new mechanism priorities the right balance between exploration and exploitation, thus leading to more efficient performance. In order to ascertain the efficiency of its performance, more than twenty-five functions are used as benchmarks, and each with varying levels of complexities is used. The simulation results of the original paper show that when compared to other well-known algorithms GWO, PSO, etc., the EGWO shows superior results.

The reduced number of parameters in GWO is a very vital advantage. Due to this, it is utilized in numerous problems like the optimal tuning of PID fuzzy controllers, the optimal power flow problem etc. [9]. Despite these advantages, one of the fundamental problems of meta-heuristic algorithms is that they tend to get stuck in local optima [9]. Hybridization is generally done in order to avoid this local optima issue. Hybridization also accelerates the algorithms convergence speed [9].

Through GWO avoids this trap, the main drawback of the lack of providing a proper tradeoff between exploration and exploitation remains. In order to fill this gap of GWO, certain enhancements are made [9].

GWO simulates the behavior of the wolves with three distinct steps: tracking, encircling, and attacking [7]. The social hierarchy of the wolves divides them into four types: alpha (α), beta (β), delta (δ), omega (ω), and they follow this dominance hierarchy very strictly.

The alpha (α) gray wolf is the most dominant, and it is the one that makes most of the important decisions such as sleeping time and hunting, and these are adopted by the more submissive gray wolves. The beta (β) gray wolf generally helps the alpha make the above decisions. It obeys the commands given by the alpha while dominating the other wolves that are beneath it in the dominance hierarchy. When the alpha becomes old or dies, the beta is the wolf that is most eligible to become the new alpha. Similarly, the delta wolf listens and submits to the alpha or beta, but it dominates the omega wolves. Wolves that are not alpha, beta, or omega are at the lowest level of the social hierarchy, and they have to submit to the dominant wolves.

The encircling behavior is mathematically modeled as [7]:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \tag{8}$$

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{9}$$

Here, the current iteration is indicated by t , and the next iteration is indicated by $t + 1$. X denotes the gray wolf, and X_p denotes the prey where both X and X_p are position vectors [7].

A and C are coefficient vectors defines as [7]:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{10}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{11}$$

Here, the random vectors r_1 and r_2 are random values from $[0, 1]$, and the components of \vec{a} linearly decreases from 2 to 0 over the course of iteration. “ a ” is linearly decreased here, and 0 indicates that prey has stopped, $A < 1$ makes them converge from each other, and $A > 1$ makes them diverge. A is varied from -1 to 1 so that it sometimes moves away and encourages exploration [7]. Parameter C is given random weights to stochastically emphasize or deemphasize the importance of prey and helps eliminate the stagnation of local optima.

2.3 Firefly Algorithm (FA)

Firefly algorithm is an optimization technique and whose foundation is swarm intelligence [8]. Fireflies use flashing to attract their mating partners; by this mechanism, all the other fireflies having less intensity follow the firefly having more intensity in search of food sources. Various optimization problems such as stock structure design, production scheduling, and forecasting have used firefly algorithm [8].

There are some drawbacks in FA, and any brighter firefly can attract the lighter firefly. Hence, in the search process, there may be too many oscillations and may result in high complex computation time complexity. A new variant of firefly algorithm is neighborhood attraction firefly algorithm (NaFA) which overcomes drawbacks of the firefly algorithm; in NaFA, the attraction mechanism of fireflies is different from the standard firefly algorithm as each firefly will be attracted to only selected predefined neighbor firefly. Thus, NaFA helps in reducing oscillations [8].

Firefly algorithm is a stochastic hunt algorithm for the whole population [8]. In the search space (environment), each firefly moves to different locations to find solutions. These solutions may be referred to as candidate solutions. The attractiveness of a firefly is determined by the intensity of light emitted by them which is determined by the fitness value of a firefly [8].

The attractiveness between X_i and X_j fireflies where $i \neq j$ is given by [8]:

$$\beta(r_{ij}) = \beta_0 e^{-\gamma r_{ij}^2} \quad (12)$$

$$r_{ij} = \|X_i - X_j\| = \sqrt{\sum_{d=1}^D (x_{id} - x_{jd})^2} \quad (13)$$

Here, the separation between X_i firefly and X_j is measured by r_{ij} , and the d th dimensions of the fireflies are given by x_{id} and x_{jd} . At the distance $r = 0$, the attractiveness is given by β_0 and light absorption quantity is γ [8].

Each firefly is being attracted to other firefly due to the brightness, so each firefly X_i will be compared to each of the other firefly X_j , where $j \neq i$ and \neq . If a firefly X_i is attracted to X_j and then X_i move toward X_j because X_j is brighter (better) when compared to X_i , then this movement is expressed by the following equation [8]:

$$x_{id}(t+1) = x_{id}(t) + \beta_0 e^{-\gamma r_{ij}^2} (x_{jd}(t) - x_{id}(t)) + \alpha \epsilon_i \quad (14)$$

where $\alpha \in [0,1]$ is a step factor and ϵ is considered as a random value in the range $[-0.5, 0.5]$ [8]. There has been a lot of improvement and enhancement in FA, and many better algorithms have been proposed one such as neighborhood attraction firefly algorithm (NaFA). Each firefly in NaFA is attracted to only k -neighborhood out of the whole population. In FA, there is a full attraction model applied as a brighter firefly can attract any firefly who's brightness is less, and due to this, it may lead to

almost full attraction between each and every firefly. Hence, to overcome the high oscillations and avoid premature convergence, the NaFA uses a better mechanism by restricting the attraction of fireflies to only a limited number of k -nearest neighbors. In NaFA, depending on the value of the k , the attraction differs because k is the value of neighbor, or number of neighbor to which the firefly might get attracted, or the attractiveness parameter may be applied when the k value satisfies for a particular firefly. K can be dynamically adjusted depending on the search space and the problem.

2.4 Particle Swarm Optimization (PSO)

Swarm describes about collective behavior of particles [4]. The particles may be seen as a group birds or fishes. Optimization is how efficiently we are going to utilize the resources to acquire the best possible result. The population of the swarm moves in the search environment (search space) for finding the food source, and this is stimulated by the particles in this algorithm [4]. It has quick converging rate and is simple to implement when compared to other algorithms, and which is why, many researchers are concerned regarding this algorithm [9].

Every particle in this algorithm works on the problem independently in the environment (search space) to obtain the finest optimal answer, while doing so, every particle finds a solution which is known as candidate solution and is represented as a point in the N -dimensional space [4]. Each and every particle is considered as a vector $x_i = (x_{i1} \cdot x_{i2} \cdot \dots \cdot x_{iN})$, and they have velocity which also considered as a vector $v_i = (v_{i1} \cdot v_{i2} \cdot \dots \cdot v_{iN})$ [4].

Algorithm

```

For each particle
    Initialize particle (say velocity and basic parameters of node position)
Do
    For every object
        Determine its fitness or health value
        If the fitness value is effective than the  $P_{best}$  in the past
            Set new  $P_{best}$  as the present value .
    END FOR
    Choose the object with the good health value among all objects as the  $G_{best}$ .
    For each object
        Work out object velocity:
             $V_i(t+1) = \omega V_i(t) + c_1 r_1 (P_{best(i)} - h_i(t)) + c_2 r_2 (G_{best} - h_i(t))$ ,
        Update object location:
             $h_i(t+1) = h_i(t) + V_i(t+1), i = m + 1, m + 2, \dots, n$ 
    END FOR
While minimum error or maximum iterate is not reached.
    
```


where V_i denotes the velocity of i th object, h_i is its estimated coordinates, $P_{\text{best}(i)}$ is the best result of the i th object or particle, G_{best} is the global best solution for the group of object which is present, ω specifies inertia weight, c_1 and c_2 are constant terms, and t is the time of iteration. The ω can be random number varying from 0 to 1 [4].

For both space distance condition and geometric topology condition as a function, multi-objective particle swarm optimization constructs a multi-objective model. It solves Pareto optimization issues by considering it as a localization problem.

In order to achieve greater correctness, consume less space for storage, and to establish greater converging rate, there are three methods or mechanisms we must make sure about [12]. They are:

1. To make less incorrect localization, the geometry topology condition should be considered.
2. For storing the good Pareto answers, the historical documents are maintained dynamically and in limited maximum capacity.
3. To speed up the converging rate for getting the global best answer, we have to apply the method of proportion of selection.

Consider the group of nodes or particles deployed in a region of surface area. The nodes may know the distance of the target spot or destination but not the exact location. So, each node moves in some or the other way in search of a destination (target) in a geographical area. If any node which is nearer to the target at any instant, the other nodes follow this particular node, and hence, each node needs to update its fitness value or position with respect to this particular node. The key features of multi-objective algorithm include maintaining historical documents, globally obtaining best selection, velocity, and updating of localization in each iteration. This is the special thing which is included in this algorithm compared to older localization algorithms [12].

1. The P_{best} value of every particle is chosen between its older P_{best} and current particles location. It works on the individual best value obtained based on Pareto efficiency or optimality [12].
2. Next comes global best (optimal selection). It works on proportion of selection and distance for each Pareto solution [12].
3. For every object, the updating of localization and velocity should happen [12].

Internode Communication in MOPSOLA:

Let us say if there are K nodes which are deployed or placed at two dimensional surface areas.

If there are A anchor nodes which have GPS facility through which we get the position of this node, then there are $K-A$ other nodes with $A < K$.

If the nodes i, j which are in communication radii, then distance of internode can be calculated as follows,

$$d_{ij} = r_{ij} + e_{ij}, \quad (15)$$

Here, the d_{ij} is obtained through RSSI technology [12].

Where $r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ describes true length between the nodes, and e_{ij} specifies the RSSI's range error.

The space distance condition's objective function can be described as [12].

$$f_1 = \sum_{i=m+1}^n \left(\sum_{j \in N_i} (\widehat{d}_{ij} - d_{ij})^2 \right) \quad (16)$$

where d_{ij} shows the distance between i and j [12],

$$\begin{aligned} \widehat{d}_{ij} &= \sqrt{(\widehat{x}_i - x_j)^2 + (\widehat{y}_i - y_j)^2} \text{ if } j \text{ is an anchor node,} \\ &\sqrt{(\widehat{x}_i - \widehat{x}_j)^2 + (\widehat{y}_i - \widehat{y}_j)^2} \text{ otherwise.} \end{aligned} \quad (17)$$

Here, $(\widehat{x}_i, \widehat{y}_i)$ and $(\widehat{x}_j, \widehat{y}_j)$ specify coordinates of other unknown nodes i, j . The geometric layout condition with its objective function can be expressed as

$$f_2 = \sum_{i=m+1}^n \left(\sum_{j \in N_i} \delta_{ij} + \sum_{j \in \overline{N}_i} (1 - \delta_{ij}) \right) \quad (18)$$

The geometric layout condition tells us connectivity condition which is not satisfied with the present estimated position of other nodes which is not anchor.

$$\delta_{ij} = \{1 \text{ if } \widehat{d}_{ij} > R, 0 \text{ otherwise} \} \quad (19)$$

Both space distance conditions along with the geometric layout condition tell us about the correctness of the nodes with specified coordinates. The greater correctness of other unknown nodes leads to less values of objective functions. For getting the efficient optimum solution of multi-objective problems, the coordinates of other unknown nodes can be designed by decreasing the values of both f_1 and f_2 objective functions [12].

2.5 Salp Swarm Algorithm (SSA)

Salps are barrel-shaped and from the Salpidae family [4]. Salp swarm algorithm (SSA) is based on the behavior of salps moving by forming a group in oceans and is a meta-heuristic technique. They look like jelly fish in texture. The living environment of salps is hardly accessible, and performing a biological research in the

laboratory environment is difficult. Salps form a chain which provides better movement and better search vision for food, and this chain is called salp chain [10]. In the mathematical model, salp's population is divided into two groups, the head position is the leader and this is the first slap of the chain which leads the group of salps, and the rest of them are the followers. The salp's location is determined by a salp-based method. The leaders are the ones who lead a particular swarm, and these swarms move in search of food sources.

Whenever the leader moves in the environment, the location of the leader is updated by [4]:

$$\begin{aligned} x_j^1 &= F_j + c_1((ub_j - lb_j)c_2 + lb_j), c_3 \geq 0, \\ F_j - c_1((ub_j - lb_j)c_2 + lb_j), c_3 < 0, \end{aligned} \quad (20)$$

Here, the position of the head is given by x_j^1 , the coefficients are random numbers, they are c_1 , c_2 , and c_3 , the upper bound and the lower bound are given by ub_j and lb_j , the food source given by F_j , and the subscript 'j' in associated with all the variables is to refer the j th dimension [4]. The position of the head is updated relative to the location of the food source. The exploration and exploitation are balanced by the coefficient c_1 and are calculated as [7],

$$c_1 = 2e^{-(4l/L)^2} \quad (21)$$

Here, L and l are for maximum iterations and current iteration, the random numbers c_2 and c_3 are generated in the range of [0, 1], and they help in determining the step size and whether the subsequent position should be closer to positive infinity or negative infinity in the j th direction [7].

The position of the follower salps is updated by the following equation [4]:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \quad (22)$$

Here, we consider $i \geq 2$, and the i th follower's j position in the j th dimension is given by x_j^i [4].

All the nodes in wireless sensor network (WSN) are deployed in the environment (search space), and every node cannot be provided with GPS as mentioned. Hence, the nodes are divided into two categories, one called the anchor nodes; they are provided with GPS; when compared with salp swarm, they make up the leader or the first salp. The second are the unknown nodes; the nodes which do not have GPS within and, hence, make up to be the follower nodes. In the search space where all the nodes are deployed and are moving around to find the optimal solution, there is a requirement of localizing the nodes because any nodes may find the maximal optimal solution (i.e., food source in terms of salp swarm).

3 Experimental Analysis

The performance in terms of locating nodes accurately and the time for computation are evaluated by performing a localization experiment on different swarm-based algorithms such as SSA, BOA, GWO, FA, and PSO. The computations is performed on MATLAB R2018b and on the system of Intel Core i7 CPU, running Windows 10 OS with 8 GB RAM. Table 1 contains the values and the parameters used in the deployment area [4].

For BOA, c is the sensory modality and set to 0.01, and the power exponent's initial value is taken as 0.1 [4]. For PSO, initial value of $c1$ and $c2$ is taken as 1.494, and ω is 0.7. For GWO, the parameter C linearly increases from 0 to 2 and the parameter " a " in the interval of $[2-0]$ [11]. For SSA, we take $c2, c3$ are random numbers generated in the range $[0, 1]$ and are taken as $c2 = 0.7$ and $c3 = 0.3$ [4]. For FA, γ is taken as 1.0, β is taken as 1, and the randomization parameter is set to 0.25 [4].

The swarm-based algorithms are run under different conditions such as distinctive quality of target nodes, distinctively quantity of iterations, distinctively quantity of anchor nodes. These algorithms are run and evaluated under different parameters such as computing time, localization error, and number of localized node. The obtained results are shown in Table 2. For all of the localization algorithms, it is seen that increasing the quantity of iterations increases the computation and the quantity of localized nodes while reduces the localization error. As the number of iterations increases, the computation also increases as a lot of higher amounts of computation is required, and increasing the quantity of iterations reduces the localization error as there is a chance of finding a better solution.

Hence, it is noticed that for all the algorithms, computational time (T_s) is increased by increasing the quantity of anchor nodes and target nodes [4]. When the table is seen with respect to the number of localized nodes (N_L) initially, all the algorithms are quite slow in localizing a large number of nodes, but as time passes, the number of nodes localized increases rapidly, and out of which, we find that SSA localizes a large number of nodes when compared with other algorithms with same time. The error in localization (E_L) is also less for SSA when compared with all other algorithms. The localization parameter increases as the anchor density increases.

Table 1 Parameters setting of simulation environment

Parameters	Values
Sensor nodes	Varies on $\sum_{i=1}^6 i * 25$
Anchor nodes	Varies on increment $i = i + 5$
Node transmission range (R)	30 m
Deployment area	100 m * 100 m
Maximum number of iterations	100

Table 2 Performance metrics of different localization algorithm

Target nodes	Anchor nodes	No. of iterations	PSO			BOA			FA			GWO			SSA		
			E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L
25	10	25	0.818	0.40	16	0.232	0.37	22	0.265	0.6	19	0.744	0.22	20	0.465	0.35	22
		50	0.812	0.40	15	0.225	0.38	23	0.262	1.2	20	0.741	0.41	21	0.462	0.35	23
		75	0.803	0.41	18	0.223	0.39	25	0.258	1.5	21	0.741	0.54	21	0.458	0.36	23
50	15	100	0.792	0.41	18	0.221	0.40	25	0.251	1.8	19	0.740	0.79	23	0.451	0.37	24
		25	0.419	0.71	41	0.338	0.84	47	0.477	1.6	46	0.690	0.42	44	0.477	0.67	43
		50	0.426	0.73	47	0.332	0.85	48	0.473	1.9	49	0.688	0.63	45	0.472	0.69	47
75	20	75	0.429	0.76	46	0.326	0.86	49	0.465	2.5	49	0.686	0.81	46	0.468	0.69	48
		100	0.434	0.76	48	0.323	0.86	49	0.465	3.5	48	0.682	0.98	48	0.464	0.70	50
		25	0.735	1.31	73	0.257	1.49	66	0.519	2.9	71	0.641	0.72	72	0.519	0.90	69
100	25	50	0.728	1.32	74	0.257	1.49	66	0.513	3.8	72	0.641	0.95	72	0.513	0.92	72
		75	0.728	1.33	74	0.255	1.50	70	0.504	4.7	73	0.638	1.3	73	0.504	0.95	73
		100	0.724	1.35	75	0.253	1.52	72	0.503	5.2	73	0.635	1.4	74	0.503	0.96	75
125	30	25	0.661	2.10	97	0.355	2.40	97	0.711	3.8	98	0.611	1.1	95	0.511	1.31	98
		50	0.658	2.16	97	0.355	2.44	99	0.709	4.2	98	0.606	1.5	97	0.509	1.33	98
		75	0.642	2.17	99	0.333	2.47	100	0.702	5.6	99	0.602	1.8	98	0.502	1.36	99
100	25	100	0.641	2.20	100	0.331	2.50	100	0.704	6.3	98	0.602	2.1	98	0.504	1.37	100
		25	0.754	4.87	120	0.549	3.84	122	0.829	2.7	122	0.589	1.5	122	0.529	1.67	123
		50	0.748	4.86	121	0.548	3.85	123	0.824	4.5	123	0.580	2.2	123	0.524	1.68	124
100	25	75	0.750	4.89	122	0.534	3.86	124	0.822	5.9	125	0.580	2.8	123	0.522	1.70	125
		100	0.752	4.95	125	0.534	3.89	124	0.822	6.5	124	0.572	3.3	125	0.522	1.37	100

(continued)

Table 2 (continued)

Target nodes	Anchor nodes	No. of iterations	PSO		BOA		FA		GWO		SSA						
			E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L	E_L (m)	T (s)	N_L			
150	35	25	0.625	5.41	145	0.766	5.88	147	0.911	2.5	149	0.559	2.8	148	0.511	2.12	149
			0.622	5.42	146	0.765	5.61	148	0.909	4.2	150	0.547	3.6	149	0.509	2.14	149
			0.619	5.44	148	0.763	5.64	149	0.904	6.4	150	0.523	4.3	150	0.504	2.16	150
			0.616	5.45	150	0.763	5.69	149	0.904	7.2	149	0.523	4.8	150	0.504	2.18	150

4 Conclusion


Node localization has various applications in WSN; in this paper, some of the well-known bio-inspired node localization algorithms are being evaluated and compared under different conditions and parameters such as the quantity of iterations and number of nodes. In this paper, the working of different bio-inspired algorithms works and how it may help in node localization is discussed. When all the algorithms are compared, we find that the SSA algorithm performs better when the anchor nodes and target nodes are increased, and when there are less target nodes and anchor nodes, we see that BOA performs better than all the other algorithms. There are a lot of research work taking place in this field, and many new hybridized algorithms and many modified algorithms are enhancing the traditional node localization problems.

References

1. Kulkarni RV, Venayagamoorthy GK, Cheng MX (2009) Bio-inspired node localization in wireless sensor networks. In: Proceedings of IEEE international conference in systems, man, and cybernetics (SMC), San Antonio, TX, USA, Oct 2009, pp 205–210
2. Lavanya D, Udgata SK (2011) Swarm intelligence-based localization in wireless sensor networks. In: International workshop on multi-disciplinary trends in artificial intelligence. Springer, Berlin, pp 317–328
3. Wang J, Ghosh RK, Das SK (2010) A survey on sensor localization. *J Control Aeory Appl* 8(1):2–11
4. Kanoosh HM, Houssein EH, Selim MM (2019) Salp swarm algorithm for node localization in wireless sensor networks. *J Comput Networks Commun* 2019, Article ID 1028723
5. Strumberger I, Tuba E, Bacanin N, Bekoy M, Tubaz M (2018) Monarch butterfly optimization algorithm for localization in wireless sensor networks, 978-1-5386-2485-2/18/\$31.00 ©2018 IEEE
6. Wang G-G, Deb S, Cui Z (2015) Monarch butterfly optimization. In: Neural computing and applications, May 2015, pp 1–20
7. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
8. Wang H, Wang W, Zhou X, Sun H, Zhao J, Yu X, Cui Z (2017) Firefly algorithm with neighborhood attraction. *Inf Sci* 382–383:374–387
9. Joshi H, Arora S (2017) Enhanced Grey Wolf optimization algorithm for global optimization. *Fund Inform* 153:235–264. <https://doi.org/10.3233/FI-2017-1539>, IOS Press
10. Anderson PA, Bone Q (1980) Communication between individuals in Salp chains. II. *Physiology. Proc R Soc Lond Ser B Biol Sci* 210(1181):559–574
11. Ahmed MM, Houssein EH, Hassanien AE, Taha A, Hassanien E (2017) Maximizing lifetime of wireless sensor networks based on whale optimization algorithm. In: International conference on advanced intelligent systems and informatics. Springer, Berlin, pp 724–733
12. Sun Z, Tao L, Wang X, Zhou Z (2015) Localization algorithm in wireless sensor networks based on multiobjective particle swarm optimization. *Int J Distrib Sens Netw* 2015, Article ID 716291, 9 pages. Hindawi Publishing Corporation. <https://doi.org/10.1155/2015/716291>

Artificial Intelligence for Automated Engineering Design



Durgaprasad Janjanam 

1 Introduction

Artificial Intelligence specifically enable the computer to assist in solving ill-structured and non-deterministic problems such as design activities including synthesis, evaluation, modeling and decision-making process. Industrial production problems and the technology's economic attractiveness are currently taking center stage in India. There are many factors inspiring these issues. Some of these are economic, social and political issues and others are technological. Here, the importance is on technological and scientific developments which promise to have a strong impact on improving productivity.

Approximately, seventy percent of the life cycle costs of major engineering systems are determined during the early stages of design of those systems [1]. The existing commercial CAD market concentrates on the detailed design phase which is dominated by drafting and layout activities. Conceptual design, on the other hand, is dominated by engineering analysis of performance, cost, reliability and executability activities. No significant computer tools are available for conceptual design that provide the flexibility.

Conceptual design for an artifact can be modeled and represented at different levels of abstraction for visualization, analysis, evaluation, etc., based on the scope of the task. One way to describe an artifact is by enumerating its parts (decision constraints). A second way to describe an artifact is in terms of its structure, that is, how the parts are connected to form the artifact's topology (topological constraints). A third way to describe the object is by drawing or geometric model representing the object's geometry (geometrical constraints). A fourth way an artifact can be represented is in terms of the underlying physics that govern its behavior; usually

D. Janjanam (✉)
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: durgaprasad.j@nmit.ac.in

in terms of the algebraic representation of the governing physical laws (physical constraints). Constraint satisfaction techniques have been used to solve topological, geometrical and physical constraints problems [2]. This kind of constraint satisfaction problems can be seen in all engineering disciplines. Perhaps cooking itself is a design process, hence no two cooks can produce the same taste. This also involves many constraints to be satisfied to complete the task.

Basically, design is a collaborative process involving designers, consumers and other actors, through ongoing meetings, consultations, clarifications and assessments, establishing and refining a common sense of requirements and possible solutions. This shared sense, crystallized as the artifact of design and made permanent as shared memory forms the foundation of the cumulative knowledge drawn upon by subsequent designs [3].

Understanding differences between novice and experienced designers and the nature of the tools required for conceptual design will have considerable impact on the conceptual design stage as well as on the overall design automation effect. The impact of constraint management tools will be of great importance in two areas of design theory and methodology: first it will help close the gap between novice designers and experienced designers, and second it will provide support to the early stages of design.

A large body of research exists on solving constraint satisfaction problems. These include ThingLab [4]—a constraint-based simulation laboratory; SOCLE [5]—a hybrid structured object and constraint representation language for knowledge representation; IDEAL [6]—a software for easy layout of complex figures and uses constraint satisfaction to allow the positions and sizes of objects to be stated as relationships; Magritte [7]—an interactive graphical layout system same as ThingLab; FREEDOM [8]—an object-oriented and constraint-based knowledge representation system for design object modeling; CONCEPT MODELER [9]—a constraint management in conceptual design using graph-based representation of constraint networks.

Frequently, perhaps always, at some point of the development of the design, a candidate solution will be found to violate one or more constraints of the current problem specification. The management of these constraints during evaluation of design is a non-trivial task. The constraints are often numerous, complex, contradictory and vague. There are many sources of constraints ranging from “soft” constraints imposed by intuition, “hard” constraints imposed by physical laws and further, the boundaries of the design space are initially unknown, and the first task is to identify the bounding constraints. The process is evolutionary and iterative.

In the face of above complexities, present paper explains the tools and system that helps in the conceptual phase of design.

2 Description of the System

The proposed system is viewed as a sketch pad for experimenting the conceptual thought process of the designer. Using this sketch pad, designer can either draw the design parameters as nodes (marks) and connect them with edges (links) based on the constraints or give the list of constraints. If the designer gives the list of constraints in the form of equations, then these equations can be converted into a set of nodes and edges and can connect them based on their dependency using standard data-structure that is a parsing technique. Dynamically, the designer can designate some of the design parameters as known (input) parameters based on his conceptualization. Then, the system takes this information from the sketch pad (computer screen) and generates the code for graph-based representation of constraints. This graph-based representation can help in managing the constraints to check for consistency and to arrange parameters systematically to simulate the flow of design process based on their order of dependency. If there is any inconsistency or conflict between constraints then the system can give the feedback to the designer and the designer can take suitable decisions, else it can simulate the design process that is given by the designer's interest. After proper representation of constraints, the design process starts. In the design, due to inappropriate values for some of the design parameters, some of the design constraints may be violated. In a normal design situation, the system would prompt that the design has failed. But, the use of possibility theory, contemplated in this paper, will help by giving appropriate suggestions.

This is organized by identification and characterization of some of the major problems; an identification of necessary research activities and tasks; and a set of final outcomes to solve some of these problems. Figure 1 shows a flow diagram that connects these problems to a set of tasks and outcomes. The tasks have overlapping methodologies and objectives.

Task 1: Graph-Based Representation of Constraints

Graph is a set of nodes and edges. These nodes are connected by edges. Basically, any process can be represented by activities that are connected based on the flow of dependency and constraints. A constraint is a relation among several parameters. Graph (network) is used to represent the constraints of a problem. This graph will contain the cycles. This is known as a non-directed graph. The outcome of the task is a simulation of the design option which is in the mind of a designer based on his constraints and parameters.

Task 2: Algorithms for Effective Constraint Management

Basically, the parameter dependency in a graph establishes the flow of information and allows to detect anomalies in the network. In order to evaluate a network or test for consistency, it is necessary to determine the interdependencies among the various parameters. These issues are solved using bipartite matching. This bipartite matching helps in eliminating cycles in the network, which intern converts non-directed graph into a directed graph. This directed graph can be rearranged in the form of a tree structure. Breadth first search is applied on this tree to arrive at the solution.

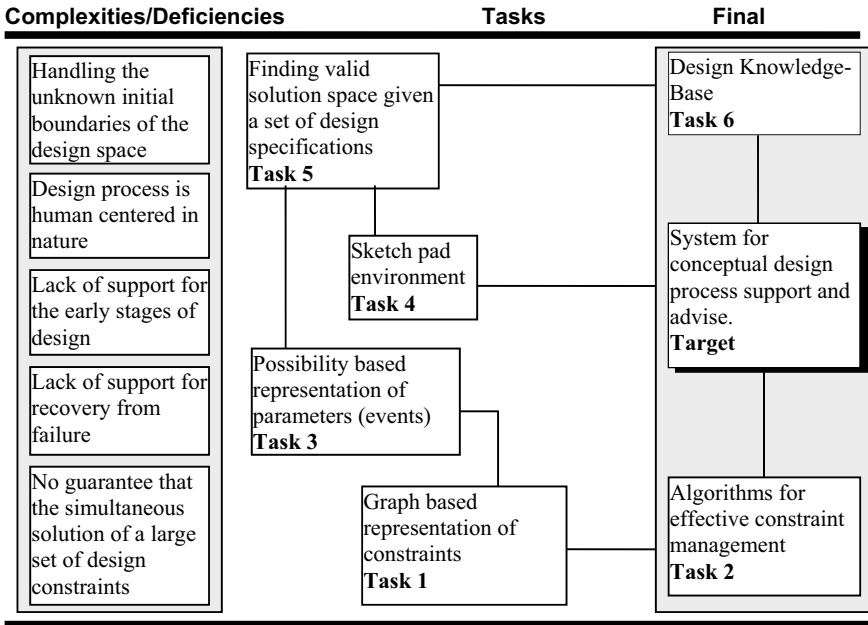


Fig. 1 Schematic view of the system

The primary task of a constraint manager is to find whether a valid solution space exists given a set of design specifications. If a solution exists, then it can be explored in order to determine trends among the design parameters. Exploring the solution space increases the designer’s knowledge of the design space and supports decision making. This entire process is explained with a simple example problem for better understanding about how flow of information is being established.

Example Problem

A simple cantilever beam problem is chosen to demonstrate the graph-based representation of constraints and constraint management. Parameters (σ , B, H, F, Y, I, K, L, M, and E) are connected as a network (graph) based on their corresponding constraints c1, c2, c3, c4 and c5 as shown in Fig. 2c.

Once network is established, designer can designate some of the parameters σ , B, H, F, Y, I, K, L, M, and E as knowns and rest of the parameters as unknowns.

Let,

Known parameters: F, L, H, B, E;

Unknown parameters: σ , M, I, Y, K;

Figure 2c shows how variables are connected through constraints. Using this figure, a systematic way of identifying how variables are connected to constraints and will help to establish flow of design parameters based on their order of dependency. The bipartite graph showing matching is shown in Fig. 2a. This graph gives the importance of unknown/known parameters in the constraints, which will finally affect

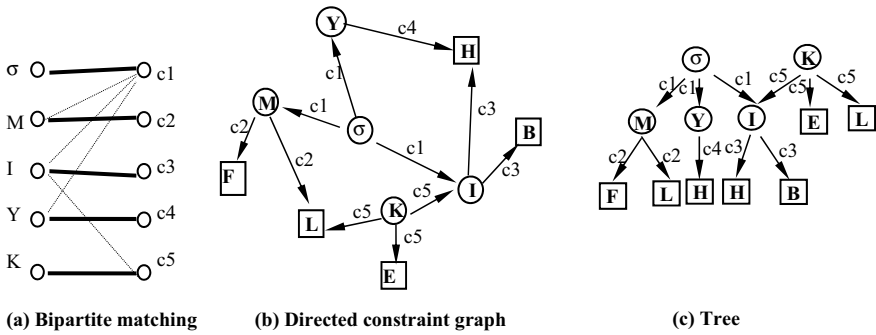


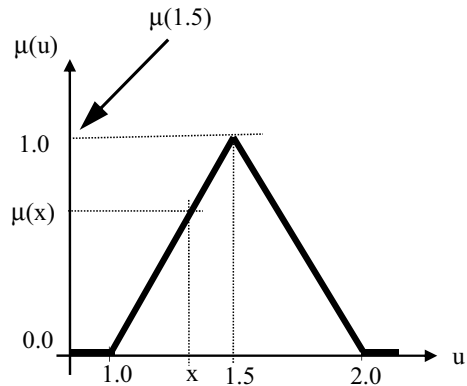
Fig. 2 a–c Solution steps of the problem

the design considerations. Maximum matching parameters to constraints are marked with thick line. Based on Fig. 2c and using the maximum matching in Fig. 2a, a directed constraint graph is obtained in Fig. 2b. This will aid the designer to identify the flow among the various known and unknown design parameters. This directed graph is converted to a standard data-structure that is a tree as shown in Fig. 2c. This tree indicates the flow of design parameters in a structured order based on the order of dependency and generated dynamically based on the choice of the designer’s input and output design parameters.

Task 3: Possibility Based Representation of Parameters

Present existing methods of constraint satisfaction techniques cannot provide any information for recovery upon failure of constraints satisfaction. Possibility based representation has the advantage of providing advice for recovery when it fails to satisfy the constraints. As, in nature, the process of design is human-centered; “... traditional system analysis techniques are not well suited for dealing with human-centered processes because they do not come to grips with the reality of the fuzziness of human thought and behavior.” [10]. Therefore, these issues can be dealt with by incorporating theory of possibility to reflect the designer’s intuitive inputs during conceptual design [11]. A center of solution strategies for treating and propagating variables with exact values are the latest research efforts. Sadly, most of the engineering constraints are expressed as inequalities [12–15]. The essence of the constraints is such that they do not prescribe particular values for design parameters but rather prescribe potential ranges for values of different membership (μ) degree. Conventional parameter assignment, e.g., $L = 1.5$ cm, can be expressed as assignment to a possible range of values with different degree of membership ($\mu = 1$ is maximum, $\mu = 0$ is minimum) as shown in Fig. 3, e.g., $L = 1.5: \mu(1.5)$ cm, where $\mu(1.5)$ is the degree of assignment ($\mu(1.5) = 1$). Relations between the parameters can also be represented using fuzzy concept. This is focused on how design parameters are viewed as distributions of possibilities. This kind of representation of values generalizes the notion of equality assignment, offers a framework for dealing with

Fig. 3 Possibility distribution



tolerances and adds versatility, enabling the conceptual process of a design to catch incompleteness and ambiguity.

Task 4: Sketch Pad Environment

Sketch pad is an iconized graph-based tool where the designer can draw network for constraints. The system can take the network from the computer screen and will generate the appropriate source code automatically.

Task 5: Finding Valid Solution Space

Generate all feasible paths of various possible successful designs in a iterative manner and forms a knowledge-base which is a valid solution space.

Task 6: Design Knowledge base

Design knowledge base is created and maintained from the obtained valid solution space, and also from past successful designs.

3 Conclusions

Application of Artificial Intelligence in conceptual design using graph theoretic techniques is demonstrated. Design is a collaborative process involving designers, consumers and other actors, through ongoing meetings, consultations, clarifications and assessments, establishing and refining a common sense of requirements and possible solutions. Toward this, a beam design problem is used that specifically demonstrate how the computer can assist in solving ill-structured and non-deterministic problems such as design activities including synthesis, evaluation, modeling and decision-making process. It is a environment for the designer to experiment with logical thought process. The same is illustrated using a beam problem architecture.

References

1. Rockwell Palo Alto Laboratory, USA, Design Sheet, from INTERNET, 1991
2. Wm L (1988) Constraint programming languages. Addison-Wesley Publishing Company
3. Subrahmanian E, Konda SL, Levy SN, Reich Y, Westerberg AW, Monarch I (1993) Equations are not enough: informal modeling in design. *AIEDAM* 7(4):257–274
4. Borning A (1979) Thing lab—a constraint oriented simulation laboratory. Technical report, Xerox Palo Alto Research Center
5. Harris DR (1986) A hybrid structured object and constraint representation language. *Proc AAAI-86* 986–990
6. Van Wyk CJ (1980) A language for typesetting graphics. Ph.D. thesis, Stanford University, Palo Alto, California, June 1980
7. Gosling J (1983) Algebraic constraints. Technical Report CMU-CS-83-132 (Ph.D. thesis), Carnegie Mellon University, Pittsburgh
8. Yokoyama (1989) An object-oriented and constraint-based knowledge representation system for design object modeling. Research Report TM-0809, Institute for New Generation Computer Technology, Japan, Oct 1989
9. Serrano D (1987) Constraint management in conceptual design. Ph.D. thesis, Dept. of Mechanical Engineering, MIT., USA
10. Zadeh L (1973) Outline of a new approach to the analysis of complex systems and decision process. *IEEE Trans Syst, Man, Cybern* SMC 3(1)
11. Durgaprasad J (1994) A new approach to engineering design process support and advisory system using possibility theory applied to wind engineering design problem. Research report under SERC/UNDP, Carnegie Mellon University, Pittsburgh
12. Krishnan V, Navinchandra D, Rane P, Rinderle JR (1990) Constraint reasoning and planning in concurrent design. Report CMU-RI-TR-90-03, Carnegie Mellon University, Pittsburgh
13. Rodgers PA, Mazzarella F, Conerney L (2020) Interrogating the value of design research for change. *Des J* 491–514
14. Scott Swan K (2019) Design roots: culturally significant designs, products and practices. *Des J* 873–878
15. Wang D, Oygur I (2015) A heuristic structure for collaborative design. 355-371

Comprehensive Strategy for Power Quality Improvement of Inverter Based Distributed Generation Systems



Shreeram V. Kulkarni, Shruti Gatade, N. Samanvita, and D. N. Gaonkar

1 Introduction

The distributed generation (DG) concept is considered as a small-scale power unit in the power system network, which can operate in grid-connected mode and as well in island- mode. There are several DGs mainly categorized into non-renewable and renewable resources, such as photovoltaic systems, wind arms, micro-turbine, fuel cells, geothermal, micro-hydro, bio-mass, tides, and ocean waves. Among these micro-sources, photovoltaic and wind plants are prominent sources suitable for integration through the power electronic converters. The IEEE 1574 standards address the standardized interconnection of microsources to grid and IEEE Std 1409TM-2012 address guide for application of power electronics for improving power quality on a distribution systems [1, 2]. To achieve an imperishable growth power system network, it is necessary to cope up with the energy demands of the consumers. Power quality plays a vital role while coping up with consumer's demands. With the development of advanced sensitive devices and the ignorance of power quality may lead to production loss, blemish electrical appliances, increased power loss, etc. The varying quality of electric power is predominantly due to the current and the voltage harmonics because of the widespread utilization of static power electronics converters and unbalanced load reactive power, voltage swell, voltage sag, flicker, or voltage interruption, etc. Hence, it is predominant to have a sustained standard power quality.

S. V. Kulkarni (✉) · S. Gatade · N. Samanvita
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: shreeram.kulkarni@nmit.ac.in

D. N. Gaonkar
National Institute of Technology, Surathkal, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_14

1.1 Power Quality

The power quality issues are harmonics, sag, and swell which are suppressed by the active filters by the unified power quality conditioner (UPQC) [3]. UPQC has the following facilities for nonlinear and voltage-sensitive loads [4].

1. It diminishes the harmonics in the supply current.
2. Necessitate the reactive power requirement of the load.
3. Keeps up load end voltage at the rated value even in the presence of supply voltage sag.

In power system network, different types of faults and increased nonlinear loads usually give rise to power quality variation. The term power quality (PQ) is the combination of voltage quality and current quality [5]. In consideration of the quality of the current is dependent on the loads connected, which draws the power from supply system. Subsequently, the standards in the PQ quality are utilized to keep the flexibly voltage within the limits. The supplied AC power is intended to work at a sinusoidal voltage with a stable frequency of 50 or 60 Hz. Disturbances in PQ have occurred at the point when there is a huge change in the supply voltage, supply frequency, as well as waveform deviation because of different sorts of loads, the sudden change in the loads, power converters, and various types of faults [6].

Most consumers are not recognizable of PQ unbalance influences that typically occur in the electrical power system. The PQ issues may make insecurities, which in turn shortens the equipment lifetime [7]. Consequently, it is basic to perceive and arrange PQ unbalance causes and avoid harmful effect due to PQ variations in the electrical power systems. The current VSIs regardless, use electrolytic capacitors that have remarkable disappointment modes and have been seemed to show quickened mileage in parched climatic conditions. In this manner, the electrolytic capacitors are regularly viewed as the bottleneck parts of a force electronic framework, putting extreme confinements on the general assistance life of a force electronic gadget.

The presence of harmonics in the power lines corresponds to issues like increased power loss in distribution, issues of electromagnetic intervention in communication systems and failure electronic types of equipment leading to failure of industrial processes. Because of these issues, the quality of the electrical energy dispatched to the end-users, like never before, is a point of concern. Passive filters have been utilized as a standard method to solve current harmonic issues, yet they have certain limitations resonance, electromagnetic inference, bulkiness, etc. To overcome these disadvantages, rigorous efforts have been made in this article for the advancement of active power filters (APF) during the power disturbances [8]. An adequate adaptive neural fuzzy interface system supervised PID controlled shunt active power filter is intended to smother the undesirable harmonics in nonlinear load applications [9]. To compensate for harmonic currents, power factor and nonlinear load unbalance a shunt active filter has been proposed with Variable Leaky Least Mean Square (VLLMS)-based controller [10]. The paper proposes a survey of the control advancements of active filters with their fundamental highlights [11]. The power rating of the

active power filter is a crucial factor to evaluate its potential to handle the power quality improvement task. Active power filters with large power ratings are best when it comes to deal with large nonlinear loads, but may not be an economical solution. The APF has widely utilized terminology in the area of power quality improvement, and in [12], a self-tuning filter unit vector generator (STF-UVG) with a shunt active power filter in the UPQC controller shows the significance in the overall efficiency of the grid power system. Different tools are available for simulation which can be for modeling and analysis of various power system problems. The most generally utilized tools in academic research are PSCAD/EMTDC, ATP/EMTP, and MATLAB/SIMULINK with SimPower System. The topology of improvement in PQ with APF in MATLAB/Simulink and Typhoon hardware in loop is dictated in the next subsequent section.

2 System Under Study

The distribution generation source (DG) is connected at the point of common coupling (PCC) through converters, step-up/step-down transformer, and an isolating switch. And then the DG connection from PCC is given to the medium voltage network (MV) for the grid integration. Where the operation mode of the PE-based DGs is decided by the isolating switch, which will decide the operation in grid-connected mode or stand-alone mode. The power electronics-based DG sources are capable of operating in both grid-connected and islanded mode of operation, and this type of converters are more feasible to integrate in the microgrid [13]. Unified power quality conditioner (UPQC) is an incredible asset to resolve the issues related to power quality, and when series and shunt active power compensators are joined into a single device, then it forms a unified power controller [14]. The general setup

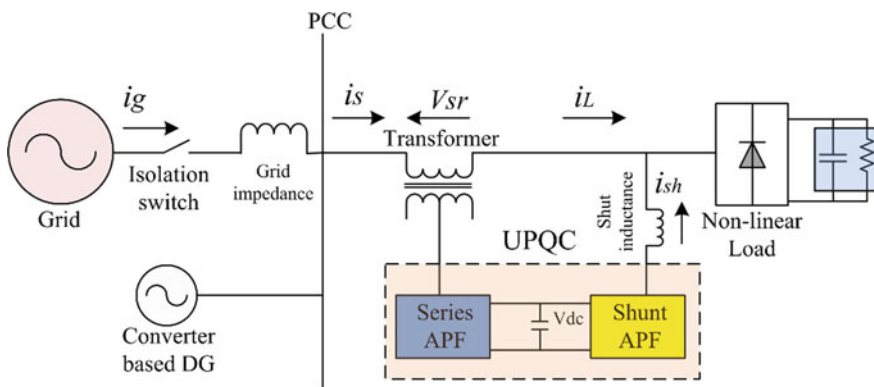


Fig. 1 General block diagram of grid integrated DG system

Table 1 Table captions should be placed above the tables

Parameter	Value	Parameter	Value
R_s	0.01Ω	K_{pv}	0.0263
C_{dc}	$8 e^{-4}F$	K_{iv}	53.6
L_f	$9.26e^{-3}H$	K_{pi}	19.26
R_f	1Ω	K_{ii}	1100
C_f	$2.104e^{-4}F$	m_{p1}	$2.5e^{-4}$
L_c	$0.01H$	n_{q1}	$1.33e^{-3}$
R_c	1Ω	f_n	50 Hz
L_v	$1 e^{-3} H$	f_c	5 Hz
R_v	-1Ω	V_{dn}	400
P_{max1}	$8e^3W$	Q_{max1}	$6e^3kVAr$
$\tau_v \& \tau_i$	$10e^{-3}s \& 1e^{-3}s$	f_s	$8e^3Hz$

of the UPQC is demonstrated is shown in Figure 1, where i_{sh} and V_{sr} are the compensating voltage and current from the series APF and shunt APF, respectively, which is illustrated in Fig. 1.

The general design of the UPQC has potential of improving power quality at the point of establishment on industrial power systems or power distribution systems. The UPQC, in this way, is required to be one of the most remarkable solutions to supply voltage unbalanced for large capacitive loads [15]. The parameters of the system used to study the system responses are listed in Table 1.

2.1 Control Scheme

The voltage source inverters are commonly used for grid integrated operation mode. There are two control parts, namely power part and control part. The power part includes converter, filter, and lines, and the generation of control signals for the power electronics based converter is generated shown in Fig. 1. The inner voltage and current controller, the reference voltage is feedback to the outer loop of the inner voltage controller, and the output of controller is given to the current controller. The switching signals for the voltage source inverters are generated by the standard PWM, and the input to the PWM is given by the current controller. The block diagram of control scheme is shown in Fig. 2.

The voltage and current controller are employed with the standard PI controllers, feedback and feed-forward gains are shown in Fig. 6 and Fig. 7. Equations (1) and (2) they are algebraic equations for the voltage controller, and where the V_{vd} is the reference virtual impedance voltage included in the voltage controller outer loop for balancing the output impedance characteristic to improve, though reducing the circulating currents [13].

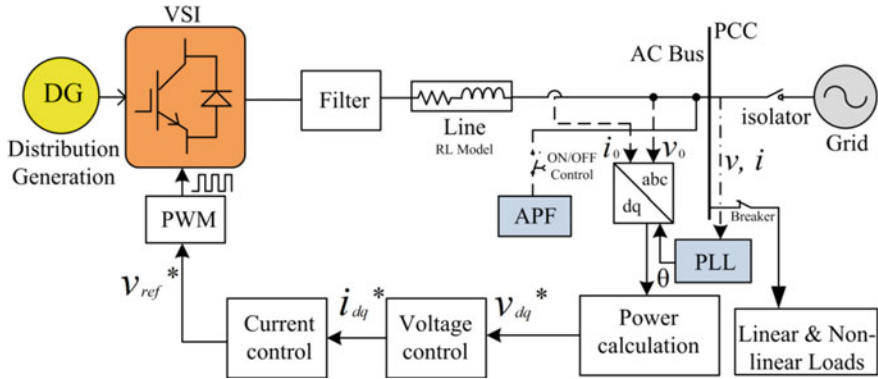


Fig. 2 Control diagram for grid integrated DG system with APF controller

$$i_{ld}^* = F i_{od} - \omega_n C_f V_{oq} + K_{pv} (V_{od}^* - V_{od} - V_{vd}) + K_{iv} \phi_d \quad (1)$$

$$i_{lq}^* = F i_{oq} - \omega_n C_f V_{od} + K_{pv} (V_{oq}^* - V_{oq} - V_{vq}) + K_{iv} \phi_q \quad (2)$$

Equations (3) and (4) are the output of the current controller and input for the standard PWM inverter to generate the control signals to the stand-alone voltage source inverter.

$$V_{id}^* = -\omega_n L_f i_{lq} + K_{pc} (i_{ld}^* - i_{ld}) + K_{ic} \gamma_d \quad (3)$$

$$V_{iq}^* = \omega_n L_f i_{ld} + K_{pc} (i_{lq}^* - i_{lq}) + K_{ic} \gamma_q \quad (4)$$

The DG control technique in grid integrated mode incorporates a constant current controller and a Phase-Locked Loop (PLL) generator [16]. PLL are generally utilized to track the frequency and phase angle of the grid for synchronization and resynchronization. In the grid-connected mode of operation, it is easy to conclude that the total load demand of the system is being supplied by both the DG system and the grid shown in Eqs. (5)–(7).

$$V_{PCC} = V_{grid} \quad (5)$$

$$P_{load} = P_{DG} + P_{grid} \quad (6)$$

$$Q_{load} = Q_{DG} + Q_{grid} \quad (7)$$

where V_{PCC} is voltage at point of common coupling, V_{grid} is grid voltage, P_{DG} , P_{grid} , P_{load} and Q_{DG} , Q_{grid} , Q_{load} are inverter, grid, load active powers in kW and

reactive powers in kVAr, respectively. If the DG is generating the excess power which is more than the required microgrid load power, then the excess will be exported to the grid and is shown in the following equations.

$$P_{DG} > P_{load}; P_{DG} = P_{load} + P_{grid} \quad (8)$$

$$Q_{DG} > Q_{load}; Q_{DG} = Q_{load} + Q_{grid} \quad (9)$$

Here, we proposed unified power quality conditioner (UPQC), which belongs to the family of active power filter (APF). In this paper, UPQC used is formed by combining series and shunt APF. The series APF goes about as a voltage source and infuses a compensating voltage so as to have a sinusoidal voltage. It is appropriate for voltage related issues. What's more, the shunt APF goes about as a current source and remunerating current flows and is appropriate for compensating the issues identified with the current [8].

3 Results and Discussion

The proposed system model depicted in Fig. 1 is simulated in MATLAB/Simulink environment. 10 kVA, 230 V/50 Hz inverter-based DG system is tied with the 230 V/50 Hz grid. And the step change in linear and nonlinear loads is connected in between grid and DG system. The series RL load is fed by three phase diode rectifier (nonlinear) load. The performance of the APF under the grid disturbance is depicted in this paper. The simulation results show the better performance under the power quality disturbances.

The distortion voltage across the load is introduced, and we can clearly see the variations or distortion in the load voltage. Figure 3 shows the PCC voltage, source voltage, and the load voltage before connecting series APF, which in turn affect the source and PCC voltages. As shown in Fig. 3, the source and PCC voltages are sinusoidal but contains the distortions (window: 0.42–0.58).

We need to compensate the distortion present in the sinusoidal source and PCC voltages, so we introduced the series APF to compensate them. After connecting series APF, the supply voltages are sinusoidal, normal, phase line with the voltages even distorted load condition, is shown in Fig. 4.

Figure 5 shows the source current, PCC current, and the load current before connecting shunt APF. The current harmonics are introduced in the system due to the nonlinear load. We can see that the source and PCC currents are affected by the load harmonics, and the window of 0.42 to 0.58 is depicted in Fig. 5 for the clear vision.

To compensate the reactive currents or the harmonics in the current waveform due to nonlinear load in the system, we introduced the shunt APF to compensate them. The compensated waveforms of PCC current and source current are shown in Fig. 6.

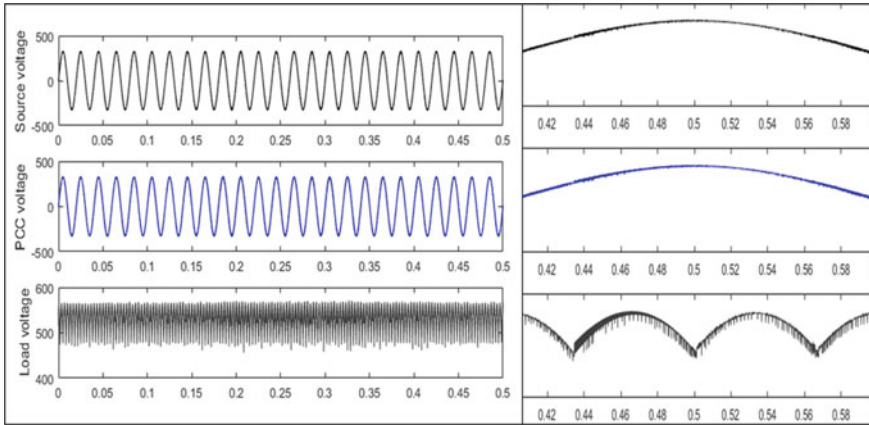


Fig. 3 Source voltage, PCC voltage, and the load voltage before connecting APF controller in the system

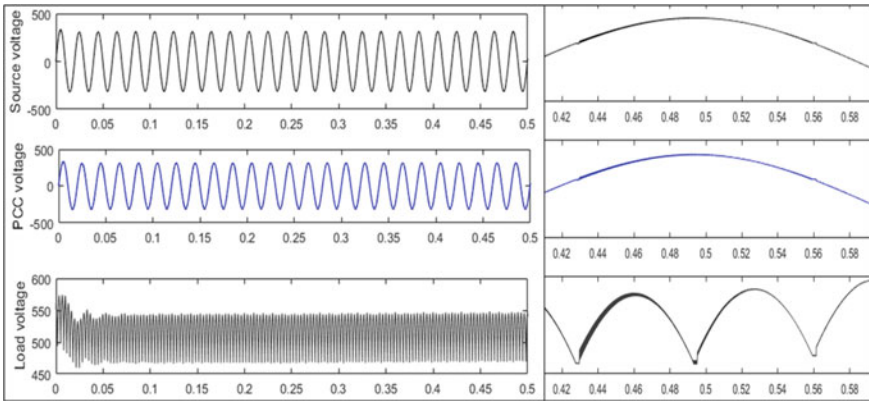


Fig. 4 Source voltage, PCC voltage, and the load voltage after connecting APF controller in the system

To analyze the performance of the shunt APF and series APF, the Fast Fourier Transform (FFT) analysis is carried out. Figures 7, 8, 9, and 10 show the FFT analysis of source voltage and source current before and after connecting the APFs.

After the shunt APF is made to work in the framework, the source current THD is upgraded from 25.83 to 2.96%. Furthermore, the source voltage THD is upgraded from 2.96 to 2.59%, later the series APF is made to work for compensating. The series and shunt active filters are begun to compensate for the voltage and current harmonics, affirming that the source voltage and current get no distortion. Here, from Figs. 7, 8, 9, and 10, we can see that both the voltage and current THD are inside the range as shown by the IEEE standards [1].

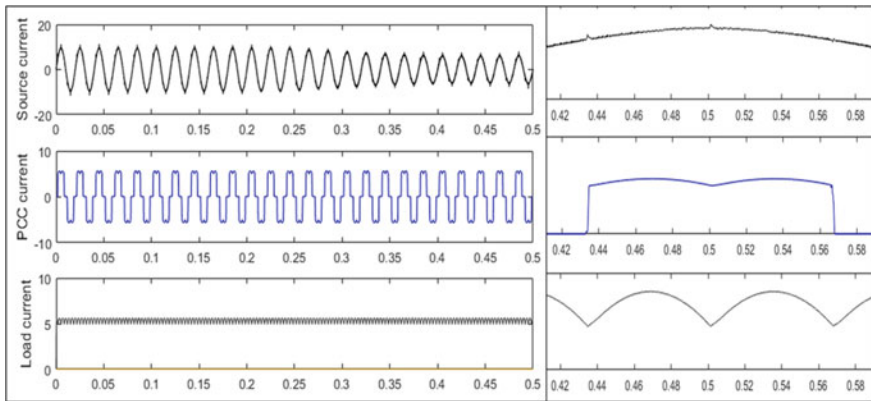


Fig. 5 Source current, PCC current, and the load current before connecting APF controller in the system

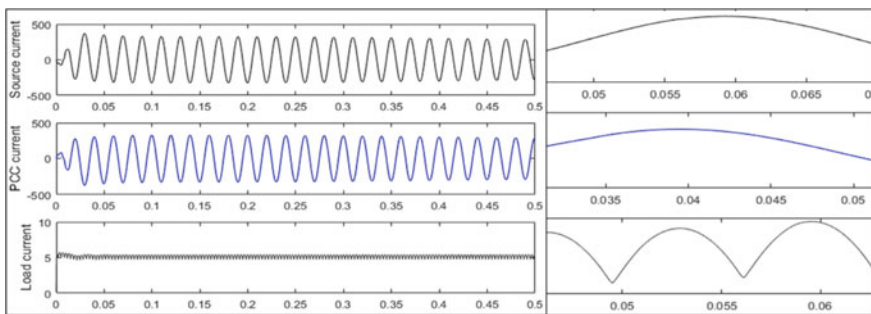


Fig. 6 Source current, PCC current, and the load current after connecting APF controller in the system

Fig. 7 FFT analysis of source voltage before connecting series APF controller in the system

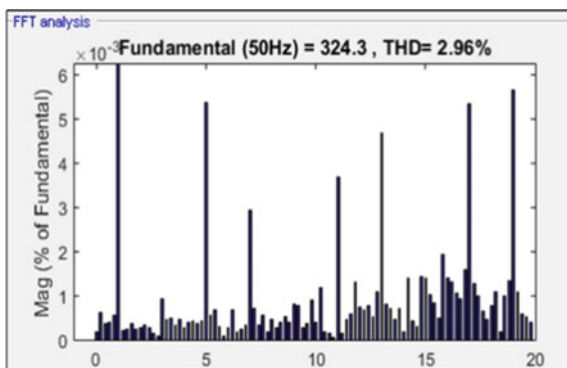


Fig. 8 FFT analysis of source current before connecting shunt APF controller in the system

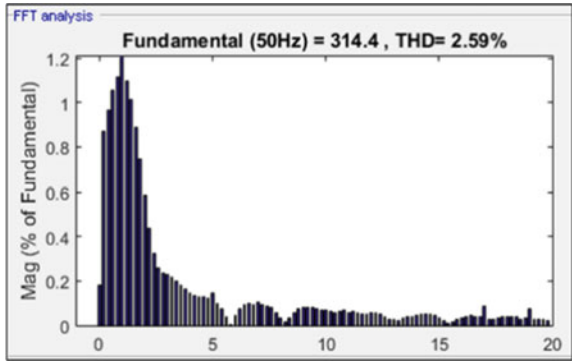


Fig. 9 FFT analysis of source voltage after connecting series APF controller in the system

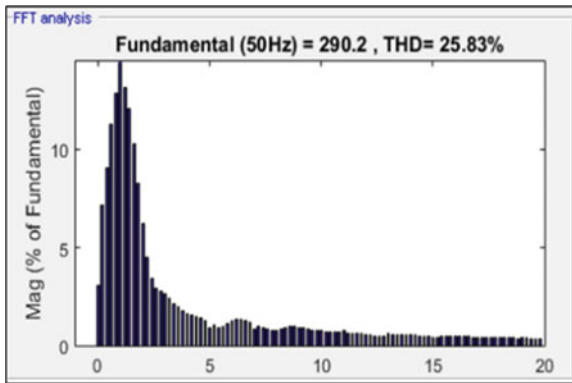
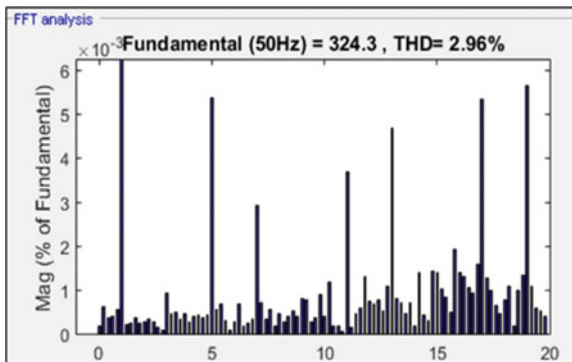


Fig. 10 FFT analysis of source current after connecting shunt APF controller in the system



4 Conclusion

The improvement in power quality has acquired additionally encouraging demand in power system network due to increasing nonlinear loads and power electronics

converters. This paper considers the upsides of the UPQC (series APF and shunt APF) for the power quality improvement during the load voltage distortion in the distribution system. The reference signal for the DG is been achieved by the combination of the power controller, the voltage controller, and the current controller and was given as input to the APF module. Thorough simulation tests utilizing MATLAB/Simulink have been studied. In this manuscript, the proposed UPQC is modeled for limiting the voltage distortion, reactive current, and the total harmonic distortion in the distributed framework. Simulation results show that the strategy is viable, which compensated the distortion in the voltages and compensates for the reactive current or the harmonics. And also exhibited good performance for total harmonic distortion which are within the prescribed limits as per IEEE standards. Different parameters like negative-sequence current, sag, and swell will be considered in the future work.

References

1. DD Sabin, A Sannino (2003) A summary of the draft IEEE P1409 custom power application guide. In: 2003 IEEE PES transmission and distribution conference and exposition (IEEE Cat. No. 03CH37495), vol 3. IEEE, pp 931–936
2. Photovoltaics DG, Storage E (2009) IEEE Application Guide for IEEE Std 1547™. In: IEEE standard for interconnecting distributed resources with electric power systems
3. Khadkikar V (2011) Enhancing electric power quality using UPQC: a comprehensive overview. *IEEE Trans Power Electron* 27(5):2284–2297
4. Dharmalingam R, Dash SS, Senthilnathan K, Mayilvaganan AB, Chinnamuthu S (2014) Power quality improvement by unified power quality conditioner based on CSC topology using synchronous reference frame theory. *Sci World J*
5. Dhakulkar SD, Manwar VA, Banole SG, Rakhonde BS (2017) Inspection of voltage sags and voltage swells incident in power quality problems-a review. *Int Res J Eng Technol* 4(1):1734–1736
6. Kumar D, Zare F, Ghosh A (2017) DC microgrid technology: system architectures, AC grid interfaces, grounding schemes, power quality, communication networks, applications, and standardizations aspects. *IEEE Access* 5:12230–12256
7. Beleiu HG, Beleiu IN, Pavel SG, Darab CP (2018) Management of power quality issues from an economic point of view. *Sustainability* 10(7):2326
8. Patel A, Mathur HD, Bhanot S (2018) An improved control method for unified power quality conditioner with unbalanced load. *Int J Electr Power Energy Syst* 100:129–138
9. Goswami G, Goswami PK (2020) Power quality improvement at nonlinear loads using transformer-less shunt active power filter with adaptive neural fuzzy interface system supervised PID controllers. *Int Trans Electr Energy Syst*
10. Ray PK (2018) Power quality improvement using VLLMS based adaptive shunt active filter. *CPSS Trans Power Electron Appl* 3(2):154–162
11. Rajagopal R, Palanisamy K, Paramasivam S (2018) A technical review on control strategies for active power filters. In: 2018 International conference on emerging trends and innovations in engineering and technological research (ICETIETR). IEEE, pp 1–6
12. Ochoa-Gimenez M, Garcia-Cerrada A, Zamora-Macho JL (2017) Comprehensive control for unified power quality conditioners. *J Mod Power Syst Clean Energy* 5(4):609–619
13. Kulkarni SV, Gaonkar DN (2017) Operation and control of a microgrid in isolated mode with multiple distributed generation systems. In: 2017 international conference on technological advancements in power and energy (TAP Energy). IEEE, pp 1–6

14. Mansor MA, Hasan K, Othman MM, Noor SZM, Musirin I (2020) Construction and performance investigation of three-phase solar PV and battery energy storage system integrated UPQC. IEEE Access
15. Choudhury SR, Das A, Anand S, Tungare S, Sonawane Y (2018) Adaptive shunt filtering control of UPQC for increased nonlinear loads. IET Power Electron 12(2):330–336
16. Jayalakshmi NS, Gaonkar DN, Jain SK (2019) Power smoothing method of PMSG based grid integrated wind energy conversion system using BESS/DSTATCOM. Int J Power Electron Drive Syst 10(4):1969

Industry 4.0 Researchers Computer Numerical Control Machine Tool to Manufacture Calligraphy Board



**B. Praveen, S. U. Abhishek, P. B. Shetty, J. Sudheer Reddy,
and B. A. Praveena**

1 Introduction

The current ongoing transformations in traditional manufacturing and industrial practices combined with smart technology is the basis of latest revolution in industry, also called as Industry 4.0. Industry 4.0 promotes large-scale use of IoT (Internet of Things), artificial intelligence, and machine learning in manufacturing industry.

CNC technology has been changed from human-machine interface logical programmer-oriented scenario to artificial intelligence energetic control model [1, 2]. In this, the role of personal computer, networking and decision making along with resource allocation, and rescheduling is widespread. In present CNC machines, various formulas are pre-fed into the controller to perform the necessary calculation for controlling the electronic controlled parts to perform simultaneous actions to achieve the necessary processes [3]. The use of various electronic devices as well as open source technology has boosted the range of ideas for researchers which can be implemented to build mini, affordable CNC machines. The GRBL files is an open-source firmware option for logic controllers which makes the research work possible because of its ability to be flexible and re-programmability. The basic commands for programming in the GRBL hex files are in C++ command format. 3-D model of the products can be designed in various open-source CAD/CAM software programs. It also has in-built features to obtain G-codes directly for various operations

B. Praveen · S. U. Abhishek · P. B. Shetty · J. Sudheer Reddy · B. A. Praveena (✉)
Department of Mechanical Engineering, Nitte Meenakshi Institute of Technology, Yelahanka,
Bangalore, Karnataka 560064, India
e-mail: praveen.ba@nmit.ac.in

P. B. Shetty
e-mail: pb.shetty@nmit.ac.in

J. Sudheer Reddy
e-mail: sudheerreddy.j@nmit.ac.in

that can be performed by the CNC machine [4]. The latest modifications in designing software programmes provide more effective tool-path selection criteria for reduced finishing time with greater efficiency in finishing the product.

This paper incorporates the latest developments and applications of Computerized Numerical Control to build a mini-CNC. It has the flexibility to redesign for education and research which can be used for any other Industry 4.0 manufacturing applications like router, grinder, and possible research in calligraphy. The basic calligraphy knowledge is a very essential for pre-school children as a preliminary skill requirement for school competence [5]. It provides a firm foundation for students for effective learning in later years as adults to quickly comprehend words. This knowledge is reinforced from a young age in the form of various activities, of which the most effective approach is through those involving fun for the children [6, 7]. The calligraphy board can be considered as an effective tool for achieving proper letter knowledge for pre-schoolers while maintaining the guise of a fun play tool.

The present mini-CNC with an open structure facilitates researchers to learn the construction of the machine and maintain the machines on their own skill without any difficulty. In this, CNC machine inheritance of specific machine compatible controllers are replaced with network-based controllers. Young researchers can further develop this by incorporating the technology of Internet of things of industrial engineering [2, 8–11]

2 Methodology

To build the low-cost CNC milling machine, three stepper motors and their respective stepper motor drivers, the poly vinyl chlore frame parts, driving leads, leader rods, and electronic devices and accessories are procured. The technical details of the assembled CNC are *X*-axis travel, *Y*-axis and *Z*-axis travel which is as per the requirement of researcher. The stepper motors are of ks42sth40. The lead screw is of stainless steel with copper nut. The spindle motor is 300 W/12000 rpm. The power supply required is basic 24 V/10A/input 100 ~ 240VAC/50/60 Hz. The stepper motor drivers are TB6560/2.9A –3.4A, 24 V/DC/ 1/16 stepping and decay devices. The assembled CNC machine is depicted in Fig. 1.

The stepper motors move in fractions of a revolution which can be varied as per requirement to generate as smooth motion as required keeping research requirement of accuracy in mind. This is known as micro-stepping. The amount of precision in in the order of 1.8 degree divided by 256. This micro-stepping gives stable torque in manufacturing [12–16]. Figure 2 shows the TB 6560 stepper motor which is used in the machine for controlling each individual axis of motion. Figure 3 shows the TB 6560 stepper motor driver [17–22].

The driver has six step switches, wherein SW1 is used to control the amperage from 0.3 A to 3.0 A, SW2 controls the stop current at 20% in closed switch position and 50% in open switch position, SW3 and SW4 gives the stimulation mode, and SW5 and SW6 varies the decay setting. The +24 V and ground points on the driver

Fig. 1 Mini, built-in CNC machine



Fig. 2 Stepper motor



is used to supply power in DC form. Motor stage A windings are restrained from A positive and A negative points. In the case of stage B windings, the B+ and B- points on the driver are accessed. The stepping motion is controlled using the CLK+ and - pins which is also known as pulse in other drivers and is presented as PUL, while the direction is controlled using the CW pins (+ve for anti-clockwise and -ve for

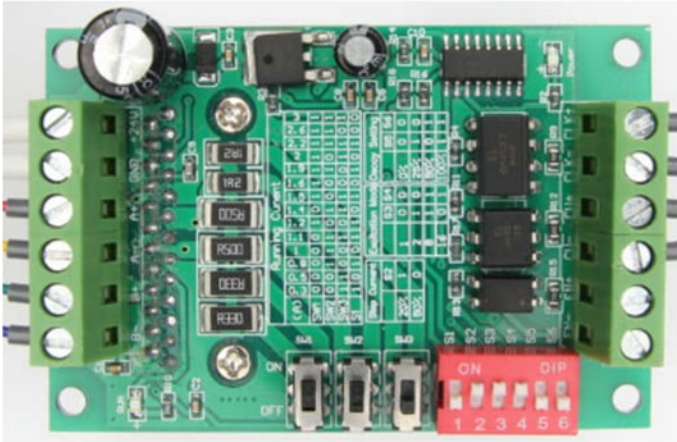


Fig. 3 TB 6560 stepper motor driver

clockwise). The EN+ and EN– are not used in the machine as the control is Arduino based. Figure 4 shows the CNC control electronics arrangement.

Once all the electronic components are connected, the spindle motor is mounted on the z-axis. The DC spindle motor shown in Fig. 5 is powered by direct current (DC) electricity via switched-mode power supply (SMPS) and mounted

Fig. 4 CNC control electronics arrangement

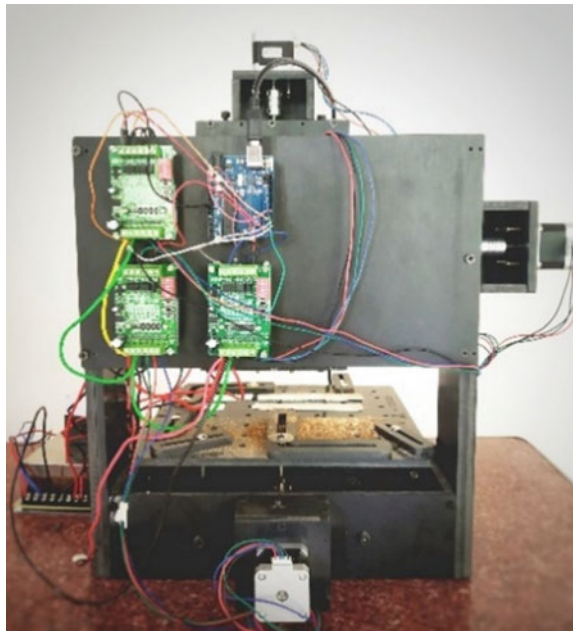


Fig. 5 DC spindle motor



on the z-axis of the machine. The speciality of this motor is that it is brushless. The brushless motors have high power-to-weight ratio, high speed, good electronic control, and low upkeep cost [23, 24].

For calligraphy demonstration a free and open source vector, graphics editor called Inkscape is used. It offers a rich set of topographies and is broadly used for creative calligraphy. It uses direction graphics to allow for piercing and renderings outputs at unrestrained resolution. Inkscape utility window is shown in Fig. 6.

The saved SVG file format is imported to any CAD/CAM software. In the CAD/CAM software, a stock is created on top of which the traced bitmap is copied. All the parameters related to the calligraphy like feed rate, tool type, and orientation are appropriately submitted. The tool path is created followed by G-Code generation.



Fig. 6 Inkscape window

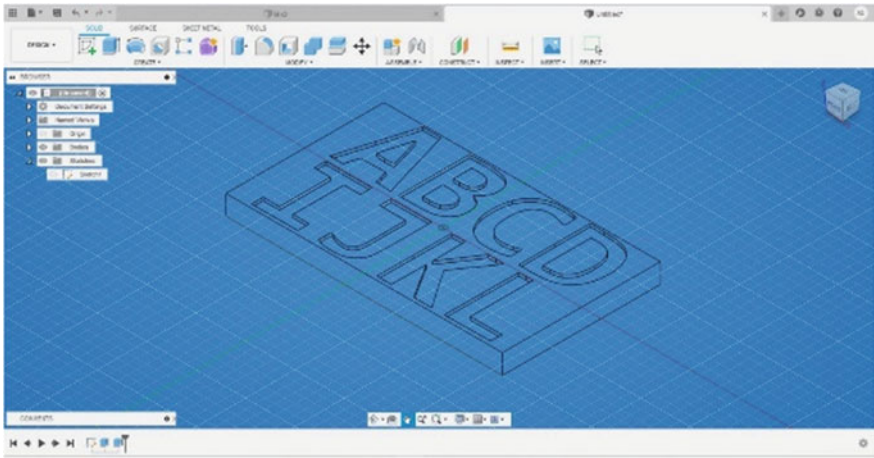


Fig. 7 CAD/CAM software window

The tool movement can be visualized in the software as shown in Fig. 7 to make sure that the operation takes place as desired.

Calligraphy is more communicating than just giving the output. Once the modelling is completed, the G-Code is sent to the CNC machine through application control software UG-C sender. The CNC machine itself runs this firmware which can interpret what the machine application software is saying and consequently drives the stepper motors to operate the mill. Figure 8 shows each stages of connectivity in the calligraphy.

G-Code files once generated, it is then visualized using a CAD/CAM software which then acts as a link for the controller for the next processing. Meanwhile hex file is uploaded on to the controller using Xloader. Xloader is a program that can be used to upload hex file to the microcontroller. Figure 9 shows the application of Xloader window.

UGC is a self-contained G-Code podium used for networking with CNC controllers like grbl. UGC sender is an autonomous and well compatible to Java application for UGS rendition. As of writing this, v1.1 is the latest stable build of UGS. Once the program is loaded, connection parameters are set and by clicking 'Open' connection to the mill is achieved. Baud is set to 115,200. Figure 10 is a sample of G-Code podium to upload Universal G-Code.

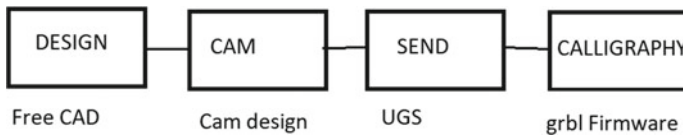


Fig. 8 Calligraphy steps

Fig. 9 Xloader window

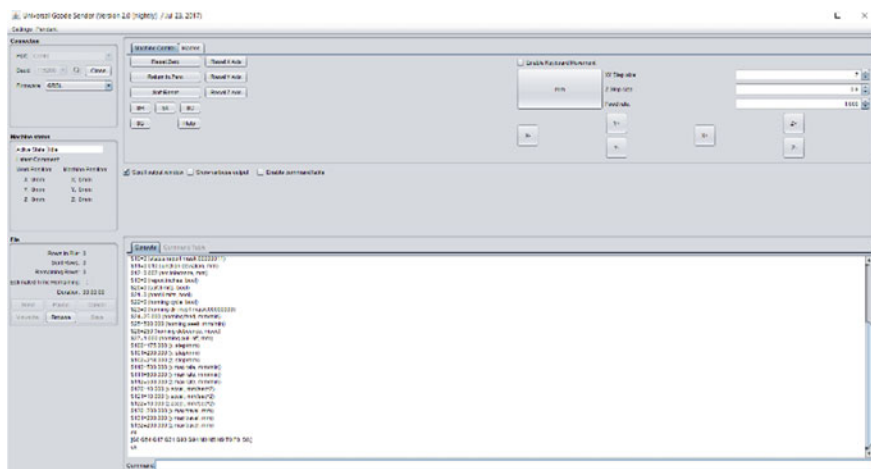
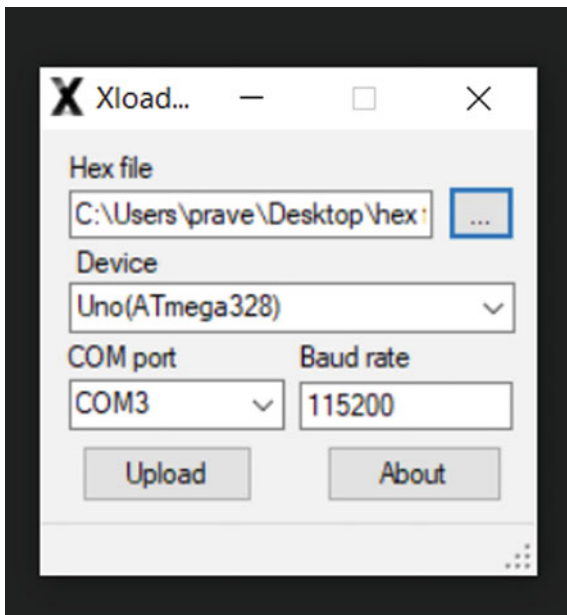


Fig. 10 Universal G-Code podium

The top layer labels allow us to enter random G-Code, upload files, operate the calligraphy command interactively, and also facilitates to send function command strings. The bottom layer terminal output and headway G-Code file as a sequence of instructions. By typing \$\$ in command line, we get access to the settings saved in the controller, and it allows amendment. The settings are directly saved into the

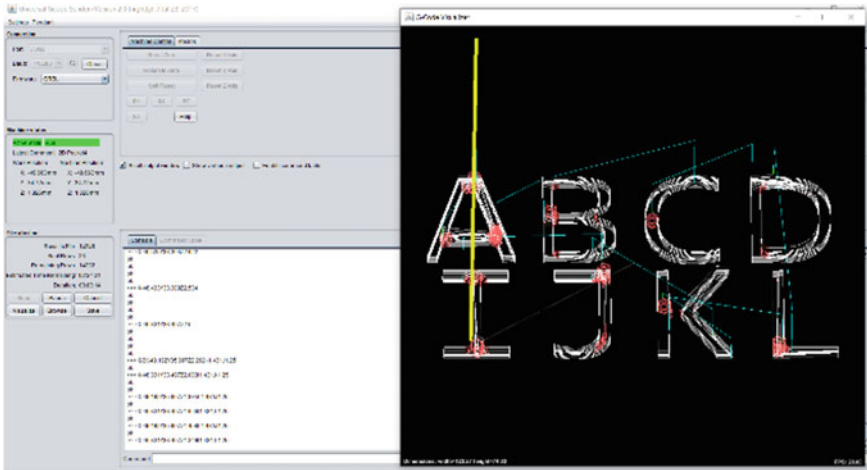


Fig. 11 G-Code visualizer window

ATmega328. Before starting the calligraphy operation, we can always visualize the tool path in the universal G-Code sender. This helps to double check that our program is correct is shown in Fig. 11.

Prior to starting the actual calligraphy operation, the machine has to be calibrated for accuracy. It is calibrated using the concept of travel in mm corresponding to the rotation of the motor. It is calculated based on the following formula, Steps per mm = [(steps per/rev)/mm per rev]/micro-step fraction. For instance, with a steps per rotation as 180 and the motor that travels 38 mm for one rotation having 1/2 micro stepping will have, steps per mm = $(180/38)/(1/2) = 9.47$ steps/mm. The machine control layer in UGC facilitates the instructions to move the mill to the calculated millimetre distance. The distance travelled is physically measured with a scale and compared with the input given. If it is satisfied, calibration is done, and if not, the mill is reset, and the procedure is repeated. Now the original setting is modified by entering the respective axis (step/mm) in command and changing its pre-set value. This procedure needs to be done for each axis.

The completion of calibration indicates that the CNC machine is ready to use. The calligraphy letters on the board is milled by following all the procedure from the tracing of alphabet to the generation of G-Codes. These generated G-Codes are sent to the machine using universal G-code sender. A soft and lightweight wood piece is preferred. The size of the wood is within the working dimensions of the CNC machine working platform. The wood piece is tightly secured and firmly clamped to the bed of the machine before the beginning of the operation. The program is run, and the final calligraphy board as per the design is produced.

This digitalised product was possible through the combination of latest electronic devices and smart algorithms for data analysis as per the concepts of Industry 4.0.

3 Results and Discussion

The latest advances in the field of electronics and software has led to the possibility of compaction of huge, expensive ventures like the CNC into any kind of manufacturing machine tool to reinforce research in Industry 4.0. It provides a flexibility for potential entrepreneurs and micro-enterprises to design and manufacture products of their choice. The demonstrative CNC machine presented in the paper with its application to calligraphy is built at a considerably lower price while maintaining the accuracy of the finished product. The conceptual calligraphy board, a teaching aid, is aimed at pre-school students to improve their letter identification and writing skills by consistent repetition of tracing the letters on the board. This product can be custom made for any calligraphy including the artistic Persian calligraphy very quickly. On continuous usage of this calligraphy technique, the children develop a better pincer grasp and strengthen small muscles such as lumbricals and interosseous muscles of the hand.

4 Conclusion

This research work addresses the perception of customized CNC machines for any kind of production machines or robots using the technology of Industry 4.0 at an affordable price for researchers. The presented concept in this paper includes three steps, namely the electronic devices assembling, the software integration, data acquisition, and finally the model realisation. In the future work, it is possible to extend the Industry 4.0 model to embrace more types of machining processes. In addition, it is possible to include artificial intelligence concepts with decision-making algorithms.

References

1. Vijayakumar R, Ramesh C, Ezhilarasan T, Praveenkumar P, Mohan R (2020) An interactive approach towards the development of portable laser cutting machine. *Mater Today: Proc* 1–4. <https://doi.org/10.1016/j.matpr.2020.04.377>
2. Sathyakumar N, K. Prasath Balaji, R. Ganapathi, SR Pandian (2018) A build your own three axis PCB Milling machine. *Mater Today Proc* 5(2018):1–10. 24404-24413
3. Fisher D, Hofmann R (2007) CNC machining plastic injection mold plates in the classroom. In: *Proceedings—frontiers in education conference, November 2007. S2A-12 .37th ASEE/IEEE Frontiers in education conference S2A*, pp 12–17. <https://doi.org/10.1109/FIE.2007.4417886>
4. Suhada R, Ariyanti S Fajar A, Komalasari A (2018) Training autodesk fusion 360 for teenage of senior high school graduates in improving ability in disasters. *ICCD 1:285–289*. <https://doi.org/10.33068/iccd.Vol1.Iss1.43>
5. Fears N, Lockman J (2020) Case- and form-sensitive letter frequencies in children’s picture books. *Early Child Res Q* 2020:370–378
6. Park B, Chang H, Park S (2018) Adoption of digital devices for children education: korean case, telematics and informatics, pp 2–28

7. Park H-S, Qi B, Dang Duc V, Park D (2018) Corrigendum to Development of smart machining system for optimizing feedrates to minimize machining time. *J Comput Des Eng* 5:299–304
8. Pandian S, Raj Pandian S (2014) A low-cost build-your-own three axis CNC mill prototype. *Int J Mech Eng Rob (IJMER)* 2(1):5–11. ISSN (Print): 2321-5747
9. Madne AN, Hande S, Radake R, Manthan Shambharkar (2020) Design and fabrication of 3 Axis CNC PCB milling and drilling machine—a review-. *Int J Sci Dev Res (IJSDR)*, 5(1):147–149. ISSN: 2455-2631
10. Gouveia RM, Silva FJG, Reis P, Baptista APM (2016) Machining duplex stainless steel: comparative study regarding end mill coated tools. *Coatings* 6:51–81
11. Zhong RY, Xua X, Klotz E, Newman ST (2017) Intelligent manufacturing in the context of industry 4.0: a review. *Engineering* 3:616–630
12. Santosh Kumar DS, Praveen BA, Kiran Aithal S, Kempaiah UN (2015) Development of pineapple leaf fiber reinforced epoxy resin composites. *Int Res J Eng Technol* 2(3)
13. Praveena BA, Shetty BP, Vinayaka N, Srikanth HV, Singh Yadav SP, Avinash L (2020) Mechanical properties and water absorption behaviour of pineapple leaf fibre reinforced polymer composites. *Adv Mat Proces Technol* 6(4):1–16. <https://doi.org/10.1080/2374068X.2020.1860354>
14. Praveena BA, Shetty BP, Sachin B, Singh Yadav SP, Avinash L (2020) Physical and mechanical properties, morphological behaviour of pineapple leaf fibre reinforced polyester resin composites. *Adv Mat Proces Technol* 6(4):1–14. <https://doi.org/10.1080/2374068X.2020.1853498>
15. Praveena BA, Shetty BP, Akshay AS, Kalyan B (2020) Experimental study on mechanical properties of pineapple and banana leaf fiber reinforced hybrid composites. *AIP Conf Proc* 2274(1):030015. <https://doi.org/10.1063/5.0022381>
16. Shashidhar S Patil, Praveen BA, Dr. Kempaiah UN, Dr. Adarsha H (2017) Fabrication and characterization of Kevlar/Jute Reinforced Epoxy. *Int Res J Eng Technol (IRJET)* 4(9)
17. Anilthota HG, Praveena BA (2017) Processing and characterization of green composites using sisal and palm fibers. *Int J Eng Sci Comput IJESC* 4262–4265
18. Korisidda, Mahadev Prasad C, Mahantesh B, Sheregar SS, Praveena BA (2018) Study on mechanical properties of bio-composites. *Int J Sci Res Develop* 6(2):1–5
19. Praveena BA, Vijay Kumar S, Manjunath HN, Sachin B, Singh Yadav SP, Lochan BR, Arun Kumar GL, Reddy JS (2021) Investigation of moisture absorption and mechanical properties of natural fibre reinforced polymer hybrid composite. *Mater Today Proc* 45(9):8219–8223. <https://doi.org/10.1016/j.matpr.2021.04.254>
20. Singh Yadav SP, Kumar SV, Avinash L, Buradi A, Praveena BA, Vasu VK, Vinayaka N, Kumar K (2021) Development of 3D printed electromyography controlled bionic arm. In: *Sustainable Machining Strategies for Better Performance. Lecture Notes i Mechanical Engineering*. Springer, Singapore. https://doi.org/10.1007/978-981-16-2278-6_2
21. Nagaraja S, Kodanda R, Ansari K, Kuruniyan MS, Afzal A, Kaladgi AR, Aslfattahi N, Saleel CA, Gowda AC, Bindiganavile Anand P (2021) Influence of heat treatment and reinforcements on tensile characteristics of aluminium AA 5083/silicon carbide/fly ssh composites materials. *Materials* 14(18):5261. <https://doi.org/10.3390/ma14185261>
22. Aftab SG, Faisal A, Hussain H, Sreedhara B, Babu NR, Praveen BA (2021) Structural analysis of human femur bone to select an alternative composite material. *Mat Today Proc*. <https://doi.org/10.1016/j.matpr.2021.08.197>
23. Praveena BA et al (2021) *IOP Conf Ser.: Mater Sci Eng* 1013 012004
24. Praveena BA (2021) *IOP Conf Ser.: Mater Sci Eng* 1013 012006

ANN-Based Wheat Crop Yield Prediction Technique for Punjab Region



Nishu Bali and Anshu Singla

1 Introduction

Agriculture is an important occupation in almost all the countries of today. The ever growing population of the world and unexpected changes in climatic trends is mounting pressure on global food supply to assure food for everyone. Thus, efforts are made to find ways of increasing the production which in turn requires correct and timely estimation of crop yield.

Crop yield is dependent on different environmental parameters which vary nonlinearly and in turn make the estimation of yield a complex procedure. Two general approaches are generally followed for crop yield estimation—crop growth model and statistical models. Crop growth models are quite efficient in estimation of yield, but they require thorough knowledge of crop physiological behaviour at different stages of growth, and their results are sometimes not transferrable or reusable on the fields due to varying environmental conditions. In statistical models, empirical equations are formulated with yield as dependent variable and factors affecting the yield as independent variables. The emergence of machine learning in the field of agriculture has given strength to the statistical method of yield estimation.

N. Bali · A. Singla (✉)

Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India

e-mail: anshu.singla@chitkara.edu.in

N. Bali

e-mail: nishu.bali@chitkara.edu.in

2 Machine Learning in Agriculture Domain

In machine learning techniques, machine is made to learn through the given data. On the basis of learnings attained by the machine, it predicts or classifies the unseen data. Machine learning has shown its role in various fields of agriculture such as crop selection, in assessing the effects of various climatic and soil parameters, crop disease detection and prediction, precise and only required use of water for irrigation and many more. Among these, yield estimation is one of the most important field in which machine learning has shown an appreciable contribution. There are broadly two types of machine learning algorithms: supervised and unsupervised. In supervised machine learning techniques, the machine is trained for already known outputs based on some inputs. The examples are linear regression, logistic regression, support vector machine, decision tree, Bayesian logic and neural networks. In unsupervised techniques, the machine is given raw data and is made to identify the patterns and classify the data according to the identified patterns or trends. These include clustering, KNN and Apriori algorithm. Whether supervised or unsupervised, the machine is first trained on some training data and is made to use that learning for prediction or classification on unseen data. Both types of techniques have contributed in the study of crop yield estimation. The present study is done with an objective to use ANN technique for wheat crop yield estimation for specific region of Punjab and to compare its efficiency with multivariate regression technique.

In next section, we present a brief literature on the works already done in the field by the researchers. This will be followed by dataset and methodology and experimental results in fourth and fifth section, respectively.

3 Related Work

The effect of climatic variations on yield of wheat crop was studied using various machine learning techniques. Support vector regression was compared with other approaches used with NDVI index and results proved that SVR outperformed the latter approach with $R^2 < 0.46$ [1]. In another study, linear regression model was used for quantifying the effect of different meteorological parameters on the rice yield in district Raipur, India. It was found that maximum temperature increase had not much detrimental effect at tillering stage of plant growth but had widespread effect at flowering stage. Minimum temperature was within the cardinal limits so was not much affecting the yield. Rainfall and sunshine were found to be prominent parameters affecting the yield [2]. Remote sensing data was used for finding the efficiency of machine learning methods for predicting yields and results concluded that a combination of sensor technology along with machine learning techniques can give even better results [3]. The use of deep learning techniques like CNN was explored in a study on orchards where fruit bearing capability of bitter melon crop was analysed based on the leaves of plant gathered from Ampalaya farms [4]. In yet another study

on orchards, two BPNN models were explored for two phases of season, opening and ripening period, for estimation of yield of fruit crops based on image analysis. Satisfactory results obtained proved the efficiency of proposed approach in area of yield estimation [5]. A comparative analysis of four machine learning techniques for corn yield estimation was done in Iowa State. Results gave good results especially for deep learning which showed most stable results [6]. In another study, spiking neural network technique was used for spatio temporal analysis of data for crop yield evaluation. The study made pre-harvest yield prediction six weeks prior to harvest with an accuracy of 95.4% and average error of prediction of 0.236 t/ha and correlation coefficient of 0.801 using a nine-feature model [7]. In another study, three machine learning techniques, counter-propagation artificial neural networks (CP-ANNs), XY-fused networks (XY-Fs) and Supervised Kohonen Networks (SKNs), were compared in performance for finding the variations of wheat yield based on multilayer soil data and satellite imagery crop growth characteristics. Results showed that in low yield class varieties, the accuracies obtained were 91%; whereas for average and high pitched yield varieties, the accuracies were 70% and 83%, respectively. Among the three machine learning models, SKN showed highest accuracy of 81.65% proving it to be the best model [8]. ANN was explored in another study for efficiency in rice yield prediction for the years 1998–2002. The results gave high accuracies of 97.5% with a sensitivity of 96.3 and specificity of 98.1 [9]. The effect of customization of ANN model on its efficiency for wheat yield estimation was studied. The customized model was compared with default ANN model and MLR technique. Significant improvement in efficiency was found in customized ANN model with higher R2 statistics and lower percentage errors [10]. Another extensive study was done to compare various machine learning techniques for crop yield estimation of multiple crops. Results favored M5-Prime and KNN techniques with lowest error values [11]. In yet another study, the architecture of ANN model was varied by wavering number of hidden layers used, and the effect of variations on efficiency of model was evaluated for finding the effect of various predictor variables related to soil and climate on yields of various crops [12]. A new hybrid approach based on modern representation learning ideas was proposed to predict county-level soybean crop yield. A new dimensionality reduction technique was used to compensate for lack of sufficient training data. Deep learning architectures like CNNs and LSTMs were used to predict the crop yield. Experimental results showed that proposed model had outperformed the customary remote sensing centered techniques in efficiency [13]. Crop yield prediction in area of greenhouse operations was studied using an intelligent system called EFuNN (Evolving Fuzzy Neural Network) for yield estimation of tomato crop. Results gave weekly prediction with an accuracy of 90% [14]. Customization of ANN models was explored in yet another study in which 11 varied ANN models with different number of neurons in hidden layers were tried, and optimum model was selected. ANN-MLP model based on conjugate gradient back propagation algorithm reported lowest MAPE making it the preferred or optimum model [15].



Fig. 1 Map showing Ludhiana Region of Punjab [16]

4 Dataset and Methodology

4.1 Study Area

The present study is focused on one of the main agriculture-based districts of Punjab, Ludhiana. The district is spread across a geographical area of 3767 km² and has 3 lakh ha of net sown area out of which almost 100% is doubly cropped and in some cases, three crops are sown in a year (Fig. 1).

Ludhiana has always been a role model for other districts as far as adoption of advanced techniques in agriculture is concerned. Wheat is one of the most important crop sown in the area. Around 2.57 lac ha of the area is devoted to wheat cultivation which contributes to 50.26 qt/ha of productivity of the crop from district. The data used in the study is mainly collected from statistical abstract of Punjab issued by Economic advisor to Government, Punjab. An extensive data of 43 years from 1970 to 2010 has been used for the study. The climatic data was obtained from meteorological department of Punjab.

4.2 Methodology

The data obtained from various sources was pre-processed. The processed data was then partitioned into train and test data. Although there is no exact protocol to divide the data into test and train, but we used 1:9 ratios, i.e., 10% data was taken as test data and 90% was taken as train data. Model after being trained with the training data was tested for the accuracies of prediction.

4.2.1 Data Pre-processing

The data obtained from various sources was scrutinized to find the occurrence of any null values in the data which were substituted with appropriate statistical values. As machine learning techniques can only work on numeric data, the features selected for the study were examined to find any non-numeric parameters in the study. Out of the annual data obtained from reports, the data for specific months actually used for the wheat cultivation ranging from October (sowing period) to April (harvest period) was selected and compiled. Environmental parameter values (maximum and minimum temperature, maximum and minimum relative humidity, rainfall and evaporation) pertaining to this period were selected and stored in an excel sheet. The data was normalized and scaled.

4.3 Machine Learning Techniques

The crop yield of wheat in Punjab region has been performed by employing artificial neural network. Present study compared the results obtained with the machine learning technique-multivariate linear regression. Sections 4.3.1 and 4.3.2 briefly describe both the techniques.

4.3.1 Multivariate Linear Regression

Linear regression is a supervised machine learning technique in which target value is determined based on some independent variables related to the target variable. Regression technique is mostly used for finding the relationships between the target and independent variables. As this technique deals with linear relationships between the variables, it is called linear regression. The function of a linear regression is defined as:

$$y = \theta_0 + \sum_{i=1}^n \theta_i * x_i \quad (1)$$

where x is input variable and y is output or target variable.

θ_0 : intercept.

θ_i : Coefficient of x_i .

The model is first trained using training data and during training, the best line that fits the data values is accepted. The model gets the best regression line by varying the values of θ_0 and θ_i .

4.3.2 Neural Networks

Artificial neural networks is a machine learning technique in which machine is made to behave and think like a human brain. Like human brains, an artificial neural network consists of neurons which are spread out in different layers. Broadly, any ANN consists of an input layer through which data is fed to the network, an output layer at which the output in the form of prediction or classification is obtained and a hidden layer which is may or may not be the part of network. The data is fed on input layer where each input is given some weight that signifies the importance of that input parameter to the study. The weighted mean of all the inputs is passed to the next layer. Here comes the task of activation function. An activation function acts like a filter to remove unnecessary information from the previous layer and pass on only the necessary or required part to the next layer for further processing. It can be taken as a simple step function to switch on or off a neuron output. Mostly, nonlinear activation functions are used in neural networks so that they can deal with complex problems and data such as images, voice and data with high dimensionality. Also, nonlinear activation functions can deal with backpropagation which is important for the improvement of the network and is difficult to be dealt by linear activation function. Finally, the output is generated at the output layer. Number of neurons in each layer and number of hidden layers used in the network are decided on the basis of number of inputs and the type of problem need to be solved (Fig. 2).

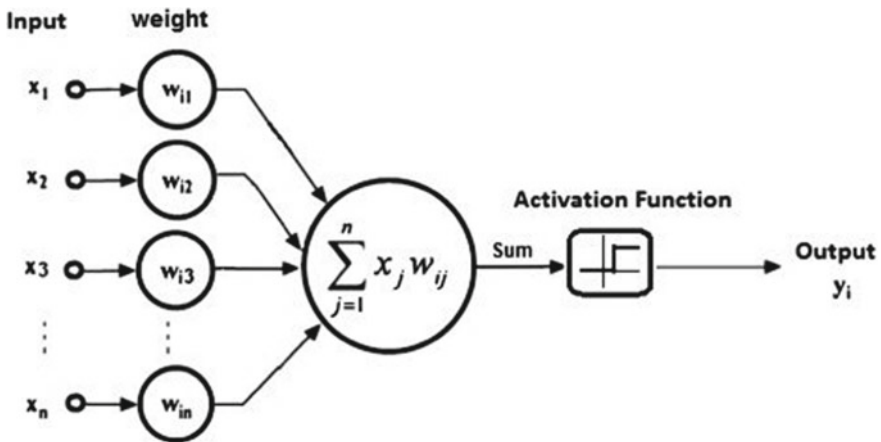


Fig. 2 Layout of a neural network [17]

5 Experimental Results

In the present study, ANN was employed on the data obtained from various sources of Punjab for Wheat crop yield prediction. For comparison purposes, another machine learning technique, multivariate linear regression was applied on the same data. The results obtained on applying both the models are discussed in Sects. 5.1 and 5.2.

5.1 Multivariate Linear Regression Model

In linear regression model, environmental parameters, were taken as independent variables, whereas yield obtained was the dependent variable. 43 climatic features were considered as the independent variables, whereas yield to be determined was taken as the dependent variable. The data was randomly selected by the model during training and testing, and the predicted and actual values obtained for the test data for various years are as shown in Table 1 and Fig. 3.

Table 1 Predicted and actual yield values for linear regression model

Years	Predicted	Actual
1971	2178.697	2800
1997	1987.084	4563
2000	4640.336	4563
2010	2610.989	4724
2009	5396.709	4689
1991	2931.266	4563
1986	4026.307	3715
1988	4537.306	3720
1973	1910.199	273

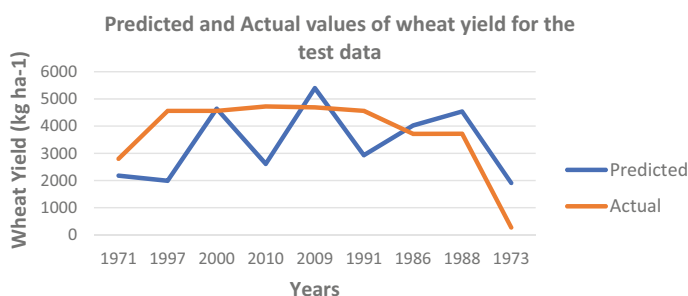


Fig. 3 Graphical variation of predicted and actual values for regression model

Table 2 Values of evaluation metrics for linear regression model

Technique	R-Square	Adjusted R-Square	RMSE	MAE
Linear regression	1.0	1.0	1337	1075

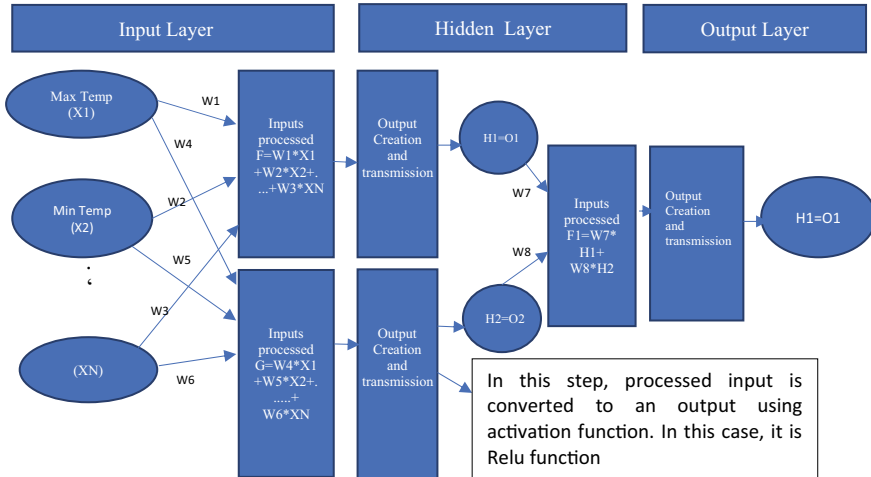


Fig. 4 Artificial neural network model used in study

The values of various evaluation metrics like R-square, adjusted R-square, RMSE and MAE are as shown in Table 2.

5.2 Artificial Neural Network

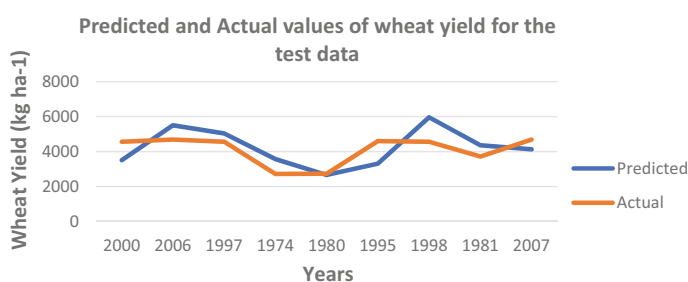
Artificial neural network model used in the study has been shown in Fig. 4. There are two ways of initializing a neural network model—defining each layer one by one or defining a graph. We used the sequential function of python library with no parameters to design the model layer by layer manually. The model was designed with three layers, one input layer, hidden layer and an output layer. Stochastic gradient descent algorithm was used for training the model and rectified activation function (Relu), and one of the most widely used activation function for nonlinear problems was used in all the layers. As the number of features was 43, so the input layer was fed with 43 neurons. The number of nodes in hidden layer was calculated as a mean of neurons in input and output layer and was taken as 19. Model was made to run for 2500 epochs in a batch size of 10.

The predicted and actual values of wheat yield as obtained from the ANN model are as shown in Table 3 and Fig. 5.

The values of various evaluation metrics obtained are as shown in Table 4.

Table 3 Predicted and actual values of yields for ANN model

Years	Predicted	Actual
2000	3506.285	4563
2006	5505.499	4690
1997	5041.7134	4563
1974	3569.511	2720
1980	2664.287	2730
1995	3315.5383	4600
1998	5961.345	4562
1981	4357.9854	3716
2007	4127.3887	4688

**Fig. 5** Graphical variation of predicted and actual values for ANN model**Table 4** Values of evaluation metrics for ANN model

Technique	RMSE	MAE
ANN	819	744.3

Results obtained clearly indicate that ANN technique has shown much closer predictions as compared to multivariate linear regression technique. Also, the values of evaluation metrics have shown that the RMSE values obtained in case of ANN technique are quite less than those in linear regression.

6 Conclusions

The present study employed ANN technique for wheat crop yield prediction in an area of Punjab. The closeness of predicted values obtained in results to actual yield values have shown good prospects for ANN as a crop yield prediction model. For comparison purposes, on examining the values of evaluation metrics, RMSE and MAE, obtained in ANN and linear regression technique, it is clearly visible that ANN has shown much less error as compared to regression technique. This further

proves that neural networks can be a better choice when dealing with nonlinear behaviours which are inherent in the study. As this study pertains to the areas of Punjab, application of various machine learning techniques in this area still needs to be explored. For future scope, many other climatic parameters related to wind and soil are not included in study which can be further investigated in future studies. Also the advanced techniques of ANN in the form of deep learning can be explored in hybridization with other techniques of AI in the said area.

References

1. Kamir E, Waldner F, Hochman Z (2020) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J Photogramm Remote Sens* 160:124–135
2. Jain A, Chaudhary JL, Beck MK, Kumar L (2019) Developing regression model to forecast the rice yield at Raipur condition. *J Pharmacogn Phytochem* 8(1):72–76
3. Chlingaryan A, Sukkariéh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput Electron Agric* 151:61–69
4. Villanueva BM, Salenga MLM (2018) Bitter melon crop yield prediction using machine learning algorithm. *Int J Adv Comput Sci Appl* 9:1–6
5. Cheng H, Damerow L, Sun Y, Blanke M (2017) Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. *J Imaging* 3(1):6
6. Kim N, Lee YW (2016) Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. *J Korean Soc Surv Geod Photogramm Cartogr* 34(4):383–390
7. Bose P, Kasabov NK, Bruzzone L, Hartono RN (2016) Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Trans Geosci Remote Sens* 54(11):6563–6573
8. Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM (2016) Wheat yield prediction using machine learning and advanced sensing techniques. *Comput Electron Agric* 121:57–65
9. Gandhi N, Petkar O, Armstrong LJ (2016) Rice crop yield prediction using artificial neural networks. In: 2016 IEEE technological innovations in ICT for agriculture and rural development (TIAR), July 2016. IEEE, pp 105–110
10. Shastry KA, Sanjay HA, Deshmukh A (2016) A parameter based customized artificial neural network model for crop yield prediction. *J Artif Intell* 9:23–32
11. González Sánchez A, Frausto Solís J, Ojeda Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction
12. Dahikar SS, Rode SV (2014) Agricultural crop yield prediction using artificial neural network approach. *Int J Innov Res Electr, Electron, Instrum Control Eng* 2(1):683–686
13. You J, Li X, Low M, Lobell D, Ermon S (2017) Deep Gaussian process for crop yield prediction based on remote sensing data. In: Thirty-first AAAI conference on artificial intelligence, Feb 2017
14. Qaddoum K, Hines EL, Iliescu DD (2013) Yield prediction for tomato greenhouse using EFuNN. *ISRN Artificial Intelligence*
15. Ghodsi R, Yani RM, Jalali R, Ruzbahman M (2012) Predicting wheat production in Iran using an artificial neural networks approach. *Int J Acad Res Bus Soc Sci* 2(2):34

16. Ludhiana District (2020, June 29). Retrieved from https://en.wikipedia.org/wiki/Ludhiana_district
17. Activation Function (2020, June 29). Retrieved from <https://www.quora.com/What-is-meant-by-activation-function>

Trace Element Analysis of Some Vegetables by PIXE Technique



S. Abdul Sattar , B. Seetharami Reddy , and K. Ramnarayana 

1 Introduction

In developing nations like India, vegetables form a part of staple diet in practically all the families. While the necessary amount of vegetables in our day by day diet must be 300–350 gm for each individual [1], Butt and Haq [2] have assessed that only 80–90 gm of vegetables are being consumed daily by each individual. Accepting that an individual consumes one of the vegetables like brinjal, potato or gourd every day, it is interesting to examine what segment of recommended dietary allowance (RDA) of significant elements like Zn, Mn, Se and so on are met from this source.

Certain toxic elements are known to be hazardous even at very low concentrations. Depending upon the nature of the soil and pollutants in the area where the vegetables are grown, the trace elements may be present in them above accepted levels. Hence, there is a need to check whether any of them are present in these vegetables in unsafe levels.

Numerous analytical techniques like atomic absorption spectrometry, voltammetry, inductively coupled plasma optical emission spectrometry, neutron activation analysis and proton induced X-ray emission spectroscopy are utilized for estimation of trace element concentrations. In the present investigation, PIXE technique was employed for the analysis of vegetables because of its multi-elemental capability with excellent sensitivity and detection limits (1ppm or less) across a wide range of atomic numbers. It has been broadly utilized in different fields.

S. Abdul Sattar (✉)
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: sheik.abdul.sattar@nmit.ac.in

B. Seetharami Reddy
Andhra University, Visakhapatnam, India

K. Ramnarayana
Sri A.S.N.M. Government College, Palakol, India

Some investigations of trace element concentrations in vegetables are found in literature [3–10]. Mauro et al. [3] explored the level of trace elements in vegetables and fruits cultivated in Southern Italy. Hana et al. [4] determined the concentrations of essential elements and toxic heavy metals from Saudi Arabian fruits and vegetables by ICP-OES. Rahman [5] employed PIXE technique to quantify trace elements in some vegetables and fruits. Fahad et al. [6] analyzed the concentrations of 18 different elements in five selected vegetables through PIXE technique. Basha et al. [7] analyzed trace metals in vegetables and fruits cultivated around the surroundings of Tummalapalle uranium mining site, Andhra Pradesh, India, using inductively coupled plasma-mass spectrometer (ICP-MS). Oyekanmi et al. [8] measured the contents of trace metal in some green leafy vegetables cultivated in some selected areas of Osun state using atomic absorption spectrophotometer. Rahaman et al. [9] measured trace elements in vegetables of southwestern Bangladesh using PIXE. Stefania Papa et al. [10] determined the concentrations of six different trace metals in various fruits and vegetables by using AAS.

Thus, it can be seen that in spite of the fact that PIXE is an accurate and sophisticated analytical technique, it has not been widely used, especially in India, for the investigation of trace element levels in vegetables, though it is important from the view point of public health. Hence, to fill this gap, in the present study, concentration of trace elements was measured in some commonly used vegetables and tested whether any of them are present in toxic levels utilizing PIXE procedure. This work was carried out using 3MV Pelletron accelerator facility at the Institute of Physics, Bhubaneswar. In the current investigation, trace element analysis of nine vegetables, tomato, little gourd, ridge gourd, potato, brinjal, carrot, beet root, bitter gourd and lady finger was under taken.

2 Experimental Details

Eighty samples of eight vegetable species were gathered from around Amalapuram town, E. G.Dt., A.P. The dried samples, after cleaning thoroughly with double distilled water, were ground into fine powder by using agate mortar and blended with graphite powder of high purity (99.999%) in the ratio 3:2 (120 mg of sample powder and 80 mg of graphite powder). The graphite powder is added to monitor the beam current. The target pellets were made from this blended mixture by using hydraulic press of 10 ton and were placed on target ladder. The samples on the ladder kept in the highly evacuated chamber (10^{-6} torr) were excited by 3 meV proton beam of beam current 20 nA. The detector was kept at an angle 90° with beam direction and at 45° with the target. The emitted characteristic X-ray spectra from each excited sample were recoded with a high resolution Si(Li) detector for a sufficiently long time to maintain good statistical accuracy. For every sample, the average beam current and total charge collected were recorded.

3 Data Analysis

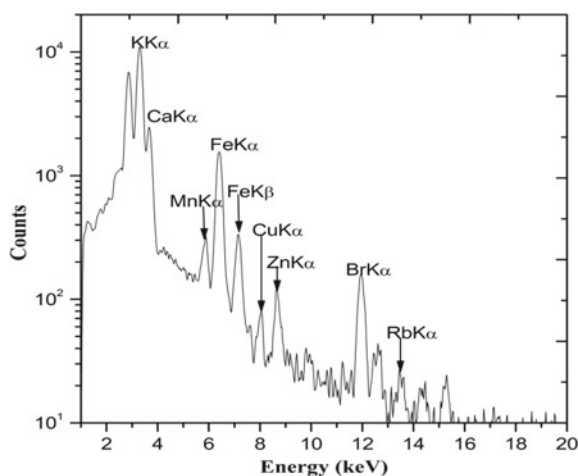
The acquired PIXE spectra of each vegetable sample were analyzed utilizing the GUPIXWIN software package, which has the facility to convert the intensities of X-rays into element concentrations using a standardization technique. It involves fundamental parameters like solid angle, the charge collected, etc. The concentration of elements in every sample were estimated in ppm (parts per million; $\mu\text{g/g}$).

Ten pellets of each sample were exposed to investigation, and variations in concentrations from sample to sample were within the standard deviation. The correctness of the technique in the estimation of concentrations of low atomic number elements was verified by analyzing the NIST apple leaves sample (1515). The results obtained including certified values are furnished in Table 1 and the corresponding spectrum is shown in Fig. 1. Thus, the technique is validated.

Table 1 Comparison of measured element concentration (ppm) values of Standard Reference Material, —NIST-1515 with certified values

Element	NIST apple leaves (1515)	
	Measured value	Certified value
K(%)	1.48 ± 0.05	1.60 ± 0.02
Ca(%)	1.61 ± 0.06	1.53 ± 0.02
Mn	48.5 ± 2.40	54 ± 3.00
Fe	88.1 ± 4.50	83 ± 5.00
Cu	5.3 ± 0.40	5.6 ± 0.20
Zn	12.9 ± 0.70	12.5 ± 0.03
Se	0.06 ± 0.01	0.05 ± 0.01
Rb	9.3 ± 1.00	10.2 ± 1.50
Pb	0.54 ± 0.08	0.47 ± 0.02

Fig. 1 PIXE spectrum of standard reference material; NIST-1515



4 Results and Discussion

The typical X-ray spectra of two vegetables are shown in Figs. 2 and 3. The average concentrations of different major, minor and trace elements present in each of these nine vegetables and the corresponding standard deviations are furnished in Table 2. The trace metals like iron, copper and manganese are essential for normal life processes.

In human body metabolism, iron is an important element which acts as a catalyst and when compared with any other trace element present in higher amount.

Fig. 2 PIXE spectrum of tomato

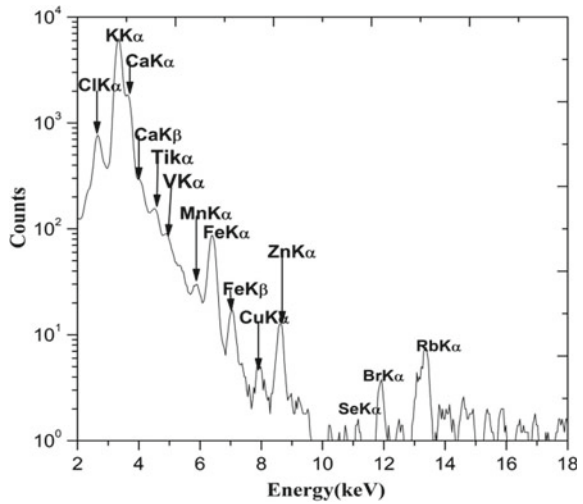


Fig. 3 PIXE spectrum of carrot

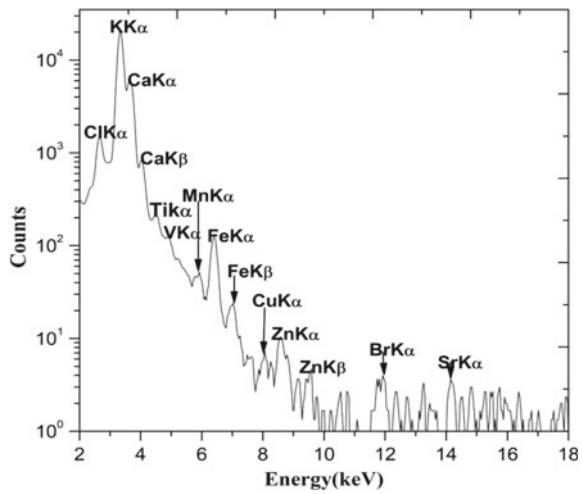


Table 2 Concentrations (ppm) of trace elements in the vegetables under study

Element	Tomato	Little gourd	Ridge gourd	Potato	Brinjal	Carrot	Beet root	Bitter gourd	Lady finger
Cl	3200 ± 100	1100 ± 60	1000 ± 110	2000 ± 180	1600 ± 60	2000 ± 175	6350 ± 850	800 ± 160	1950 ± 80
K	10,500 ± 150	7500 ± 800	6300 ± 750	7400 ± 680	7800 ± 650	12,800 ± 1770	9800 ± 950	6100 ± 1150	9050 ± 170
Ca	1300 ± 50	1950 ± 50	2250 ± 210	400 ± 40	1250 ± 140	3450 ± 480	2200 ± 250	1550 ± 150	3650 ± 265
Ti	60 ± 10	31.5 ± 5.7	44.1 ± 9.3	42.6 ± 7.6	52.3 ± 6.8	43.2 ± 8.1	40.8 ± 8.8	41.2 ± 7.7	36.4 ± 7.8
V	14.2 ± 6.1	6.3 ± 3.5	ND*	ND*	ND*	ND*	4.2 ± 1.1	ND*	2.6 ± 2.2
Mn	ND*	3.9 ± 1.9	6.0 ± 1.8	ND*	11.2 ± 2.3	8.5 ± 3.0	30.3 ± 3.8	7.3 ± 2.5	8.7 ± 2.8
Fe	48 ± 5	37.5 ± 2.8	45 ± 4	66.7 ± 3.8	46.5 ± 3.5	52.7 ± 5.1	42.5 ± 5.2	41.9 ± 3.1	35.5 ± 4.1
Cu	ND*	2.3 ± 1.0	ND*	1.8 ± 0.7	6.0 ± 1.4	ND*	1.5 ± 0.6	1.4 ± 0.6	ND*
Zn	10.9 ± 2.7	5.1 ± 1.4	10.5 ± 2.4	9.6 ± 2.0	6.4 ± 1.6	7.8 ± 1.9	13.3 ± 2.4	ND*	12.3 ± 2.3
Se	3.0 ± 1.0	1.5 ± 0.9	ND*	2.6 ± 1.1	ND*	ND*	4.0 ± 2.7	12.3 ± 2.1	ND*
Br	10.1 ± 4.6	ND*	ND*	ND*	ND*	ND*	ND*	ND*	ND*
Rb	12.9 ± 4.2	ND*	ND*	ND*	9.3 ± 4.7	ND*	ND*	ND*	ND*
Sr	ND*	ND*	ND*	ND*	10.5 ± 4.8	10.0 ± 5.9	ND*	ND*	ND*

* ND not detected (Lower than the detection limit)

According to an estimate, in human body, 57.6% iron is in Hemoglobin and 8.9% in myoglobin; approximately 33% in non-heme iron complexes, including ferritin and hemosiderin and 0.5% in the cytochrome enzyme. But at the same time, it is known that infectious microorganisms and viruses also require iron and other micronutrients for replication and endurance. Hence, access of high amounts of iron to these organisms is not desirable. Also, the chance of overabundance of iron which may initiate free radical interceded harm is to be avoided [11]. Yet to keep up appropriate concentration of iron so that the host can mount an optimum immune response, National research council suggested the daily dietary allowance of iron for male as 10–12 mg and for female 15 mg and the recommended upper limit of iron is up to 30 mg [12]. Vegetables may compensate to some extent and other diet provides the deficient intake of iron.

In the five of currently examined vegetables, copper is found in small amounts little gourd (2.3 ppm), potato (1.8 ppm), brinjal (6 ppm) and beet root (1.5 ppm). Copper is an essential trace element essential for normal biological functions of amino acids and tyrosinase enzymes. Excess intake of copper may cause hemolysis. The RDA for copper is 1.0–3 mg [13].

Manganese is indispensable for normal growth, skeletal formation and normal reproductive function. It helps in neutralizing of free radicals, forestalling diabetes and necessary for normal nerve function. Manganese deficiency causes some diseases and presence of excess leads to exterminating of central nervous system. Ingestion, absorption or inward breath may cause manganic pneumonia [14]. The estimated safe and adequate dietary intake in adults is 11 mg/day [12]. The maximum concentration of manganese was found in beet root (30.3 ppm) and minimum in little gourd (3.9 ppm).

In the human body, the second abundant element is zinc. All cells of the body contain zinc, and 1.4–2.3 mg is present in normal human body. This is a component of DNA, protein and insulin synthesis. It is essential to normal cell functioning including protein synthesis, metabolism of carbohydrates, cell growth and cell division. The RDA for zinc is 12–15 mg [12]. Zinc was found in all vegetables analyzed except cabbage and cucumber. The concentration of zinc in these vegetables ranged from 5.1 ppm (little gourd) to 13.3 (beet root).

Calcium is important for bone formation and maintenance, teeth growth and safe gum production. It is needed in clotting of blood, stabilizing many physiological functions and is thought to help in prevention of bowel cancer. It has a natural calming and relaxing effect and is required to maintain a normal heart beat and nerve impulses transmission. It helps to reduce cholesterol, to develop muscles, to avoid muscle cramps. This also provides strength, breaks down fats, preserves proper permeability of the cell membrane, assists in neuromuscular function and helps maintain healthy skin. Deficiency of calcium results in osteoporosis, hypertension and colon cancer. The RDA for calcium is 800–1200 mg [13]. In all the vegetables analyzed in the present study, calcium was found in appreciable amounts. The maximum was found in lady finger (3650 ppm) and minimum in potato (400 ppm). Thus all the vegetables analyzed contain some beneficial trace elements in varying quantities and all of them in safe levels.

Here, it is important to point out that the presence and concentrations of different elements in different plants depend to some degree on the composition of the soil, water and fertilizers used, as well as the permeability, selectivity and absorbability of plants to absorb those elements.

5 Conclusions

Nine vegetables commonly consumed in Asian countries are analyzed for determination of trace element concentrations. All the vegetables contain beneficial elements Ca, K, Cu, zinc and iron in various quantities. None of vegetables contains any of these trace elements above accepted levels and are free from toxic elements like As, Cd, Pb and Hg. Vegetables may compensate to some extent and other diet provides the deficient intake of essential trace elements.

References

1. Anonymous (1998) Health guideline for the use of wastewater in agriculture and aquaculture. Tech. Rep. Sci. 778, 10, Report of WHO Science Group, World Health Organization, Geneva, Switzerland
2. Butt SJ, Imran-Al-Haq M (1993) Vegetable cultivation—secret of manifold prosperity's science. *Technol Dev* 12(3):46–51
3. Esposito M, De Romaa A, Cavallo S, Miedicob O, Chiaravalle E, Soprano V, Baldia L, Gallo P (2019) Trace elements in vegetables and fruits cultivated in Southern Italy. *J Food Compos Anal* 84:103302–103309. <https://doi.org/10.1016/j.jfca.2019.103302>
4. Hana RA, Hope K, Emmanuel A, Tsdale M, Austin JT, Carol MB, Sayo OF (2017) Determination of macro, essential trace elements, toxic heavy metal concentrations, crude oil extracts and ash composition from Saudi Arabian fruits and vegetables having medicinal values. *Arab J Chem* 10:906–913. <https://doi.org/10.1016/j.arabjc.2016.09.012>
5. Rahman MR (2016) Elemental analysis of some selected fruits and vegetables samples in Bangladesh. *Am J Biol Environ Stat* 2(1):1–6. <https://doi.org/10.11648/j.ajbes.20160201.11>
6. Fahad SM, Mahmudul Islam AFM, Ahmed M, Nizam U, Alam MR, Alam MF, Khalik MF, Hossain MS, Hossain ML, Abedin MJ (2015) Determination of elemental composition of *Malabar spinach*, lettuce, spinach, hyacinth bean, and cauliflower vegetables using proton induced X-ray emission Technique at Savar Subdistrict in Bangladesh. *BioMed Res Int* 10. (Article ID 128256) <https://doi.org/10.1155/2015/128256>
7. Basha AM, Yasovardhan N, Satyanarayana SV, Reddy GV, Kumar AV (2014) Baseline survey of trace metals in ambient PM₁₀ at Tummalapalle uranium mining site. *Toxicol Rep* 1:505–512. <https://doi.org/10.5094/APR.2014.068>
8. Oyekanmi AM, Farombi AG, Adebayo OR (2014) Determination of trace element in raw leafy vegetables grown in selected local government of Osun State. *Am J Chem* 4(1):38–41. <https://doi.org/10.5923/j.chemistry.20140401.06>
9. Rahman MR, Shariff MA, Rahaman MO, Uddin MS, Shafiq AKM, Shameem MA, Hasan MM, Hasan SJ, Huq MA (2014) Studies of essential and trace elements in some fruits and vegetables of Southwestern Bangladesh by PIXE Technique. *Pak J Nutr* 13(2):62–66. <https://doi.org/10.3923/pjn.2014.62.66>

10. Papa S, Cerullo L, Di Monaco A, Bartoli G, Fioretto A (2009) Trace elements in fruit and vegetable. *EQA—environmental quality/Qualité de l’Environnement/Qualità ambientale*, 2:79–83. <https://doi.org/10.6092/issn.2281-4485/3819>.
11. Oppenheimer SJ (2001) Iron and its relation to immunity and infectious disease. *J Nutr* 131:616S-635S. <https://doi.org/10.1093/jn/131.2.616S>
12. Kaplan LA, Pesca J, Kazmierczak SC (1993) *Theory, analysis, correlation in clinical chemistry*, 4th edn. Published by Mosby, p 707
13. *Recommended dietary allowances by the National Academy of Sciences*. National Academy Press, and Washington (1989)
14. Underwood EJ (1977) Academic Press Inc., New York, p 545

Fully Automated Waste Management System Using Line Follower Robot



V. Geetha , Sanket Salvi , Sampat Kumar Ghosh, Shaikh Sahil Ahmed, and Rohit Sunil Meshram

1 Introduction

With the increasing population, it is becoming more important to address the problem of waste management. Urban India generates 62 million tons of waste annually, and 43 tons of municipal solid waste is collected annually, out of which 31 million is dumped in landfill sites and just 11.9 million is treated [1]. Improper waste management may lead to several health hazards including cholera, dysentery, typhoid fever, etc. The municipal solid waste can be classified into recyclable material, composite wastes, biodegradable wastes, inert waste, and domestic hazardous waste, and toxic waste [1]. The segregation process of these wastes at collection point is very important, and among all these waste types, the collection of biodegradable waste must be very timely as these are perishable. With the advent of new technologies such as Cloud computing and the Internet of things, we can design smart systems to handle waste management issues especially at waste collection points in more effective and efficient ways. This domain which addresses issues of waste management with the help of new technologies is called smart waste management. This paper proposes an efficient smart waste collection system for the timely collection of waste from various locations in the urban area. We propose a two-part subsystem for the smart waste collection system: smart waste bin (SWB) with garbage level indicator and autonomous waste collection vehicle (AWCV).

V. Geetha (✉) · S. Salvi · S. K. Ghosh · S. S. Ahmed · R. S. Meshram
National Institute of Technology Karnataka, Surathkal, Mangalore, India
e-mail: geethav@nitk.edu.in

S. Salvi
e-mail: sanket.177it001@nitk.edu.in

The remainder of the paper is organized as follows: in Sect. 2, we have discussed the related works which inspired the design of the proposed system. Section 3 highlights the proposed system followed by its implementation. And finally, the results and conclusion are discussed.

2 Related Work

This section provides literature survey on various smart waste management systems and smart waste collection systems.

The proposed system by Adam [2] monitored the container contents using sensors, and based on the garbage overflow frequency of bins, the distribution of the containers is optimized. The containers are categorized into three-level, empty, half, and full. If the containers get full, the information is sent to the Web page via TCP/IP and is stored in the database. For this type of implementation, if the containers are nearly going to be full, it will not send any information. After putting more, waste will overflow the container which is a disadvantage for this implementation.

Papers [3, 4] proposed a way that allows the municipal corporations to observe the bin status remotely over the Web server and keep the areas/cities clean by optimizing time and cost required for it. As soon as the bin has been filled its maximum level, an alert message has been passed by the department of waste management through the GSM module so the department can initiate the waste collector vehicle to the respective place for gathering the garbage. The study helps in detecting the smart garbage management systems that can be used to make the area/city clean and hygienic. Until the bins are full, it will not send any data to the coordinator through the GSM module.

Paper [5] proposed a model where smart bins are located in the urban area are interfaced with microcontroller-based systems using ultrasonic sensors and Wi-Fi modules where the level of the smart bin is detected by the ultrasonic sensors and then sends the signal. The microcontroller receives the signal and passes the signal to the central system through the Internet. The data will be received, analyzed, and processed which displays the status of the garbage in the smart bin on the dashboard. Using the separate android application for the smartphone is less effective, as we can make the same whole system as an automated system just by using an autonomous waste collection vehicle (AWCV).

Hong [6] proposed an IoT-based smart garbage system (SGS) to reduce the quantity of food waste. In an SGS, battery-based smart garbage bins (SGBs) interchange information with each other using wireless mesh networks, and a router and server collect and examine the information for service provisioning. This system requires more maintenance cost, and it is very complex than our proposed system.

An IoT-based solid waste management system that permits routing of garbage collector trucks in a smart city, garbage bin monitoring, and dynamic scheduling is proposed in paper [7–9]. Garbage bins are supplied with low-cost embedded devices that are located at various locations in the entire city. A mobile application was

built by them that facilitates the waste collection drivers to go to the garbage bins using dynamic and shortest routes. The data is sent to the authority, but the garbage collection system is not automatic. So due to this, a disadvantage is that there may be some traffic that may be faced by the truck driver while going for garbage collection.

Tambare [10] proposed a system in which multiple dustbins are located around the city, and the dustbins are provided with a low-cost embedded device that helps to track the level of the garbage bins. A unique ID is provided for every dustbin in the city to identify the level of bins. As the level reaches the threshold limit, the device will notify the level with the unique ID provided so that immediate action can be made to clean the dustbins.

Inspired by these works, our proposed system uses an autonomous dedicated waste collector vehicle which requires no manual intervention. Thus, this system can be operational 24×7 and avoids additional risks of infections due to manual handling of waste disposal. The advantage of our proposed system provides real-time tracking, and the garbage is collected automatically by the robot as the smart bins send the real-time data continuously and the action is taken instantly.

3 Proposed System

The main motivation of the proposed system is to find a way of collecting and effectively disposing of the waste. The proposed smart waste collection system (SWCS) is divided into two subsystems which are smart waste bin (SWB) with garbage level indicator and autonomous waste collection vehicle (AWCV) as shown in Fig. 1.

In the AWCV subsection, Arduino Yun is considered with a 5 V power supply as a microprocessor for controlling all the connected sensors. Two motors are used

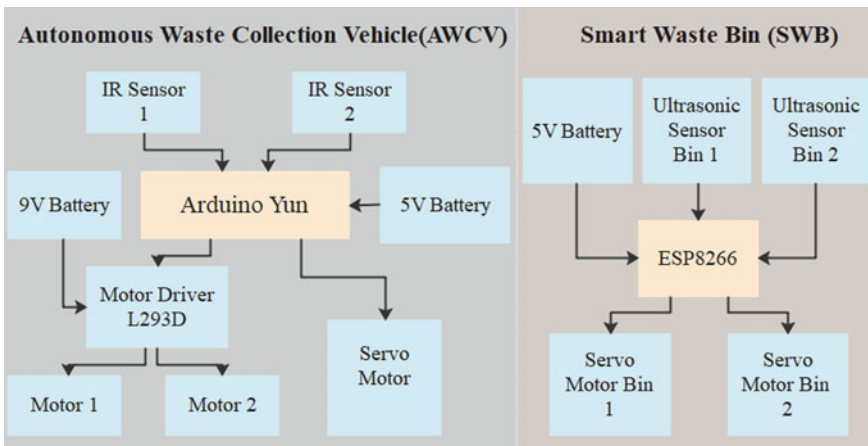


Fig. 1 Block diagram of the proposed system

for physical movement of the AWCV, and 9 V power is supplied to it via L293D motor driver. Two IR sensors are mounted for guiding the AWCV along the road. One servo motor is used for emptying the bin of the AWCV. The Arduino Yun runs the whole AWCV by collecting and sending signals to all the sensors according to the program.

In the SWB subsection, ESP8266 with a 5 V power supply is used as a micro-controller for controlling the connected sensors. Two ultrasonic sensors are used to keep the track of garbage level in each SWB. Two servo motors are used to open the bottom lead of each SWB for transferring the garbage in the AWCV's bin.

The functional flow of the proposed system starts with garbage level detection in SWB. An ultrasonic sensor connected to ESP8266 will be used to continuously monitor and update the garbage level on ThingSpeak Cloud. Once the level reaches up to a specified threshold, a value in the form of the SWB number will be updated over a separate channel field. This channel field will be subscribed by the AWCV, thus as soon as the value is updated, the AWCV will know the destination SWB number.

The AWCV will start moving using connected motors, and it will be continuously guided by IR Sensors connected to Arduino Yun. Based on the SWB number, the AWCV will understand how many crosses in the path have to be skipped. Once the AWCV reaches destined SWB, the AWCV will open the collection unit lid and SWB will open the disposal lid, thus completing the garbage collection. A time delay is provided to accomplish this task. Once the SWB is empty, the value is updated over respective ThingSpeak fields, and AWCV will return to the initial place again using the IR guided AWCV unit, thus completing the proposed workflow.

This system is mainly designed to reduce the overhead of organizations like the municipal corporation and can be also used by various other organizations and individuals. For example, in schools, restaurants, gardens, or any other public places where the waste management requires daily. These types of prototypes can be used where the waste products are the frequent event to effectively manage the waste.

4 Requirement Analysis and Implementation

First, the system is booted and then gradually all the components will start working according to the design. After the system is ready the sensors are activated and sensed data is transmitted to ThingSpeak through NodeMCU. This data is fetched by the Arduino which is mounted on the AWCV robot. After the analysis of the data, it will start moving to all the places where it is programmed to go and visit all the possible bins which are about to go full. It will collect the waste from bins to the container which is mounted on the robot. After collecting the waste, the robot will go to the waste dumping area and dump the waste.

Table 1 shows the features and the usage of the used components. Figure 2 shows the flowchart of the system. The AWCV will be at the dumping site initially, and it will check whether any bin is about to be full or not by reading data from ThingSpeak

Table 1 Features and usage

Name	Features	Usage
Arduino Yun	The Arduino is a microcontroller board which is basically a combination of Arduino Leonardo with a Wi-Fi system on a chip	It is used for controlling the waste collection vehicle as well as to connect to Cloud
NodeMCU esp8266	It is an esp8266 family programmable board with onboard Wi-Fi capability	It is used for sensing the wastage level in the bin and update the same over Cloud
L298D Motor Driver	L298D is a 16-pin integrated circuit allows the DC motor to drive in both the direction according to our requirements	It is used for controlling the DC Motor connected to waste collector
DC Motor	12-V DC motor is used to rotate the wheels of the robot car whose speed is 300 RPM and rated torque value is 1.2 kg-cm	It takes input from microcontroller and provides motion to the waste collection vehicle
Ultrasonic Sensor	Ultrasonic sensors uses the ultrasonic waves for measuring the distance	It is used to monitor the waste level in the waste bin
Servo Motor	An electrical device which is used to pull, push, or rotate an object	It is used for opening and closing the waste bin lid
Arduino IDE	A programming tool for Arduino and NodeMCU	It used to program waste collect and bins
ThingSpeak	A Cloud platform for IoT applications	It is used to update waste level values
IR Sensor	It is an infrared electronic sensor device that measures and captures infrared radiation in its surroundings. It is also able to detect motion and can measure the heat radiating from the object	It is used for identifying black strips and helps to navigate the waste collection vehicle. It also helps waste bin to open the lid

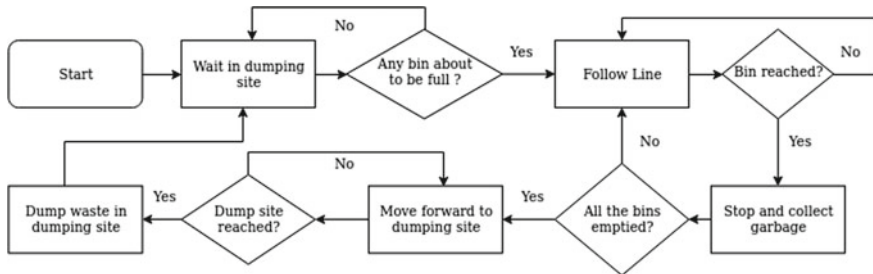


Fig. 2 Flowchart of proposed system

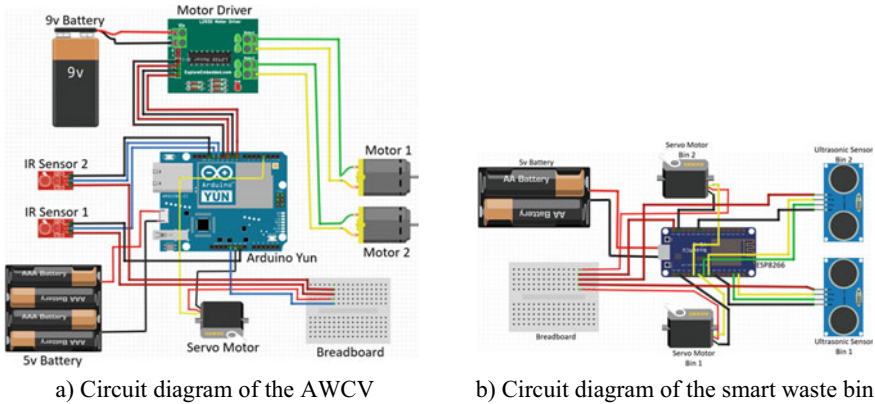


Fig. 3 Circuit diagram of the whole system

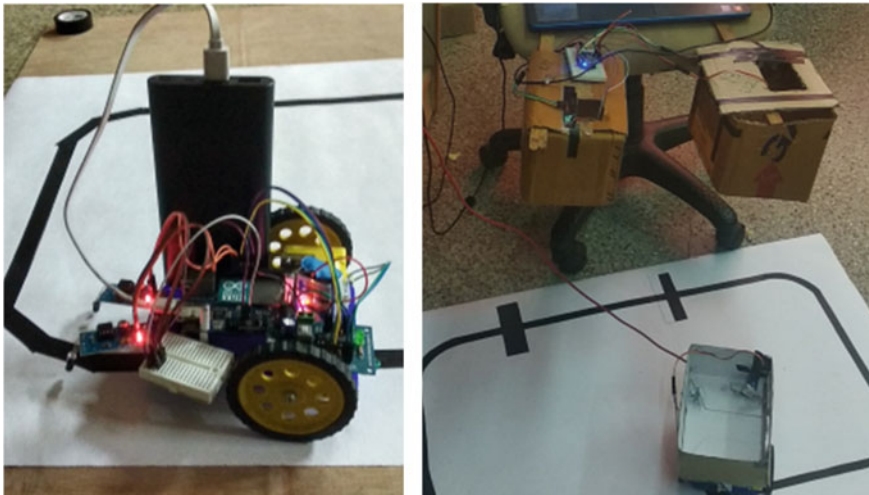
which is sent by the NodeMCU to ThingSpeak Cloud. If one or more bins are about to go full, the Arduino will send a signal that the robot is coming to collect. After that, the AWCV moves forward until it reaches the first bin that is about to go full. After reaching the bin, the lid of the bin is opened with the help of the servo motor, and the waste is dumped onto the robot. After the bin is emptied the robot moves forward to the next bin which is supposed to be emptied. If all bins are emptied, then the robot moves forward to the dumping site, where the waste inside the container of the robot is emptied and remains there.

Figure 3a shows the circuit diagram of the AWCV. Two IR sensors were attached in front of the robot which was used to follow the line. A servo motor was used for opening and closing the lid of the bin which is attached on top of the robot. Two motors were attached to the wheels to control movements and they were connected to the motor driver which was connected to the Arduino Yun.

Figure 3b shows the circuit diagram of the smart waste bin. In this experiment, two smart waste bins were used. But the number of smart bins can be increased according to the requirement of the place. Each bin has an ultrasonic sensor and a servo motor. The ultrasonic sensor is used to check the amount of garbage inside the bin, and the servo motor is used to open and close the lid of the bin to dump the waste onto the bin of the robot.

5 Experimental Setup

Figure 4 shows the experimental setup with two smart waste bins and an AWCV. Figure 4a shows the AWCV sensing the path and moving forward while testing. Later on, as shown in Fig. 4b, the path consists of three stations where two stations have a smart waste bin above it and one of the stations is a dumping area.



(a) The AWCV

(b) Smart Waste Bins

Fig. 4 Experimental setup

The AWCV gets the power supply from the power bank to operate Arduino Yun and one 9 V battery to operate the motors. The bins get a power supply of 5 V from the laptop to operate the ultrasonic sensors and servo motors.

6 Results

A set of electronic components like ultrasonic sensors and IR sensors are used in the implementation of the waste management system. The waste collector robot does an efficient job of collecting all the waste from the desired bins and dumps it to the dumping area. Having two-way communication with the waste bins via ThingSpeak, it continuously fetches the data, whether one or more bins are about to go full or not. If yes, the waste collector robot does the same work and gives repetitive updates to the waste bins via ThingSpeak. After completing its job, dumps the waste and remains in the dumping area.

In this experiment, threshold value of 75% is considered. This means, whenever one smart bin has waste more than 75% of its capacity, a signal will be sent to the robot via ThingSpeak, and it will start moving on its route to empty only the smart bins which have waste more than the threshold. If none of the smart bins are above the threshold, the robot stays at the dumping yard. The percentage change of waste in the bins is shown in Fig. 5.

We used a static algorithm for collecting the waste where the robot visits all the smart bins irrespective of their current threshold as long as at least one of the

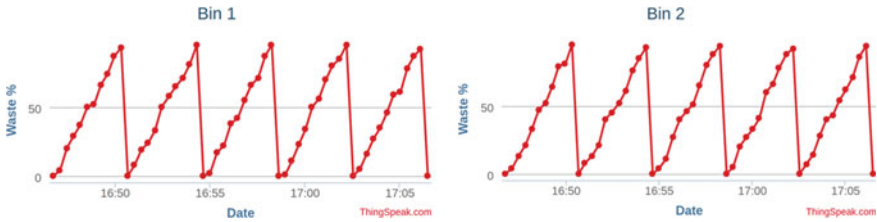


Fig. 5 Data obtained from ThingSpeak showing percentage of waste inside each bin

smart bins sent the signal of having waste over the threshold. There are two main components in this system, the waste bins and the waste collector. The waste collector moves from bin to bin to collect the waste. In this section, the different modules of the system are described to give an overall idea of how the system is formed.

7 Conclusion

The main aim of doing this paper was to find an idea in which the waste can be collected and is managed effectively. By implementing the automated waste collector (AWCV), we can maintain a clean environment with fewer efforts using the technology we have discussed. This waste disposal technique can be used effectively and easily in real-life situations. But the system also has its limitations. The waste container units must be properly aligned with the white and black lines of the path before the usage. The collector moves over the black line. Therefore, the black line has to be perfectly drawn. The system is fully dependent on power supply. The system should be equipped with proper power supply at all times in order to work efficiently.

The system is not designed to operate on multiple stories, but it can collect waste at a ground level, which is one of the limitation of this system.

8 Future Work

In the future, focus must be given to segregation of waste, which is one of the most important steps in waste management and environmental preservation. It allows for the better disposal of waste and encourages recycling and reuse of matter. The segregation technique must work efficiently which can segregate waste without human presence. Each bin will be categorized into three sections. One section will contain all the recyclables (plastic, paper, and glass), the second one will contain metals and the third one for biodegradable waste. When the waste is identified, the waste collector moves the waste to its appropriate dumping area. Further improvement on design of robot with improvement with respect to smart intelligence can be considered.

References

1. <https://www.recycling-magazine.com/2020/05/06/waste-management-crisis-in-india/>
2. Adam M, Okasha ME, Tawfeeq OM, Margan MA, Nasreldeen B (2018) Waste management system using IoT. In: 2018 International conference on computer, control, electrical, and electronics engineering (ICCEEE), pp. 12–14
3. Loganayagi S, Jeyabharathi C (2019) Development of an Iot System for efficient classification and management of solid waste in indian cities—a research. *Int J Innov Technol Explor Eng (IJITEE)* 8(12). ISSN: 2278–3075
4. Teja PS, Krishna MM, Kolluru VR (2019) Development of IoT based garbage management system using NodeMCU. *Int J Eng Adv Technol (IJEAT)* 8(4)
5. Muhammeth A (2018) IOT based waste collection monitoring system using smart phones. In: 8th international symposium—2018, south eastern university of Sri Lanka, at south eastern University of Sri Lanka. University Park, Oluvil, Sri Lanka.
6. Hong I, Park S, Lee B, Lee J, Jeong D, Park S (2014) IoT-based smart garbage system for efficient food waste management 646953
7. Chaudhari S, Bhole V (2018) Solid waste collection as a service using IoT-solution for smart cities. In: Conference: 2018 international conference on smart city and emerging technology (ICSCET).
8. Srikanth CS, Rayudu TB, Radhika J, Anitha R (2019) Smart waste management using Internet-of-Things (IoT). *Int J Innov Technol Exp Eng (IJITEE)* 8(9)
9. Aravindaraman BA, Ranjana P (2019) Design of a monitoring system for waste management using IoT. In: Conference: 2019 1st international conference on innovations in information and communication technology (ICIICT)
10. Tambare P, Venkatachalam P (2016) IoT Based waste management for smart city. *Int J Innov Res Comput Commun Eng* 4(2):8

Mathematical Model for Application of Natural Language Description in the Creation of an Animation



Shradha Kannan and M. K. Vathsala

1 Introduction

Natural Language is the means by which humans communicate with each other. Today's world runs on the wheels of automation. The field of artificial intelligence (AI) where a computer is made to understand the Natural Language and to process it to simplify day-to-day functions is known as Natural Language Processing (NLP). By combining the power of artificial intelligence (AI), computational linguistics and computer science, NLP helps machines 'read' text by simulating the human ability to understand language. NLP relies on Machine Learning to derive meaning from human languages. It's not an easy task to teach a machine to understand how we communicate. Artificial intelligence is found in every aspect of the functional and engineering domain where the prime objective is to achieve automation.

Some examples of NLP in real life are information retrieval, information extraction, machine translation, text simplification, sentiment analysis, text summarization, spam filter, auto-predict, auto-correct, speech recognition, question answering, natural language generation, etc.

In NLP, algorithms are applied such that the natural language is identified and extracted, and the unstructured language data is converted into a form that the computer understands.

The potential of a machine to process a natural language and generate an animation has evolved over the years and has achieved significant progress.

Researchers in this field have been motivated to explore and understand the underlying challenges and to find various approaches to the mapping between the natural language and the video like:

S. Kannan (✉) · M. K. Vathsala
MSRIT, MS Ramaiah Nagar, Bengaluru 560054, India

M. K. Vathsala
e-mail: vathsalamk@msrit.edu

- Generating text from videos.
- Generating videos from text.

Generating a video from a text is a challenging task. An initiation was done by the Texas Lab, Austin which created a motion capture dataset to generate an animated video from the text.

KIT Germany contributed to provide a large, open and extensible motion language dataset which is a great contribution to the research community. However, standardized and openly available datasets do not exist to support the development and evaluation of such systems [1, 2].

2 Existing Approaches

In the literature, many works are found where the problem of video to the language is extensively studied, and the use of deep neural network is mostly suggested. Few of the significant works are:

- (1) Hierarchical Recurrent Neural Encoder for Video Representation with Application to Captioning by Pingbo et al. [3],
- (2) Sequence-to-Sequence (S2S)—Video to Text by Subhashini et al. [4] and
- (3) Video Paragraph Captioning using Hierarchical Recurrent Neural Networks by Haonan et al. [5].

Very less work is found in the literature to deal with text to video generation and few related works include:

- (1) Video Generation from Text by Yitong et al. [6],
- (2) To Create What You Tell: Generating Videos from Captions by Yingwei et al. [7],
- (3) Encoding Spatial Relations from Natural Language by Tiago et al. [8],
- (4) Generative Adversarial Text to Image Synthesis by Scott et al. [9] and
- (5) StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks by Ha et al. [10].

In all these models, the pixel level images of low quality are used, whereas the concise 3D models that can be flexibly rendered into a variety of high-definition visual content using standard computer graphics techniques. Few of the significant work found in the literature are:

- (1) Semantic parsing for text to 3d scene generation by Angel et al. [11],
- (2) Learning spatial knowledge for text to 3D scene generation by Angel et al. [12],
- (3) Text to 3D Scene Generation with Rich Lexical Grounding by Angel et al. [13],
- (4) Text2Shape: Generating Shapes from Natural Language by Learning Joint by Embeddings [14],

- (5) Body Talk: Crowd shaping Realistic 3D Avatars with Words by Stephan et al. [15].

The limitations in all the above works are that they focus on the static scenes rather than the animation. The closely related work found in the literature where the work of the authors towards mapping of natural language descriptions of human activities to motion sequences using mocap data is:

- (1) Learning a Bidirectional Mapping Between Human Whole-Body Motion and Natural Language using Deep Recurrent Neural by Matthias et al. [16] and,
- (2) Paired Recurrent Autoencoders for Bidirectional Translation Between Robot Actions and Linguistic Descriptions by Tatsuro et al. [17].

These approaches have the following limitations:

- They do not specifically focus on generating animated videos for graphics applications.
- And do not evaluate the quality of the generated animations using human judges.

The research problem in this domain is categorized into three types of problems as below:

- i. Video Captioning.
- ii. Animation Synthesis from Text.
- iii. Direct Animation Synthesis.

Video Captioning: There has been a growing amount of recent work on generating NL descriptions of videos. In the work of Venugopalan et al., a seq2seq model is designed for the purpose of video captioning using a technique of convolutional neural nets (CNNs) for encoding the frames and RNN to map the sequence of frames to a sequence of words [18]. This method is further evolved using language model [19], pretrained word embedding, attention and hierarchical modelling [3, 20].

Animation Synthesis from Text: Here, the main aim is to generate the animation using text, and the popular method used is data representation [21]. One such work by Angel et al. considers the position of the hip of the character having a fixed 3D location and the character moves in place not along the floor [11]. The authors Matthias et al. developed a seq2seq model to map the text to a series of Gaussian Distribution in order to represent the joint angle of the character [22]. The work of Tatsuro et al. uses a model that is incomplete as it involves a method of Autoencoder where the description of the motions often includes information of how the character moves in the global coordinate frame with a shared latent space to generate animations from text [23].

Direct Animation Synthesis: Since collecting motion capture data is expensive, generating new animation sequences from existing data has a long history in the computer graphics community. Recently, researchers have used deep learning techniques to tackle this problem. One line of work has been on synthesizing animations under user constraints such as foot placement locations and times [24–26]. There have also been several works on synthesizing the continuation of an animation sequence given a few frames at the beginning of the animation [27–29].

Research Gap: In contrast to the proposed solution approach proposed in the paper, the input and output domain are the same in these methods, therefore their networks only need to model the uncertainty and wide range of futures, rather than also modelling the gap between the text and animation domains.

3 Proposed Model

This paper deals with the multi-domain mapping of NLP and computer visual graphics based on the sequence classification and the relation between the text feature using automated and learning models.

The natural language provided by the user is analysed by the system via neural networks, which identify the keywords and produce an outcome as a motion vector feature.

The implementation of the model is carried out based on the hybrid modelling of NLP technique and artificial neural network (ANN) for generating animation from the natural text.

The architecture of the proposed model consists of a NLP unit, a sequence-to-sequence (S2S) processing unit and a Physics Processing Unit (PPU).

The S2S processing unit consists of Long Short-Term Memory (LSTM) which is used for word embedding and Gated Recurrent Units (GRU) to generate pose in form of vectors. The PPU processes the vectors and then generates a video as output.

A human motion dataset is adopted for the feature learning in both LSTM and GRU. Training on both learning model is carried out independently and then connected to perform together to generate the output.

In prediction module, the user provides an input as natural language specifically in written text which goes to the three core function modules, i.e., NLP and S2S.

The NLP module takes the input as text and processes it to understand the text description and extracts significant text features such as verb and adjective from the natural description. The natural text also goes to LSTM module where natural text is converted into work vector using word embedding module so that input given by the user can be understood by the computational learning model. The word vector goes to the learning model to perform text description sequence classification LSTM. LSTM units predict training set behaviour determined by recognizing their underlying patterns. Once the processing is done it classifies as unseen samples and estimates the likelihood of the best possible next word as output, i.e., text features and frame features. Then, the result of LSTM is fed to S2S module where GRU process the text features and frame features and generates motion feature-(pose). The result of the GRU is then fed to the physics processing module to create a trajectory of the motion feature which after processing turns out in motion vector. The motion vector is expressed as transition between pose to pose representations with movement over time. These motion vectors are then fed to animation software to animate a realistic character visual scene (Fig. 1).

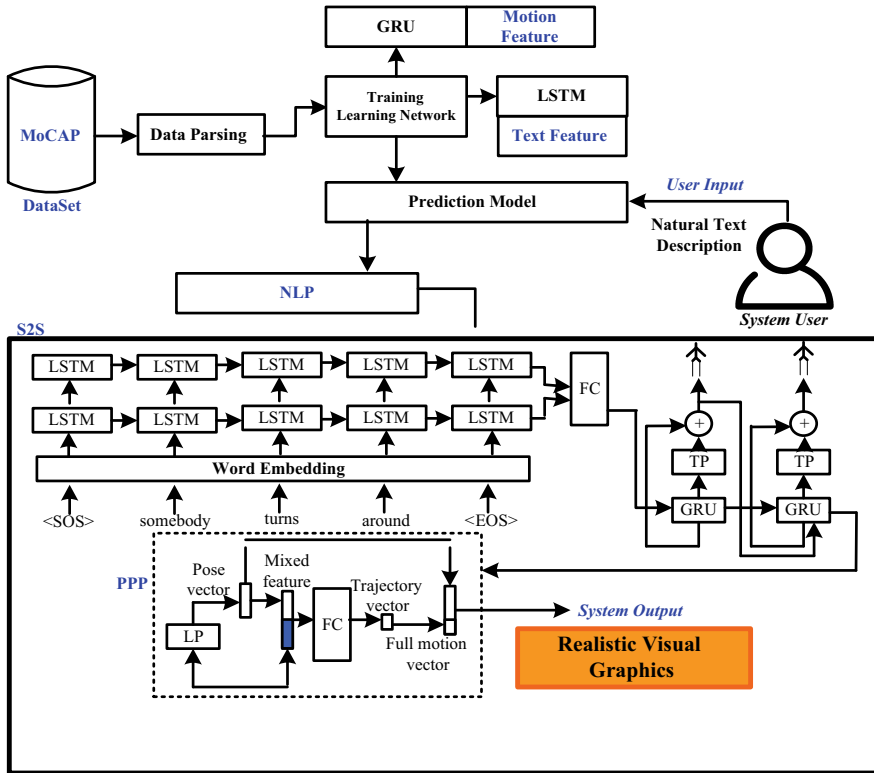


Fig. 1 System architecture

The input data given is in XML, JSON and C3D form. The data is arranged in Master Motor Mapping (MMM) format which is in the form of nodes and connections. It consists of metadata which is the description; raw data which is the actual data and annotation which is the description.

Algorithm-1 Natural Description to Visual Graphic

Input: Natural Description-(T)

Output: Visual Graphics of human motion Activity-(V_G)

Start

1. Initialize T, D-(Dataset)
2. $D_f \rightarrow f_1(D)$
3. $D_{set} \rightarrow \text{Word2Vector}: f_2(D)$
 - a. NLP: fs(Statement)
 - b. Get \leftarrow Verb, Frame
4. Model $\rightarrow f_3(\text{LSTM, GRU})$
5. LSTM $\leftarrow f_4(D_{set})$
 - a. Input = $f_5(\text{shape}: [1])$
 - b. L1 = $f_5(\text{size, connect} \rightarrow \text{input})$
 - c. L2 = $f_5(\text{size, connect} \rightarrow \text{L1})$
 - d. Output = $f_5(\text{shape}: [1,3], \text{connect} \rightarrow \text{L2})$
6. GRU $\leftarrow f_4(\text{LSTM:Output})$
 - a. Input = $f_5(\text{shape}: [1,3, \text{shape}(\text{frames})[0]])$
 - b. L1 = $f_5(\text{size, connect} \rightarrow \text{input})$
 - c. L2 = $f_5(\text{size, connect} = \text{L1})$
 - d. Output = $f_5(\text{shape}(\text{frames}), \text{connect} \rightarrow \text{L2})$
 - e. $M_{feat} = \text{LSTM}(\text{Data}), \text{GRU}(\text{Data})$
7. Train $\leftarrow f_6(\text{GRU, LSTM})$
8. Execute Model
 - a. $D_{read} \rightarrow f_5(D_{path})$
 - b. Input $\leftarrow T$
 - c. Word2Vector $\rightarrow f_2(T)$
 - d. NLP $\rightarrow f_5(T)$
 - e. LSTM $\rightarrow [T_{feat}, F_{feat}]$
 - f. GRU $\rightarrow [M_{feat}]$
 - g. $M_{vec} \rightarrow f_7(M_{feat})$
 - h. $V_G \leftarrow f_8(M_{vec})$

End

4 Implementations

The programming language used in the development of the model and its implementation has multiple functions and dependencies on packages.

The packages considered are (Table 1):

Table 1 Packages used

S No.	Package name	Purpose
1	argparse	Parsing command-line Input argument
2	os	Enables the user to interact many O/S functions
3	JSON	Conversion of JSON string to dictionary structure
4	XML.etree.cElementTree	Allows interaction with the XML file of the entire dataset as a tree
5	logging	Allows status messages to be written to the output stream
6	NumPy	Processing multi-dimensional array objects
7	From tool import perform action	A user defines a module for generating visual graphics

Table 2 Package descriptions

Package Name	Signature	Function	Arguments
argparse	argparse.add_argument(arg)	add_argument()	Nargs
os	os.function_name(arg)	join(), exists()	path, filename
json	json.function_name(arg)	load()	filename
XML.etree. cElementTree	et .function_name(arg)	parse()	pathname
logging	logging.function_name(arg)	warn()	message string
numpy	np.function_name (arg)	array()	filename

Package descriptions are as follows (Table 2):

5 Pseudocodes

Pseudocode for generating visual animation from module perform_action

```

# define a function perform_action() and pass arg
#     where arg: the sentence is an input given by the user
# create an empty vector file
# initialize a variable path with a string Dataset
# for each directory path, directory name, and file name in the dataset
#     for each file name
#         check annotation in each file
#         append all the files having annotation
# create an empty cell for a directory name
# create a vector for sentence
# create an empty vector for action
# for each file
#     Open JSON file and read it and store output obtained in an object
#     for each object
#         Append sentence
#         Append action by splitting with \ and _
#         Display data
# call function industrial-strength NLP() for English language and
assigned to a variable NLP
# definition of a function: find verb() and pass arg
#     where arg: Sentence to NLP()
#     return: a list of filtered verbs from sentence
# to get input from the user to system
# define a function get sentence() with argument x
# call a function() list out the mapped object with a string argument
# return string value with first indexing in the list
# from table data take verbs
#     read the index in sentence randomly definite to action
#     store obtained outcome as trace
#     where trace: motion vector
# render the motion vector for physics processing

```

6 Performance and Evaluation

The KIT, Germany, has contributed to provide a large, open and extensible motion language dataset which is a great contribution to the research community because though linking human motion and natural language is of great interest for the generation of semantic representations of human activities as well as for the generation of robot activities based on natural language input. However, while there have been

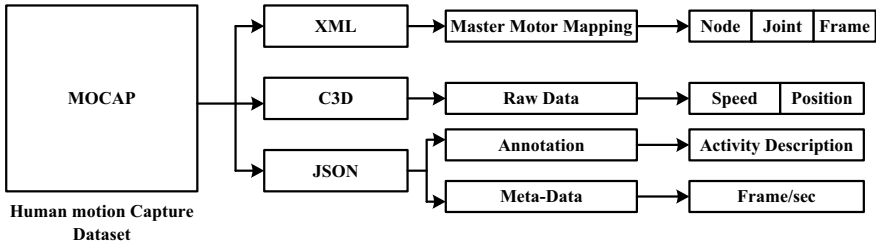


Fig. 2 Description of dataset adopted in model

years of research in this area, no standardized and openly available dataset exist to support the development and evaluation of such systems.

Figure 2 shows the description of dataset called MOCAP which is a huge collection of human motion or movement activity sensor captured dataset. It can be seen that the dataset MOCAP consists of three different files formats, namely XML, C3D and JSON. The XML format is subjected to the master motor mapping (MMM) file which contains description of human pose into the set of nodes, joints and frames. The nodes file represents a simple dot point in 3D space, joint is the link or connection between the nodes and the frame holds the information of changing position of the nodes (i.e., trajectory). The file with C3D format is raw data that holds information about the human movements such as speed and position over time. The file in JSON format consists of two different set of information, namely Annotation and Metadata. The annotations are the description of the motion activity and Metadata is the frame rate per second that describes variable activity position with respect to time.

7 Software Testing

The model that was developed was subjected to the following testing methods:

- i. Unit Testing.
- ii. Integration Testing.
- iii. System Testing.

7.1 Unit Testing: Each Component in Model Was Tested Individually and the Interface

Also, the interface between components was tested. Unit test cases are depicted in Table 3.

Table 3 Unit testing

Test Case	Name	Item	Sample input	Expected output	Actual output	Remarks
UTC-1	User input as natural text	Argument parser	Running sentence: example: the person is jumping	Word vector	Same as expected output	Test successful
UTC-2	Natural language processing	LSTM	Word vector	Text feature	Same as expected output	Successful
UTC-3	Motion vector processing	GRU	Frame feature and text feature	Fully connect mapped frame	Same as expected output	Successful
UTC-4	Physic processing	Motion trajectory	Fully connect mapped frame	Motion vector	Same as expected output	Test successful

7.2 Integration Testing

Even though the components are designed and trained independently, they are integrated and can have an impact on the other. The interface is tested and the test cases are presented here (Table 4).

Table 4 Integration testing

Test case	Name	Item	Sample input	Expected output	Actual output	Remarks
ITC-1	NLP	NLPU interface	Natural language description	Data structure compatible for S2SU	Same as expected output	Test successful
ITC-2	S2SU	S2SU interface	Text feature and frame feature	Data structure compatible for PPU	Same as expected output	Test successful
ITC-3	PPU	PPU interface	Fully connect mapped frame	Data structure compatible for graphics rendering	Same as expected output	Test successful

Table 5 System testing

Test case	Name	Item	Sample input	Expected output	Actual output	Remarks
STC-1	Development platform installation	Installation of development and deployment platform	Compatibility of python SDK on virtual box for Unix platform on windows. Linux. MAC OS	Execution of Test Program as Printing Hello World written in Python	Same as expected	successful
STC-2	Dataset readability	Reading files from the dataset	Path of the dataset	Display the list path of each files (JSON, XML.C3D)	Same as expected	Successful
STC-3	Third party package import compatibility	JSON, element tree, Numpy, os, argparse	Log into sys. modules	Logging entry	Same as expected	Test successful

7.3 System Testing

All components in the system are integrated and tested. System testing has great impact in project design and delivery. So, each component is tested to ensure proper working of the system. The test cases are displayed in Table 5.

8 Analysis of Results

This section summarizes the outcome of the proposed model. Natural language is given as input resulting in a realistic human motion depiction.

The visual outcome obtained from the text input of six different instances by the user is displayed here with (Figs. 3, 4, 5, 6, 7 and 8).

Graph of input versus accuracy of output is depicted below.

The input is tested with output. Figure 9 represents the graph between input that is code and the output obtained (Table 6).

9 Conclusions

This paper introduces a hybrid algorithm design for predicting text. In particular, an efficient computational prediction model based on LSTM and GRU is designed

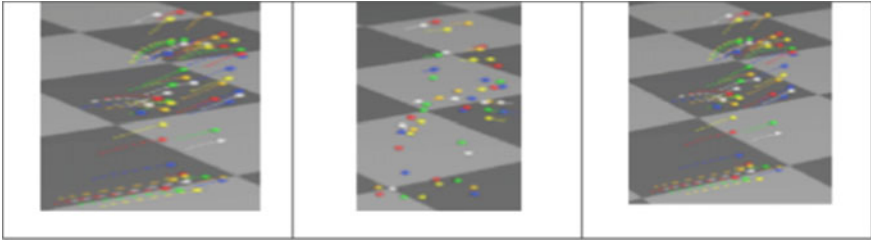


Fig. 3 Person is running

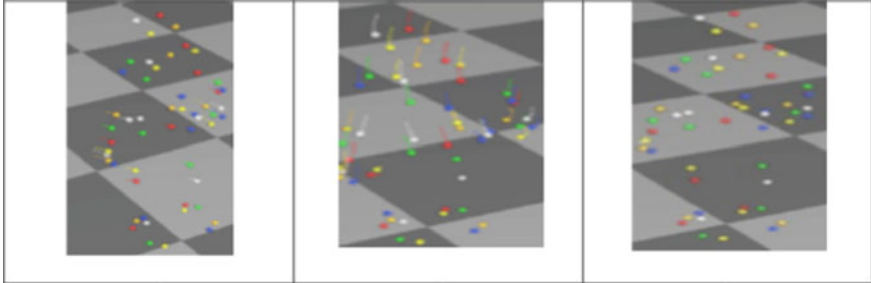


Fig. 4 Person is jumping

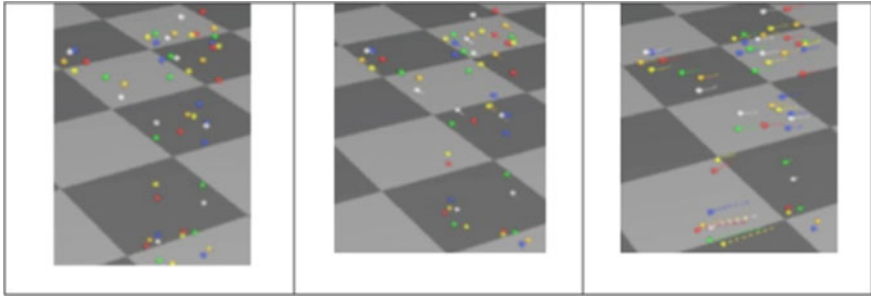


Fig. 5 Person is dancing

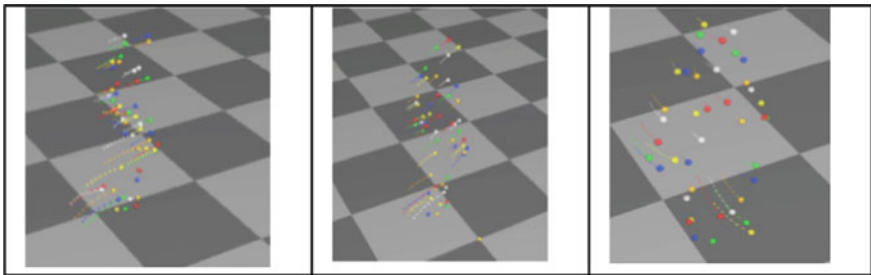


Fig. 6 Person is walking

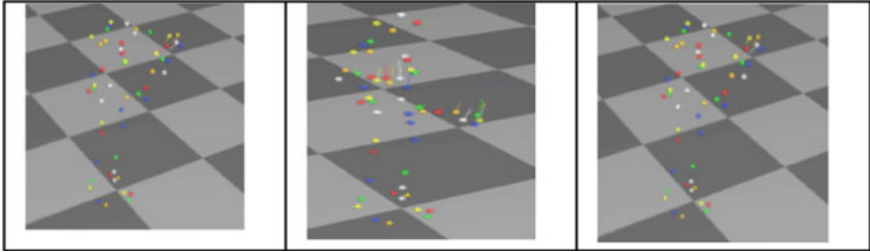


Fig. 7 Person is playing tennis

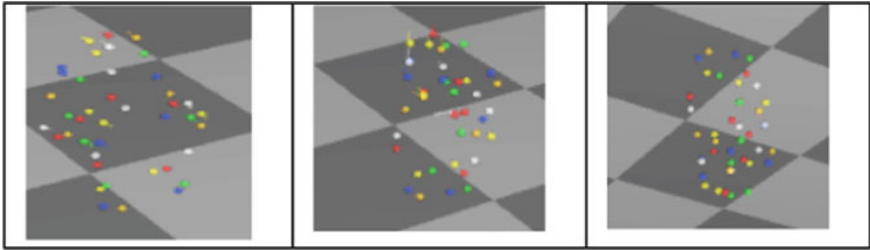


Fig. 8 Person is sitting

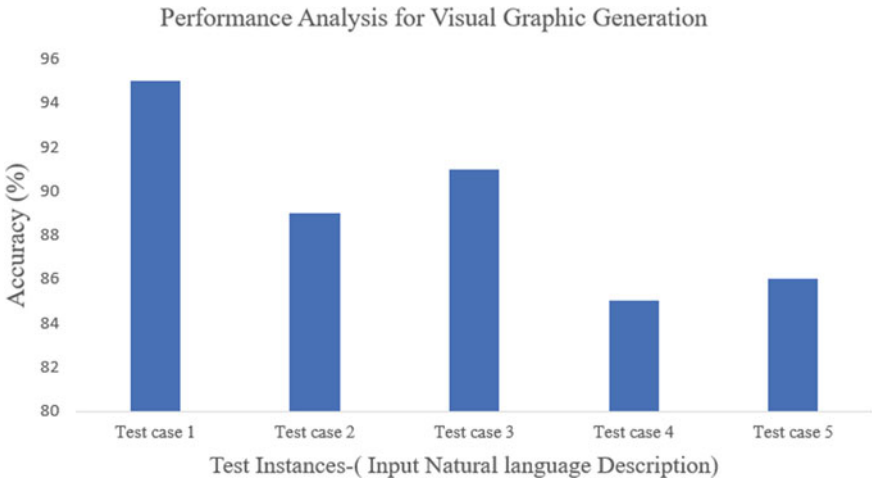


Fig. 9 Graph between test instances and accuracy

to map NLD to the real visual graphics space. Both the models are independently trained and executed together to generate output for animation. Compared with the existing system, the learning model introduced in this project has achieved good performance on MOCAP dataset. The model provides high-resolution animation

Table 6 Description of test instance

Test instance	Description
Test instance-1	A person is Walking
Test instance-2	A person is running over around
Test instance-3	A man is dancing
Test instance-4	A person is running and jumping
Test instance-5	A man is playing tennis

graphics with less processing time. The scope of the model is not limited to the current implementation and design modules. Therefore, it can be extended to future work, which considers the integration of additional functions to obtain input directly from the user's voice instead of natural text to generate visual graphics.

References

1. Plappert M, Mandery C, Asfour T (2016) The KIT Motion-Language Dataset. *Big Data* 4(4):236–252. <https://doi.org/10.1089/big.2016.0028>
2. <https://motion-annotation.humanoids.kit.edu/dataset/>
3. Pingbo P, Xu Z, Yang Y, Wu F, Zhuang Y (2016) Hierarchical recurrent neural encoder for video representation with application to captioning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1029–1038
4. Subhashini V, Rohrbach M, Donahue J, Mooney R, Darrell T, Saenko K (2015) Sequence to sequence–video to text. In: *Proceedings of the IEEE international conference on computer vision*, pp 4534–4542
5. Haonan Y, Wang J, Huang Z, Yang Y, Xu W (2016) Video paragraph captioning using hierarchical recurrent neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4584–4593
6. Yitong L, Min MR, Shen D, Carlson D, Carin L (2018) Video generation from text. In: *Proceedings of AAAI-2018*
7. Yingwei P, Qiu Z, Yao T, Li H, Mei T. To create what you tell: generating videos from captions. In: *Proceedings of the 2017 ACM on multimedia conference*. ACM, pp 1789–1798
8. Tiago R, Kocisk'y T, Besse F, Eslami SM, Melis G, Viola F, Blunsom P, Hermann KM (2018) Encoding spatial relations from natural language. *arXiv preprint arXiv:1807.01670*
9. Scott R, Akata Z, Yan X, Logeswaran L, Schiele B, Lee (2016) Generative adversarial text to image synthesis. In: *Proceedings of the 33rd international conference on machine learning (ICML)*, pp 1060–1069
10. Han Z, Xu T, Li H, Zhang S, Huang X, Wang X, Metaxas D (2017) StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. In: *Proceedings of the international conference on computer vision (ICCV)*, pp 5907–5915
11. Angel C, Savva M, Manning C (2014) Semantic parsing for text to 3d scene generation. In: *Proceedings of the ACL 2014 workshop on semantic parsing*, pp 17–21
12. Angel C, Savva M, Manning CD (2014) Learning spatial knowledge for text to 3d scene generation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp 2028–2038
13. Angel C, Monroe W, Savva M, Potts C, Manning CD (2015) Text to 3D scene generation with rich lexical grounding. In: *Proceedings of ACL*

14. Kevin C, Choy CB, Savva M, Chang AX, Funkhouser T, Savarese S (2018) Text2Shape: generating shapes from natural language by learning joint embeddings. arXiv preprint [arXiv:1803.08495](https://arxiv.org/abs/1803.08495), 2018.36
15. Stephen S, Quiros-Ramirez MA, Hill MQ, Hahn CA, Zuffi S, O'Toole A, Black MJ (2016) Body talk: crowdshaping realistic 3D avatars with words. *ACM Trans Grap (TOG)* 35(4):54
16. Matthias M, Mandery C, Asfour T (2018) Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. *Robot Auton Syst* 109:13–26
17. Tatsuro T, Matsunaga H, Ogata T (2018) Paired recurrent autoencoders for bidirectional translation between robot actions and linguistic descriptions. *IEEE Robot Auto Lett* 3(4):3441–3448
18. Venugopalan S, Rohrbach M, Donahue J, Mooney R, Darrell T, Saenko K (2015) Sequence to sequence–video to text. In: *Proceedings of the IEEE international conference on computer vision*, pp 4534–4542
19. Nicolas B, Yao L, Pal C, Courville A (2016) Delving deeper into convolutional networks for learning video representations. In: *Proceedings of ICLR*
20. Venugopalan S, Hendricks LA, Mooney R, Saenko K (2016) Improving LSTM-based Video description with linguistic knowledge mined from text. In: *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*
21. Ömer T, Ulbrich S, Mandery C, Do M, Vahrenkamp N, Asfour T (2014) Master motor map (mmm)—framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. In: *2014 14th IEEE-RAS international conference on humanoid robots (humanoids)*. IEEE, pp 894–901
22. Matthias P, Mandery C (2018) Asfour T (2018) Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. *Robot Auton Syst* 109:13–26
23. Tatsuro Y, Matsunaga H, Ogata T (2018) Paired recurrent autoencoders for bidirectional translation between robot actions and linguistic descriptions. *IEEE Robot Autom Lett* 3(4):3441–3448
24. Ikhsanul H, Holden D, Schwarz J, Yearsley J, Komura T (2017) A Recurrent variational autoencoder for human motion synthesis. *proceedings of BMVC*
25. Daniel H, Saito J, Komura T (2016) A deep learning framework for character motion synthesis and editing. *ACM Trans Graph* 35(4):1–11. ISSN-0730-0301. <https://doi.org/10.1145/2897824.2925975>
26. Daniel H, Komura T, Saito J (2017) Phase-functioned neural networks for character control. *ACM Trans Graph (TOG)* 36(4):42
27. Partha G, Song J, Aksan E, Hilliges O (2017) Learning human motion models for long-term predictions. In: *Proceedings of the 2017 international conference on 3D vision (3DV)*. IEEE, pp 458–466
28. Zimo L, Zhou Y, Xiao S, He C, Huang Z, Li H (2018) Auto-conditioned recurrent networks for extended complex human motion synthesis. *Proceedings of ICLR*
29. Julieta M, Black MJ, Romero J (2017) On human motion prediction using recurrent neural networks. In: *Proceedings of CVPR*. IEEE, pp 4674–4683

Ingenious Bracelet for Women Safety and Security Using IoT



K. Hemasagar, R. Manoj Kumar, V. Sai Manojna, and B. K. Priya

1 Introduction

Violence against women, includes sexual and gender based, is prevalent all around the world. Every day, woman is a victim of violence. Approximately 70% of rape or sexual assault victims experience moderate to severe distress. A larger percentage than it is for any other crime. 90% of adult rape victims are females. 94% of women who are raped experience symptoms of post-traumatic stress disorder. Violence against women can happen at different moments of their life and in a number of different ways. Just walking back and forth to school or going to get water can put women at risk. In some countries, violence against women is very common. India stands first globally as “the world’s most unsafe place for women”. According to national crime records bureau, a crime against women is committed every 3 min. India is considered to be the world’s most dangerous country for sexual violence against women. Rape is one of the most common crimes in India.

2 Related Work

As per the observations made in [1], body temperature is impacted based on the situation a person is in. A sudden fluctuation in the body temperature could be noticed. As cited by [2], in some circumstances, the person may have restricted body movements or fall unconscious at those times relying on sources which require

K. Hemasagar (✉) · R. Manoj Kumar · V. Sai Manojna · B. K. Priya
Department of Electronics and Communication Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Bengaluru, India

B. K. Priya
e-mail: bk_priya@blr.amrita.edu

manual authentication like clicking a button will not be of help at all times. As the pulse would vary in that situation along with the body temperature, the variations in these parameters observed with the combination of pulse rate sensor and temperature sensor is used to initiate the security system. There is a possibility that a variation in the above-mentioned two parameters is due to some other reasons. So to provide more authenticity to the validation, as mentioned in [3], a motion sensor to detect the movement of any living being in the area under its range adds on to the identification of situations under threat in an insightful way. As specified in [4], a GPS sensor module helps in determining and locating the place where the person is present and the information is forwarded to the ICE contacts. A piezo buzzer acts like an alarm system to gain help from on the nearby region. In [5], the idea of transmitting the location via GSM module is proposed. Through this, the location is communicated to the nearby police so that quick help can be got. In [6], by using a camera module, the images of the culprit are captured and are streamed to a port of Raspberry Pi. This helps in proving the culprit guilty and stands as a strong evidence in further legal procedures. The above discussed papers have used the panic button, but it is difficult to press always so integrating all the ideas in a single module to improve the efficiency and accuracy.

3 Resources

3.1 Raspberry Pi

It is a compact sized device with high computational capabilities and performs all the operations that a computer is capable of doing. This highly efficient device supports many platforms which help in the easy development of applications. The Raspberry Pi plays its role as analysing, memory storing, decision making and processing body when connected to sensors through GPIO pins and performs the actions accordingly to the specified instructions. Owing to all these above-mentioned advantages, the Raspberry Pi has found its place in the device being developed to understand the data and acts a platform of communication among the devices interfaced or connected to it.

3.2 Temperature Sensor

Temperature is a physical quantity and hence this sensor output is analogous in nature. The output analog signal is directly proportional to instantaneous temperature. LM 35 sensor in this system is used to determine the temperature of the body. The temperature of the body varies in case of sudden shock or surprise and is detected by this sensor. Hence, this is used to analyse the psychological state of the person and

there by estimating the situation. Also, its advantages like low cost, greater accuracy and no necessity for any external calibration unlike thermistor makes it popular to use.

3.3 Pulse Rate Sensor

TCRT 1000 reflective optical sensor consists of light emitting diode and light detecting resistor. The light emitting diode and the detector are arranged side by side so as to block the surrounding ambient light, which could otherwise affect the sensor performance. The heart beat is directly related to the amount of blood flowing to various regions of the body. TCRT 1000 is used for photoplethysmography. Hence, from LED, a light source is emitted into the tissue while the detector placed at the adjacent side detects the resultant light. The output is analogous in nature and is given to ADC retrieve BPM from the output. The BPM of the person can be known by this, and if it is greater than the threshold value, it acts as a factor of validation.

3.4 Motion (IR) Sensor

An IR proximity sensor helps to identify the objects or obstacles nearby. It has an IR LED which generates the IR waves which are invisible to the human eye. If the generated IR waves hits with any object or human, the reflected wave received by the receiver compares with initial voltage with the help of comparator circuit in the sensor. Hence, the motion of the body is determined. In this proposed model, the IR sensor helps us to identify the motion of person approaching nearby. Also, this acts as a second stage of validation of the situation. For suppose, the heart rate might be increased when exercised or doing something which involves physical energy. At such times, even though the person is not actually under threat, but heart rate is increased. At such cases, this IR sensor acts as an additional factor for validation.

3.5 GPS Receiver

GPS, which stands for Global Positioning System, operates on the mathematical principle of “Trilateration”. It helps to detect a particular person on ground. The location is determined by three satellites and confirmed by the fourth satellite and is sent to GPS receiver. When the proposed system detects threat, it gets activated and the current position of the victim is given by this GPS module. The corresponding coordinates of latitude and longitude are noted and sent later to pre-defined contacts.

3.6 ADC Convertor

The Raspberry Pi cannot read analog signals directly. It is a digital-only computer. Sensors which are used to measure the physical status of victim are analog in nature. That is because there is continuous monitoring of data from sensors. Hence, the MCP 3008 board, whose working principle is SAR method, performs analog to digital conversion and acts as a bridge to communicate between Raspberry Pi and the sensors.

3.7 Piezo Electric Buzzer

Piezo electric buzzer is an electronic device which produces a tone or sound. There is a piezo crystal, which is made of a special material which changes its shape when voltage is applied to it. When the crystal is pushed, it generates a pressure wave which is picked up by the human ear as sound wave. This buzzer is used in this prototype to create an alarm which helps the victim to alert the people nearby and get help from them.

3.8 Pi Camera Module

A camera module is an image sensor integrated with a lens. In camera module, the light collected by the object through lens is converted firstly into an electrical signal and then into digital image signal by an internal image processor. In this proposed system, it is used as a security system. It helps to stream the whole situation which can be later used for better analysis of the situation and also can be used for any legal actions if any. The software components used to realize the module are Raspbian OS, Twilio, Secure shell, Virtual network computing and Telegram app. Raspbian OS is a Debian-based computer operating system for Raspberry Pi. SSH&VNC viewer helps in accessing the Raspberry Pi remotely. SSH is used to access the command line of Raspberry Pi remotely while the VNC viewer enables the Raspberry Pi access in some other device. Telegram is a social networking platform. It allows the users to send text, videos to large number of users with greater security and speed.

3.9 Twilio

Twilio is a cloud-based platform where the user will be able to send the messages. In the proposed model, GSM module is replaced with Twilio to overcome the network issues. Twilio finds its advantage in remote areas where GSM module struggles to

get a proper network to transmit the message within a proper time which sometimes may lead to adverse consequences.

3.9.1 Telegram

In order to facilitate a greater outreach for help within shorter span of time, social media applications will be of greater help. Telegram is one of the open-source applications which has been inculcated in the proposed device to send alert message along with the coordinates of location of the victim. This can be further developed in various other social media platforms by creating alert notifications seeking for help.

3.9.2 Live Streaming

The contents captured by the camera module is live streamed through a private channel through You Tube and this can be accessed only by the ones who know the IP address of the Raspberry Pi of that particular device. This can be used later for any legal actions if required.

4 System Design

The design of the proposed model is mainly focused on getting help, to reach out to the person in possession of the device. The device can be activated in two ways. First way is to hold the button for few seconds to start the device. This option is useful when there is a possibility to click the button at the time of attack or when the person feels there is a chance of being attacked and feels the necessity of help to reach out to her.

Second way to start the device is based on the data sensed from three sensors, i.e. temperature sensor (LM 35), pulse rate sensor (TCRT 1000 Reflective Optical Sensor) and motion (IR) sensor when proved to be authenticated. This way is opted taking into consideration that there are chances where it may not be always possible to manually activate the device as there may be situations of restricted body movements. The temperature and pulse rate sensors are verified to authenticate when there are fluctuations or abnormalities in the readings keeping the normal human body temperature and pulse rate as base values. When a person is under threat or attack, the body temperature as well as the pulse is affected. Taking these fluctuations as parameters is one of the possible solutions to predict if the person is under the situations of attack or not. These fluctuations may happen when the person is tensed or prone to sudden shocks also. To avoid the activation of the device in such cases, a motion sensor is incorporated along with these to detect the motion of any approaching body towards the person wearing the device. When data read by all the three sensors reports abnormalities deviating from the normal readings, i.e. meeting the threshold

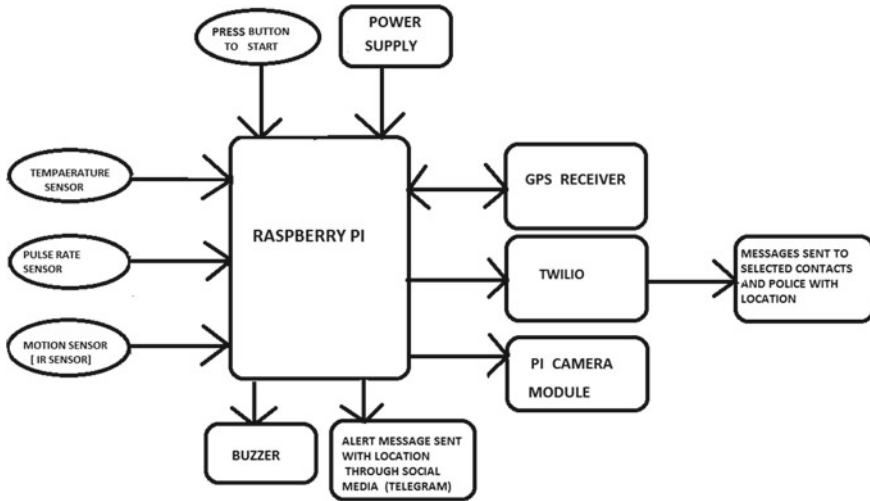


Fig. 1 Block diagram of proposed system

values of normal conditions, the device gets activated. Immediately after activation, the piezo electric buzzer produces alarming noises in the range 80–110 dB which can be heard up to a distance of around 50 feet, and it is helpful to get help from nearby people. The GPS sensor tracks the current location at which the person is present with the help of information sent through satellites to the GPS receiver. This information of current location along with the coordinates received is sent to already specified contacts and police along with an alert message using Twilio API server so that there is no much ado in sending the messages. In case of hardware components like GSM module is used, the accessibility to network in remote places becomes a factor of delay in situations where losing time may turn the situation more dangerous. Also, a camera module is incorporated which acts as a security system to capture and stream the live situation which is accessible to pre-defined people only to ensure privacy; also, this streaming of video can be of use in the legal proceedings further. The proposed idea is as shown in Fig. 1, and the flow chart for the implementation is as shown in Fig. 2.

5 Experimental Results

Initially, when the power supply is turned ON, the Raspberry Pi checks for sensors input. Firstly, the pulse of the person is monitored using pulse rate sensor. The reading can be noted when finger is pressed onto the light emitting source and light detector of the pulse rate sensor as shown in Fig. 3.

If the pulse rate sensor meets the threshold value, Raspberry Pi then checks for value from temperature sensor to validate the situation. The temperature sensor

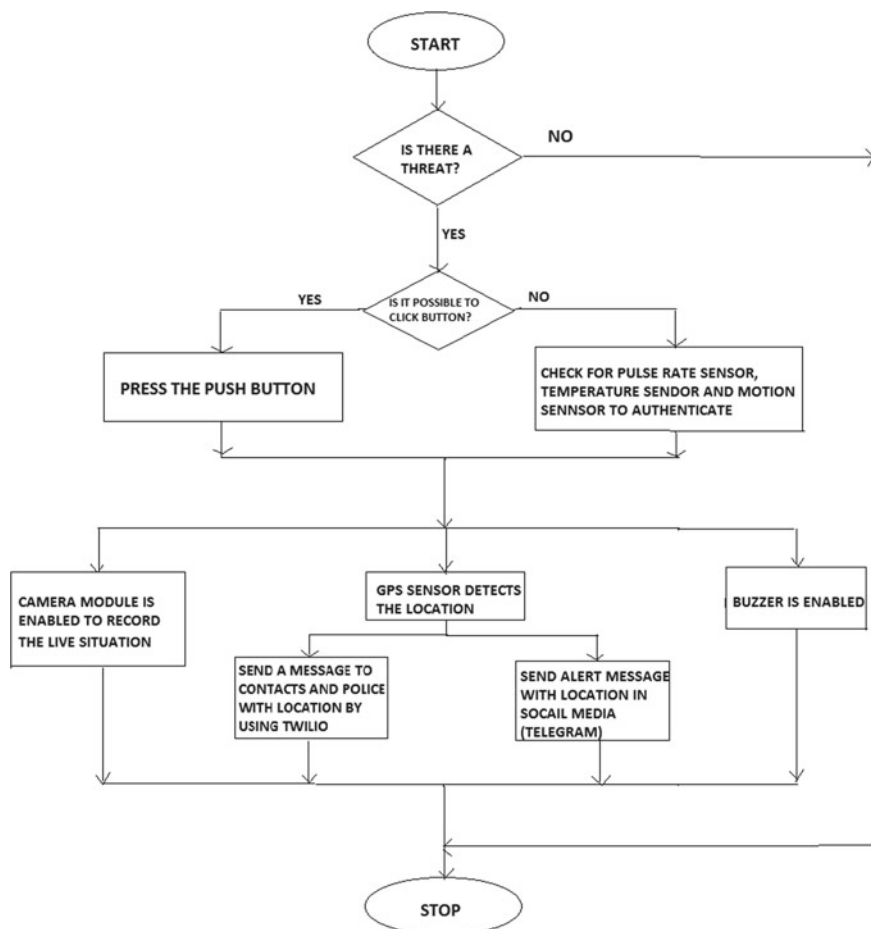


Fig. 2 Flowchart of the proposed model

Fig. 3 Pulse detection

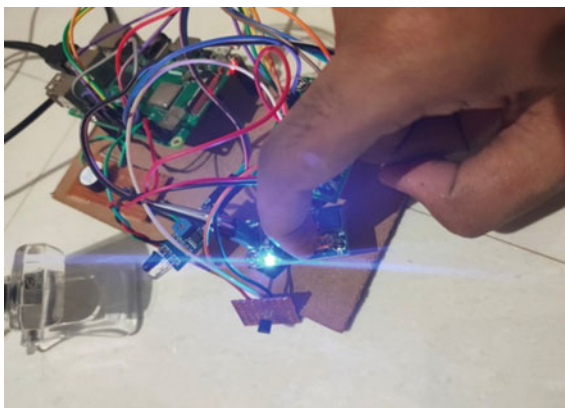


Fig. 4 Temperature detection

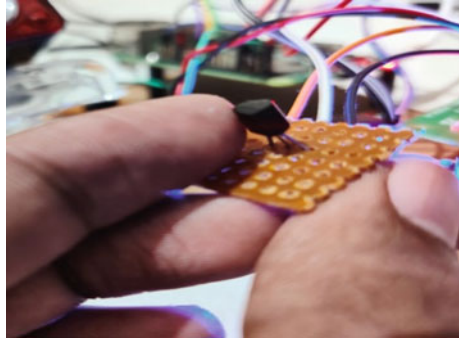
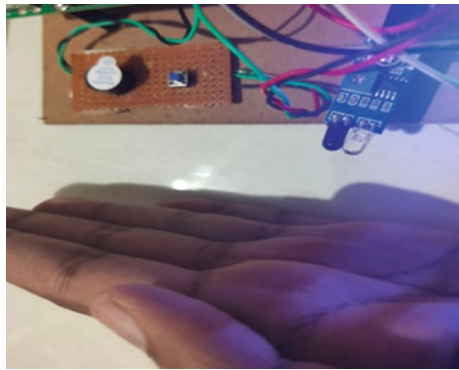


Fig. 5 Motion detection



detects the temperature when body comes in contact with the sensor. The working of the temperature sensor is as shown in Fig. 4.

If both the temperature and pulse sensor meets with threshold values, then the Pi checks for the IR sensor inputs which adds another factor to the validation process. IR sensor detects if any person is nearby as shown in Fig. 5.

The sensor readings are tabulated as shown in Fig. 6. If the inputs from all the sensors meet the threshold conditions, then immediately GPS module tracks the location at which the person is present and sends an alert message along with location coordinates to already defined contacts or police through Twilio API as shown in Fig. 7. The same message is sent through the telegram app which is an open-source tool as shown in Fig. 7.

A buzzer is also activated to seek help from people nearby as shown in Fig. 8.

Also, a camera module is activated to record the situation which helps to capture the situation and make use of it for further requirements with live streaming as shown in Fig. 9.

The complete prototype device developed is shown in the Fig. 10, and it is compared with some common devices that were developed for women safety as shown in Table 1.

Fig. 6 Sensor readings

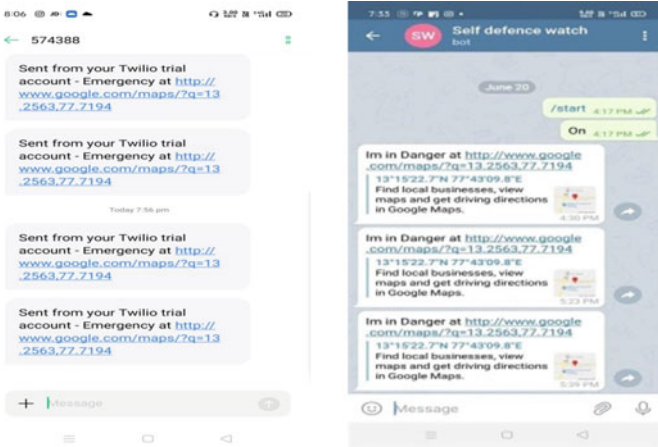
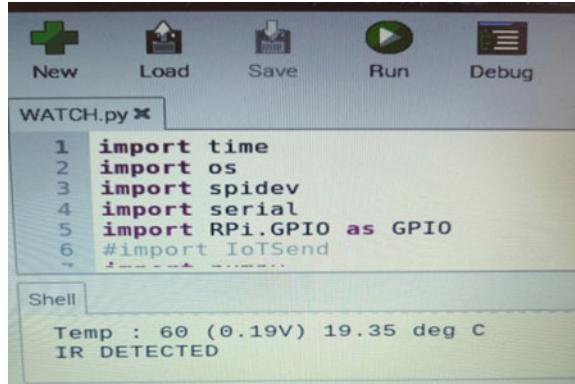


Fig. 7 GPS alert message through Twilio and Telegram

Fig. 8 Buzzer module



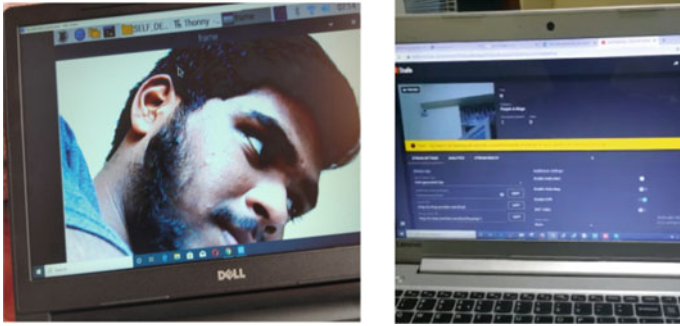
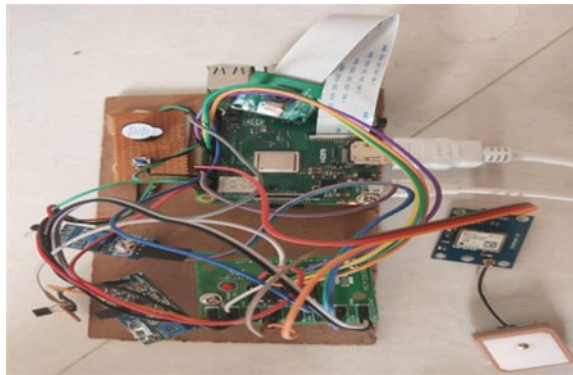


Fig. 9 Video output and live streaming

Fig. 10 Demonstration of prototype model



6 Conclusion

Various technical papers describing prototype models and methods proposed to analyse the situation and develop a device for women safety and security have been studied, and one of the possible prototype model is developed keeping in mind the requirements, constraints and feasible equipment. The results have been listed. The device is developed with the main intention of getting help from possible sources at the time of attack on the victim. It is intended to develop the device in such a way that manual activation of the device is not necessarily required as the victim may not stand a chance to have enough time or may have restricted movement. So practically activating the device at the time of threat may not be possible. In the future, the prototype developed can work better with slight modifications in the design and be of help to a better extent with the extended features.

Table 1 Comparison of the prototype model with the existing models

S. No	Existing models	Description of existing model	Drawbacks of existing model	Proposed model advantage over existing models
1	Abhaya	It is a type of SOS application, activating will send the current location to emergency contacts with the help of GPS and a phone call is made to the police	To activate the application, the victim must be able to use the phone; may not be possible at such extreme conditions	It constitutes sensors that will automatically senses the situation and activates the whole system
2	Wristwatch	It detects the location of the victim using cell tower triangulation method. On activation, an SOS signal is sent to the pre-defined contacts along with a notification to the police	It depends highly on the cellular tower density. The system may totally fail	Even in the cases of low network connectivity, the proposed model operates with the response obtained from the sensors. Also as a primary help, the buzzer is activated
3	AESHHS	In case of attack, it sends the location information	Only location is detected and sent. This may not be of a greater help to the victim	Features likes buzzer, camera module and sending messages through social media will be of great help to the victim
4	VithU app	With the click of power button of mobile phone twice, location of the victim is sent to contacts every two minutes in the form of alert messages	In situations of threat, the possibility of accessing the mobile is not always feasible	The situation is detected automatically by the sensors and performs the required action
5	Smart Belt	This activates when the pressure sensors detect pressure and a screaming alarm is activated and rings the siren	It solely depends on pressure sensors for the activation. The pressure sensors may get activated in normal situations which leads to false triggering	It has three different sensors to authenticate the situation which gives better prediction and analysis of the situation

(continued)

Table 1 (continued)

S. No	Existing models	Description of existing model	Drawbacks of existing model	Proposed model advantage over existing models
6	Design and implementation of a rescue system for safety of women	It is activated when the switch is manually pressed. ON activation, on body shock circuit, camera and audio module are activated and the corresponding location of the victim is sent through messages	It is switched ON only through manual activation. In case of poor network areas, GSM module gives significant delay in delivering the alert messages	It can be activated either manually or automatically through authentication from sensors. The delay from GSM can be countered with the use of TWILIO API, effectively reducing delay, cost and hardware. Buzzer helps to alert the people nearby to seek help and Telegram will help in greater outreach

7 Future Scope

The proposed model can be developed in the following ways.

1. A shock generator circuit can be devised along with the proposed model which acts as instant self defence mechanism.
2. An audio recorder can be incorporated which improves the analysis of the situation along with camera module.
3. An alert notification on social media can be generated so as to gain quick help from large group of people nearby.

References

1. Rai U, Miglani K, Saha A, Sahoo B, Sarobin MV (2018) Reach out smart safety device. In: 6th edition of international conference on wireless networks and embedded systems (WECON), 16–17 Nov 2018
2. Khandelwal T, Khandelwal M, Pandey PS (2018) Women safety device designed using IOT and machine learning. In: IEEE Smart world, ubiquitous intelligence and computing, advanced and trusted computing, scalable computing and communication, cloud and big data computing, internet of people and smart city innovation (Smart World/SCALCOM/UICIATC/CBD Com/IOP/SCI), 08–12 Oct 2018

3. Harikiran GC, Menasinkai K, Shirol S (2016) Smart security solution for women based on Internet of Things(IOT). In: International conference on electrical, electronics, and optimization techniques (ICEEOT)
4. Ahir S, Kapadia S, Chauhan J, Sanghavi N (2018) The personal stun-A smart device for women's safety. In: International conference on smart city and emerging technology (ICSCET), on 5 Jan 2018
5. Helen A, Fathila MF, Rijwana R, Kalaiselvi VKG (2017) A smart watch for women security based on IOT concept 'watch me'. In: 2nd international conference on computing and communications technologies (ICCCT), 23–24 Feb 2017
6. Mahajan M, Reddy KT, Rajput M (2016) Design and implementation of a Rescue system for safety of women. In: International conference on wireless communications, signal processing and networking (WiSPNET), 23–25 Mar 2016
7. Paknikar R, Shah S, Gharpure P (2019) Wireless IOT based solution for women safety in rural areas. In: 2019 international conference on communication and electronics systems (ICES)
8. Tejonidhi MR, Aishwarya, Chaithra KS, Dayana MK, Nagamma H (2019) IoT based smart security gadget for women's safety. In: 2019 1st international conference on advances in information technology (ICAIT)
9. Ruman MR, Badhon JK, Saha S (2019) Safety assistant and harassment prevention for women. In: 2019 5th international conference on advances in electrical engineering (ICAEE)
10. Sharma V, Tomar Y, Vydeki D (2019) Smart shoe for women safety. In: 2019 IEEE 10th international conference on awareness science and technology (iCAST)
11. Ramachandiran R, Dhanya L, Shalini M, A survey on women safety device using IoT. In: International conference on system, computation, automation and networking (ICSCAN)
12. Kavitha M, Sivachidambaranathan V (2018) Women self-protecting system using Internet of Things. In: 2018 IEEE international conference on computational intelligence and computing research (ICCIC)
13. Kunnath AT, Pradeep P, Ramesh MV (2012) Locating and monitoring emergency responder using a wearable device. In: Proceedings of the international conference on advances in computing, communications and informatics (ICACCI), Chennai, pp 1163–1168
14. Sandheep S, John H, Harikumar A, Vinitha Panicker J (2017) BusTimer: an android based application for generating bus schedules using crowdsourcing. In: 2017 international conference on technological advancements in power and energy (TAP Energy), Kollam, India
15. Arvind S, Anantha Narayanan V (2019) An overview of security in CoAP: attack and analysis. In: 2019 5th international conference on advanced computing and communication systems, ICACCS 2019. Institute of Electrical and Electronics Engineers Inc., pp 655–660

NL2SQL: Natural Language to SQL Query Translator



T. J. Revanth, K. Venkat Sai, R. Ramya, Renusree Chava, V. Sushma, and B. S. Ramya

1 Introduction

The high intricacy behind SQL language and database compositions has made databases questioning a moving errand to human developers. In this paper, we present our new regular language database questioning framework as an elective answer for retrieving information. Databases are successful methods for putting away and recovering huge volumes of information rapidly and effectively. There are a few database executives' frameworks which are industrially accessible and are utilized all over the world. Anyway, its anything but a simple activity to get information out of these vaults. To speak with these databases, an uncommon communication language called structured query language (SQL) is utilized. It has become progressively essential to speak with these databases. Utilizing natural language to convey between a database framework and its human clients has gotten progressively significant since database frameworks have gotten across the board, and their availability to non-master clients is attractive, if not fundamental, to encourage full utilization of the database framework.

The point of natural language to SQL translator (NLS-to-SQL) is to diminish the unpredictability in questioning databases. First it is important to utilize a language that is comprehended by anyone, regardless of whether a specialist database developer or individual with no PC information. The most appropriate language for this reason for existing is the English language. This implies NLS-to-SQL needs to

T. J. Revanth · K. V. Sai · R. Ramya · R. Chava · V. Sushma · B. S. Ramya (✉)
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore 560064, India
e-mail: ramya.bs@nmit.ac.in

V. Sushma
e-mail: sushma.v@nmit.ac.in

interpret English or normal language inquiries into SQL before recovering information from the database. In view of this, we have thought of a strategy that changes over a characteristic language explanation to its equal SQL proclamation and recover the information. To make “NLS-to-SQL Translator” progressively adaptable, a lightweight methodology is utilized to change over the natural language contribution to its SQL proportional.

2 Literature Survey

According to the research paper by Xu et al. [1] NADAQ: Natural language database querying based on deep learning has revealed parts of action between human customers and complex machine systems. NADAQ’s architecture comprises advanced SQL parsing techniques and cutting-edge deep learning techniques. It helps the program to increase the consistency of query translation and to propose innovative answers to irrelevant, unprocessable queries. The results show that NADAQ provides human clients with almost business standard support as a questioning tool when they do not understand SQL or appreciate the database diagram.

Nandhini et al. [2] mentioned a concept in their paper which is translation of Hindi speech into SQL query. This concept involves many phases, translating user input into English code, generating SQL query utilizing python libraries such as NLTK, Goslate. The proposed model consists of the following steps:

Start → User → Speech Input → Extracting data from Input → Lexical Analysis → Syntax Analysis → Semantic Analysis → SQL query → Output → Stop.

The final query is generated based on extracted elements after classifying the relations, attributes, and clauses and further removing ambiguity. The data is then retrieved from the database, and the tests are then generated appropriately.

Zhong et al. [3] have discussed how In-the-loop database execution is used to train Seq2SQL to learn a policy to produce the SQL query conditions, which is unordered and unsuitable for optimization through cross entropy loss. In this paper, they have proposed Seq2SQL, a profound neural system for making an interpretation of inquiries to SQL questions. Their model uses the structure of SQL questions to diminish the yield space of the model. It uses WikiSQL, a query server and SQL queries that are orders of magnitude order greater than equivalent data sets. This provides precision of execution from 35.9 to 59.4% and consistency of logical type from 23.4 to 48.3%.

Xu et al. [4] have proposed a method where SQLNet is used to address the issue of “order-matters” by ignoring the framework of sequence-to-sequence when the order is meaningless. This paper uses a sketch-based approach, where the sketch includes a dependency graph so that one prediction can be made by using previous predictions on which it depends. It also suggested a sequence-to-set paradigm for synthesizing the query, as well as the column attention process. Through integrating these approaches, it reveals that on the WikiSQL task, SQLNet will outperform current art through 9–13%.

Chaudhari and Pranali [5] they proposed a method which consists of 2 components that is pre-processor and post-processor.

Pre-processor-English phrases are given information. Form of question is defined, and based on it, a pre-processor provides the output English sentence containing noun clauses, substituted quantitative, and reference operators, and then this output is submitted to the post-processor as input.

Post-processor recognizes the things in the statement such as tables, values, etc. The argument is then decomposed depending on the five SQL clauses and aggregate functions. The sort of query-template is then determined by the post-processor. Use this to convert template queries into SQL equivalent. Dictionary is added which has to be upgraded, respectively.

Sathick et al. [6] they discussed a critical element is the recognition of the methodology of information extraction in the social network sector. The framework is built using Java programming language and different application software. The details are processed using an oracle database. The user's feedback is not needed in the form of WH-questions. DML and DDL queries can be executed.

This model consists of four modules, and those are morphological analysis, semantic evaluation, mapping desk, and report retrieval. Here the sentence of the natural language is given as an input by the consumer.

Morphological Analysis → Syntactic Analysis → Semantic Analysis → SQL Query. The system's data dictionary will be revised periodically with terms that are unique to the system.

Finegan-Dollak et al. [7] they found the difficulty of querying. First, by analyzing the complexity of the questions and queries, they found that human-written datasets require properties that are not yet included in the automatically generated large-scale query sets.

Second, in the way, examples are separated into training and test sets they found a problem. They implemented a slot-filling model focused on models, but cannot generalize to new requests and separating, depending on the SQL application. We also showed that state-of-the-art programs fail on query-based splits with outstanding results on conventional question-based splits. Finally, they demonstrated variable anonymization which eliminates from the system where a challenging type of uncertainty is included.

3 Proposed System

Recovering the right details from a database is extremely confusing for the people who are not familiar with the design of the databases. DBMS cannot oblige questions introduced in any language other than the standard dialects of the database. The proposed system offers a facility where a client is allowed to pose a query in English that will be taken care of by a few modules to make an undifferentiated SQL query, so as to render the recovery progressively easy and connecting with clients who lack

knowledge in SQL. There are four primary levels engaged with the transformation of natural language questions to SQL query. These are as following:

- (1) **Tokenization:** System performs tokenization for the given query by isolating it as single words. Every word speaks to a token. At that point, these words will be put into a different list and passed on to lexical analyzer.
- (2) **Lexical Analysis:** The words present in the tokenized list will be mapped with the word reference. These words will get replaced by the database words from the word reference and passed to syntactic investigation.
- (3) **Syntactic Analysis:** It analyses word reference of table names, properties, and stop words. Every tokenized word gets mapped with qualities in word reference.
- (4) **Semantic Analysis:** In semantic analysis, the system will find words that represent symbols or conditions, and the word which is found out will get mapped to the word reference. (For example: If the input query has “lesser than or equivalent to” then it will get mapped with the image “< =”).

4 Methodology

Natural language processing is a branch of artificial intelligence, stressed over joint efforts among computers and humans (standard), explicitly how to program PCs to process and separate numerous basic language information.

The steps involved in natural language processing are:

Step 1: Segmentation of sentences

The first step in the pipeline is to break the text apart into separate sentences. For example:

London is the capital and most populous city of England and the United Kingdom.

Step 2: Tokenization of the Word

The next step in our pipeline is to part the sentence into terms or tokens in isolation.

This is known as tokenization. For example:

“London,” “is,” “the,” “capital,” “and,” “most,” “populous,” “city,” “of,” “England,” “and,” “the,” “United,” “Kingdom,” “.”

Step 3: Predicting spoken parts for each Token

Next, we will take a gander at each token and try to uncover its syntactic element whether its an item, a word for action, a descriptor, and so on. Knowing the activity of each word in the phrase will help us start by understanding what the phrase is about.

Step 4: Lemmatization of a text

Lemmatization is generally accomplished by observing a table of lemma types of words dependent on their syntactic structure and most likely having certain custom principles to treat terms.

Step 5: Finding the Stop terms

English has as many filler terms as possible, like “and,” “the” and “a.” Such words create a great deal of commotion while allowing insights into the text, as they turn up much more every now and then than other words. Stop terms are typically only found by evaluating a hard-coded list of learned words. There is no uniform list of stop words appropriate for all implementations in any situation. Depending on your submission the list of words to be disregarded can differ.

Step 6a: Decoding the Dependence

The following stage is to offer significance to the manner by which all the words in a sentence communicate. This is known as dependency parsing.

Step 6b: Setting Noun Sentences

This step uses the parse tree dependency information to automatically group terms which all mean the same thing together.

Step 7: Named Recognition of Entities (NER)

The following nouns are present in the sentence is shown as:

London is the **capital** and most populous **city** of **England** and the **United Kingdom**.

Some of those nouns feature real-world terms. “London,” “England” and “United Kingdom” for example represent physical places on a map.

5 Implementation

The overview of the implementation is shown in Fig. 1. Also the steps which are involved in our implementation are as follows:

(1) Tokenize and tag:

With the help of the tokenizer, the input normal language question is broken into discrete tokens from the “NLTK” set. The tokenized succession of terms is labeled utilizing the Standard POS tagger as per the part-of-speech tagger. Every single resulting process utilizes these labeled tokens for preparing.

(2) Analyze tagged tokens:

The aggregate functions corresponding to the tokens are often represented using a pre-created corpus of words with their respective nouns. At this point, the data decision will be made, and the natural language statement represents a data recovery query (SELECT), for instance, when words like “insert” and their certain synonyms appear in the data, the query type is “INSERT” and so on. In question, such as “S” (SELECT), “F”(FROM), “W”(WHERE), “O”(ORDER

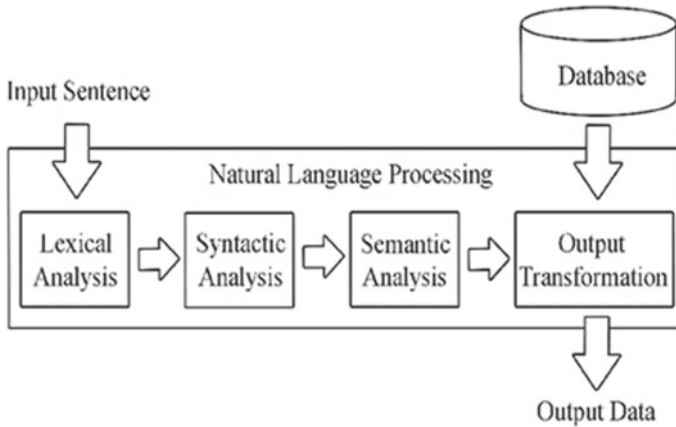


Fig. 1 Structure of the system

BY), “G”(GROUP BY) are mapped to show the clauses they belong to. We have built “data dictionaries” for exceptional clauses for this purpose. Such dictionaries include for example, the aggregate clause information dictionary is:

“number”: “COUNT,” “count”: “COUNT,” “total”: “SUM,” “sum”: “SUM,” “average”: “AVG,” “mean”: “AVG.”

So, if any of those tokens are found, to get an aggregate clause, these nouns are marked with the clause tag.

(3) Map to table names and attributes:

The noun map and verb list are used to prepare the table set, and it will hold the table that is required to form the query. This is based on whether the names of the table are nouns or verbs. The noun map is used to locate appropriate attributes in the final query. The attributes, the table associated with the attribute, and the clause tag will be stored in an attribute-table map that will be used in the final query formation stage. It is achieved using the string-matching algorithm that is implemented into the system. The terms in the input statement do not have to be precisely as they are in the database. The words are applied with stemmer and lemmatize before they are coordinated utilizing string-matching algorithms. The information gathered during this progression, for example, attribute-table map and table set is well on the way to be in the last query, and it may be altered later.

(4) Filter redundancy and complete clauses of the query:

The system has figured out which clauses are probably going to show up in the final query using the various data dictionaries identified and mapped the details to the clauses. Nevertheless, at this point, some of the details must be finalized. The information related to the GROUP BY and HAVING statement was acquired utilizing prior information and SQL’s basic laws. For example,

if the total capacity is contrasted with a steady, for example “MAX (salary) > 40,000,”, at that point “HAVING” condition can be utilized rather than “WHERE.”

As referenced in the past stage, information refinement must be finished. Here, several filter algorithms are used to delete the unnecessary tables and attributes. One of the algorithms, for instance, filters tables and its comparing characteristics which are a part of some other table in table collection. To sort out the effects and conclude the table collection and table-attribute map, a few different algorithms are applied.

- (5) Form the final query and execute:

The framework only manages MySQL queries, and the models used to shape queries would be focused on the syntax of MySQL. The appropriate template is chosen according to the type of query selected during the second stage of the process. The template shall be drawn from (Figs. 2 and 3):

For questions concerning data recovery:

```
SELECT < select proviso >
FROM < tables >
WHERE < where provision >
ORDER BY < request by condition >
GROUP BY < bunch by proviso >
HAVING < having statement >
```

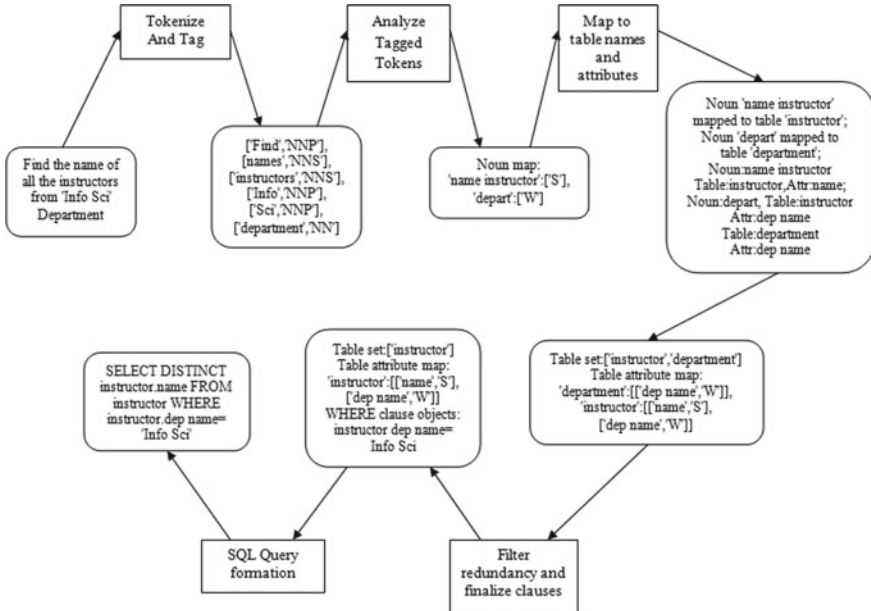


Fig. 2 Implementation

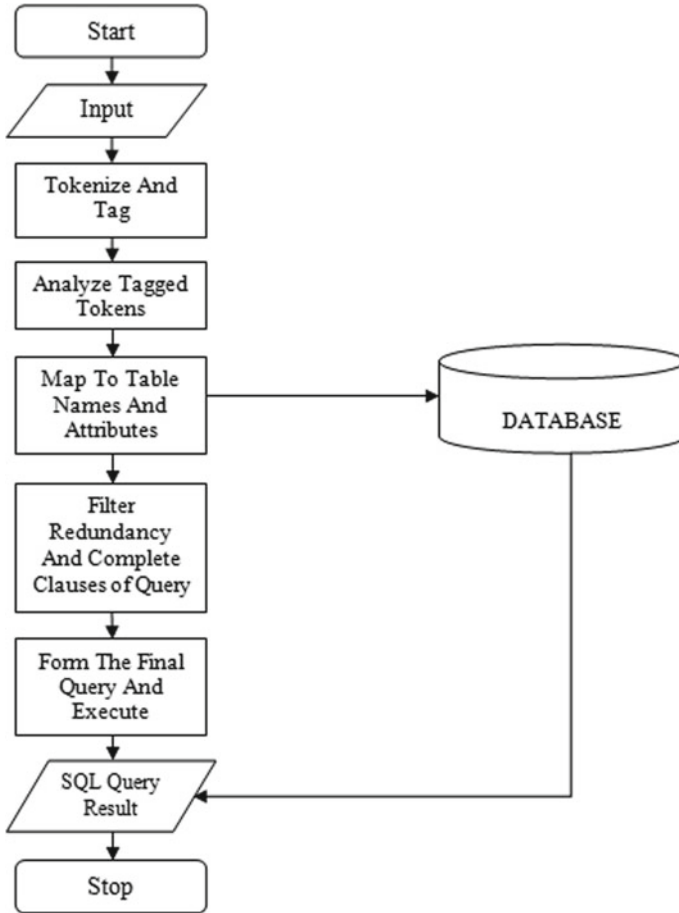


Fig. 3 Flowchart

6 Algorithm

Construction of SQL query from natural language.

Input: Natural language query in English content.

Output: SQL query.

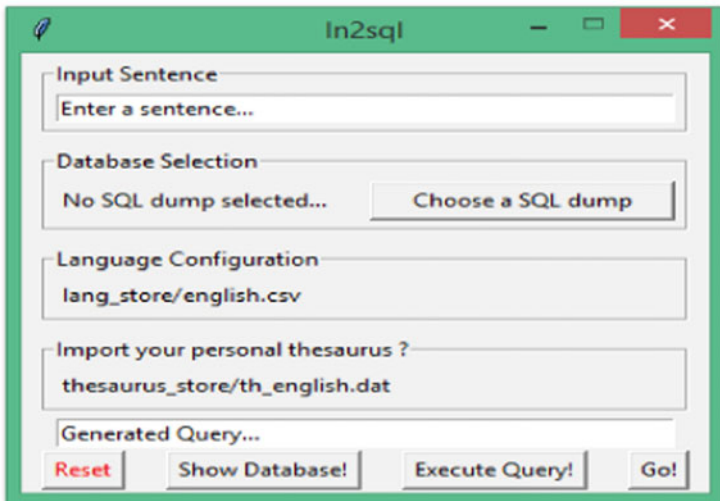
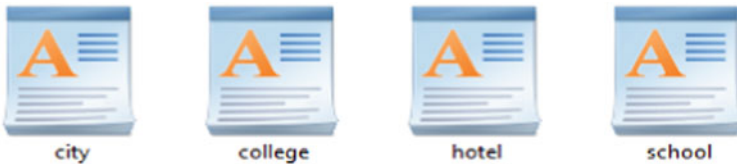
- (1) Tokenization.
Breaking down the input sentence into individual tokens or words.
- (2) Stemming and Lemmatization:
Stemming is the technique to reduce words to their root form (a canonical form of the original word).
Stemming usually uses a heuristic procedure that chops off the ends of the words.

Lemmatization is the text conversion process that converts a word form into its basic form.

- (3) Perform POS labeling.
 POS tagging is used to tag the parts of speech for individual tokens.
- (4) Parse_sentence = Parse utilizing customary articulations
- (5) If table.attribute ∈ Parsed_sentence
 - a. Formation of query
 - b. Extraction of query
 - c. SQL query execution.

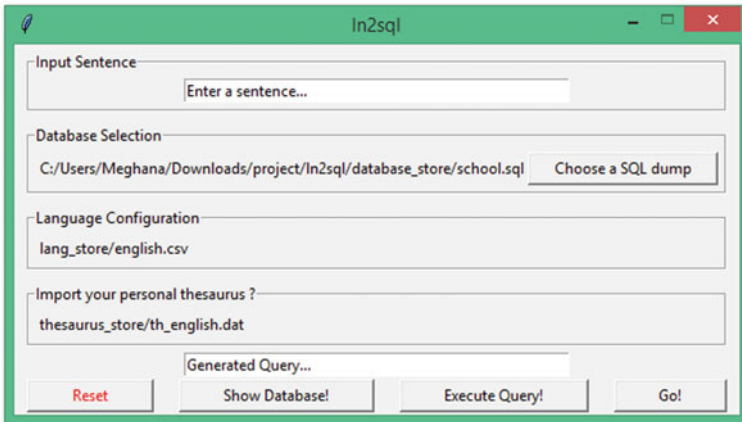
7 Output

Step 1: For database selection “Choose a SQL dump” button is present. Select the appropriate database from the SQL dump. Show database feature acts as a reference to input the sentence by knowing the tables in the dumped database.

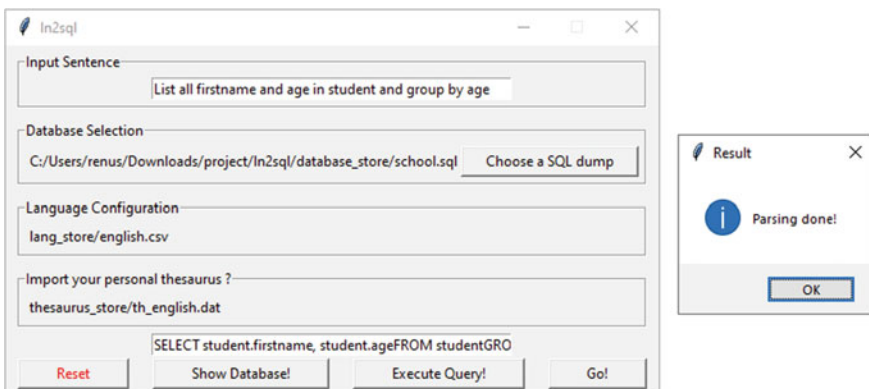


Step 2: Enter the input statement, the keywords in the input sentence will be matched with the data in the database while processing.

By default, language configuration and thesaurus are English, and these are used in removal of stop words and formation of SQL query.



Step 3: On clicking on “Go!” button the relevant SQL query for the given input sentence is generated.



Step 4: Click on the “execute query” button, to display the output of the resultant SQL query from the database.

	SELECT	udent.firstnan	
1	Jean	20	
2	Jean	21	
3	Rose	22	
4	Alain	23	
5	Sandy	24	

8 Conclusion

This project has given us a great opportunity to come up with a solution for writing tedious queries. We were able to learn and implement NLP. Although there are various methods of extracting data from the database, natural language processing (NLP) has set another standard in extracting the data.

This is first done by analyzing the input text, which is later broken into tokens, followed by formation of SQL Query, with the help of several python libraries like NLTK. After query formation, the data is extracted from the database. In order to produce an accurate SQL query from the characteristic language inquiry, various procedures such as tokenization, lemmatization, syntactic, and semantic investigation are processed.

We can add more equivalent words for column names so that the system is able to handle more queries. This research paper also plans to infer a programmed inquiry interpreter for natural language-based inquiries into their related SQL queries and gives an easy way to understand the interface between end users and the database for simple access of information. We present a way to deal with computerizing the change of natural language query to SQL query successfully. The proposed model is cost effective as it only requires initial development cost. It is also a quicker alternative to learn by students as it saves time and effort. The project objective is to provide accessibility to handle data in the database without having any specific background in computer science.

9 Future Scope

The future scope of this project will take a look at prompting the user for correct enter token to corpus token mappings to build up a data dictionary on the database and enhance its performance. It additionally does not cope with non-natural column or table names, studentName or student_name for example. In this case, we could split on camelcase and underscores, respectively, and keep the nouns in the new word

set. The input tokens can be matched against the set of words that are present in the data dictionary.

There are many steps to improve the work we have done. We can use a data dictionary to match input tokens not only to the table and column names, but their synonyms as well would greatly increase our accuracy and chances of input tokens can be matched to the correct corpus name. The input information can be in the form of audio format that can be changed into textual format and with the help of NLP process we can generate SQL Queries. Additionally, JOIN keywords can be used to retrieve the data from different tables related to each other in the database.

References

1. Xu B, Cai R, Zhang Z, Yang X, Hao Z, Li Z, Liang Z (2019) NADAQ: natural language database querying based on deep learning. *IEEE Access* 7:35012–35017
2. Nandhini S, Viruthika B, Almas S, Das SS (2019) Extracting Sql query using natural language processing. *Int J Eng Adv Technol (IJEAT)* 8(4). ISSN: 2249-8958
3. Zhong V, Caiming X, Richard S (2017) Seq2sql: generating structured queries from natural language using reinforcement learning. arXiv preprint. [arXiv:1709.00103](https://arxiv.org/abs/1709.00103)
4. Xu X, Liu C, Song D (2017) Sqlnet: generating structured queries from natural language without reinforcement learning. arXiv preprint. [arXiv:1711.04436](https://arxiv.org/abs/1711.04436)
5. Chaudhari PP (2013) Natural language statement to SQL query translator. *Int J Comput Appl* 82(5)
6. Sathick KJ, Jaya A (2015) Natural language to SQL generation for semantic knowledge extraction in social web sources. *Indian J Sci Technol* 8(1):1
7. Finegan-Dollak C, Kummerfeld JK, Zhang L, Ramanathan K, Sadasivam S, Zhang R, Radev D (2018) Improving text-to-sql evaluation methodology. arXiv preprint. [arXiv:1806.09029](https://arxiv.org/abs/1806.09029)

Cost Optimization for Migration of Data in Cloud Data Centers



S. Spoorthi, N. Rakshitha, and K. S. Chandrababha

1 Introduction

Developing broker model is to assist cloud with overhauling clients to realize which cloud specialist organization is reasonable for the necessities he has inquired. In reality situation, this sort of model does not exist as a result of this numerous clients face seller lock in circumstance where if a client needs to change cloud specialist organization he cannot as he and CSP will hold service level agreement (SLA). Consequently, in the recommended module a client can transfer, download documents through their account.

Cloud storage providers (CSPs) offer geographically data stores providing several storage classes with different prices. An important problem facing by cloud users is how to exploit these storage classes to serve an application with a time-varying work-load on its objects at minimum cost. This cost consists of residential cost (i.e., storage, put, and get costs) and potential migration cost (i.e., network cost). To address this problem, we first propose the optimal offline algorithm that leverages dynamic and linear programming techniques with the assumption of available exact knowledge of workload on objects.

Utilizing various cloud service providers: Dependence on a solitary cloud service provider brings about three-folds obstructions: accessibility of administrations, data lock-in, and non-efficient use. To lighten these deterrents, one may utilize numerous cloud services that offer figuring, determined capacity, and system administrations with various highlights, for example, cost and performance. Being propelled by these different highlights, programmed choice of cloud services dependent on their capacities and client's predetermined prerequisites are proposed to figure out which cloud service providers are appropriate in the exchange offs, for example, cost versus latency or delay and cost versus performance. Amazon S3, Google Cloud Storage,

S. Spoorthi (✉) · N. Rakshitha · K. S. Chandrababha
Siddaganga Institute of Technology, Tumkur, Karnataka, India

and Microsoft Azure as driving CSPs offer different sorts of capacity (i.e., blob, block, file, and so on.) with different costs for at least two classes of capacity administrations: Standard storage and reduced redundancy storage. Each CSP additionally gives API orders to recover, store, and erase information through system administrations, which forces in-and out-network cost on an application. In driving CSPs, in network cost is free, while out-network cost is charged and might be different for providers.

1.1 Problem Statement

Information moving among data centers of a CSP (e.g., Amazon S3) in various locales might be charged at lower rate (hereafter, it is gotten diminished out-arrange cost). This expansion assumes a focal job in the enhancement of information the executives cost in cloud situations. We target improving this cost comprises of private expense (i.e., capacity, put, and get expenses) and potential relocation cost (i.e., organize cost). The expense of data hosting the board is likewise influenced by the normal remaining task at hand of an article. Data hosting decides storage mode and the clouds that the data should be stored in.

2 Literature Survey

1. *Title of the work*: “Towards Network-level Efficiency for Cloud Storage Services”

Authors: Z. Li, C. Jin, T. Xu, C. Wilson, Y. Liu, L. Cheng, Y. Liu, Y. Dai, and Z.-L.Zhang.

Publication details: IMC, ACM.

Description: Cloud stockpiling administrations, for example, Google Drive, OneDrive provide customers with beneficial and reliable technique to keep and provide statistics from anyplace, on any gadget, and whenever. The basis of those administrations is the statistics syncing pastime which therefore maps the modifications in customers’ community record structures to the cloud by way of a development of gadget interchanges in a handy way. If now no longer established appropriately, be that because it may, the large degree of statistics sync site visitors can probable cause (monetary) torments to each expert co-ops and users. This paper has a tendency to a trustworthy but primary inquiry: Is the existing statistics sync site visitors of allotted garage advantages correctly utilized? We, to begin with symbolize a dimension named TUE to assess the Traffic Use Productivity of statistics synchronization.

2. *Title of the work*: “Proficient Batched Synchronization in Dropbox-like Cloud Storage Services”

Authors: Z. Li, C. Wilson, Z. Jiang, Y. Liu, B. Y. Zhao, C. Jin, Z.-L. Zhang, and Y. Dai.

Publication details: ACM.

Description: To restrain the framework overhead, dispersed capability agencies use matched diff, statistics pressure, and diverse gadgets while transferring updates among customers. Regardless, however, those progressions, we see that internal seeing progressive, brief updates to purchaser statistics, the framework site visitors created via way of means of conveyed capability blessings habitually suggests obsessive inefficient viewpoints. We mean this lead because the site visitors misuse issue. To deal with this issue, we suggest the update-bundled deferred synchronization (UDS) part.

3 Design

This application involves primarily three sections:

3.1 *Broker*

Name of the module: Broker.

Purpose: Broker module is the main part of our project. He is the one who administers everything which happens between the user and the cloud service provider. He channels or suggests the users to the appropriate cloud service provider according to the user needs or the requirements without having vender lock-in issue by being a channel between the user and cloud service provider without them being directly involved.

3.2 *User*

Name of the module: User.

Purpose: User or the client is the one which make use of the service provided by the cloud according to his needs and requirements without totally involving with the cloud service provider because the broker will take care of everything.

3.3 *Cloud Service Provider*

Name of the module: Cloud Service Provider.

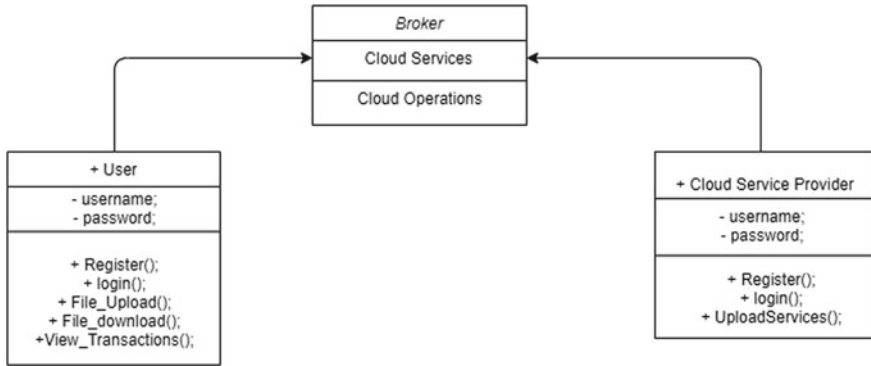


Fig. 1 Class diagram

Purpose: These are the service providers like Amazon S3, Microsoft Azure, Google, etc., which provide services like storage to the users with the help of broker without intervention of themselves.

Class Diagram

In Fig. 1, we can see how the classes are interacted. Each block in the above figure has class name, class attributes, and class methods. “+” indicates the attribute or class or method is public. “-” indicates the attribute or class or method is private. “” indicates the attribute or class or method is protected.

Architecture

In Fig. 2, user requests for the specification to the broker or he selects he files to upload on to the cloud. According to the requirements of the user like available resource, efficiency, etc., the broker suggests the services and update it to the user.

4 Implementation

Algorithm Used: Rivest–Shamir–Adleman (RSA).

It is an asymmetric block cipher algorithm utilized for encryption and decoding of data. It is used for secure data transfer. It has a private key for decryption and public key for encryption.

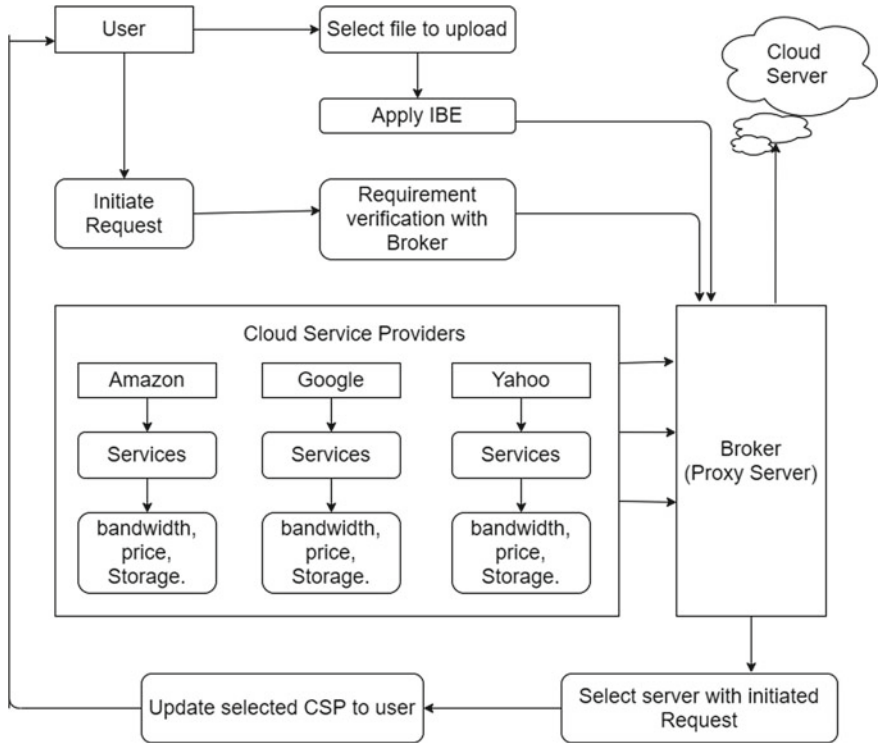


Fig. 2 Architecture

4.1 Functionalities

4.1.1 Register

In this module, admin should update the service parameter in every time interval, and this updated detail should be displayed to cloud user to get access the service throughout broker module.

4.1.2 Login and Logout

These modules enable customer to login to the application to see their nuances. The customer can login to their records by entering the genuine username and mystery word. If customer enters wrong capabilities, by then fitting slip-up message is showed up, and customer is not allowed to login. Exactly when the customer login and after all the undertakings, the customer needs to logout of the gateway.

Outsource (e.g., via AWS): *monthly cost*

- S3 costs: \$0.12 per GB month. EC2 costs: \$0.10 per CPU hour (costs from 2009) Storage = $\$ 0.12 \times 524 \times 1000 \sim \62 K
- Total = Storage + CPUs = $\$62 \text{ K} + \$0.10 \times 1024 \times 24 \times 30 \sim \136 K

Fig. 3 Cost calculation formula

4.1.3 Upload File

In this module, we can upload the quantities of data available and can be provided by service providers.

4.1.4 View Transactions

This user module, user should upload files to cloud server. In this module, service details are displayed it shows that which cloud service provider what kind of services.

4.1.5 Download File

This is a user module where user can download the uploaded files whenever user desires to and can migrate to other CSPs if they provide better required service for the user.

4.1.6 Migration of File

This is a user defined operation where the broker will suggest the user while downloading their file if they want to migrate to better CSP.

In Fig. 3, the standard AWS cost calculation formula is shown.

Implementation work flow:

- First is to construct the application. This is accomplished utilizing the eclipse IDE.
- Next, client necessities that must be given was picked in like manner.
- Then we incorporated the representative module with the client module and CSPs with specialist module.
- We have taken Amazon Web Administrations and Windows Azure as CSPs.
- According to the client necessities, specialist module proposes the reasonable CSP for the client.
- Using the above calculation formula, the price for the service will be calculated and displayed in pop-up box notification.

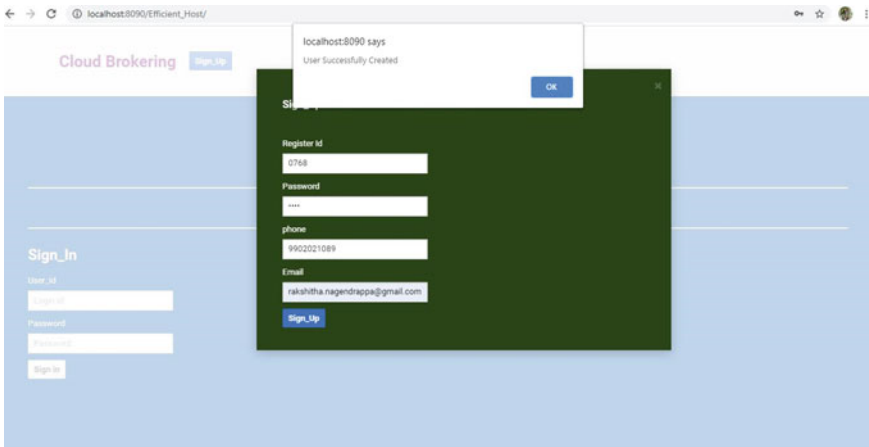


Fig. 4 Sign-up page

5 Result

- Sign in qualifications is given to sign in to the portal.
- If there does not exist a record enlistment is finished utilizing information exchange choice.
- User is given a structure to fill the essential subtleties of their records and capacity required.
- Feasibility of moving from one segment to other is given by a bar on top, which contains the accessible cloud specialist co-ops, their presentation which is plotted utilizing diagram and logout choices are given.
- Viewing the diverse CSPs is accessible.
- Logout alternative is accessible where the client can logout of their records in the wake of utilizing.

Figure 4 shows user sign-up page where he/she needs to register themselves for broker page.

Figure 5 shows the page where admin should update the service parameter in every time interval, and these updated details should be display to cloud user to get access the service thought broker module.

Figure 6 displays the performance graph of all the CSPs.

6 Conclusion

To reduce the cost of migration of data for one provider to other, we must make best and full use of the price difference between storage and network services across

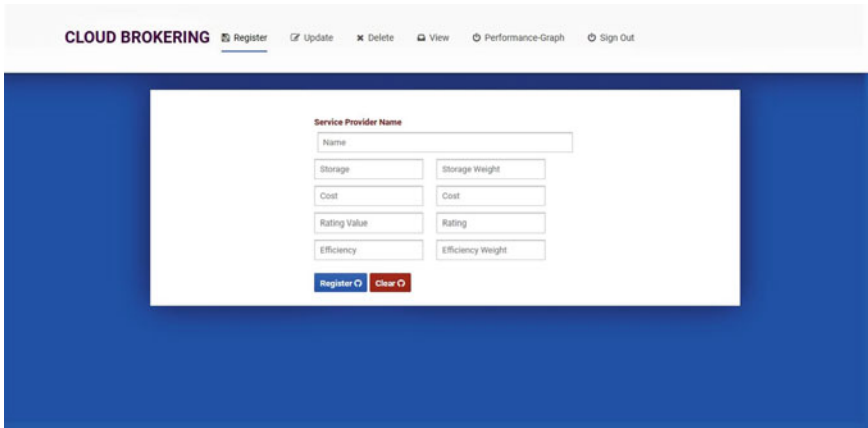


Fig. 5 CSP register page

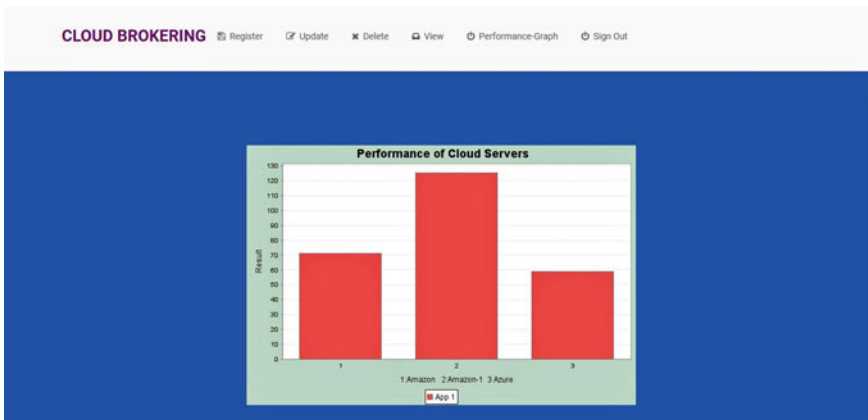


Fig. 6 Viewing performance graph

various CSPs. To achieve this, we have used algorithms, first is optimal resource selection algorithm to reduce the cost of storage, migration cost and for choosing the suitable CSP and RSA for security. To meet our projects main intention of saving the cost using the algorithms, we have used Amazon’s and Microsoft’s clouds storage services Amazon Web Service (AWS) and Azure as services.

We achieved proposing a cost-efficient data hosting scheme with high accessibility in heterogeneous multi-cloud with limited money-related expense and ensured accessibility. Specifically, we merge the two abundantly used mechanisms, i.e., replication and erasure coding. Data security for redistributed data in cloud server is achieved. We were successful in preventing Vendor lock-in situation. We can successfully migrate files from one service provider to another.

As a comprehensive storage framework, there are a few different components to be thought of, for example, cache systems and so on. Notwithstanding, we just aim at minimizing the cost while meeting adaptable accessibility necessities. In spite of the fact that we have thought about the multifaceted nature and attainability when structuring this methodology, the framework configuration is out of the extent of our project right now, and we put the point by point framework plan of multi-cloud information facilitating into future work. In future, we can enhance the project to next level by adding a greater number of cloud service providers to the web application hence gives more clarity when compared with all other clouds. Also, broker can be made using multiple servers instead of one, handling many processes at a time in faster rate with high efficiency.

References

1. Mansouri Y, Toosi AN, Buyya R (2017) Cost optimization for dynamic replication and migration of data in cloud data centers. *IEEE Trans Cloud Comput* 7(3):705–718
2. Teli P, Thomas MV, Chandrasekaran K (2016) Big data migration between data centers in online cloud environment. *Procedia Technol* 24:1558–1565
3. Son AY, Byun JY, Yong C, Huh EN, Hyun JH, Kang KK (2017) Energy efficiency oriented migration scheme in cloud data center. In: *IEEE international conference on cybernetics and computational intelligence (CyberneticsCom)*, IEEE, pp 108–113
4. Li X, Wang L, Lian Z, Qin X (2018) Migration-based online CPSCN big data analysis in data centers. *IEEE Access* 6:19270–19277
5. Shen Q, Zhang L, Yang X, Yang Y, Wu Z, Zhang Y (2011) SecDM: Securing data migration between cloud storage systems. In: *2011 IEEE 9th international conference on dependable, autonomic and secure computing*, IEEE, pp 636–641
6. Ni J, Chen Y, Sha J, Zhang M (2015) Migration from HPC-based data processing systems to cloud-computing based data mining systems. In: *2015 8th international conference on internet computing for science and engineering (ICICSE)*. IEEE, pp 181–187
7. Khalil I, Hababeh I, Khreishah A (2016) Secure inter cloud data migration. In: *2016 7th international conference on information and communication systems (ICICS)*. IEEE, pp 62–67
8. Hao W, Yen IL, Thuraisingham B (2009) Dynamic service and data migration in the clouds. In: *2009 33rd annual IEEE international computer software and applications conference*, vol 2. IEEE, pp 134–139
9. Bansal A, González-Vélez H, Chis AE (2016) Cloud-based NoSQL data migration. In: *2016 24th Euromicro international conference on parallel, distributed, and network-based processing (PDP)*. IEEE, pp 224–231
10. Sabiri K, Benabbou F, Moutachaoui H, Hain M (2015) Towards a cloud migration framework. In: *2015 3rd world conference on complex systems (WCCS)*. IEEE, pp 1–6

Leveraging X-Ray and CT Scans for COVID-19 Infection Investigation Using Deep Learning Models: Challenges and Research Directions



Vidyadevi G. Biradar, H. C. Nagaraj, and H. A. Sanjay

1 Introduction

COVID-19 pandemic is threatening the world, and health care bodies are facing challenges to treat coronavirus infected patients. COVID-19 symptoms include fever, cough, and shortness of breath with possible abdominal symptoms and exhibits asymptomatic symptoms in many cases. The radiology is a helpful technology in analyzing chest X-ray and CT image to inspect lung infection. Chest CT scans and X-ray play significant role in detecting and assessing the severity of COVID-19 infection and evaluating the patient response for the treatment given [1, 2]. The COVID-19 is contagious and therefore posed tough challenges to mitigate the spread of infection. The corona virus infection is proven to be life threatening and inflicted great fear in the minds of people across the world. In the context of Indian culture, people across the country offered prayers together and manifested their belief in facing coronavirus threat through customs lighting the lamps, and thanksgiving was offered to care takers by hand clapping by entire nation.

Figure 1 shows various strategies to mitigate corona virus infection, its impact on the workload of health care and law enforcement professionals and outlines the scope for deep learning techniques for COVID-19 pneumonia infection diagnosis.

Artificial intelligence (AI) support in analyzing chest X-ray and CT scans reduces the burden on radiologist and eliminates inaccuracies induced by extraordinary workload. Therefore, there is a need for development of a tool for automatic analysis of radiographic images. This tool computes high-opacity abnormalities in lungs and relates to COVID-19 infection. The use of X-ray and CT-scan images are in debate; however, when these are combined with reverse-transcription polymerase chain reaction (RT-PCR) pathological test, it resulted in accurate diagnosis. CT

V. G. Biradar (✉) · H. C. Nagaraj · H. A. Sanjay
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: vidyadevi.g.biradar@nmit.ac.in

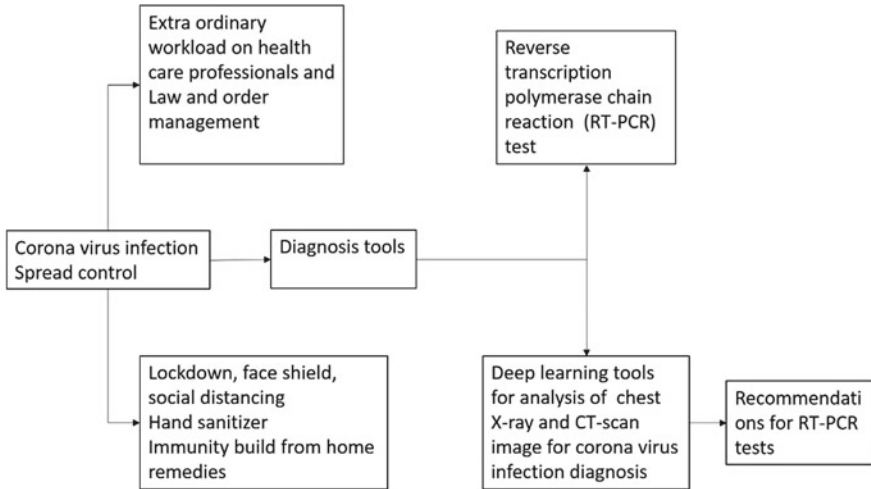


Fig. 1 Strategies to curb coronavirus infection

images are examined for opacity lesions in posterior and peripheral lungs which indicates COVID-19 infections [3]. X-ray is not a powerful tool in finding COVID-19 symptoms, and there were many cases of missing pulmonary nodules which were found in CT scans [4]. Medical professionals facing the challenges of COVID-19 diagnosis are provided with artificial intelligence support system which accelerates the assessment of coronavirus infections from chest X-ray and CT images [5].

Radiologists can investigate COVID-19 infection with the aid of CT scans and X-ray of the chest. The chest CT image demonstrates significant patterns which helps in COVID-19 infection inferences, and important features are ground glass opacities, crazy paving, vascular dilation, traction bronchiectasis, subpleural bands, and architectural distortion. These patterns report the progression of COVID-19 infection, ground glass opacities (GGO) - presence of this pattern in CT image indicates the early stage of COVID-19 infections, these patterns are multifocal, bilateral, and peripheral located in the inferior lobe of right lung, ii. Crazy paving indicates thickened interlobular and intralobular lines, presence of crazy paving in addition to ground glass opacities represents the later stage of COVID-19 infection, iii. Vascular dilation is widening of the vessels in ground glass area, iv. Traction Bronchiectasis is another finding in the areas of ground glass, v. Subpleural bands and Architectural distortion, in some patients there are architectural distortions in addition to new development of subpleural bands [6].

Chest X-rays are significantly useful in triaging and disease progression investigation although these are less informative as compared with CT scans. The COVID-19 pneumonia patterns help in deciding whether CT imaging is required [7]. Chest radiograph images (CXR) in the poster anterior views (PA) are useful in diagnosis of an acute and chronic infection in lungs. The X-rays are examined for lung opacities

which are white vague areas in the dark regions of lungs and lung opacities in X-ray of COVID-19 positive has high contrast regions as compared to the pneumonia X-ray.

The computer-aided diagnostic tools which are designed using deep convolutional neural network models have been useful in pneumonia infection investigation from chest X-ray and CT scan images. In some cases, they have outperformed expert opinions. Therefore, deep learning research community is continuously exploring the feasibility of deep learning paradigm for COVID-19 diagnosis. This paper presents a comprehensive survey on deep convolutional neural network models for corona virus infection diagnosis.

The survey presented in this paper gives an insight into suitability of pre-trained models such as SqueezeNet, VGG-16, VGG-19, GoogleNet, ResNet50, and DarkNet, etc., through transfer learning. The contributions of this work include identification of best pre-trained models, augmentation techniques for COVID-19 positive dataset, enumerates a list of possible dataset with necessary description that are used in research work on analysis COVID-19 chest X-ray and CT-scan images using deep convolutional neural networks.

The paper is organized into different sections, and Sect. 2 gives a brief survey on COVID-19 investigation methods using chest X-ray. Section 3 presents survey on COVID-19 pneumonia detection using CT-scan images. Section 4 gives survey on COVID-19 database, and Sect. 5 presents augmentation techniques, and Sect. 6 discusses on possible research directions for COVID-19 diagnosis using deep convolutional neural network, and Sect. 7 provides conclusion.

2 COVID-19 Infection Diagnosis Using Chest X-Ray Images

The chest X-rays allow quick triaging, X-ray equipment is available in all health care centers, X-rays will be taken in isolated room so there is a low possibility of COVID-19 transmission. The models utilize the salient information content using X-ray images.

Deep learning is a paradigm for automatic feature extraction, and convolutional neural networks are used for feature extraction and classification in various medical applications. In convolutional neural networks, earlier layers extract more primitive features whereas middle layers and end layers extract features that are more specific to image dataset. Therefore, by utilizing the knowledge of state-of-the-art pre-trained models, a novel problem is being solved [8]. Transfer learning is a technique, wherein the feature extracted by a CNN from large image dataset is utilized to solve new problem which needs analysis of altogether a new image dataset, which is small [9]. Deep learning models have limitations to learn features from small dataset from the scratch.

The investigation on COVID-19 infection analysis in [10] is carried out using chest X-ray using deep learning model. The results yielded by the model were utilized in reverse-transcription polymerase chain reaction (RT-PCR) test recommendations. The COVIDx dataset which comprises of 13,975 number of X-rays pertaining to 13,870 patients is used in this work. The database with large number of COVID-19 positive samples is created by combing the images from five open source benchmark COVID-19 databases. A novel convolutional neural network model known as COVID-Net was developed for classification of X-rays as COVID-19 positive cases, COVID -19 negative case and no infection. The authors aimed at building a robust model, and therefore, the model was trained in two stages. In first stage, the model was trained on ImageNet which contains millions of non-medical images, and then in the second stage, it was trained on COVIDx dataset. The model training is carried with hyperparameters: 22 epochs, batch size of 64, and $2e-4$ learning rate, augmentation is applied through intensity shift, translation, zoom, horizontal shift, and rotation. The re-balancing strategy is applied to balance the distribution of dataset. The model is built using Keras libraries with TensorFlow as background [10].

Researchers of [11] have developed a model for classification of X-ray images into COVID-19 positive and COVID-19 negative cases based on inception residual recurrent convolutional neural network (IRRCNN) which is pre-trained on pneumonia dataset. Another model based on NABLA-3 network was developed for segmentation of infected regions from X-ray and CT images. Then IRRCNN model is trained to classify these segmented image patches as COVID-19 infection and normal using strategic transfer-learning technique. Authors claim that the results are encouraging and to get better confidence in the models, they need to be evaluated on large dataset [11].

Authors in [12] have explored the suitability of pre-trained CNNs such as ResNet50, ResNet18, DenseNet-120, and SqueezeNet to draw corona virus infection inferences using chest X-ray images by applying transfer-learning techniques. The models were tuned to classify COVID-19 dataset by unfreezing the last layer of convolutional neural networks, and for earlier layers, pre-trained model weights are utilized. The models are evaluated on publicly available chest COVID-X-ray-5 k dataset comprising 5000 X-rays corresponding COVID-19 positive cases as identified by a radiologist. To improve the performance of these models, COVID-19 positive X-ray images were augmented during training of the model by applying flipping, adding small distortion, oversampling, and rotation operations.

A set of pre-trained models such as MobileNet V2, VGG-16, VGG-19, Inception-ResNet V2, Xception, DenseNet201, ResNet52 V2, Inception V3, and NASLetLarge are explored in [13] using transfer-learning for diagnosis of COVID-19 infections from X-ray. This work uses X-ray images from Dr. Joseph Cohen's COVID-19 dataset. The dataset is augmented on fly during model training using augmentation techniques: rotation, flip, shearing, zooming, cropping, and adding small random noise. The results show that VGG-16 and VGG-19 are superior as compared with models in discussed in this work.

Work in [14] reports the investigations on pre-trained convolutional neural networks: MobileNet v2, VGG-16, Inception, Inception ResNet v2, Xception, for

prediction of COVID-19 infection using transfer-learning. These models are trained on COVID-19 infected X-rays. Transfer learning helps in exploiting the power of pre-trained models when the dataset pertaining to new application is small. The X-ray dataset is collected from medical database which comprises of two sets. First, 1427 number of total X-ray images in which 224 images COVID-19 affected, 700 images falling in bacterial pneumonia, and 504 number of normal images. Second, a dataset comprising 224 number of images representing COVID-19 disease, 714 images of viral and bacterial pneumonia, and 504 images depicting good health conditions. The results show that VGG-16 and MobileNet v2 gave better results.

The research work in [15] explores a novel deep learning model known as DarkNet-19 for COVID-19 pneumonia diagnosis using chest X-ray images. The architecture of this model is designed by considering the architecture trend of DarkNet model as a reference. The model presented in this work follows the trend of convolutional layers from DarkNet, however, with reduced number of filters and convolutional layers as compared to the original DarkNet model. The researchers in this work present their findings stating that to analyze COVID-19 infection, and it is important that the model learns every minute detail from the image and just by designing very deep neural network model like ResNets and ResNext models may not help. The DarkNet-19 model is designed with 17 convolutional layers, batch normalization layer, and LeakyRelu layer. The model is evaluated on Joseph Cohen's dataset, and model gave an accuracy of 98.08%. Authors claim that DarkNet model offers highest classification accuracy as compared to many existing deep learning models.

The investigation presented in [16] design models based on the architectural trend of deep convolutional neural networks such as ResNet50, Inception V3, Inception-ResNetV2 models for classification COVID-19 pneumonia infection. This work uses a popular Joseph Cohen's X-ray repository. The COVID-19 datasets that are publicly available to carry out research are small, and on the contrary to exploit the power of deep convolutional neural networks, large dataset is needed. Therefore, in this work, state-of-the-art pre-trained models which have been trained on ImageNet database are utilized for prediction of COVID-19 infection using transfer-learning techniques. The models are trained in Google collaborative environment, and the parameters include random weight initialization and Adam optimizer. The results show that model based on ResNet50 gave best performance with accuracy of 98%; however, Inception-ResNetV2 gave poor accuracy of 84%.

Most of the publicly available COVID-19 positive datasets are small, investigation in [17] addresses the model overfit issue of deep neural network models when they are trained on a small dataset manifest generalization tendency. This work presents a deep convolutional neural network model known as generative adversarial network (GAN) to generate additional numbers of X-rays on contrast to data augmentation techniques. The research outcome indicates that for classification of COVID-19 pneumonia, a very deep convolutional neural network is not beneficial rather a scheme for detecting salient information from chest X-ray of COVID-19 is needed. The pre-trained models; AlexNet, GoogleNet, and ResNet18 comprise of comparatively a smaller number of layers. Therefore, in this research, these pre-trained models

are used for analysis of COVID-19 dataset using transfer-learning techniques. The models are evaluated on COVID-19 dataset. The performance achieved by AlexNet and GoogleNet is 80.6% and 99.9% of accuracy, respectively. The results indicate that the GoogleNet has yielded highest accuracy. This has an added advantage of reduction in time and memory complexity and suitable for real time analysis.

The work in [18] also explores the usefulness of X-rays in COVID-19 infection detection. The dataset comprising of 122 COVID-19 cases collected from the GitHub. The pre-trained VGG-16 model is adopted for X-ray analysis due to small dataset and therefore applied transfer learning. The last layer is substituted by a fully connected network with 64 neurons, and global average pooling is used for subsampling instead of max pooling. Training is carried out with RMSprop loss function. The model achieved an accuracy of 96.1% with 100% accuracy for COVID-19 classification.

3 COVID-19 Infection Diagnosis Using Chest CT-Scan Images

A large number of pre-trained models with transfer learning are evaluated for classification coronavirus infections, which include SqueezeNet, GoogLeNet, Inception-v3, DenseNet-201, MobileNet-v2, ResNet-18, ResNet-50, ResNet-101, Xception, Inception, ResNet-v2, ShuffleNet, NasNet-Mobile, NasNet-Large, AlexNet, and VGG-16.

The models were trained and evaluated on COVID-19 CT scans which are publicly available [19]. The outcome of this research indicates that there is no need for segmentation of region of interest in CT images and models predict from entire chest CT image. Further, results also indicate that these models gave better results without image augmentation on contrary with augmented image dataset in general [20].

The research investigation in [21] comes out with a deep neural network model for corona virus infection progression analysis. A model known as COVID-19Net is implemented whose architecture is designed by following the architectural trend of DenseNet. This work utilizes a database of 5372 number of CT images in which 4016 CT images are used for training pre-trained deep learning model. The model COVID-19Net is trained on the image patches that corresponds infectious areas in chest CT-scan images. These patches are automatically segmented using pre-trained DenseNet121-FPN mode using transfer-learning techniques. COVID-19Net architecture includes multiple convolutional layers followed by batch normalization and ReLu activation layers, and feature reduction in the last convolutional layer is carried out using global average pooling method. COVID-19Net model performance as indicated by AUC and achieved $AUC = 0.87$ for COVID-19 infection and $AUC = 0.86$ for viral infection. The results of COVID-19 infection classification also showed concurrency with the opinion of radiologist.

In this investigation, a deep learning model is developed by consulting doctors for the recommendation for chest traverse- section CT scans for COVID-19 diagnosis.

The pathological RT-PCR requires longer time for findings of COVID-19 infection analysis. Therefore, doctors expressed that there is a need for AI system which assists in COVID-19 infection detection. The methodology adopted in the development of deep learning model is, traverse-section CT images are preprocessed, and then a 3D CNN is used for segmentation of multiple images cubes which are then classified as infectious and non-infectious. In this work, VNET-IR-RPN deep learning model, a variant of VNET pre-trained model which was used in tuberculosis detection in earlier work is utilized for segmentation. Then, the Bayesian function is used for computation of total score from individual classification results. The model was evaluated on a COVID-19 CT-scan dataset collected from an affiliated hospital in Wuhan city. As the performance of deep learning models demand large dataset, researchers have enlarged the original COVID-19 positive cases dataset by applying image augmentation techniques such as random clipping, left translation and right translation and flipping and mirroring operations [22].

The model detects COVID-19 pneumonia from CT scans by observing the patterns such as crazy paving, ground glass opacities, architectural distortion, and multifocal organizing pneumonia [23]. The AI-based deep learning models are proven to be better than an expert radiologist [24]. A deep learning model which is designed with U-Net architecture as a reference is developed for segmentation of infected area in CT images and inception model as encoder [25]. The COVID-19 CT dataset is collected from Indian Hospital which contains 275 CT scans of COVID-19 patients and total 5212 number of CT images. The model gave 0.964 sensitivity and 0.884 specificity.

A set of pre-trained models modified with transfer learning are used for classification of chest CT scans as corona virus infection and normal cases. The models under consideration were ResNet18, ResNet50, ResNet101, and SqueezeNet. In this work, limited dataset issue was addressed by image augmentation using wavelets followed by conventional augmentation techniques such as rotation, translation, and shear operations. The results showed that ResNet18 have achieved highest testing accuracy of 99.4%. Further, the abnormality in CT scan images are visualized using feature map. Other contributions in this work is a brief survey on COVID-19 diagnosis using X-ray, and CT scans is presented [26].

In this work, a combined model using on recurrent neural network (RNN), and CNN known as ProgNet model is presented for prognosis of COVID-19 pneumonia from a multi-temporal chest X-rays. Time-series analysis is carried out using this model. The RNN model is designed for analysis of chest X-ray radiology time series, and CNN model is for classification of X-ray with COVID-19 and non_COVID-19 infection. The classifier is implemented using pre-trained ResNet50 and LSTM models. This combined model gave an accuracy of 92% for prognosis of X-ray with COVID-19 infection [27].

Three different models are implemented in [28] for classification of corona virus infection, and their performances were compared. The first model is the base model, a convolutional neural network implemented from scratch with five convolutional layers and ReLu function, each layer with 32 filters with 3×3 as kernel size. The base model does use pooling layers and batch normalization layers. The second

model is derived from pre-trained models ResNet50 and DenseNet121 which are fine-tuned using transfer learning. The third model known as COVID-CXNet is implemented using CheXNet which is pre-trained on largest pneumonia and hernia dataset. All these three models were trained on a common COVID-19 dataset. This dataset includes 738 number of chest X-rays diagnosed as COVID-19 positive from RT-PCR tests are collected from various hospitals. The interpretation of these deep learning models is visualized by using gradient weighted class activation mapping (Grad_CAM) and Local Interpretable Model-Agnostic Explanations (LIME) techniques. The performance of these three models are compared, and it is reported that COVID-CXNet gave highest F1 score of 0.96, and base model gave 0.94 [28].

A set of pre-trained convolutional neural network model tweaked for COVID - 19 dataset using transfer learning is used for feature extraction and support vector machine for classification. Pre-trained models VGG-16, GoogleNet, and ResNet50 were used for feature extraction, and features extracted from these models are fused. In this work, total of 3000 image patches of size 16×16 and 32×32 with different gray levels are taken from CT scan images for COVID-19 diagnosis. These patches are labeled as COVID-19 and non-COVI-19 cases. The redundancy in the features set is reduced using t-test which also reduces the memory complexity. These features are classified into COVID-19 infection and normal using SVM classifier. This model outperforms with an accuracy of 95.60% as compared to the results obtained by the individual pre-trained models [29].

4 Benchmark Databases for COVID-19 Diagnosis Using Deep Learning Models

Table 1 is the comprehensive list of databases of X-ray and CT-scan images that are used for COVID-19 diagnosis using deep learning models. This table gives the details about which databases are selected by various researchers to generate a database to incorporate large amount of sample for normal, pneumonia, and COVID-19 pneumonia. These databases contain X-ray and CT-scan images.

The deep learning models give dependable outcomes when they are subjected to large dataset and existence of individual small COVID-19 datasets pose a question on reliability of prediction accuracy of these models. Therefore, all these databases need to be combined to create a large database which helps in tuning-up model performance. Figure 2 shows sample COVID-19 X-ray images from Dr. Joseph Cohen's dataset which are annotated expert radiologist, this is a popular database and widely used in academic research.

Table 1 Benchmark databases for COVID pneumonia infection diagnosis

S. No	Authors	Dataset name	Dataset description
1	Mohammad Rahimzadeha, Abolfazl Attarb [30] [2020]	Covid chest X-ray dataset	This contains 180 number of X-rays from COVID-19 positive cases, 42 for pneumonia
		Rсна pneumonia detection challenge	This contains 6012 number of pneumonia infection X-rays and 8852number of X-rays of normal persons
2	Enzo Tartaglione, Carlo Alberto Barbano, Marco Calandri, Claudio Berzovini and Marco Grangetto [31] [2020]	COVID-Chest X-ray	It comprises of total 287 CXR and CT images with COVID-19 and other types of pneumonia infections Out of 137 CXRs, 108 COVID-19 positives and 29 COVID negatives
		CORDA: Gathered from website and publications	It contains 447 chest X-rays from 386 patients, 150 samples COVID-19 negatives and 297 positives
		Montgomery County X-ray Set	This dataset contains 138 X-ray images are collected under tuberculosis control program 80 images are from normal patients and 58 are from abnormal patients
		Shenzhen Hospital X-ray Set	This database contains total 662 X-rays, 326 X-rays of healthy patients, and 336 X-rays of abnormal patients
		Chest X-ray	This dataset comprises of 1583 normal patients, 2780 affected by bacterial pneumonia, 1493 affected by viral pneumonia
		RSNA	This dataset comprises of total 20,672 normal chest X-rays and 6012 X-rays of pneumonia case
3	Gianluca Maguolo, Loris Nanni [32] [2020]	NIH dataset/Chest x-ray 8 database	This database comprises of 108,948 X-rays taken from 32,717 patients, and these X-rays are classified into eight categories of infections
		CHE dataset	This dataset contains 224,316 number of X-rays acquired from 65,240 number of patients which are categorized into 14 classes
		KAG dataset	There are 5863 pediatric images
		COV dataset	There are 144 frontal x-ray of corona virus positive patients

(continued)

Table 1 (continued)

S. No	Authors	Dataset name	Dataset description
4	Mohammad Rahimzadeh, Abolfazl Attar, Seyed Mohammad Sakhaei [33] [2020]	COVID-CTSet	This dataset comprises of total 63,849 CT images from 377 patients. There are 15,587 images of 95 number of corona virus patients, and 48,260 CT images of 282 healthy persons
5	Linda Wang, Zhong Qiu Lin and Alexander Wong [10] [2020]	COVIDx dataset	COVIDx is a benchmark dataset with highest COVID positive cases. It contains of a total number of 13,975 CXR images from 13,870 patients This database combines image dataset from five different databases The images of COVID-19 cases are taken from three databases. These include databases: ActualMed COVID-19 Chest X-ray dataset initiative, COVID-image data collection and chest X-ray dataset initiative Dataset for usual pneumonia cases are taken from RSNA dataset and COVID-19 radiography database
6	Ali Narin, Ceren Kaya, Ziyne Pamuk [16] [2020]	Dr. Joseph Cohen dataset	This database contains Chest X-ray and CT-images from the patients suffering from acute respiratory distress syndrome (ARDS), COVID-19, Middle East respiratory syndrome (MERS), pneumonia, severe acute respiratory syndrome (SARS). It contains 50 corona virus positive patients and 50 normal chest X-ray Chest X-ray images Kaggle database
7	Asmaa Abbas · Mohammed M. Abdelsamea · Mohamed Medhat Gaber [34] [2020]	JSRT dataset and Dr. Joseph Cohen's chest X-ray image dataset	It comprises of 80 chest X-rays of normal cases in Japanese Society of Radiological Technology Dr. Joseph Cohen's dataset then contained 105 total images with 11 images with COVID-19 and SARS
8	Umut Ozkaya, Saban Ozturk, Mucahid Barstugan [35] [2020]	Societa Italiana di Radiologia Medica e Interventistica	Total 53 CT images from COVID_19 positive patients are taken from Societa Italiana di Radiologia Medica e Interventistica to create a database of CT image patches. In this work generates two sets of data Set-1 comprises of 16 × 16 size patches, 3000—COVID-19 patches and 3000—normal images Set 2 comprises of 32 × 32 size patches, again 3000 number of patches from COVID-19 positive images and 3000-normal images

(continued)

Table 1 (continued)

S. No	Authors	Dataset name	Dataset description
9	Md Zahangir Alom, Shamima Nasrin, M M Shaifur Rahman, Mst, Tarek M. Taha, and Vijayan K. Asari [36] [2020]	Dr. Joseph Cohen's dataset	Dr. Cohen's dataset is used for collecting COVID-19 positive dataset [37] Kaggle repository chest X-ray pneumonia images, which contains, the total number of 5216 images with 1341 normal and 3875 pneumonia infected images Kaggle repository lung segmentation image chest dataset with masks A total of 420 CT images collected which contains 247 normal and 178 images are COVID-19 affected
10	S. Tabika, A. Gómez-Ríos, J.L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Areac, D. Chartea, E. Guiradod, J.L. Suárez, J. Luengo, M.A. Valero-González, P. García-Villanovab, E. Olmedo-Sánchez, F. Herrera [38] [2020]	Blend of 6 databases	COVID-19 image data collection is used for COVID-19 dataset. This dataset contains 76 COVID positive cases and 26 COVID-19 negative cases The database of non-COVID-19 is generated from five databases, they are SNA Pneumonia CXR challenge dataset, COVID-19 Chest X-ray dataset initiative, Chest X-ray8, MIMIC-CXR dataset, and PadChest dataset
11	Rohit Lokwani, Aniruddha Pant, Ashrika Gaikwad, Viraj Kulkarni, Amit Kharat [39] [2020]	Indian Hospital database	This work uses CT scan images from COVID-19-positive and non-COVID data from GitHub and consolidation and healthy CT scans from Indian hospital. This data contains 275 CT scan COVID-19-positive findings from 143 patients. These images are validated from ground truth results obtained from RT-PCR tests
12	Chiou-Jye Huang, Yung-Hsiang Chen, Ping-Huan Kuo, Yuxuan and Ma [40] [2020]	Surging News Network and World Health Organization (WHO)	Database is created from the data collected from Hubei Province, Zhejiang, and Guangdong Province cities
13	Aayush Jaiswala, Neha Gianchandania, Dilbag Singha, Vijay Kumarb and Manjit Kaurc [41] [2020]	SARS-CoV-2 CT scan dataset	The dataset comprises of a total of 2492 CT-images. 1230 are negative for SARS-CoV-2 infection, and 1262 are positive for SARS-CoV-2 infection

(continued)

Table 1 (continued)

S. No	Authors	Dataset name	Dataset description
14	Sakshi Ahuja, B.K.Panigrahi, Nilanjan Dey, Venkatesan Rajinikanth, Tapan Gandhi [42] [2020]	Covid-ct-dataset UCSD-AI4H/COVID-CT https://www.medrxiv.org/ https://www.biorxiv.org/	This work makes collection 349 CT scans COVID-19 positive across 216 patients and 397 CT images from non-COVID patients by combining dataset from of COVID-ct-dataset, images from medrxiv.org and biorxiv.org websites

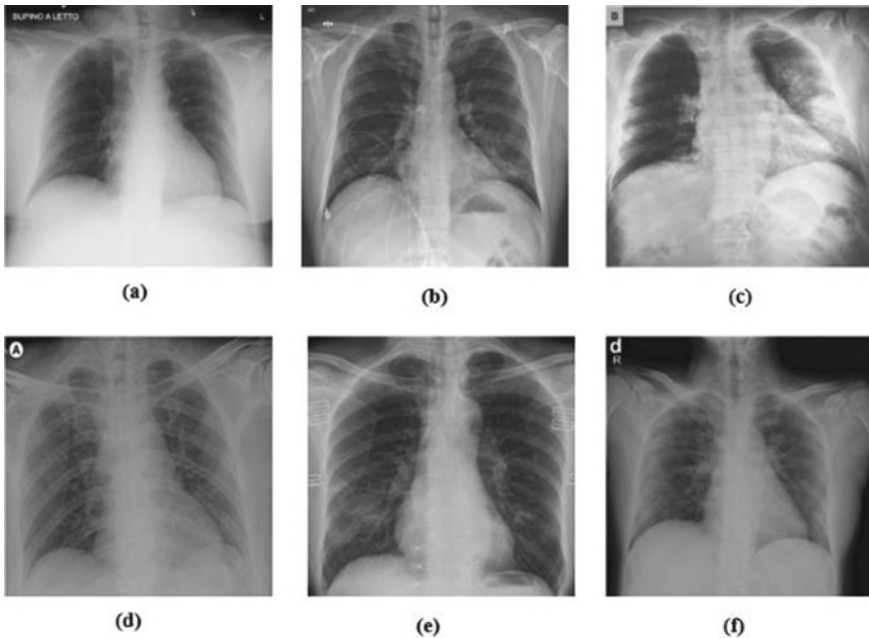


Fig. 2 Dr. Joseph Cohen’s dataset, **a** cardio-visual shadow, **b** increased left basilar opacity indicating development of infection, **c** consolidation and progressive infiltrate, **d** ground glass opacities in lower lobes and small consolidation in upper lobe, **e** right infrahililar opacities demonstrating infection, and **f** progression of bilateral perihilar infiltration with ill-defined patchy opacities [37]

5 Image Augmentation for COVID-19 X-Ray and CT-Scan Image Dataset

Image augmentation techniques are needed to enlarge the dataset and thus addressing the issue of imbalanced dataset. The augmentation enhances inter-class variance within training and testing dataset and, therefore, reduces model generalization error [43].

The COVID-19 X-ray images are augmented by mirroring the X-ray image, and then multiple images are derived by applying various operations on original and mirrored image. These operations include, addition of Gaussian noise, shearing images with affine transform, decreasing, and increasing the brightness of images [44]. Dataset balancing is done through image augmentation on X-rays in normal, viral pneumonia, and corona virus X-ray image datasets. Augmentation on normal and pneumonia dataset is carried out once, whereas six times on COVID-19 dataset. Rotation by 5, 10 and 15 both in clockwise and anticlockwise directions is applied to normal, pneumonia, and COVID-19 positive images. Translation by -5% and $+5\%$ is applied to only normal and pneumonia images [45]. The work presented in [46] uses images augmentation techniques: flip vertical, horizontal, and vertical plus horizontal, rotation by $-90, 60, 90$ and 150° , translation by 10 and 50, brightness by 0.5 and 1.5, flip vertical, horizontal and vertical plus horizontal with change in brightness between 0.5 and 1.5. Experimental results showed improved model accuracy.

In the paper [47], the researchers augment the COVID-19 X-ray dataset on fly during model training with parameters such as rotation, shearing, and zooming by 20%. Image augmentation is applied to counter the COVID-19 positive dataset. During training, each image is randomly picked, and augmentation is done by performing horizontal flip, zooming at the center, rotation, and translation operations, and authors have reported that this has increased number of positive predictions for truly positive cases [48]. The work in [49] focuses on the preprocessing steps when COVID-19 X images from different databases need to combined into one large database, however, to enlarge the database, further image augmentation techniques are applied by adjusting contrast, saturation, flip left right, flip up down, flip up down left right, and rotations.

Generative adversarial network (GAN) [50] is combined with variational autoencoder to generate additional fine grain X-rays from COVID-19 database. The resulting image set was validated by doctors rating them between 0 and 5. Then based on the average from the rating by 5 doctors, X-rays with score lower than 4.5 are eliminated. Finally, image augmentation is applied on the resulting image dataset [51]. The investigation in [52] states that GAN is not suitable for generation of high-resolution images and, therefore, explored a variant of GAN known as Auxiliary classifier GAN(AC-GAN) for COVID-19 X-ray dataset augmentation through synthetic image generation. The AC-GAN takes as an input latent vector noise and class label as inputs and generates an X-ray image that falls in the category of COVID-19 positive case. The performance convolutional neural network which designed in this work for COVID-19 X-ray image classification is evaluated on original dataset and augmented dataset, and they observe there is 10% increase in the model accuracy on augmented dataset.

To generate a balanced CT scan image dataset for COVID-19 analysis, a technique is presented which combines conventional augmentation approach with synthetic CT image generation to mimic COVID-19 positive cases. In this work, synthetic chest CT images are generated using conditional generative adversarial network (CGAN). The final database included original COVID-19 positive CT images, augmented images,

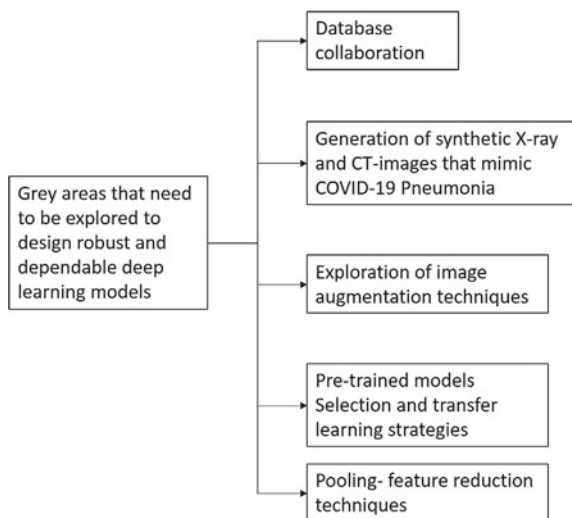
and images generated by CGAN. With these techniques, the original validation set for COVID-19 positive and non-COVID-19 is increased from 60 to 870 and 58 to 846, respectively. The conventional augmentation techniques include operations: rotate 45 degrees vertical flip, rotate by 45 degrees, vertical flip, and horizontal flip operations. The pre-trained models AlexNet, VGGNet16, VGGNet19, and ResNet50 are explored for COVID-19 diagnosis using CT scan images dataset resulting from conventional augmentation and CGAN. AlexNet, VGG-16, and ResNet50 have shown considerable improvement in classification accuracy, and VGGNet19 and GoogleNet have shown moderate improvement [53].

Investigations in paper [26] are aimed at exploiting the information extraction capability of wavelets, the CT scan COVID-19 and non-COVID-19 training set images are decomposed up to three levels by applying Haar and Daubechies wavelets, then augmentation techniques such as shear, translation, and rotation are applied on resulting images. The pre-trained models ResNet18, ResNet50, ResNet101, and SqueezeNet are using for COVID-19 diagnosis using these CT images with transfer learning. The test accuracy of 99.4% is achieved by ResNet18 model.

6 Research Directions and Challenges

The survey presented in this work on deep convolutional neural networks which are used for the diagnosis of corona virus lung infection provides an insight into various challenges faced by the researchers. This study also gives directions to carry out further research. The various research challenges are represented in Fig. 3, and they are

Fig. 3 Research directions in COVID-19 pneumonia diagnosis



Deploying COVID-19 positive case X-ray or CT scan image database by collaborating open dataset facilitated by medical organization, GitHub repository, and datasets from Kaggle competitions dataset.

There is an issue in publishing the dataset by health organizations due apprehension of negative impact on the society, and therefore, there is a need for generation of synthetic images and explore credibility of deep learning models in COVID-19 infection diagnosis as these models demand large amount of data. Therefore, there is need for exploration of techniques for generation of synthetic X-ray and CT-images.

The Keras library provides ImageDataGenerator which helps in image augmentation, and therefore, in a similar manner, there is a need to explore deep learning platforms that helps in image augmentation in consultation with radiologist.

The potential of deep learning models can be exploited by designing right architecture for a model so that model shows success in capturing salient information relevant to corona infection from X-ray and CT-images. Therefore, there is need to explore the strategy for fixing the architectural parameters of deep convolutional neural networks. Typical parameters include number of convolutional layers, number of filters in each layer, size of the filters, weight initialization strategies for filters, size of stride, and pooling strategy for dimension reduction. Therefore, there is need to carry out huge experimental research and conclude which parameters are the best.

The performance of convolutional neural networks also depends upon the strategy adopted for feature reduction in pooling layer, survey indicates that global average pooling preferred on contrary with max pooling. Also, some researchers have explored wavelets-based feature reduction techniques. Therefore, there is need to carry out detailed research to explore different types of wavelets to implement pooling layers.

7 Conclusion

In this paper, a critical survey and analysis on various deep learning models used for analysis of COVID-19 infection are presented. This survey gives an insight into the scope for exploiting the knowledge of benchmark pre-trained models for COVID-19 pneumonia infection diagnosis on contrary to building a deep learning model from the scratch, which demands high computational power and rigorous experimentation. Other contributions include recommendations on combining publicly available COVID-19 positive dataset into single large database to counter the problems posed by imbalanced dataset. The work also presents a thorough investigation on suitable augmentation techniques and other synthetic image generation techniques as applicable to COVID-19 X-ray and CT-scan images.

References




1. Orsi MA, Oliva AG, Cellina M (2020) Radiology department preparedness for CO-VID-19: facing an unexpected outbreak of the disease. *Radiology*, 201214
2. Wong HYF, Lam HYS, Fong AH et al (2019) Frequency and distribution of chest radio-graphic findings in COVID-19 positive patients. *Radiology*, 201160
3. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, Schluger NW, Volpi A, Yim JJ, Martin IBK, Anderson DJ, Kong C, Altes T, Bush A, Desai SR, Goldin J, Goo JM, Humbert M, Inoue Y, Kauczor HU, Luo F, Mazzone PJ, Prokop M, Remy-Jardin M, Richeldi L, Schaefer-Prokop CM, Tomiyama N, Wells AU, Leung AN (2020) The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society. *Radiology*
4. Kate MY (2020) X-ray may be missing COVID-19 cases found with CT. *Korean J Radiol*
5. Researchers find untapped potential for AI-enhanced chest X-ray imaging in COVID-19 diagnosis, *HospitMedica International staff writers*, Posted on 08 Jun 2020
6. Zhou S, Wang Y, Zhu T, Xia L (2020) CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China. *AJR* 214:1287–1294
7. Yi PH^{*†}, Kim TK^{*}, Lin CT (2020) Generalizability of deep learning tuberculosis classifier to COVID-19 chest radiographs. *J Thor Imaging* 35(4):W102–W104
8. Deng L, Yu D et al (2014) Deep learning: methods and applications. *Found Trends® Signal Process* 7(3–4):197–387
9. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. *J Big data* 3:9
10. Wang L, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. [arXiv:2003.09871](https://arxiv.org/abs/2003.09871)
11. Alom MZ et al (2020) COVID MTNet: COVID-19 detection with multi-task deep learning approaches. In: *arXiv preprint* [arXiv:2004.03747](https://arxiv.org/abs/2004.03747)
12. Shervin M, Rahele K, Milan S, Shakib Y, Ghazaleh JS (2020) Deep-covid: predicting covid-19 from chest x-ray images using deep transfer learning. *arXiv preprint* [arXiv:2004.09363](https://arxiv.org/abs/2004.09363)
13. Antonios M, Ioannis K, Konstantinos T. COVID-19 detection from chest X-ray images using deep learning and convolutional neural network
14. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 1
15. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of covid-19 cases using deep neural networks with x-ray images. *Comput Biol Med* 103792
16. Narin CK, Z Pamuk (2020) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. [arXiv:2003.10849](https://arxiv.org/abs/2003.10849)
17. Loey M, Smarandache F, Khalifa, NEM (2020) Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry* 12:651. <https://doi.org/10.3390/sym12040651>
18. Hall LO, Paul R, Goldgof DB, Goldgof GM (2020) Finding Covid-19 from chest X-rays using deep learning on a small dataset. *arXiv preprint* [arXiv:2004.02060](https://arxiv.org/abs/2004.02060)
19. COVID-CT. <https://github.com/UCSD-AI4H/COVID-CT>. Accessed 05 May 2020
20. Tuan DP, A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks
21. Wang S, Zha Y, Li W et al (2020) A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J*. In press. <https://doi.org/10.1183/13993003.00775-2020>
22. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Chen Y, Su J, Lang G, Li Y, Zhao H, Xu K, Ruan L, Wu W (2020) Deep learning system to screen coronavirus disease 2019 pneumonia, pp 1–29. *arXiv preprint* [arXiv:2002.09334](https://arxiv.org/abs/2002.09334)
23. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L (2020) Correlation of chest ct and rrt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology* 200642

24. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpan-skaya K et al (2017) Chexnet: radiologist-level pneumoniadetection on chest x-rays with deep learning. arXivpreprint [arXiv:1711.05225](https://arxiv.org/abs/1711.05225)
25. Rohit L, Ashrika G, Viraj K, Aniruddha P, Amit K (2020) Automated detection of COVID-19 from CT scans using convolutional Neural Network
26. Ahuja S, Panigrahi B, Dey N, Gandhi T, Rajinikanth V (2020) Deep transfer learning—based automated detection of COVID-19 from lung CT scan slices. <https://doi.org/10.36227/techrxiv.12334265.v1>
27. Fakhfakh M, Bouaziz B, Gargouri F, Chaari L (2020) Prognnet: Covid-19 prognosis using recurrent and convolutional neural networks. medRxiv
28. Arman H, Mahdiyar M, Ko S-B (2020) COVID-CXNet: detecting COVID-19 in frontal chest X-ray images using deep learning
29. Ozkaya U, Saban O, Mucahid B (2020) Coronavirus (COVID-19) classification using deep features fusion and ranking technique. arXiv preprint [arXiv:2004.03698](https://arxiv.org/abs/2004.03698)
30. Rahimzadeh M, Attar A, Sakhaei S (2020) A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset. <https://doi.org/10.13140/RG.2.2.10063.92320/2>
31. Tartaglione E, Barbano CA, Berzovini C, Calandri M, Grangetto M (2020) Unveiling COVID-19 from chest x-ray with deep learning: a hurdles race with small data. [arXiv:2004.05405](https://arxiv.org/abs/2004.05405)
32. Maguolo G, Nanni L (2020) A critic evaluation of methods for COVID-19 automatic detection from X-ray images. [arXiv:2004.12823](https://arxiv.org/abs/2004.12823)
33. Rahimzadeh M, Attar A, Sakhaei SM, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset. In-Press. <https://doi.org/10.1101/2020.06.08.20121541>
34. Abbas A, Abdelsamea MM, Gaber MM (2020) Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. arXiv preprint [arXiv:2003.13815](https://arxiv.org/abs/2003.13815)
35. Zturk S, Ozkaya U, Barstugan M (2020) Classification of coro-navirus images using shrunken features. medRxiv
36. Alom MZ et al (2020) COVIDMTNet: COVID-19 Detection with multi-task deep learning approaches. In: arXiv preprint [arXiv:2004.03747\(2020\)](https://arxiv.org/abs/2004.03747)
37. Cohen JP, Morrison P, Dao L (2020) COVID-19 imagedata collection. In: arXiv 2003.11597(2020). <https://github.com/ieee8023/covid-chestxray-dataset>
38. S. Tabika, A. Gómez-Ríosa, J.L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Areac, D. Charrea, E. Guiradod, J.L. Suárezza, J. Luengoa, M.A. Valero-González, P. García-Villanovab, E. Olmedo-Sánchez, F. Herreraa, COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on Chest X-Ray images. [arXiv:2006.01409v1](https://arxiv.org/abs/2006.01409v1)
39. Lokwani R, Gaikwad A, Kulkarni V, Pant A, Kharat A. Automated detection of COVID-19 from CT scans using convolutional neural networks. [arXiv:2006.13212](https://arxiv.org/abs/2006.13212)
40. Huang C-J, Chen Y-H, Ma Y, Kuo P-H (2020) Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. medRxiv
41. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. <https://doi.org/10.1080/07391102.2020.1788642>
42. Ahuja S, Panigrahi BK, Dey N, Rajinikanth V, Gandhi TK. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. [techrxiv.org](https://arxiv.org/abs/2006.13212)
43. Rajaraman S, Antani S (2020) Weakly labeled data augmentation for deep learning: a study on COVID-19 detection in chest X-rays. *Diagnostics (Basel)* 10(6): E358. Published 2020 May 30. <https://doi.org/10.3390/diagnostics10060358>
44. Ucar F, Korkmaz D (2020) COVIDiagnosis-Net: deep bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images [published online ahead of print, 2020 Apr 23]. *Med Hypotheses* 140:109761. <https://doi.org/10.1016/j.mehy.2020.109761>
45. Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al Emadi N, Reaz MB (2020) Can ai help in screening viral and covid-19 pneumonia?, arXiv. preprint [arXiv:2003.13145](https://arxiv.org/abs/2003.13145)

46. Hu R, Ruan G, Xiang S, Huang M, Liang Q, Li J (2020) Automated diagnosis of COVID-19 using deep learning and data augmentation on chest CT. <https://doi.org/10.1101/2020.04.24.20078998>
47. Tsiknakis N, Trivizakis E, Vassalou EE, Papadakis GZ, Spandidos DA, Tsatsakis A, Sánchez-García J, LópezGonzález R, Papanikolaou N, Karantanas AH, Marias K (2020), Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays. *Exp Therapeutic Med* 20:727–735. <https://doi.org/10.3892/etm.2020.8797>.
48. Kim O, McCourt M, Wang L, Parametrizing data augmentation in COVID-Net, advanced optimization techniques, convolutional neural networks, data augmentation, deep learning, healthcare, multimetric optimization. <https://sigopt.com/blog/parametrizing-data-augmentation-in-covid-net-development/>
49. Hansen C, Using deep learning to take on the COVID-19 virus, creating models to make good predictions on new, unseen data. <https://developer.ibm.com/technologies/artificial-intelligence/articles/using-deep-learning-to-take-on-covid-19/>
50. Bao J, Chen D, Wen F, Li H, Hua G (2017) CVAE-GAN: fine-grained image generation through asymmetric training. In: Proceedings of the IEEE international conference on computer vision, pp 2745–2754
51. Albahli S, Efficient GAN-based chest radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. <https://doi.org/10.7150/ijms.46684>
52. Waheed A, Goyal M, Gupta D, Khanna A, Al- F, Pinheiro PR (2020) CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. *IEEE Access* 8:91916–91923. <https://doi.org/10.1109/ACCESS.2020.2994762>
53. Loey M, Manogaran G, Khalifa NE (2020) A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. Preprints 2020040252

Predicting Diabetes Using Id3 Algorithm



Pratibha Dandin, D. B. Srinivas , H. Lakshmi , and K. M. Deepika 

1 Introduction

Data mining plays a decisive role for diagnosing and predicting disease in medical field. Various data mining algorithms are applied medical data which are accessible for comprehensive and deeper awareness of data providing evaluation of medical pivot in order to detect early stage disease and prevent disease. Nowadays, researchers used many data mining methods to detect various diseases, example diabetics, cancer, thyroid, etc. Today, because of lack of exercise and obesity diabetes is increasing rapidly. In human body, insulin is most essential hormone that regulates the sugar level, and if the insulin is not produce, a large amount of sugar will be produced and leads to all type of diabetic diseases. In this, synergic and other classification approaches are proposed for diabetes diagnosis. Ensemble classifiers are seized into account as definite in prediction and performance id disease.

Diabetes mellitus will be classified into two types: 1. Type one and 2. Type two. The cause for type one is pancreas cannot produce insulin in the body of person, and in type, two insulin will be released by pancreas, but the body will take very less amount. The problem allied to this is absence of initial detection of disease, understand it, that is maybe little slow for treatment.

P. Dandin · D. B. Srinivas (✉) · H. Lakshmi · K. M. Deepika
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: srinivas.db@nmit.ac.in

H. Lakshmi
e-mail: lakshmi.h@nmit.ac.in

K. M. Deepika
e-mail: deepika.km@nmit.ac.in

The reason for diabetes will be relies upon two sorts Type 1 and Type 2. Type 1 of diabetes is somewhat acquired, after which initiate by methods for positive contaminations, with a couple of proof pointing at Coxsackie B4 infection. It is inconsequential to way of life. Type 2 Kind 2 diabetes is expected the truth of style of abiding variables and history. The manner of thinking of way of life factors is essential to the improvement of Type 2 diabetes, alongside dried nonattendance of physical development like not exactly heavenly eating daily practice, stress, and urbanization.

In our proposed system, we apply the data mining algorithms such as decision trees, rule-based such as Id3 and ensembles classifiers for the diagnosis of diabetes mellitus. Results obtained from experiments show that id3 decision trees have better accuracy and efficiency.

The paper is organized as follows. Section 2 contains related work. Section 3 describes proposed architecture. Section 4 defines id3 algorithm. Section 5 describes experimental results, and Sect. 6 concludes the paper.

2 Related Work

A framework for the detection of diabetes is presented in [1]. Framework processes the patient data and predicts whether the affected person in query is vulnerable to diabetes mellitus or not. Article [2] centers around diabetes in pregnant ladies. Proposed architecture uses decision tree and Naïve Bayes model to predict diabetes in pregnant women. Authors in article [3] categorized decision tree and the decision policies. Decision tree algorithms are very well-known in that. For the reasonable class of the articles with the given properties, inductive methodologies utilize those calculations basically and are extremely essential inside the class of the items. Article [4] sums up different methods, characterization, and usage utilizing different sorts of programming instruments and strategies. Using artificial neural network, the diagnosis of diabetes can be done.

Authors in [5] find out regionally periodic diseases which include heart disease, lung, and breast cancer. They focus on finding regionally common patterns in terms of cost using clustering, classification, regression, association rule mining, and CART are broadly applied data mining techniques. In [6], authors have presented mining relationship in diabetes data for efficient classification. Characterization pace of 91% was acquired for C4.5 calculation. Future improvement uses C4.5 calculations to upgrade order rate to achieve higher precision in arrangement. Authors in [7] broke down numerous classification calculations on Diabetes dataset. Class foreseeing a specific result principally dependent on a given info. Calculation advancement a preparation set containing a lot of trait and the particular result, regularly called objective or reason or forecast trademark. Human's elderly greater than 65 years are majorly suffering from diabetes. In [8], authors utilized CART technique to anticipate exactness and an accuracy of 96.39 and 100.00% in distinguishing Diabetes.

Artificial neural networks, decision trees and logistic regression were applied in [9] for prediction of diabetes among sets of people beings. Fuzzy ontology applied to the diabetes knowledge and establishes relations based totally on the structure of fuzzy diabetes ontology. In [10], authors have presented decision support system for detecting of disease utilizing three diverse man-made reasoning classifier calculations to be specific multilayer perceptron and Naive Bayes classifier and J.48. In [11], authors utilized information digging procedure for medicinal services the executives to anticipate the affliction sickness and analysis of the diabetes. Paper [12] utilizes the probabilities of gaining a coronary ailment. Most machine learning techniques inclusive of Naïve Bayes and support vector machines are applied.

3 Proposed Architecture

Overall proposed architecture is as shown in Fig. 1. Proposed architecture contains different levels. Figures 2, 3, and 4 show level 0, level 1, and level 2, respectively. We

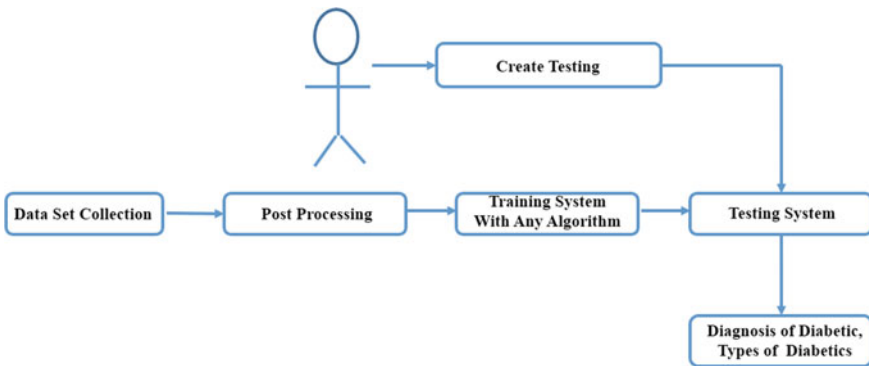


Fig. 1 Overall architecture

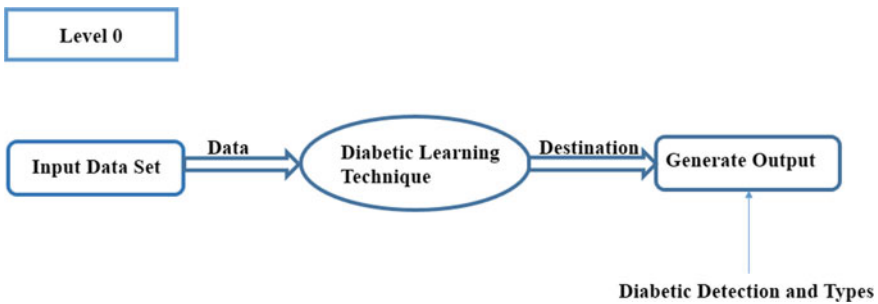


Fig. 2 Flow Diagram of process level 0

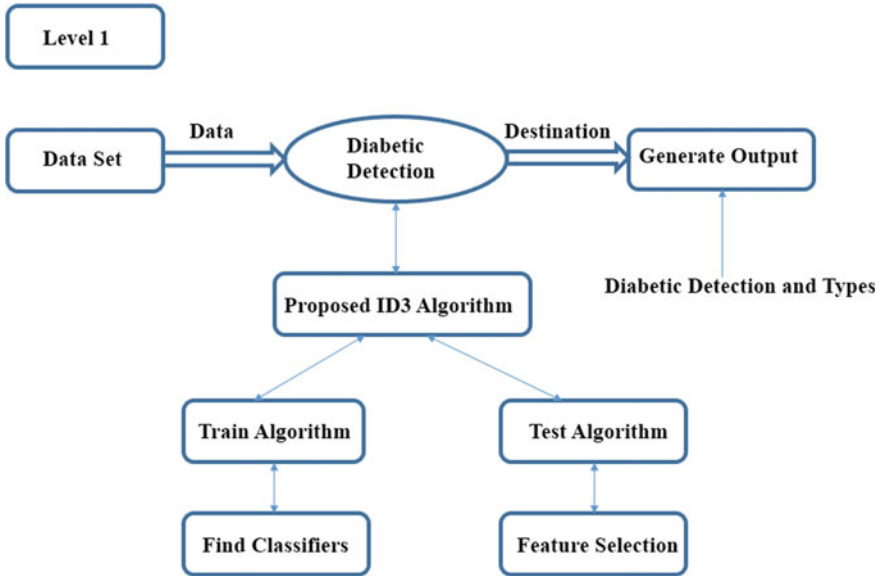


Fig. 3 Flow diagram of process Level 1

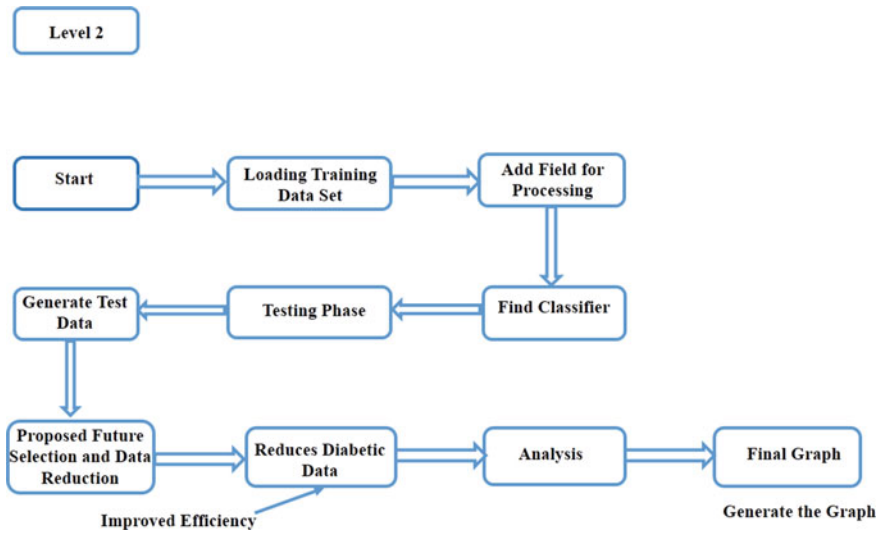


Fig. 4 Flow diagram of process level 2

input diabetic dataset to the proposed work and apply diabetic learning techniques to classify and generate the output. Here we have taken pima dataset and train our classifier using ANN/ SVM.

3.1 Rules of Classifying

Class is a supervised learning method in that it requires a classified training fact that generates rules to categories the test data. According to classification rules, test data is classified into classes. Id3 makes use of information gain and entropy for type. If entropy is 0, it is miles a homogeneous node. If entropy is 1, it is miles a heterogeneous node. Class 1 represents as tested positive, and Class 0 represents as tested negative.

3.2 Formulas Used

- Entropy is calculated by using formulae:

$$\text{entropy}(s) = (p(+)) * \log_2(p(+)) - (p(-) * \log_2(p(-))) \tag{1}$$

Gain is calculated by using formulae:

$$\text{Gain}(S, A) = \text{entropy}(s) - \sum_v \frac{|S_v|}{|S|} \tag{2}$$

where

$S \rightarrow$ Training data set.

$A \rightarrow$ Specific attribute.

$|S_v| \rightarrow$ Number of elements in S_v .

$|S| \rightarrow$ Number of elements in S .

$v \rightarrow$ All possible values of an attribute

$$p(+)/p(-) \rightarrow \text{proportion of + ve and - ve examples in } S$$

Entropy will be zero in case of whole homogeneity. If entropy value one, then it is divided equally. Decrease in entropy offers the information gain. We choose for splitting attribute because the characteristic having highest information gain. Then based on discovery of the maximum most homogeneous branch, a choice tree can be built.

4 ID3 Algorithm

Id3 algorithm utilizes a top down ravenous way to deal with assemble a decision tree. Using **top down** approach algorithm starts building a tree from a root node (top) and uses a **greedy** approach to select a best feature through iteration. Id3 calculation utilizes factual measure called data gain (how well a given property isolates the preparation ascribes as indicated by their objective characterization) to locate the best element. Feature with highest information gain is selected as best one.

id3(objects, T_Attribute, Attributes).

// Input: Objects are the training objects, T_Attribute attribute esteem anticipated by the tree and Attributes are the rundown of characteristics which might be tried by the educated choice tree.

// Output: classified decision tree.

1. If it is correctly classifying given objects, then returns a decision tree
2. First create a root node for the decision tree
3. If all objects are positive, then return a single-node root tree and label with +
4. If all the objects be negative, then come back a tree root with single-node with label with -
5. If attributes are null, then return single-node tree root with =
6. Otherwise Start
7. Assign best characteristic of example $\rightarrow A$
8. Decision attribute for Root $\rightarrow A$
9. For each value v_i of A
10. Insert a new tree branch below tree root and corresponding to the test $A = v_i$
11. Let $items(v_i)$, be the subset of object have the value v_i for A
12. If $objects(v_i)$ is empty
13. Update a leaf node with label = most common target value in the objects below this new branch
14. Update the subtree $id3(objects(v_i), T_Attribute, Attributes - \{A\})$ below this new branch
15. End
16. Return Root

5 Results

Result analysis of id3 and existing approach (SVM) is measured with respect to efficiency and accuracy. Tabulated results in Tables 1 and 2 show efficiency and accuracy of an id3 and SVM algorithms, respectively. Similarly, Figs. 5 and 6 show the graphical the representation of efficiency and accuracy of ID3 algorithm and SVM algorithms.

Figures 7 and 8 show the efficiency and accuracy of an id3 algorithm, respectively. So efficiency and accuracy of an id3 algorithm is far better than better than existing system.

Table 1 Efficiency and accuracy of ID3 algorithm in detection of diabetes

Efficiency	Accuracy
0.5	0.54
0.6413	0.67
0.6847	0.71
0.7934	0.81
0.8913	0.9
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1

Table 2 Efficiency and accuracy of existing system in detection of diabetes

Efficiency	Accuracy
0.5	0.5133
0.5342	0.5466
0.5342	0.5466
0.5342	0.5466
0.5513	0.5633
0.5513	0.5633
0.5513	0.5633
0.589	0.6
0.589	0.6
0.6232	0.6333
0.6232	0.6333
0.6232	0.6333
0.6232	0.6333
0.6575	0.6666
0.6575	0.666

6 Conclusion

By using ID3 algorithm, we can predict diabetes present in a patient or not. Our proposed algorithm makes use of information gain to predict. The attribute having

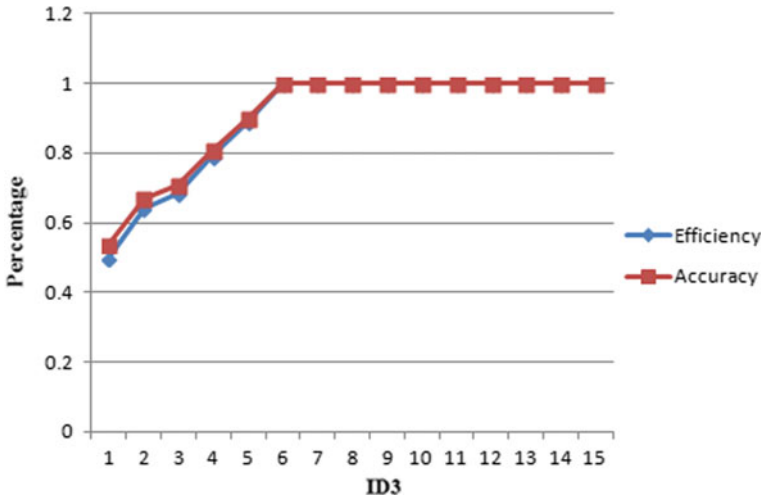


Fig. 5 Efficiency and accuracy of Id3 in detection of diabetes

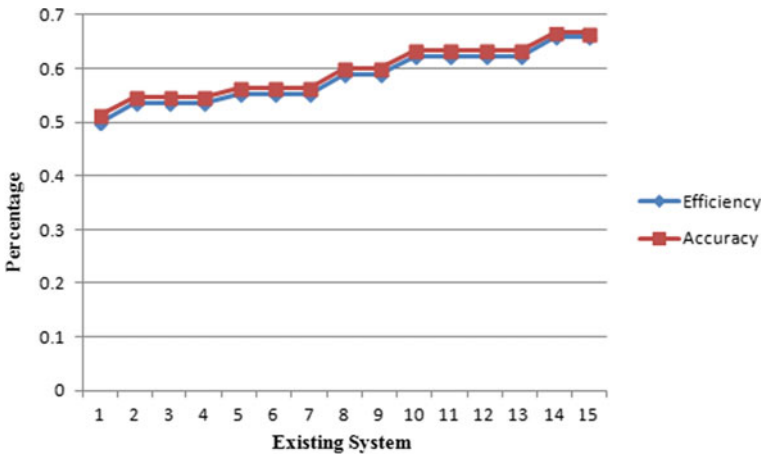


Fig. 6 Efficiency and accuracy of existing system in detection of diabetic

the highest information gain is selected for classification. The proposed system takes the less time to compute the detection of diabetics and its types and more efficient compare to existing system. The id3 algorithm detects the diabetics using retrieval of data with respect to relevant fields. Thus, outcome of id3 algorithm more accurate as compare to existing classification techniques.

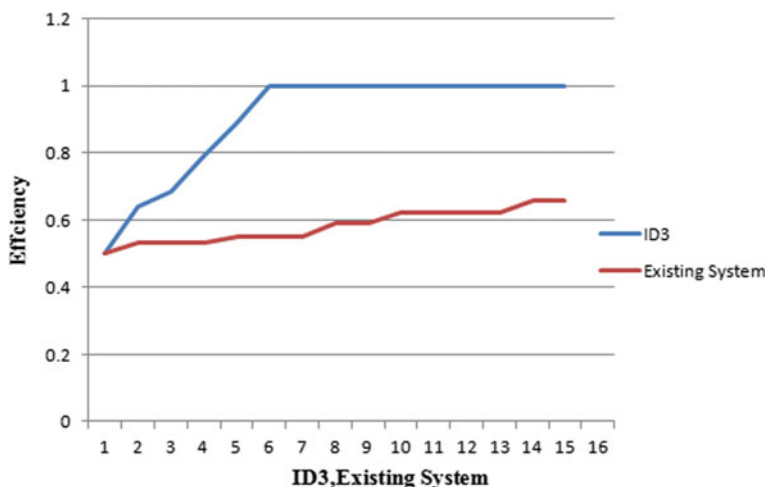


Fig. 7 Comparison of efficiency of id3 algorithm and existing system in detection of diabetic

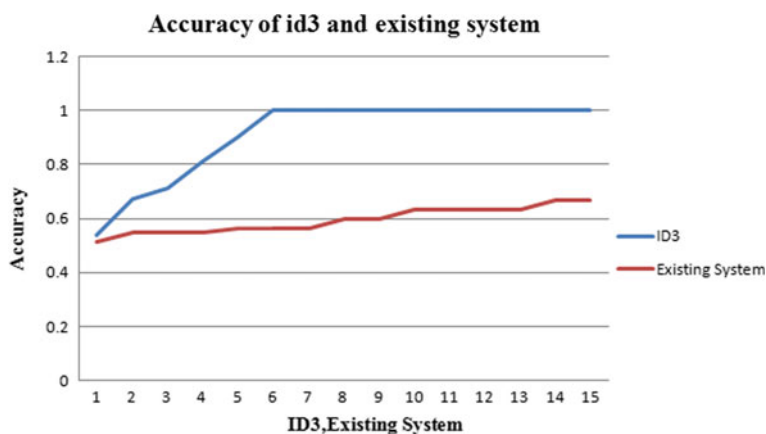


Fig. 8 Comparison of accuracy of id3 algorithm and existing system in detection of diabetic

References

1. Adekunle YA (2015) The prediction, diagnosis and treatment of diabetes mellitus using an intelligent decision support system framework. *Int J Adv Res Comput Sci Softw Eng* 5(3)
2. Iyer A, Jeyalatha S, Sumbaly R (2015) Diagnosis of diabetes using classification mining techniques. *Int J Data Mining Knowl Manage Process (IJDMP)* 5(1)
3. Rupali B, Sonia V (2013) Implementation of ID3 algorithm. *Int J Adv Res Comput Sci Softw Eng* 3(6)
4. Agrawal P, Dewangan AK (2015) A brief survey on the techniques used for the diagnosis of diabetes-mellitus. *Int Res J Eng Technol (IRJET)* 02(03)

5. Khaleel MA, Pradham SK, Dash GN (2013) A survey of data mining techniques on medical data for finding locally frequent diseases. *Int J Adv Res Comput Sci Softw Eng* 3(8)
6. K. Rajesh, V. Sangeetha “Application of Data Mining Methods and Techniques for Diabetes Diagnosis” *International Journal of Engineering and Innovative Technology (IJEIT)* Volume 2, Issue 3, September 2012.
7. L. Cynthiya Juliyet1, Mr.K.Mohamed Amanullah2 “The Surveillance on Diabetes Diagnosis Using Data Mining Techniques” *IJSART* - volume 1 Issue 4 –APRIL 2015.
8. Kavitha K, Sarojamma RM (2012) Monitoring of diabetes with data mining via CART method. *Int J Emerg Technol Adv Eng* 2(11). ISSN 2250-2459
9. Thirumal PC, Nagarajan N (2015) Utilization of data mining techniques for diagnosis of diabetes mellitus—a case study. *ARPN J Eng Appl Sci* 10(1)
10. Murat K, Yavuz U (2013) Analysis of a population of diabetic patients databases with classifiers. *Int J Med Health Biomed Pharm Eng* 7(8)
11. Singh S, Kaur K (2013) A review on diagnosis of diabetes in data mining. *Int J Sci Res*
12. Parthiban G, Srivatsa SK (2012) Applying machine learning methods in diagnosing heart disease for diabetic patients. *Int J Appl Inf Syst* 3(7). ISSN: 2249–0868. (Foundation of Computer Science FCS, New York, USA)

Expert System: A Fault Diagnosis Expert System for High-Power Industrial Production Platform



S. Jagdeesh Patil

1 Introduction

Developing and deploying expert systems has undergone a sea change right from its inception. Such an expert system can store the knowledge of several experts in the field and can greatly be useful for solving breakdown problems immediately. Keeping in mind, an attempt has been made to conceive, design, and develop an expert system for high-power digital broadcast transmitter system. Major aspects of the expert system are presented in this paper.

2 System Architecture

In designing the expert system for high-power broadcast digital transmitter, the following guidelines were kept in view:

- Faults in transmitter are best solved by applying expert knowledge and by using interactive methods
- The more narrowly and specifically the fault is defined, and the more powerfully and accurately can the knowledge base be applied to solve problems
- All decision-making elements of an expert system must be incorporated to speed up fault diagnosis.

The basic components of any general expert system are shown in Fig. 1, and following this model, a block schematic of broadcast digital transmitter is shown in Fig. 2. As seen in here, the expert system consists of the following functional building

S. Jagdeesh Patil (✉)

Nitte Meenakshi Institute of Technology, Bengaluru NMIT, Bangalore, India

e-mail: jagadish.patil@nmit.ac.in

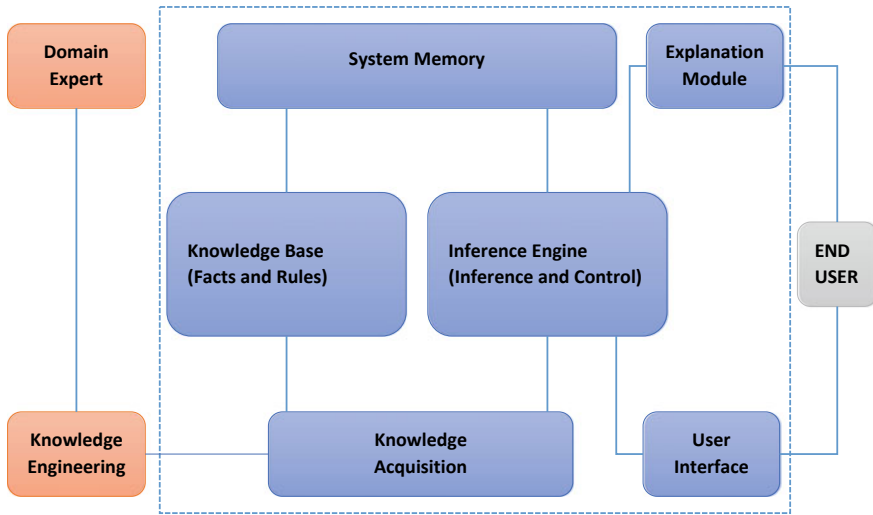


Fig. 1 Components of an expert system

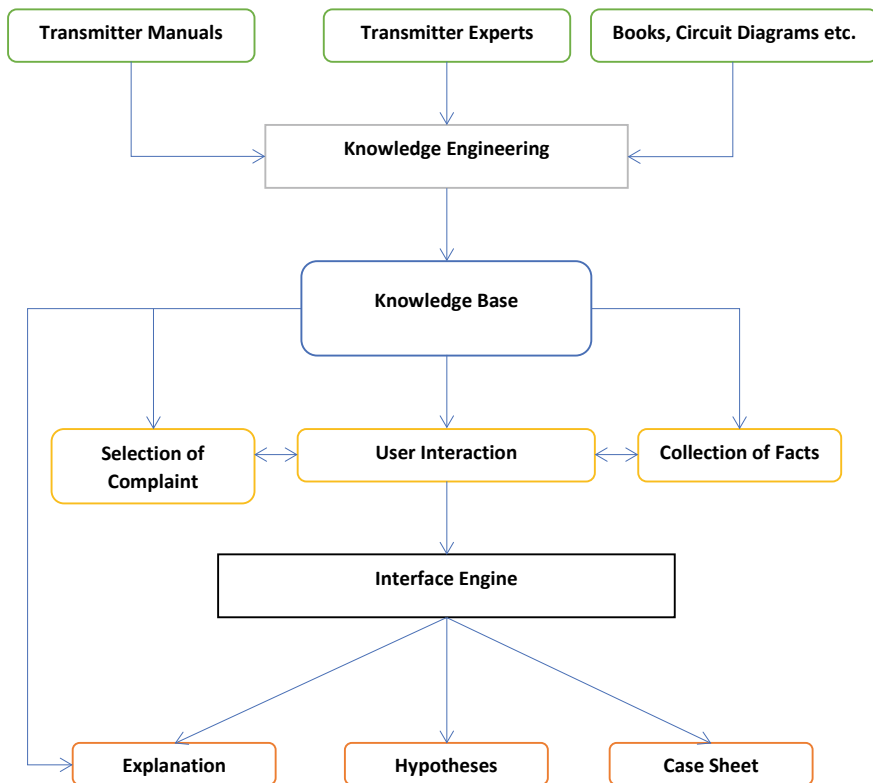


Fig. 2 Block schematic of broadcast transmitter expert system

blocks: (a) knowledge base module, (b) inference engine module, (c) explanation and case sheet module, and (d) user interface module. A brief description of these modules is given below.

2.1 Knowledge Base Module

This module stores the information gathered from transmitter manuals, transmitter experts, and other sources in the form of production rules. These parameter values, logical assertions, and text are centred around this. The process of organizing information in the knowledge base is called here as knowledge engineering. A production rule has two component parts: the left-hand side (LHS) is known as the antecedent, premise, condition, or situation, and right-hand side (RHS) is known as the consequent, conclusion, action, or response. In simple form, the LHS is known as IF part and RHS as the THEN part of the rule. The knowledge base the transmitter mainly comprises the following items:

- (i) **Complaints:** The component of a complaint are its name and explanation. Each complaint has asset of hypotheses
- (ii) **Hypotheses:** The constituents of the hypothesis are its name, a set of actions as well as explanations
- (iii) **Parameters:** All parameters related to a hypothesis are collected, and the values of these parameters are passed through the input/output interface by the use of menus. The detailed descriptions also include prompt messages to collect the data, options for the values, ranges of the values, and help messages to assist in answering queries for the data value, and explanations for collecting data.

2.2 Inference Engine Module

The inference engine of the transmitter expert system accepts user input queries and responses to the questions through the input/output interface and uses this dynamic information together with static knowledge stored in the knowledge base. The inference process is carried out in three stages—(i) match, (ii) select, and (iii) execute. When a rule is executed, it places new facts in the working memory. Now it may request for additional information from the user, or simply stop the searching process and result in a conclusion to be found and reported to the user. Both exhaustive and depth-first search techniques are incorporated in the software.

2.3 Explanation and Case Sheet Module

The explanation module provides a user with the explanation of the reasoning process when requested. It gets its input from the inference engine. The information provided by this module allows a user to determine if the expert system's reasoning process is sound or not. With the help of the explanation's mechanism, the different inferencing steps can be followed at any time during a consultation process. The case sheet function of this module provides a list of hypotheses checked and their confidence factors.

2.4 Input/Output Interface

This provides the user a means of communicating with the system in a more natural way through a simple selection menu. This also permits the user to select the parameter values in three different modes: -

- i. **Batch Mode:** In this mode, all parameters for the selected problem are collected initially, and then, the inference engine is invoked
- ii. **Interactive mode:** Here, the inference engine is active throughout the consultation process. As soon as a new parameter is available to it through the input/output interface, the inference at this stage is displayed and the user can stop the consultation process at that stage or continue further.
- iii. **Batch interactive mode:** In this mode, the user has the options to start the consultation session starting from any parameter, which is often the case in the transmitter fault diagnosis process.

The parameter selected by the user may appear in more than one hypothesis. The inference engine then collects the other relevant parameters of the hypothesis before reaching a conclusion. A block schematic of a typical system fault diagnosis sequence is shown in Fig. 3. As seen here, for every hypothesis, the expert system goes through a sequence of inference steps before arriving at conclusions and providing explanations. Also, if a hypothesis is not confirmed, the knowledge base is updated accordingly for use in fault diagnosis.

3 Software Implementation

The software package is using C++ programming language. The choice of C++ made considering its many advantages as compared to LISP and PROLOG languages, even though these languages were used extensively while implementing expert systems, for their ability to process language texts. On the other hand, expert systems written in C++ are faster and can be easily interfaced with other APIs. Also, object-oriented

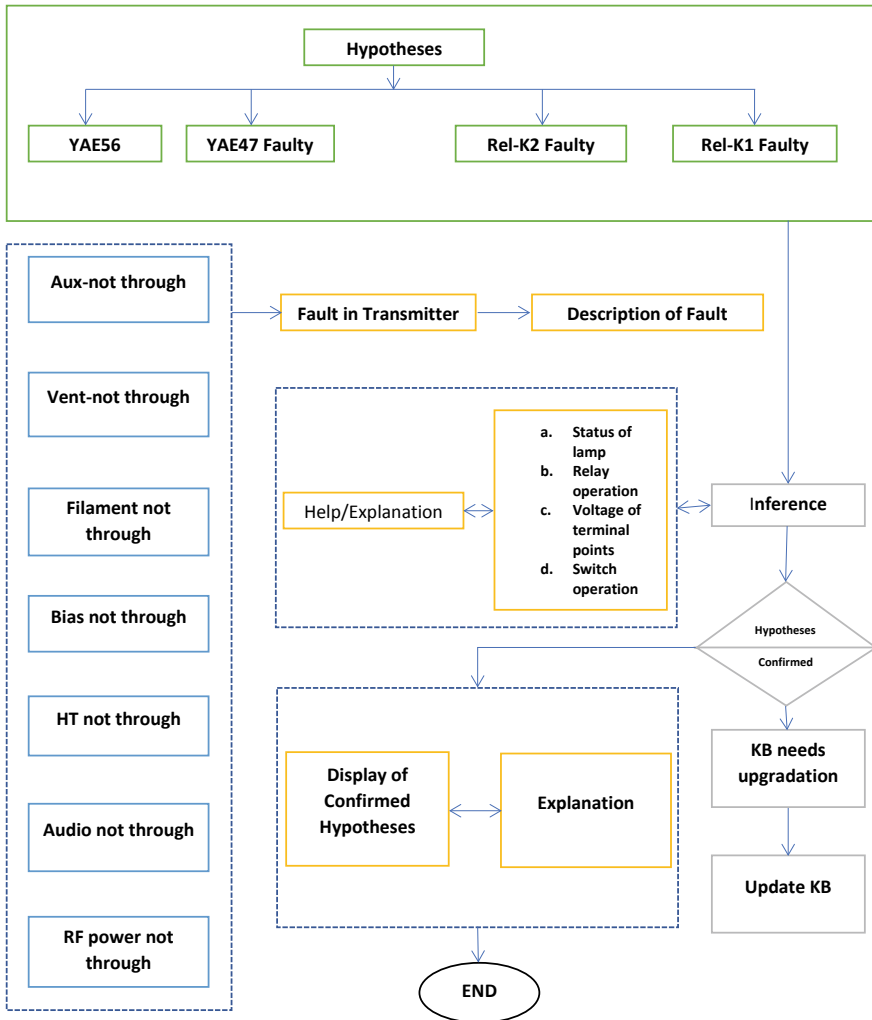


Fig. 3 Typical broadcast transmitter fault diagnosis system

programming, low-level programming, and ease of upgradation are other advantages in C++ in this application.

The software has been developed using the modular approach with individual program modules for knowledge base, inference engine, and user interface. The major programs in the package are as follows:

1. Aux-ON class /function definition program
2. Vent-ON class /function definition program
3. HT-ON class /function definition program
4. Aux-ON rule / parameter function program

5. Vent-ON rule /parameter function program
6. HT-ON rule / parameter function program
7. Knowledge base program
8. Inference engine program
9. User Interface program
10. Main program.

In addition to above programs, several test programs have been provided to check the different functions of the expert system. Also, security program is incorporated to prevent unauthorised access to the package.

4 Field Trials

The major modules of the transmitter were subjected to field trials at super power transmitter service at Bangalore. The expert system is loaded on location center computer and was operated by an engineer on duty at the transmitting station. Faults were created in the transmitter without the knowledge of engineer on duty and who was asked to clear the faults with the help of installed expert system. The engineer was able to diagnose all the faults successfully, and the inferences drawn and explanation given by the expert system were up to the satisfaction of the engineer. Some preliminary results / observations from the field trails of the expert system are given below:

1. Comparison of time taken by the engineers in clearing faults with and without the use of expert system provides a means of evaluating its performance, Usually, the engineers have to go through entire circuit diagrams, and after applying the parameter values, derive the conclusions. Normally, the time taken for this process is from few hours to days. In contrast, the expert system takes few seconds or less than minute for parameter value entry and provides valuable diagnostic information in much shorter time.
2. Sometimes, it may be found that the faults confirmed by expert system are not exactly same as in actuality. In such cases, expert system takes a user very near to the actual fault. However, in the background, the information about the faults are used to update/improve the knowledge base of expert system. In future, when the same faults are handled by the expert system, they are confirmed exactly. Thus, refinement and enlargement of expert system is a continuous process.

The field trials have shown that many complicated faults in the transmitter can be diagnosed and cleared within a matter of less than minute with the user-friendly environment of expert system.

5 Conclusion

The rapid evolution of industry has made problems related to maintenance, and particularly fault diagnosis, of very complex system such as broadcast transmitters and satellite launching stations difficult to solve. The use of expert system presented in this paper can greatly simplify fault diagnosis in such systems. The major features of our expert system are summarized below:

- Our expert system is a modular software package with many user-friendly features
- Its knowledge base can be modified or augmented without affecting the structure of the entire program
- It can be easily adopted to different types of transmitters or any other type of engineering systems
- It can be used in training technical staff for acquiring the knowledge of broadcast transmitters and for generating useful database of transmitter faults.

The results of field trails of our expert system in a transmitting station have clearly shown its practical utility in broadcasting stations.

References

1. Jagdeesh Patil S, 'BTEX: a fault dianosis expert system for broadcast transmitter. Indian Institute of Science Bangalore
2. Edmunds RA (1998) Guide to Expert Systems. Prentice Hall
3. 3 MW – (500KW x 6) Brown boveri transmitter manuals, vols 1 to 5 - super power transmitting station, Bangalore

Real-Time Hand Gesture Detection Using Convex Hull and Contour Edge Detection



Smriti Amatya, Ishika, M. V. Manoj Kumar, B. S. Prashanth, H. R. Sneha, Likewin Thomas, and Vishnu Yarlagadda

1 Introduction

Computer vision is an exciting domain of research, with new advancements made by researchers rigorously. It is an interdisciplinary scientific field which deals with how computers can gain useful insights and inferences from digital images or video streams. From the technical perspective, the domain aims at emulating and automating tasks like that of the human visual system.

Recognizing hand gestures has been a vital topic of interest and research in computer vision. The very basic research in computer vision primarily concentrated on emulating the activity of our human visual system using computers. The primary goal of our work is to develop a virtual assistant for the deaf and dumb or physically handicapped people that enables them to use digital equipment at ease for their daily activities.

Modeling the hand gestures and tying them to meaningful messages is quite a tedious process. Hand gestures are an old way of communication that involves the use of one's fingers and hand position where each position has a different meaning and it conveys a message [1]. This method has been a common mode of communication

S. Amatya (✉) · Ishika · M. V. M. Kumar · B. S. Prashanth · H. R. Sneha
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore, India

H. R. Sneha
e-mail: sneha.hr@nmit.ac.in

L. Thomas
Department of Computer Science and Engineering, PES Institute of Technology and
Management, Shivamogga, India

V. Yarlagadda
School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT),
Vellore, India

information exchange among dumb and deaf people in the real world. Efforts were made to automate the above task using computers and many works related to it have been proposed and implemented. Some of the work involves wearable gloves and mechanical equipment of similar kind, but the technology has proved to be costlier to be brought into the market for consumer usage. Designing a Gesture detection model that is efficient and cost-effective is a daunting task. The following are the issues that affect the design decisions,

1. Choosing a suitable machine learning method for real-time processing of gestures which is efficient and gives a good prediction
2. Defining useful messages which can be tagged for gestures which get detected by the model
3. The hardware interface and sensors associated should be minimal even if it is considered
4. Cost of designing the model and NRE cost should be minimal.

By considering all the factors and issues, this work proposes a cost-efficient gesture detection product that can be brought into the market with ease and can be deployed in real-time production. The intended end-users of this product are physically challenged, deaf and dumb people for whom the proposed work on implementation becomes quite useful. The proposed work uses Convolutional Neural Networks for detecting hand gestures and attaching meaningful messages to them [2].

The upcoming sections of this paper are organized as follows. Section 2 presents related work. Framework followed in the paper is briefed in Sect. 3. Section 4 highlights the mathematical background and experimental setup used for gesture detection. The brief algorithm for the method of gesture detection is described in Sect. 5. Results of the experimental study are presented in Sect. 6. The open research issues are detailed in Sect. 7. This paper concludes with a concise conclusion in Sect. 8.

2 Related Work

Efforts were made to bring in machine learning and Artificial Intelligence into the setup so that to be able to recognize the hand gestures effectively. Some of the extensively used algorithms for gesture detection are *K-Nearest Neighbor (KNN)* and *Convolutional Neural Networks (CNN)*. Though the resultant models were proven effective and yield good results, the models require a huge number of samples for training the model, which is difficult to ensemble quickly.

“Various computer vision-based man–machine interface research were developed by using cameras of the single-lens [3], multi-lens [4], depth perception lens [5], or infra-red lens [6].” We have explored many models that were already implemented for the same and gained insights on the advantages and disadvantages of each method. This article discusses a novel approach for hand gesture recognition based on OpenCV library. The proposed work helps us to improvise the interface between users and computers, in turn, contributes to human–computer interaction

(HCI) technology progression [2]. Some researchers proposed using the electrical activity of muscle tissue and its representation as a visual display and signal [7]. Efforts are made to recognize the hand gestures using a monochrome glove which works even if the background light is darker [8]. Further, the group of Haar-like patterns was trained to go up, down gestures. Static hand gestures were extracted by the face-based adaptive skin model [9]. There are some attempts to detect hand gestures based on contact-based and vision-based methods using deep learning [10]. Combinational feature method for the region of hand gesture recognition was found to be promising as it achieves 97.4% accuracy [11]. AEPI method for the hand gesture for varying background and by blurring the image. AEPI stands for accurate endpoint identification which focuses on mobile computing and image processing by using morphological computation to produce the output [12].

3 Framework

The framework of the proposed method for hand gesture detection is given in Fig. 1. Initially, the accumulation of input images from the real-time camera with good pixel quality (0.9 Mega Pixels and above) is done. The image captured is processed to cleanse the noise and background blurring, it helps to concentrate more on the *Region of Interest* (ROI) in the image.

Image is used to obtain *Convex Hull*, which is used to segment the hand portion in the image. Along with these images, we select skin color to get the characteristics of the hand. The next step is to get the estimated hand state to extract the different hand features for defining a deterministic process of finger recognition. After the hand portion is segmented from the background, the counter vector is used to compute a series of coordinates on the edges of the hand.

Further processing of the counter vector gives the location of the fingertip. The convex hull implementation starts by calculating the points with minimum and maximum x and y-coordinates and joins all those points. It helps us to define a border rectangle, within which the hull is contained.

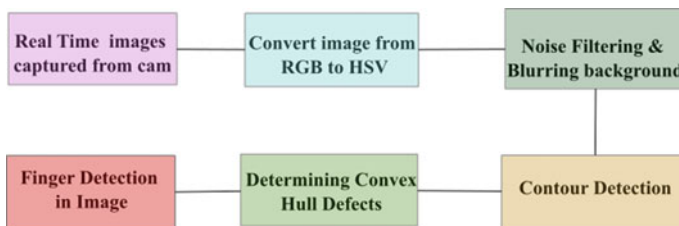


Fig. 1 Framework for hand gesture detection

The defect points are most likely to be the center of all the gap between fingers and then we gave each point a text section to make the machine understand what it means and using gestures the machine works.

To perform preprocessing, gesture capturing, feature extraction and gesture recognition, OpenCV library is used, which provides rich support with pre-defined libraries for real-time image processing [13]. OpenCV is widely used in HCI, robotics, biometrics, etc., and is written using C++, but provides interfaces for python and other languages as well.

4 Algorithm

The algorithm for gesture detection is briefed in this section, as to the input for the algorithm is a video of hand gestures, and the output is recognized hand gestures with appropriate actions. The detailed steps are given in algorithm 1.

Algorithm 1: Detailed steps for gesture recognition

Input:	Activate the Webcam and initiate the video capture process to capture the hand gestures put forth against the camera	
Output:	Set different message according to the unique area ratio and number of defects	
Start		
	Step 1	Define the ROI (Region of Interest) with dimension range of 300*100
	Step 2	Convert the image from its RGB color format to HSV color format
	Step 3	Apply Making for Skin color range with 0(black) and 1(white) as value
	Step 4	Reduce the noise by blurring and dilating the captured image
	Step 5	Find the maximum contours in ROI which represents the hand portion in the image
	Step 6	Apply Convex hull for the image and compute the area of the convex hull, contour and area ratio
	Step 7	Count the number of defects which are less than 90 degrees and highlight it with a blue circle
	Step 8	Compute the number of fingers in the image which is one more than the number of defects
	Step 9	Output the gesture if detected
End		

5 Experimental Setup and Mathematical Background

The rest of this section gives an elaborated view of how the gestures are identified from the preprocessed images captured via the camera. The input is taken in the form of a video from the front camera, and the area of the region is defined by setting a rectangular frame. Part of the hand which comes in the frame region will get recognized.

The subsystem to produce the output has to finish the three important steps, firstly the hand region needs to be detected, then, the features need to be extracted/constructed from the detected hand region, and finally, the features are processed to interpret the hand gesture. The detailed steps are presented in the following subsections.

5.1 Hand Region Detection

The input which comes in the defined frame is used for gesture detection. Firstly, all the RGB color will get converted into HSV color and skin color range is determined. This operation will make the system acknowledge the skin color feature in white color. The rest of the area is acknowledged as black. To make the provided hand in the frame, a more recognizable mask will be used to neglect the noise which comes in the frame. The resulting image after removing noise is used for calculating the contour.

For identifying the hand which comes under the area of interest, the RGB color space will be converted to the HSV color range. HSV will be defined in two ranges, i.e., lower skin range (0, 20, 70) and upper skin range (20, 255, 255). The HSV color range is as shown in Fig. 2.

Fig. 2 HSV color range

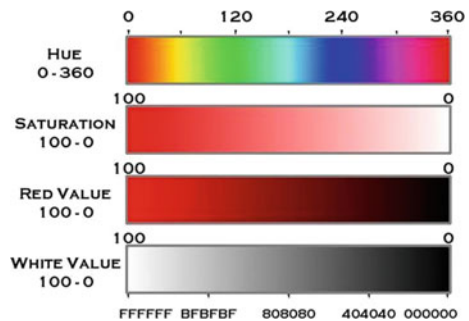




Fig. 3 Processed image after applying contour detection and identifying convex hull

5.2 Feature Extraction

The contour detection method is used to extract the features. It is the most used method for edge detection in images. Contour is basically an outline around the hand, and maximum contour value will be calculated for it.

Contour algorithm gives better results for background separation and to specify the region boundary, but it cannot detect the object of skin color with similar color background [14]. After the boundary has been identified, a convex hull is created around it, where a convex hull is a compact convex bound around the shape. The sample outcome of contour detection method is shown in Fig. 3.

5.3 Gesture Recognition

Using contour detection, the region of interest is identified. Region of interest identification is predominantly decided by the shape of the convex hull. The detailed steps for identifying the region of interest are detailed below.

The area ratio of ROI is calculated using Eq. 1.

$$\text{Area Ratio}(AR) = \left\{ \frac{\text{area}_{\text{underhull}} - \text{area}_{\text{underhand}}}{\text{area}_{\text{underhand}}} \right\} * 100 \quad (1)$$

where

$\text{area}_{\text{underhull}} \rightarrow$ area of the Convex Hull

$\text{area}_{\text{underhand}} \rightarrow$ area under the hand portion

Every hand gesture shown above will have its own area ratio (AR).

Now, we will go back to calculate the defects in the convex hull. The defects are the space which has not been occupied by the hand portion in the image. The number of defects can be counted by applying the cosine rule. For the defect regions in the image with a focus on ROI, compute the interior angle by using the formula,

$$\text{Angle} = \arccos\left(\left\{\frac{b^2 + c^2 + a^2}{2bc}\right\}\right) = 57$$

If the *Angle* computed above is less than *90-degree* increment the defect count and make a blue circle over the defects.

Number of fingers \rightarrow Number of defects + 1

Number of defects \rightarrow 1

For the different number of fingers, a different message will be tagged and displayed.

6 Results and Discussion

The proposed work uses Python language using Anaconda IDE with OpenCV 3.3.0 library. The system which has been used throughout the process has an arrangement of 8 GB Primary Memory, Intel i5-7200U @2.5 GHz. The OpenCV library comes under the Python library and is mainly focused on computer vision platforms. OpenCV modules used in the experiment are contour and Haar-like feature extraction, Convex hull.

An attached Web Camera of low resolution consisting of 0.9MP is used for seizing the shots of hand in a real-time environment. The captured images were used for further processing that detects the skin color, reduces the noise. The detection of gesture is performed by using methods like convex hull algorithm, and the count of fingers is evaluated by availing convex defects. 5 Different signs were identified based on the number of fingers. All gestures were recognized through the live video. An identification of the sign was successfully performed within 0.1 to 0.4 s. Throughout the research, various problems were identified which are the noise and the lighting that affects the detection of hand gesture.

The clarity of the captured image was improved to reduce the noise of the background and skin color. To make the system identify the hand signs, it is always necessary to have a person's hand gestures to be within the area of interest in the proper environment for further processing and giving out the results.

The results of the experimental study are shown in Fig. 4a–d. Where depending on different hand gestures shows the successful recognition of the one-finger hand gesture as the assigned tagged word “Alexa.” Figure 4b shows the successful recognition of the two-finger gesture along with its tagged word “what.” Figure 4c illustrates the three-finger gesture which is tagged to the keyword is. Four finger gesture

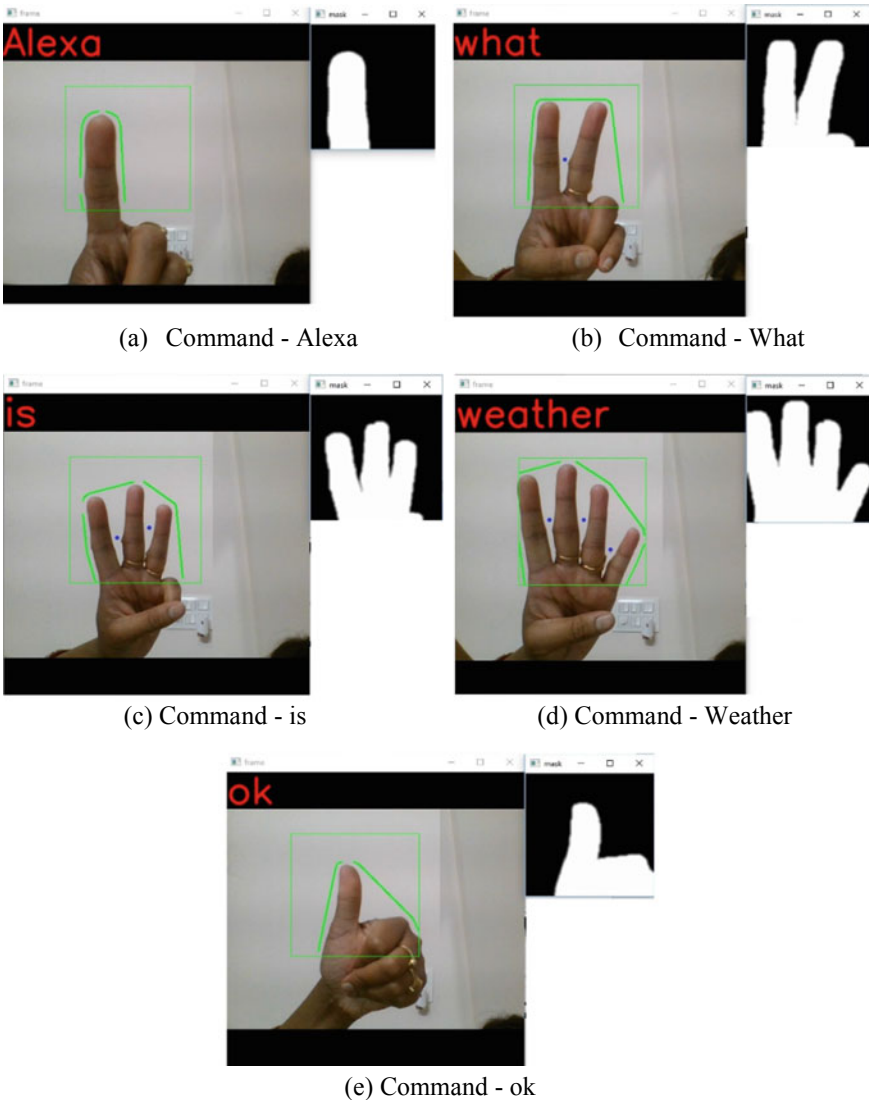


Fig. 4 a Hand Gestures being identified with the Tagged commands “Alexa” b What? c is for two fingers d weather for four fingers e ok for thumb finger (with the defects labeled as a blue circle in the image)

Table 1 Confusion matrix for evaluating the accuracy of the proposed system

		Predicted	
		Correct	Wrong
Actual	Correct	600	150
	Wrong	50	200

shown in Fig. 4d is tagged to command “weather.” The last gesture 4d illustrates the command “ok.” Each of these commands tagged to different hand gestures is recognized. Further, these commands can be linked/interfaced to control personal assistant devices such as Amazon Alexa, Apple Siri, or Google’s Google Assistant.

To evaluate the accuracy of the system, the repeated trials are carried on the proposed system, the confusion matrix for the repeated trials is given in Table 1. It resulted in overall 80% accuracy in correctly identifying the hand gestures.

7 Future Research Directions and Open Research Problems

Gesture recognition has been the base technology in touchless sensing market. Touchless sensing technology uses body gestures, postures and movements to control the activity of the computer without the aid of keyboard, mouse and screen. Gesture recognition has also been a key player in the domain of automotive infotainment centers, video game consoles, smart homes and automated automobile safety systems. The work can be extended to build an education system using which the dumb and deaf people who are not aware of the sign language can learn about it quickly.

Some of the open research issues that need to be addressed in the domain of gesture recognition are,

1. **Noise in input:** Noises are natural in the processing of images and are a challenging topic to research on.
2. **Using low-resolution video/image source:** Gesture recognition heavily relies on the resolution of the camera used, coming up with a system which can even work for low-resolution cameras would be another horizon to explore further.
3. **A generalized gesture set is yet to be defined** which is tied to generally accepted messages would be another work which needs to address
4. **Considering the other detail in the captured image:** The bounding box we defined for the gesture detection would limit the interpretation of gestures since

the background is subtracted during the processing step. Additional information can be extracted by considering the other objects in the background.

5. **Using CNN/ DL/ Other architectures:** Coupling the gesture recognition paradigm with the standard frameworks to come up with a standard CNN/DL/NN trained model specifically for the hand gesture detection and recognition is another research problem that needs more attention.
6. **Gestures can be implemented in smart homes** or any public places where these physically handicapped people do not have to feel insecure about themselves.

8 Conclusion

The research work detailed in this paper demonstrated a real-time Gesture recognition technique. It uses a convex hull to approximate the area of interest, and a convex hull combined with contour is used to detect the exact gesture. The paper illustrated the recognition of 5 different hand gestures. Each of these gestures is linked to different commands such as—is, weather, ok, what, Alexa, etc. The proposed method resulted in overall accuracy of 80% in correctly identifying the gestures with their tags.

The real-time mapping of these gestures with the commands can be changed depending on the application requirement. The proposed method of gesture detection is extremely useful in controlling the personal assistants by physically challenged individuals. It has a vast variety of applications. This paper also proposed a list of open research issues to address if anyone is interested in to further explore the topic.

References

1. Bhatt R, Fernandes N, Dhage A (2013) Vision based hand gesture recognition for human computer interaction. *Int J Eng Sci Innov Technol (IJESIT)* 2(3):110–115
2. Murthy GRS, Jadon RS (2009) A review of vision-based hand gestures recognition. *Int J Inf Technol Knowl Manage* 2(2):405–410
3. Du W, Li H (2000) Vision based gesture recognition system with single camera. In: WCC 2000-ICSP 2000. 2000 5th international conference on signal processing proceedings. 16th World computer congress 2000, vol 2, pp 1351–1357. IEEE
4. Segen J, Kumar S (1998) Human-computer interaction using gesture recognition and 3D hand tracking. In: Proceedings 1998 international conference on image processing. ICIP98 (Cat. No. 98CB36269), pp. 188–192. IEEE
5. Liu Y, Jia Y (2004) A robust hand tracking and gesture recognition method for wearable visual interfaces and its applications. In: Third international conference on image and graphics (ICIG'04), pp 472–475. IEEE
6. Kim D, Lee S, Paik J (2009) Active shape model-based gait recognition using infrared images. In: International conference on signal processing, image processing, and pattern recognition, pp 275–281. Springer, Berlin, Heidelberg
7. Jaramillo AG, Benalcázar ME (2017) Real-time hand gesture recognition with EMG using machine learning. In: 2017 IEEE second ecuador technical chapters meeting (ETCM), pp 1–5. IEEE

8. Ishiyama H, Kurabayashi S (2016) Monochrome glove: a robust real-time hand gesture recognition method by using a fabric glove with design of structured markers. In 2016 IEEE virtual reality (VR), pp. 187–188. IEEE
9. Hsieh CC, Liou DH, Lee D (2010) A real time hand gesture recognition system using motion history image. In: 2010 2nd International conference on signal processing systems, vol 2, pp V2–394. IEEE
10. Hussain S, Saxena R, Han X, Khan JA, Shin H (2017) Hand gesture recognition using deep learning. In: 2017 International SoC design conference (ISOCC), pp. 48–49. IEEE
11. Dehankar AV, Jain S, Thakare VM (2017) Using AEPI method for hand gesture recognition in varying backgrounds and blurred images. In 2017 International conference of electronics, communication and aerospace technology (ICECA), vol 1, pp 404–409. IEEE
12. Yu C, Wang X, Huang H., Shen, J., & Wu, K. (2010, October). Vision-based hand gesture recognition using combinational features. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (pp. 543–546). IEEE
13. Dardas NH, Georganas ND (2011) Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Trans Instrum Meas* 60(11):3592–3607
14. Chen Q, Georganas ND, Petriu EM (2008) Hand gesture recognition using Haar-like features and a stochastic context-free grammar. *IEEE Trans Instrum Meas* 57(8):1562–1571

Mathematical Behavior of a Pandemic Using Models and Growth Curves



Harsh Singh , Ashish Kumar , and N. Nalini 

1 Exponential Growth

Exponential growth is quite a common topic yet; human intuition has a hard time to wrap its head around it [1]. We can anchor on a sequence of small seeming numbers but become surprised when it shows substantial growth. In simpler words, exponential growth means as you go from one day to another it involves multiplication by a constant which is greater than one [2]. Viruses are a textbook example of exponential growth because what causes new cases are, the existing cases. That is the reason that initially all the pandemic appears to be in an exponential curve and it can be quantified by algebraic equations.

$N_d \Rightarrow$ Number of cases on a given day

$S \Rightarrow$ Average number of susceptible people exposed to infected people each day.

$p \Rightarrow$ probability of each exposure becoming an infection.

$$\Delta N_d = S.p.N_d \quad (1a)$$

$$N_{d+1} = N_d + S.p.N_d \quad (1b)$$

H. Singh (✉)

University of Alberta, Edmonton, Canada

e-mail: harsh6@ualberta.ca

A. Kumar

University at Buffalo, Buffalo, USA

e-mail: akumar59@buffalo.edu

N. Nalini

Nitte Meenakshi Institute of Technology, Bangalore, India

e-mail: nalini.n@nmit.ac.in

$$N_{d+1} = N_d(1 + S.p) \tag{1c}$$

As we can notice, N_d itself is directly proportional to the change in the number of cases between two days and its multiple $(1 + S.p)$ is a constant greater than 1 [2]. Hence, growth is exponential. However, in a system where the total sample space is fixed, i.e., the total population of the area is under consideration, the growth cannot be exponential for long. The term p will decrease with time as the person who is already infected cannot be infected again. So if we denote p by:

$$p = \left(1 - \frac{N}{\text{population}}\right)$$

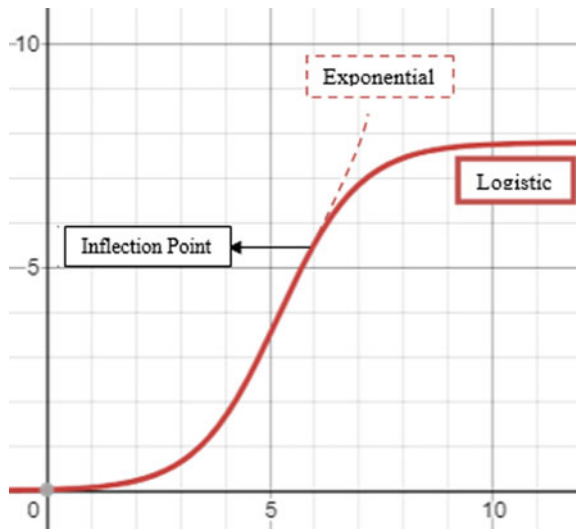
Now, if we factor p in equation 1a, we'll get:

$$\frac{dN}{dt} = c \left(1 - \frac{N}{\text{population}}\right)N$$

where c is some constant. Upon plotting this equation, we will get a logistic curve that is indistinguishable from an exponential curve in the beginning but ultimately levels out before reaching population size after a point known as an inflection point. After the inflection point, the number of new cases remains roughly constant before it starts decreasing (Fig. 1).

One term which is quite commonly used for tracking inflection point is the growth factor. It is the ratio of the number of new cases one day (ΔN_d) to the number of cases the previous day (ΔN_{d-1}).

Fig. 1 A sample graph demonstrating logistic curve, exponential curve, and inflection point



$$\text{Growth factor} = \frac{\Delta N_d}{\Delta N_{d-1}}$$

when the value of the growth factor is greater than 1, it suggests that we are still on the exponential part of the graph. When it reaches one, it suggests that we have reached the inflection point. It is quite a significant point as you may speculate that total cases will max out at two times from this point. During this time, strict actions can be taken by administrative authorities such as imposing lockdowns so that the curve starts to flatten swiftly. Importance of knowing inflection point and growth factor can be understood by below example:

Taking our equation 1c and modifying it for N_d , we get:

$$N_d = (1 + S.p)^d . N_0$$

For an N_0 of 100,000 and estimating cases for 15 days at the inflection point,

$$N_{15} = (1)^{15} . 100000 \Rightarrow 100000$$

Now, if a strict measure is taken by the administrative authority such as lockdown for 15 days and it brings down the growth factor by a minimum of 0.10 then,

$$N_{15} = (0.90)^{15} . 100000 \Rightarrow 20589$$

With just the decrease in 0.10 in growth factor, there was a 79.4% decrease in the number of cases.

2 Plotting the Curve

Plotting total cases against time gives us an exponential curve, which is the most popular way of reporting a pandemic [3]. However, this approach does not create an accurate picture of the situation. The first issue with that is, with a logistic curve, it is hard to visualize the inflection point and it makes it difficult to compare it with the graph of other countries as shown in Fig. 2. As we can see in the graph, it indicates that the United States is performing poorly in handling the coronavirus cases and India is doing a decent job. Although the real story is India has 767,296 cases, which is a huge number. Along with that if we assume that India or Brazil is doing better than the US then also we'll be wrong. Although the number of cases in the US is still more than that of India and Brazil, it is important to note that these countries are still on the exponential part of the curve and there is a possibility that they may cross that mark eventually as the curve of the US flattens.

Now, this brings us to the second issue with the orthodox way of plotting the pandemic graph. As the current time, i.e., the day or month or a year is not a factor that affects a pandemic in any way so plotting graph using time does not make any

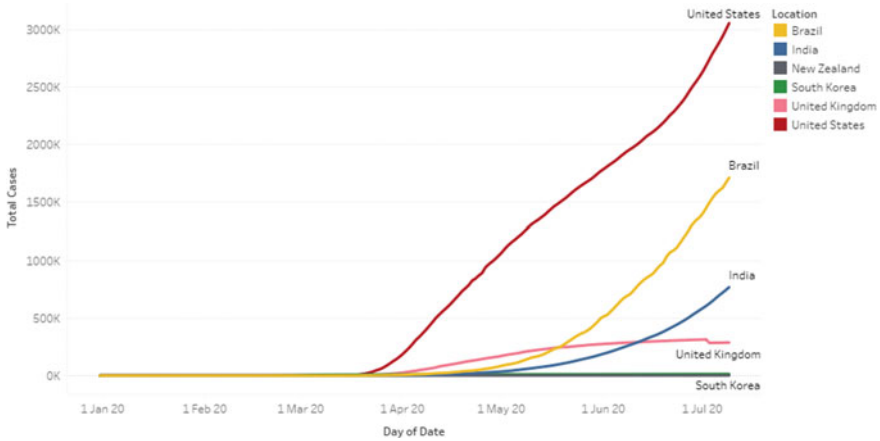


Fig. 2 Graph demonstrating the growth of COVID-19 cases of various countries with respect to time

sense. The increase in cases with respect to time is a correlation. Also, using total cases as the y-axis distracts us from what we should be focussing on. When we use total cases, it does not tell about when has the cases stopped being exponential or slowing down [4]. Besides, it is not the right metric to compare the condition of the two countries. Using new cases on the other hand makes it easier to gauge the situation of the pandemic. Using new cases vs total cases gives us the right trend to monitor as existing cases are the causation of new cases. Since everyday cases can result in many small and regular fluctuations, we can take a cumulative sum of new cases for the last 7 days to reduce the fluctuation. Now, overcome the exponential growth shown on the graph, we can use a logarithmic scale on both the axis so that we can get a linear graph which is very easily fitted with linear regression and we can easily visualize when we hit the inflection point and when the case starts to decrease (Fig. 3).

As we can observe in Fig. 2, India and Brazil are following the trends of the US while countries such as the UK, New Zealand, and South Korea have already met the inflection point. This graph helps us understand that countries like India and Brazil are not doing better than the US but are a few days behind it. When we understand that we acknowledge the gravity of the situation and act accordingly.

However, there are a couple of demerits of this graph as listed below:

- (1) The logarithmic scale distorts the scale. 1000 cases appear very close to 10,000 and that might make the pandemic look less serious than it is.
- (2) Since we are taking weekly new cases on our y-axis, it results in slightly delayed trends. Although this is for the better as it does not show us minute reduction of cases but focuses on the inflection point and makes us look at the big picture.

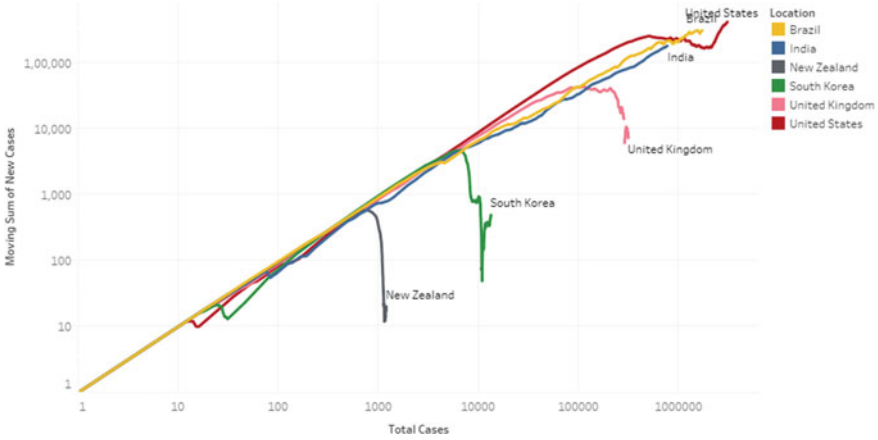


Fig. 3 Graph demonstrating an increase in new COVID-19 cases of various countries with respect to total cases on a logarithmic scale

3 Pandemic Models

It is broadly classified into two types namely mechanistic and phenomenological models.

3.1 Mechanistic Models

In these models, we use processes that might have given rise to the data to hypothesize a relationship between the variables in the dataset. Let’s dwell deep into some of the commonly used mechanistic models.

SIR: It is the most basic model used for understanding a disease outbreak. The other mechanistic models also use the SIR model as their fundamental principle and then build upon it [5]. Consider the following key terms associated with the model:

S => number of people susceptible to the virus.

I => number of people already infected by the virus.

R => number of people who were infected but now have recovered.

Now, using differential calculus we can model the dynamics of any pandemic using the following system of differential equations.

$$\frac{dS}{dt} = -\alpha SI \tag{2a}$$

$$\frac{dI}{dt} = \alpha SI - \beta I \tag{2b}$$

$$\frac{dR}{dt} = \beta I \quad (2c)$$

where α is the transmission rate per individual, and β is the recovery rate. Here, the rate of change of susceptible people per day is a negative of αSI , as these are the fraction of people who have transitioned from susceptible to the infected state. In 2b, there is a negative component βI as these are the fraction of infected people who have recovered and transitioned to the recovered state.

SEIR: In this model, it has an extra phase called exposed. This is the fraction of the population who are infected but not infectious. It can be defined by the following system of differential equations.

$$\frac{dS}{dt} = -\alpha SI \quad (3a)$$

$$\frac{dE}{dt} = \alpha SI - \lambda E \quad (3b)$$

$$\frac{dI}{dt} = \lambda E - \beta I \quad (3c)$$

$$\frac{dR}{dt} = \beta I \quad (3d)$$

where λ is the rate at which people leave exposed (E) and become infected (I). In this model rather than directly moving from susceptible to the infected stage, people move to the exposed state first and then from the exposed state to the infected state and at last to the recovered stage.

SIR with Vital Dynamics: In this model, we consider the SIR model along with vital dynamics. We consider the birth rate and death rate in this which makes these equations further complicated. Given below is the system of differential equations defining the model.

$$\frac{dS}{dt} = -\alpha SI + \mu N - \sigma S \quad (4a)$$

$$\frac{dI}{dt} = \alpha SI - \beta I - \sigma I \quad (4b)$$

$$\frac{dR}{dt} = \beta I - \sigma R \quad (4c)$$

where μ is the birth rate and σ is the death rate.

3.2 Phenomenological Models

Phenomenological or statistical models are the ones that describe the empirical relationship of the disease to each other. These models to some extent are consistent with fundamental theory but do not emphasize the process through which the said relationship came to be [6]. Let’s go through a couple of these models:

Logistic model: The logistic model is the simplest model which shows an initial exponential curve that is followed by slowing down the growth factor resulting in saturation [5]. It can be modeled using a previously mentioned equation,

$$\frac{dN}{dt} = c \left(1 - \frac{N}{\text{population}} \right) N$$

where c is the exponential growth rate. As N increases, $\frac{dN}{dt}$ keeps decreasing until it saturates when N becomes almost equal to the population.

Richards model: These models are the more flexible version of logistic models [5]. The logistic model shows a fixed exponential growth rate slowing down. To add flexibility, we can use this model for the cumulative incidence curve. It can be modeled using the following equation.

$$\frac{dN}{dt} = c \left(1 - \left(\frac{N}{\text{population}} \right)^\delta \right) N$$

where parameter δ controls the steepness of the curve. It is interesting to note here that the logistic model is a special case of the Richards model where $\delta = 1$.

Among all these models including mechanistic and phenomenological models, SIR tends to provide the best fit to data for the entire pandemic curve. The other models with these extra parameters become complex and do not provide a good fit and hence fail to give us a true representation of the pandemic as shown in Fig. 4.

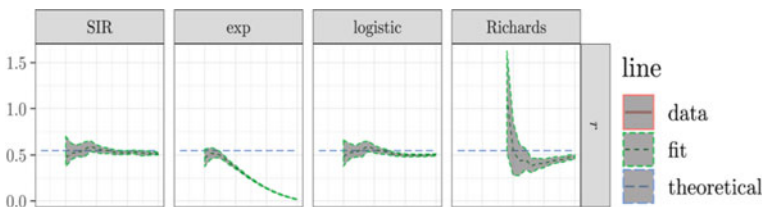


Fig. 4 The comparison of the results of fitting the SIR, exponential, logistic, and Richards models to a simulated weekly incidence curve, as a function of the endpoint of the fitting window. [Source: Estimating epidemic exponential growth rate and basic reproduction number – Junling Ma]

4 Analysis

To ascertain if a viral outbreak will result in a pandemic, we need to examine the second equation of SIR in detail during the initial outbreak [7].

$$\frac{dI}{dt} = \alpha S_0 I_0 - \beta I_0$$

where S_0 and I_0 are the initial fraction of susceptible and infected people. If $\frac{dI}{dt}$ is greater than 0 then it means we have a higher number of susceptible people getting infected than infected people getting recovered. Hence, there is a higher chance of disease spread transitioning into a pandemic. If it is less than zero, then it suggests that the recovery rate of people is more than the rate at which people are getting infected. Hence, cases might increase but will eventually die down. Simplifying this equation further.

$$\begin{aligned} \frac{dI}{dt} &= \alpha S_0 - \beta < 0 \\ R_0 &= \frac{\alpha S_0}{\beta} < 1 \end{aligned} \tag{5}$$

The constant R_0 that we just derived is commonly known as Reproduction Number [1]. It is the number of people each infected person can infect on an average. So, to prevent the exponential increase in cases and the possibility of a pandemic, all we have to do is find a way to make this term less than one as fast as possible.

One way to do that is by reducing the transmission rate (α) as it is directly proportional to R_0 . It means we need to take measures to reduce the spread of the virus when an infected person comes in contact with susceptible people. Few ways of achieving this are by washing hands regularly, keeping workplaces well-sanitized, and wearing masks.

Reducing the number of susceptible people (S_0) will also result in lowering R_0 and in turn, getting to inflection point quickly. We can do this by doing social distancing. Many governments have tried to implement this by conducting lockdowns. Although in doing so, as we have experienced during COVID-19, has a huge impact on the economy and also affects the mental health of people. Hence, it is not an ideal way to go.

Increasing the recovery rate (β) will also help us in reducing R_0 . Although it is an arduous task since we have to either vaccinate more people or come up with a medication that cures the patients faster.

5 Proposed Model for Handling a Pandemic

Let's start with a definite population of a city where the virus has just started to spread. Now, in the initial stage of a pandemic, there will be a few infected people denoted by 'X' and more susceptible people denoted by 'O' as shown in Fig. 5.

As discussed in the previous section, for controlling the increase in cases, we must reduce the transmission rate (α) and the number of susceptible people (S_0). For reducing the transmission rate (α).

The government can issue guidelines to citizens like wearing masks, washing hands, etc. For reducing the number of susceptible people government must impose restrictions on the movement of people and as discussed in the previous section the most common way of doing that is by imposing lockdowns. If implemented properly, it creates a significant reduction in S_0 . However, it has its limitations, a couple of which are as follows:

- (1) It creates a huge dent in the economy and the recovery from it will take more and more time as the virus continue to exist.
- (2) With continuously staying inside the house and lack of resources make people develop anxiety, irritation, isolation, depression, etc.

A better way to deal with it is by creating social bubbles. Creating a social bubble means restricting the movement of people to a definite area. We can divide the population of a city into small strategically planned bubbles as shown in Fig. 6 where we have divided the same population from Fig. 5 into 4 different bubbles. Each bubble will be allocated with a hospital, police station, grocery vendors, and other essential services. In this model, the restaurants and shops can still be open but only for deliveries and takeaways within the bubble. In this way, people who can work from home and small businesses can still operate. This model will not keep the economy running as it was before the restrictions but it will surely help in mitigating the impact of restrictions.

Fig. 5 Diagram demonstrating a finite population which contains infected people (X) and Susceptible people (O)

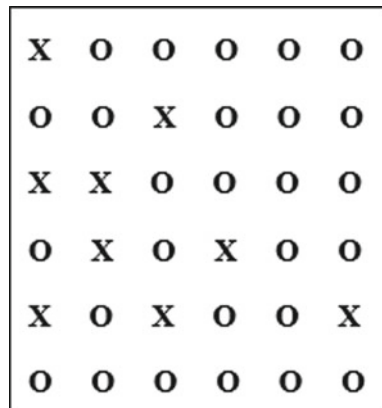
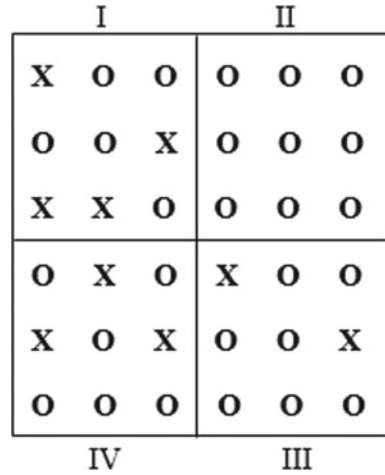


Fig. 6 Diagram demonstrating a finite population divided into 4 bubbles I, II, III, and IV which contains infected people(X) and Susceptible people(O)



However, it still doesn't counter the mental effects caused due to lockdowns. To counter it, we have a smaller bubble between 5 and 6 households, commonly known as a support bubble where people can interact without social distancing (Fig. 7).

It is an approach that is quite popular in New Zealand and the United Kingdom while handling COVID-19. There must be certain guidelines issued by the government on how the bubble can be formed [8]. Some of the obvious ones can be:

- (1) It cannot consist of more than 8 members so that it makes it easier to do the group testing.
- (2) It should be within the social bubble which they are currently in.

Fig. 7 Diagram demonstrating a finite population divided into 4 bubbles I, II, III, and IV which are further divided into support bubbles and contains infected people (X) and Susceptible people (O)

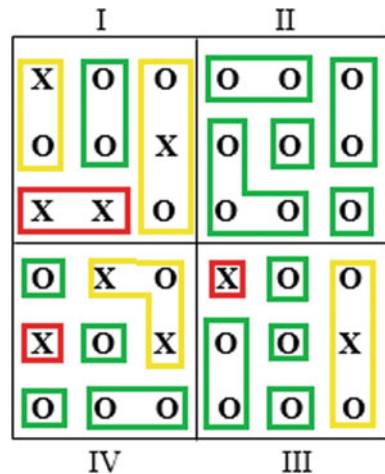
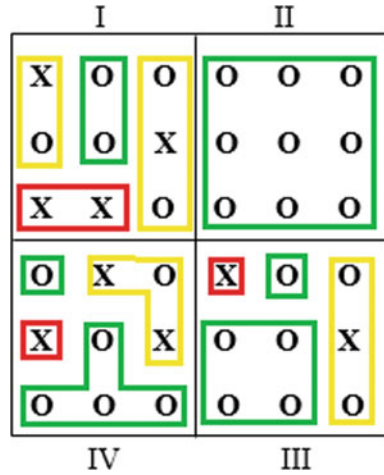


Fig. 8 Diagram demonstrating a finite population divided into 4 bubbles I, II, III, and IV which are further divided into support bubbles and contains infected people (X) and Susceptible people (O)



- (3) It can be restricted to people who are staying alone and single parents with kids less than 18 years of age.
- (4) When any person in the group shows symptoms of the disease all the members of the group should self-isolate themselves.
- (5) The people within the support bubble should be tested separately and the remaining people can be group tested, etc.

Group Testing: It is a method of taking samples from a group of people and mixing them. Now, the mixture as a whole is tested for the virus. If the result is negative, then the whole group can be cleared as negative and in turn, we save that many testing kits and time. If the result is positive, depending on the size of the group it can be divided again and tested. It is a great way of testing more people with a lesser number of kits. Unfortunately, it is viable only until the prevalence of a given population is around 20% [9]. After that, it starts getting more and more expensive.

As shown in Fig. 8, in this way, we can try to contain the virus within support bubbles which makes the group testing more viable. In the worst-case scenario, if the virus spreads to the majority of the people, the curve will soon become logistic since the population of the social bubble is not much. Along with that, there is a chance of developing a herd immunity within that social bubble which intern will help in reducing the restrictions for that bubble. Although with numerous merits, there are a few demerits to this model as well:

- (1) It will still affect the people who cannot work from home and whose workplaces are outside their social bubble.
- (2) It becomes very difficult to divide the social bubbles where the ratio of place to essential services is very low.
- (3) It cannot minimize the number of susceptible people (S_0) as well as lockdown does.

Expanding the Bubble: For this model to yield the maximum benefit, we need to continuously make the bubble larger consisting of people/bubbles that are not infected. This way we will help the economic activities return to their prime. One way to quickly expand the bubble is by conducting extensive group testing and start expanding and declaring the bubbles as not infected. As shown in Fig. 8, the bubble II had no infected person, so we expanded their support bubble and now bubble II is free of the virus. Similarly, for bubble III and bubble IV, we have expanded certain support bubbles that were free of the virus. In any unfortunate event, if anyone in the bubble gets infected, we can return to the earlier stage of the support bubble. This time, requiring group testing of only the people of the support bubbles that came in contact with the infected person. To reduce the number of testing, the government can make use of certain exposure notification apps, which can help narrow the bubbles to be tested.

References

1. Wallinga J, Teunis P (2006) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Royal Soc B Biological Sci* 274:599–604
2. Khan Academy, Exponential Growth Functions | Exponential and logarithmic functions | Algebra II | Khan Academy. <https://www.youtube.com/watch?v=6WMZ7J0wwMI>
3. Sanderson Grant, Exponential Growth and Epidemics. <https://www.youtube.com/watch?v=Kas0tIxDvrg&t=29s>
4. Reich Henry, How to tell if we're beating COVID-19. <https://www.youtube.com/watch?v=54XLXg4fYsc&t=102s>
5. Bazett Trefor, The Math of the Epidemics! Into to the SIR Model. <https://www.youtube.com/watch?v=Qrp40ck3WpI&t=318s>
6. Frigg R, Stephan H (2015) Models in Science. In: Zalta, Edward N (ed), *The Stanford Encyclopedia of Philosophy* (Fall 2012 ed.). Retrieved 24 July 2015
7. Ma J, Dushoff J, Bolker BM, Earn DJD (2013) Estimating initial epidemic growth rates. *Bull Math Biol* 76:245–260
8. Michelle Roberts, Coronavirus bubbles: How do they work and who is in yours? Published On 06 July 2020
9. Biswas Atanu, COVID-19: A Way to Test More People with Fewer Kits. <https://science.the-wire.in/the-sciences/covid-19-a-way-to-test-more-people-with-fewer-kits/>
10. Ma Junling, IDM volume-5 (2020). Estimating epidemic exponential growth rate and basic reproduction number 2019.12.09

Technological Advances in Airborne Wind Power: A Review



A. Meghana , B. Smitha , and Shefali Jagwani 

1 Introduction

Wind originates due to unequally heated land surfaces by the sun. Wind power is amongst the most popular sources of renewable energy. According to the recent report of the World Wind Energy Association, the current wind capacity has reached 6508 GW in 2019 and is increasing considerably every year [1]. Figure 1 presents the rate of growth of wind power in the last five years.

The conventional wind turbines need more space, are costly and have many other limitations. To overcome these limitations, Airborne Wind Energy Systems (AWEs) have gained popularity in recent times [2]. AWEs use tethered flying devices to reach higher heights where the wind is steadier. Towers are not required in AWEs, thereby reducing the cost of installation considerably. Harnessing wind energy at higher heights enables it to generate four times of power as compared to small conventional wind turbines. Also, it can generate 2–3 times the power as compared to solar. In the initial investigation phase, Loyd, in the year 1970, built a towerless wind generator, tethered to the ground like a kite. Later, various AWEs were designed with the assumptions made by Loyd. The first application studies went worldwide during the early twenty-first century. Many academic experts began to work on these concepts and developed prototypes [1–14]. Researchers have proposed various technologies and the control techniques in this field [15–33]. In support of the AWE

A. Meghana (✉) · B. Smitha · S. Jagwani
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: meghana.a@nmit.ac.in

B. Smitha
e-mail: smitha.b@nmit.ac.in

S. Jagwani
e-mail: shefali.jagwani@nmit.ac.in

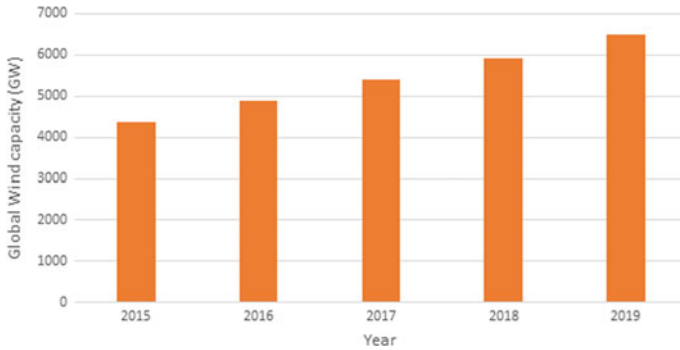


Fig. 1 Global wind capacity in GW

sector, today many academic institutions have filed patents [34–37] to startups and public companies are investing millions of dollars.

2 Emerging Technologies in AWEs

AWEs are the electro-mechanical systems which generate electricity from kinetic potential of the wind. Figure 2 shows the different types of AWEs, Ground-Gen AWEs (GG-AWEs) and Fly-Gen AWEs (FG-AWEs) [3].

In GG-AWEs, mechanical work is transmitted to the ground for the production of electricity by the generator mounted on the ground, using force from the aircraft with the help of single or multiple ropes. Again GG-AWEs can be divided into fixed ground stations where devices are rooted to ground and where devices in a moving vehicle is known as moving ground station. In FG-AWEs, aircraft produce power using wind turbines. Special cables are used to send electrical power to ground [3]. The output power in AWEs is dependent on the wind speed as shown in the equations.

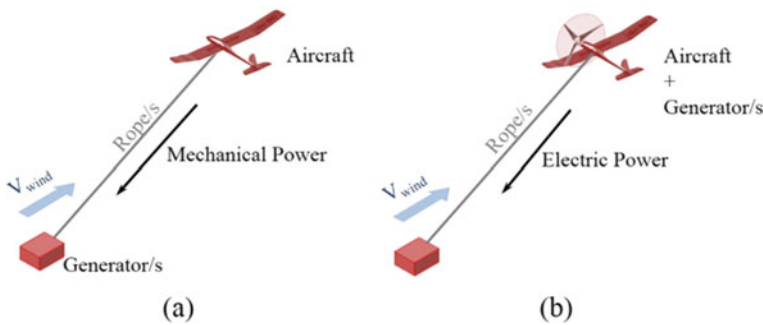


Fig. 2 Different types of AWEs- **a** GG-AWEs, **b** FG-AWEs [3]

Power calculated in crosswind turbines as simplified from Loyd's formula [3]:

$$P = \frac{2}{27} \rho_{\alpha} A C_L G^2 V^3 \quad (1)$$

where P is output power; ρ is the air density, V is the wind speed, θ is the angle that the tether makes with direction of wind, G is the effective gliding ratio, C_L is the lift coefficient and A is wing area.

Here, calculation of wind speed $V_{(z)}$ is an important requirement for these airborne systems which is done either by using observers or a wind speed model. As found in the literature, the equations used for finding wind speed are as follows [4]:

$$V_{(z)} = v_0 \left(\frac{z}{z_0} \right)^{\gamma} \quad (2)$$

or

$$V_{(z)} = v_0 \frac{\ln \ln \left(\frac{z}{z_r} \right)}{\ln \ln \left(\frac{z_0}{z_r} \right)} \quad (3)$$

where v_0 is the measured wind speed at an altitude z_0 , γ is a friction coefficient and z_r . The wind velocity is used for calculating the wind power.

3 Technological Advances in AWEs

Researchers are using different technologies for harnessing wind power at higher altitudes. Table 1 shows the classification of AWES based on different technologies. Fixed GG systems are widely used by various academicians and companies in which there are two phases: generation phase and consumption phase. In the generation phase, force due to lift is used along with the crosswind flight. In the consumption phase, the aircraft is brought to its actual position by rewinding ropes with the help of a motor [3]. The various aircrafts used in these GG systems are: Inflatable kites, Foil kites, Delta kites, Gliders, Parachute, Aerostat, Magnus effect, etc. They use lift, drag, or Magnus effect as their main force. Moving GG systems are much more complex in operation but are simple in terms of connection with the grid. These systems generate power by using traction force of ropes and not by winding or unwinding mechanisms as in case of fixed systems. These moving ground systems use vertical axis generators and rail generators.

In FG-AWEs, the power generated on aircraft is transmitted through a special rope. This rope is integrated with electrical cables and power is transmitted to the station. Electrical energy conversion is achieved by special aircrafts using crosswind and

Table 1 Classification of AWESs based on different technologies [2]

		AWE													
		Ground -Gen							Fly-Gen						
		Fixed ground Station					Moving ground Station		Crosswind	Non-Crosswind					
Technologies		Inflatable Kite (Lift)	Foil Kite (Lift)	Delta Kite (Lift)	Glider (Lift)	Semi-rigid wing (Lift)	Parachute (Drag)	Aerostat (Magnus effect)	Rail- Inflatable Kite (Lift)	Rail-Foil Kite (Lift)	Axial- Inflatable Kite (Lift)	Turbines on a tethered aircraft (lift)	Tethered quadcopter (thrust)	Turbine on an air balloon (Buoyancy)	Magnus Effect turbine Buoyancy)
Companies	Kite Gen Stem	Sky Sails	Ener kite	Ampyx Power	Kite Gen Stem	Guangdong Tech	Omni-dea	Kite Gen Rail Carousel	NTS	Kite Gen Rail Carousel	Makani Power	Sky Wind Power	Altaeros Energies	Omni-dea	
	Windlift	Ener-kite		e-Kite				Kit-energy			Joby Energy				
	Kite Power			Kite mill											
	Kit-energy			Twing Tec											
	Swiss Kite Power														

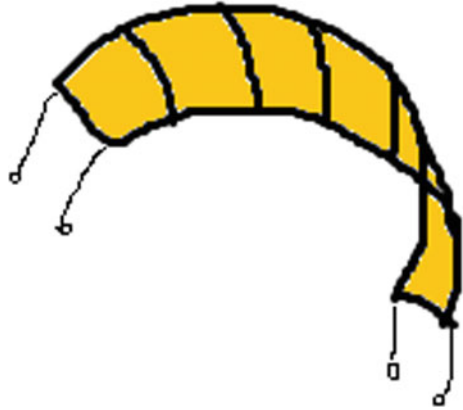
non-crosswind flights with wings lift, buoyancy and thrust as their main force. The detailed description of the companies using these technologies is given as follows.

3.1 KiteGen Stem

This European-based company uses the concept of harnessing wind from higher altitudes. It is a vertical axis wind turbine which eliminates the problems of traditional wind turbines [5]. KiteGen Research has come up with KiteGen Stem technology which has many favorable features.

In KiteGen Stem, a kite-like semi-rigid wing is wound with ropes to the pulley system which is the stem. From the stem, the top of the control station is linked to the pivot joint. The stem supports the kite structure and damps the forces in the rope due to speedy winds. The stem can move in two directions with respect to the ground. The kite is initially hung upside down and once it has taken off, the production will start. The kite makes paths, thereby the ropes are wound up and the winches will move. Mechanical energy will be transformed into electrical energy by the motor-generator sets. The main aim is to withdraw the cable using low energy using a maneuver known as ‘flagging’. This mode enables the kite to fly sideways [3]. The

Fig. 3 Semi-rigid wing design by KiteGen



power consumed is lesser than the power produced in the phase of production. There is another special maneuver which returns the lift force on the kite. It has a high and a constant speed leading to a good capacity factor which is needed for wind stability. The structure is light hence saving costs on fabrication. The generators can be placed at a lesser distance when compared to conventional wind turbines (Fig. 3).

3.2 *Enerkite*

This company was founded in Germany and they developed a 30 kW, which consists of a portable generator that generates power continuously. It is a FG-AWES, where a generator is installed using pivot joints for azimuthal rotation of the kite on a truck. Enerkite generates the power in a power and recovery phase. During the power phase, the wing flies due to the cross-wind, thereby generating power while in recovery phase wings return to the starting point using minimum energy. Enerkite is capable of producing 100–500 kW, portable and suitable for larger agricultural areas, mining companies and disaster relief as it can be operated on-grid and off-grid [3, 6].

3.3 *Windlift*

This company was started in 2006 and mainly develops Airborne Power generators (APG) and control systems for actuated kites. There are three tethers provided from the kite to the ground station. It can produce up to 12 kW power from the available wind and has the power and the recovery phase. It has applications in disaster recovery, island and remote villages [3, 7]. Due to its high-power density and lightweight, it can be easily transported and can be used in military and off-grid locations as it is easy to deploy.

3.4 Makani

Makani has developed prototypes Airborne Wind Turbine (AWT) which is a Ground-Gen AWES, with a solution for take-off and landing issues by using bimodal flight where electricity is generated at utility scale. In bimodal flight, rotors are used as engines on AWT, same as in quadcopter flight and it takes off vertically, due to propeller thrust. The AWT changes flight form into a tethered flight airplane when unrolled all the rope length. In the second flight mode, the wind itself will power the flight in circular path and power is generated from the wind by the rotors on AWT as generators and cable length will be the same during this phase, flight lands like a quadcopter. Makani company has also developed 'Wing 7', which is fully automatic operations and capable of generating 20 kW power. Makani also developed 'the M600', which is a prototype capable of generating 600 kW with eight turbines, each with five propeller blades [3, 8]. Makani can be economically and strongly applicable in deep offshore water where wind is consistent at higher altitudes between 500 and 1000 feet.

3.5 Kite Power/Swiss Kite Power

It is a company based in Switzerland which started its collaborative research in 2009 [3]. Its first prototype was on a C-kite tethered on one end and a pod on the other to control it. The main components are a load-bearing tether, a ground-based electric generator and a high-performance kite [9].

The main advantages are its lower cost, silent performance and easy upgradation. It also uses 90% less material and is twice efficient. It can go up to an altitude of 450 m and power of 100 kW.

3.6 TwingTec

TwingTec which is a Swiss-based company developed GG-AWES of capacity 100 kW. During takeoff and landing, rotors are used and it rotates perpendicular to the plane wings. TwingTec can be easily installed in remote areas by deploying multiple drones together. The company has a plan to supply power to the remote and off-grid areas with a standard 20-foot shipping container with generator and energy conversion devices [3].

3.7 *Skysails Power*

This is a German-based company developed GG-AWES for cargo vessels and yachts to generate power using kite technology, operating at altitudes between 200 and 800 m. It has developed mobile AWES with capacity of 250 kW-1 MW and offshore AWES with capacity 1–3.5 MW. Skysails used foil kites in which the control pod is used with one rope to control steering of the kite and angle of attack [3]. During takeoff, it uses natural wind force on the deflated kite with a storage compartment to compact the mast and it expands vertically to some meters above the ground level at the beginning of the launching phase. The kite will be having definite shape and stiffness during the production phase by inflating the kite. Skysails AWES can be installed on-land as a mobile or stationary installation and it can be installed offshore using conventional or floating foundations.

3.8 *eWind*

eWind solutions is a US-based company which developed a low height, rigid wing GG-AWES. It has come up with AWES that are reasonable. It uses a Tethered energy device (TED) to harness reliable wind energy from greater heights. It flies crosswind and turns on the ground generator, thereby generating electricity. The main area of application is in agriculture. It produces around 12 kW peak power and 40,000 kWh average energy which is sufficient for farming operations. eWind systems are sustainable, cost efficient and have a long-life span. It frees up acres of land for crops and has low environmental impact [10]. They also generate their own energy hence have green branding opportunities. The movements of the TED are similar to that of a bird and its shadow curbs pest's encroachment.

3.9 *Ampyx Power*

This was the first company in the Netherlands to develop a pumping glider generator. They have officially registered 2 aircrafts, the AP-2A1, 2 which are known as 'Power Planes,' which are automatically controlled. In these planes, the house is made up of carbon which contains onboard electronic devices like sensors. The ground station is connected to the glider using a rope. During takeoff, at some meters of distance, the glider will be on the ground in such a way that it faces toward the ground station [3]. When the traction force is exerting on rope, the glider will start moving on the ground and the glider takes off when the lift force is more than the weight force. The glider landing is similar to the airplane. These types of AWES are very compact to harness the energy from powerful wind at high altitudes. It reduces the cost for construction and installation as it does not require heavy foundation and tower.

3.9.1 Kite Mill

Kite mill situated in Norway started with the development of Ground-Gen AWES. It has a developed wind technology which makes use of constant winds at higher altitudes (around 400–1000 m). It uses a kite which flies in a spiral while one line is tethered to the ground station, thereby generating power. This recovery of energy is comparatively cost-effective than other forms of energy. This procedure of energy production does not require much of human help; hence, it can work in silence. It reduces the cost of energy by 50% as it uses 90% less material and also increases the energy yield. The technology access to its resources makes it financially viable for wind energy [11].

3.9.2 Sky Pull

This is a Swiss-based company started in 2013. It produces GG-AWEs with 95% less material and has high production capacity than traditional wind turbines [12]. So, it produces electrical power at lower cost than fossil fuels. It is fully automatic vertical takeoff and landing at higher altitudes and produces electrical energy using tether connected to the generator.

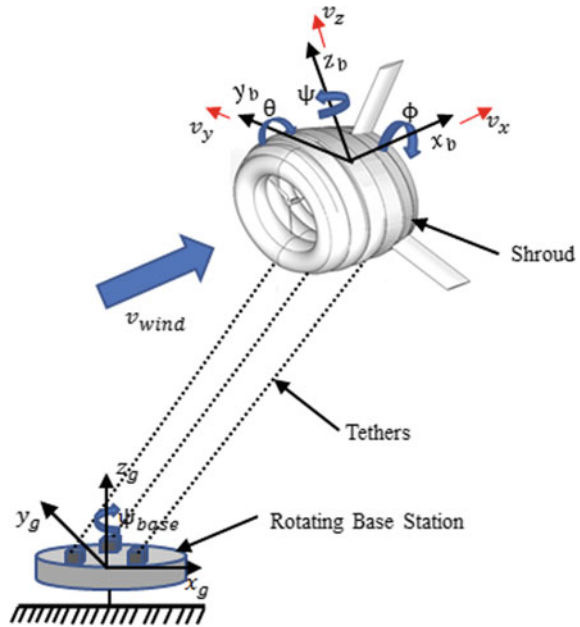
3.9.3 Altaeros

Altaeros, founded at MIT in 2010, developed the world's first advanced autonomous aerostat platform. According to the dictionary meaning, 'An aerostat is lighter than an air aircraft that gains its lift through the use of a buoyant gas.' Instead of using wings to lift, altaeros uses helium for the inflatable shell with a wind turbine inside it. Strong ropes are used to hold the Buoyant Air Turbine (BAT) at a height of 1,000 feet [3]. It generates double the wind power generated by tower-mounted turbines of the same size, reducing the cost of electricity for remote sites and microgrids. Aerostats can survive snowfalls, hurricane-level winds and have features to ensure safe landing. The BAT is aerodynamically designed so that the wind and buoyant force together support in maintaining the generator at high altitude [3]. This BAT has many applications including island and remote power generation, mining, oil and gas, agriculture and disaster (Fig. 4).

3.9.4 Sky Wind Power

Sky Wind Power was the first company whose idea was tested in 1986. It presented a special type of tethered device known as 'Flying Electric Generator' which resembles a big quadrotor with same rotors placed on an airframe. This airframe is connected to the station using a rope with inner electric cables. When this FEG reaches the high altitude, the airframe inclines itself at an angle with respect to wind. The rotors which

Fig. 4 Prototype of the blimp-based generating system developed by Altaeros energy [13]



were functioning as a motor till now, start functioning as a generator. At this position, these rotors get the wind projected parallel to their axes. This allows auto-rotation resulting in generation of electricity and thrust. The advantages of this technology include lesser area swept by the turbine, more generated power, lighter turbine, easy operation and maintenance, higher redundancy and lesser losses [14].

4 Conclusion

The generation of electricity using wind power at high altitudes between 300 m - 1000m and AWESs became the most promising resources. In the past 10 years, several companies started implementing emerging technologies in AWES. In this paper, several emerging technologies in AWES are reviewed with different concepts used by academic research and few startup companies. Some of them are at testing stage and few of them are at concept stage. There is a lot of scope in the fundamental research of these AWEs and scaling up the potential in these systems is a future challenge. Thus, this review has reflected the state-of-art technologies for flying windmills with the advantages, challenges, their current status and the future scope. These emerging technologies give the solution for on-grid as well as off-grid power for remote areas and even during natural disasters without constructing the tower and using durable and light materials, which is also cost effective.

References

1. World Wind Energy Association <https://wwindea.org/information-2/information>
2. Simon Watsona, Alberto Morob, Vera Reisb, Charalampos Baniotopoulosc, et al (2019) Future emerging technologies in the wind power sector: A European perspective. *Renewable and Sustainable Energy Reviews* 113, 109270 <https://doi.org/10.1016/j.rser.2019.109270>
3. Antonello Cherubini, Andrea Papini, Rocco Vertechy, Marco Fontana, (2015) Airborne Wind Energy Systems: A review of the technologies, *Renewable and Sustainable Energy Reviews* 51, 1461-1476 <https://doi.org/10.1016/j.rser.2015.07.053>
4. Uwe Fechner, Rolf van der Vlugt, Edwin Schreuder, Roland Schmehl (2015) Dynamic model of a pumping kite power system, *Renewable energy* 83(2015)705-716 <https://doi.org/10.1016/j.renene.2015.04.028>
5. Ippolito, Massimo & Saraceno, Eugenio. (2019). KiteGen Research High Altitude Wind Generation Tropospheric Wind Exploitation Under Structural and Technological Constraints. DOI: <https://doi.org/10.13140/RG.2.2.12701.97766>
6. Enerkite website. <http://www.enerkite.com>
7. Windlift website <https://windlift.com>
8. Makani power website <https://makanipower.com/>
9. Kitepower website <https://en.wikipedia.org/wiki/Kitepower>
10. eWind website <http://www.ewindsolutions.com/>
11. Kitemill website <https://www.kitemill.com/>
12. Skypull website <https://www.skypull.technology/news>
13. <https://iopscience.iop.org/article/10.1088/1742-6596/524/1/012079>
14. sky Wind Power website <https://www.skywindpower.com/>
15. Lorenzo Fagiano, Member, Mario Milanese and Dario Piga (2010) High-Altitude Wind Power Generation, *IEEE Transactions on Energy Conversion*, VOL. 25, NO. 1. <https://doi.org/10.1109/TEC.2009.2032582>
16. Ruiterkamp R, Sieberling S. Description and preliminary test results of a six degrees of freedom rigid wing pumping system. In: Ahrens U, Diehl M, Schmehl R, editors. *Airborne wind energy* Berlin: Springer; 2013. p. 443–58 [Chapter 26].
17. Ramya S, Tejaswini A, Sneha I (2015) Flying Wind Mill, *International Journal of Advanced Research In Science And Engineering*, Vol. No.4. Special Issue 02:753–759
18. Vermillion C, Grunnagle T, Kolmanovsky I (2012) Modeling and Control Design for a Prototype Lighter-Than-Air Wind Energy System. *American Control Conference Fairmont Queen Elizabeth*. <https://doi.org/10.1109/ACC.2012.6315434>
19. Ahmed M, Hably A, Bacha S (2012). High Altitude Wind Power Systems: A Survey on Flexible Power Kites. <https://doi.org/10.1109/ICEIMach.2012.6350170>
20. Jeevan Adhikari, S K Panda (2014) Overview of High Altitude Wind Energy Harvesting System. <https://doi.org/10.1109/PESA.2013.6828249>
21. Yutaka Terao and Norimitsu Sakagami (2014). A Feasibility Study on the Ocean Higher Altitude Strong Wind Energy Utilization System. <https://doi.org/10.1109/OCEANS-TAIPEI.2014.6964357>
22. Jeevan Adhikari, I V Prasanna and S K Panda (2015) Maximum Power-point Tracking of High Altitude Wind Power Generating System Using Optimal Vector Control Technique. <https://doi.org/10.1109/PEDS.2015.7203477>
23. Aldo U. Zraggen, Lorenzo Fagiano and Manfred Morari (2016) Automatic Retraction and Full-Cycle Operation for a Class of Airborne Wind Energy Generators. *IEEE Transactions on Control Systems Technology*, vol. 24, no. 2. <https://doi.org/10.1109/PEDS.2015.7203477>
24. Jeevan Adhikari, Prasanna I V and Godwin Ponraj (2017) Modeling, Design, and Implementation of a Power Conversion System for Small-Scale High-Altitude Wind Power Generating System. *IEEE Transactions on Industry Applications*, VOL. 53, NO. 1. <https://doi.org/10.1109/TIA.2016.2604388>

25. Michelle Kehs, Chris Vermillion, and Hosam Fathy (2018) Online Energy Maximization of an Airborne Wind Energy Turbine in Simulated Periodic Flight. *IEEE Transactions on Control Systems Technology*, VOL. 26, NO. 2. <https://doi.org/10.1109/TCST.2017.2665553>
26. Natapol Kors Prasertsak, Thananchai Leephakpreeda (2017) Comparative Investigation of Short-Term Wind Speed Forecasting Models for Airborne Wind Turbines. <https://doi.org/10.1109/ICMAE.2017.8038688>
27. von den Hoff D, Haberschusz D, Rotering N, De Doncker RW (2019). Design and Evaluation of a Battery-Supported Electric Drivetrain for Kite-Based High-Altitude Wind Energy Conversion. <https://doi.org/10.1109/CWD.2019.8679526>
28. Karthik Rajan Chandrasekaran (2020) An Alternative Design Approach for a Lighter than Air Airborne Wind Turbine Generator System. <https://doi.org/10.1109/ICSCM46742.2019.9081819>
29. Prithi Seshadri, R Dharmalingam (2016) The Next Generation Windmill. *International Journal for Scientific Research & Development* Vol. 4, Issue 09, pp. 623–626
30. P.Seshadri, R.Dharmalingam (2017) Helium Balloon Windmill. *International Research Journal of Engineering and Technology*, Volume: 04 Issue: 04, pp. 282-287
31. Loyd ML Crosswind kite power (1980) (for large-scale wind power production). *J Energy* 4(3):106–11
32. Specter M (2013) Inherit the Wind. *The New Yorker*. {available online} <http://www.newyorker.com/magazine/2013/05/20/inherit-the-wind>.
33. Maass J, Erhard M. In: Ahrens U, Diehl M, Schmehl R, editors (2013) *Software system architecture for control of tethered kites, airborne wind energy* Berlin: Springer: p. 599–611 [Chapter 35].
34. Gary Dean Ragner (2001) Axial-Mode Linear Wind-Turbine. Application.: 09/941,337.
35. Stephan Wrage, Stephan Brabeck (2007) Wind Energy Plant with a Steerable Kite. Application.: 11/694,715.
36. X Development LLC (2013) Airfoil for a Flying Wind Turbine. Application: 14/145,550.
37. Griffith S, Lynn P, Hardam C (2010) Wind power generation. US Patent Application US7847426B1.

Four-Layer Authentication with Honeypot and Cloud Data Encryption



Sanjay Chari, Harshith Umesh, Athul Sandosh, S. Suganthi, and Prasad Honnavalli

1 Introduction

In today's day and age, utilisation of cloud data storage and compute resources has become a norm for many businesses as all of the technical details, such as server geography, scaling, and memory allocation, have been abstracted through streamlined graphical user interfaces that are easy to use and support fast product development, which aids businesses in being the first-to-market in their categories. However, security loopholes in cloud data storage often prevent businesses from migrating from traditional data storage to cloud storage.

Data security is essential for private cloud as leaks of confidential data could lead to corporate espionage in the hands of malicious users. For this reason, data security on the cloud is an active area of research today.

An authentication technology or mechanism is considered "strong" if it uses multiple authentication steps rather than single-factor authentications. Single-factor authentication systems can be easily compromised: a password could be written down and misplaced or shared, an identity card could be lost or stolen and security questions are easily found out with well-crafted social engineering methods.

We propose a secure private cloud data storage service that makes use of four-layer authentication, with password, OTP, ReCAPTCHA and a unique RSA certification file, for logging in a user into the cloud. We propose the usage of AES-256 encryption for all files stored on the cloud. We make use of Amazon S3 services to store files on

S. Chari (✉) · H. Umesh · A. Sandosh · S. Suganthi · P. Honnavalli
PES University, Banashankari Stage III, 100 Feet Road, Dwaraka Nagar, Banashankari,
Bengaluru, Karnataka 560085, India

S. Suganthi
e-mail: suganthis@pes.edu

P. Honnavalli
e-mail: prasadhb@pes.edu

the cloud. In case a user fails any of the steps involved in the authentication process, he/she is redirected to a honeypot file system with irrelevant content that will be displayed to the user.

We propose a four-factor authentication system over the traditional two-factor authentication system that uses only password and OTP as two-factor authentication systems have been subject to security breaches by bots and hackers in the past. By incorporating ReCAPTCHA into the authentication system, we ensure that bots are denied access to the cloud. By generating a unique RSA certification file for every genuine user, we ensure that hackers cannot gain access to the cloud even if they gain access to the password and OTP of a user as the RSA certification file is confidential to each user. This system might cause minimal overhead in real-time applications, but when dealing with highly confidential data, the overhead is acceptable as a compromise for greater security.

2 Existing Systems

Existing cloud systems emphasise at providing integrity and the storage of data, but the problem of confidentiality still exists and is not still resolved completely. Most cloud systems use one or at most two-factor authentication techniques which can be bypassed if the victim's credentials or mobile device is obtained. Well-crafted social engineering methods can also be used to unlawfully obtain credentials.

Some examples include security questions, a question to which you have to provide the answer to every time you try to login; most answers to such questions could be looked up at public records or social media. Push notifications are another existing method where a notification is received on trying to login; this method does not work if no internet connection present and chances of accidental approves and pushes are ever present. Biometrics is another authentication method which seems unbreakable at first but has a major flaw hidden within it; a compromised biometric is compromised for life because a fingerprint or face id cannot be changed like a phone number or password.

No security technologies to avert intruders, bots and spam are in place; this leaves the system susceptible to DDOS and other attacks. Existing cloud systems implement DES or AES 128-bit encryption methods which are strong encryption techniques but with enough time and good hardware can be broken. All these systems do not ensure 100 per cent security and fail to provide complete security assurances to their cloud users.

3 Proposed System

The system proposed in this paper aims to provide better security aspects to the user. Our system will take care of all integrity, confidentiality and storage of data.

The proposed system uses the concept of advanced encryption standard 256-bits cryptography to completely secure the data saved on the cloud (Amazon S3). Once data is saved on the cloud, it can only be accessed by registered users using valid credentials during the four-factor authentication-based login; if an unintended person or bot try to access the data using wrong credentials they will receive duplicate files implemented using a honeypot. The honeypot log files could be used to alert the administrator and track details of the intruder acting like an intrusion detection system.

We propose a four-factor authentication system over the traditional two-factor authentication system that uses only password and OTP as two-factor authentication systems have been subject to security breaches by bots and hackers in the past. By incorporating ReCAPTCHA into the authentication system, we ensure that bots are denied access to the cloud. By generating a unique RSA certification file for every genuine user, we ensure that hackers cannot gain access to the cloud even if they gain access to the password and OTP of a user as the RSA certification file is confidential to each user. This system might cause minimal overhead in real-time applications, but, when dealing with highly confidential data, the overhead is acceptable as a compromise for greater security.

4 Literature Survey

Cyber security and cloud computing have significantly impacted every section of our lives be it in social media, in education, in health or financial services and businesses. Cloud security a combination of the two domains refers to securing data on the cloud, where it should be quick to access, secure and easy to modify. Securing of data should be considered by software engineers while building them in the design phase of cloud services. Not only pay attention to data redundancy and isolation but also consider data security [7]. Discusses a three-level advanced cloud computing security mechanism consisting of MAC address matching, OTP and biometric verification. The first level of security, MAC address is verified through the login page. If Mac address is not verified, then a randomly generated numeric code is sent to the user's mobile number which is registered with the user's service provider along with their registered email address which makes up the second level of security. In spite of the matching of Mac address, fingerprint verification is a compulsory step which acts as a biometric scanning verification mechanism makes up the third level of security. If all three mechanisms, the Mac address, OTP and biometric scanning, are successfully verified, only then the access is granted [4]. Paper analyses the performance of various encryption techniques for cloud security. It also analyses various multi-level encryption algorithms based on different performance evaluation parameters such as encryption and decryption time, memory usage and throughput [8]. Paper discusses the use of advanced encryption standard and data encryption standard in cloud security, their efficiencies and the respective advantages and disadvantage of the encryption techniques [9]. Paper discusses the approach and implementation

of encryption techniques to enhance data security and data migration in the cloud. The authors combine AES-256 and the information dispersal algorithm (IDA). For encoding, the original data is first encrypted using AES-256 algorithm. Then the encrypted file is divided into several separate files using IDA and stored separately. For decoding, the encrypted data in different files is reconstructed using IDA, then reconverted using AES-256 decryption to get back the original data [10]. Paper on A CAPTCHA-BASED INTRUSION DETECTION MODEL discusses the use of IDS and IPS in security. IDS and IPS (intrusion prevention system) are tools used to detect and prevent bots from getting entry into a system. CAPTCHA or completely automated public turning test is an IPS technique to tell computer systems and humans apart. To detect illegal access of the system from an attacker, IDS and IPS can be implemented with the use of honeypot technology to track their IP address and location and prevent the attacker from gaining access into the system. Considering the above related work, challenges were identified and presented. The solutions proposed by many of the researchers used (1) single-factor authentication techniques or outdated multi-factor authentication techniques, (2) old and reliable encryption algorithms, with which security can be maintained and achieved but not to the fullest and (3) honeypots using IDS and CAPTCHA to only catch spyware and bots. However, this research intends to design an intelligent cloud storage system using four-factor authentication techniques with ReCAPTCHA and the strongest encryption AES-256 using production honeypot.

5 Research Methodology

Our aim is to provide secure cloud access and data security to all the users of the cloud system, where they can login and store their files in encrypted form. Cloud access is implemented using multi-factor authentication. Multi-factor authentication refers to providing authentication after multiple stages of verification in which a user is granted access only after successfully presenting two or more pieces of evidence to a verification mechanism: knowledge (some information only the user knows), possession (something only the user has), and inheritance (something only the user is). The four authentication methods used are: (i) a password or passkey known only to the user. (ii) An OTP, or one-time password, is a dynamic password that is valid for only one login session or transaction, on a computer system, thereby making them invulnerable to reply attacks. A new OTP is generated by the server every time the service is requested. (iii) A ReCAPTCHA is a provider of human verification systems. ReCAPTCHA is designed to protect your website from spams and automated bots. (iv). PEM or privacy-enhanced mail is a file format for storing and sending cryptographic keys and certificates. PEM files have a “.pem” suffix. A PEM file may contain multiple instances like a file containing an RSA certificate, a list of trusted CA certificate etc. Any organisation or company choosing to implement the above-mentioned system will have to make a choice between ease of access versus enhanced security, but considering it will mostly be used for private cloud

environments where only limited personnel should be able to access the cloud from a local area network, the latter choice of enhanced security becomes the clear choice. For data security, multiple cryptography encryption and decryption techniques are available. There are two major types of cryptographic techniques that is symmetric and asymmetric key cryptography. Symmetric key cryptography access one single key to encrypt at sender’s side and for decryption at receiver’s side. And inverse to that in asymmetric key cryptography, one key which is called as the public key is shared with everyone the sender uses the receiver’s public key to encrypt messages and the receiver use their own private key to decrypt the message. Since no transportation of the key is involved from within our cloud system, we decided to use symmetric key cryptography specifically the AES-256 encryption. AES is the most secure, efficient and widely used symmetric algorithm. The advantages of AES are that it provides strong security and fast encryption–decryption speeds. The AES algorithm consists of four steps.

5.1 Substitute Bytes

The first step, each byte of input data is replaced with another byte from the S-box substitution table (Fig. 1).

In the SubByte step, each byte is looked up in a fixed 8-bitlookup table and replaced with the subsequent entry (Fig. 2).

$$b_{ij} = S(a_{ij})$$

Fig. 1 Substitute Box (S-Box) *Source* https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
00	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
10	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
20	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
30	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
40	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
50	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
60	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
70	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
80	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
90	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a0	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b0	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c0	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d0	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e0	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f0	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

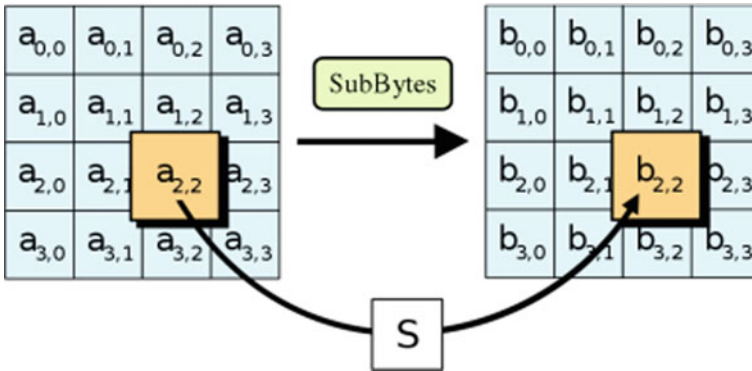


Fig. 2 Sub byte Source https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

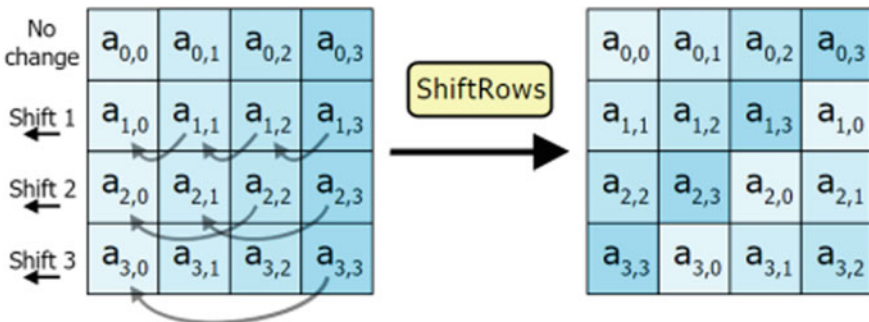


Fig. 3 Shifting rows Source https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

5.2 Shift Rows

In the second step, every byte in each row of the state is shifted in a cyclic manner to the left. The number of places each byte is shifted differs for each row (Fig. 3).

5.3 Mixing Columns

In the third mix columns step, each column of the state is taken and multiplied by a fixed polynomial $c(x)$ (Fig. 4).

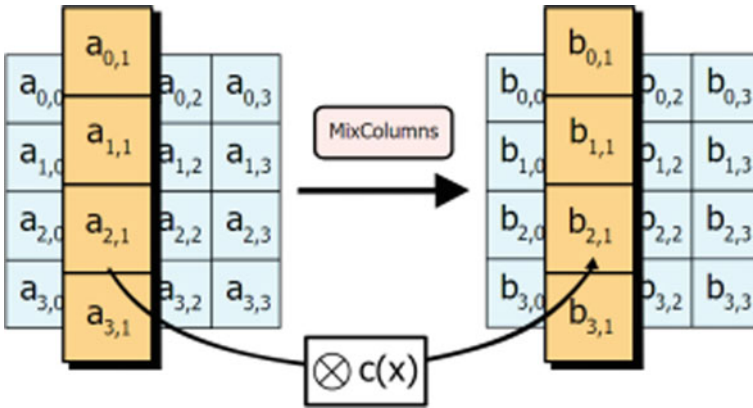


Fig. 4 MixColumn Source https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

5.4 The AddRoundKey

In the fourth add round key step, each byte of the state is XOR'ed with a byte of the round subkey (Fig. 5).

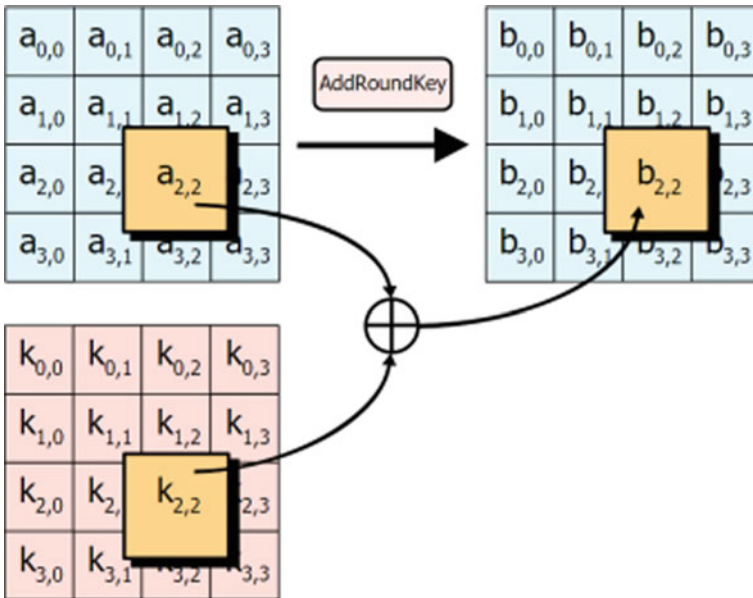


Fig. 5 Adding Round Key Source https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

The above four processes are repeated for 10, 12 and 14 rounds for 128, 192, and 256 key bits variations of the algorithm, respectively. We have decided to use AES-256 as we want maximum security.

To ensure additional security of data stored on the cloud, the concept of honeypot has been implemented. A honeypot is used to handle unauthorized access to files in a system by detecting and redirecting intruders to different locations. Inside the honeypot, the unauthorized intruders are presented with an illusion of data present at a location, which is not the actual location of the data on the cloud. The actual data is stored in encrypted form, in a different file system, which can only be accessed by authorized users of the cloud.

Honeypots can be classified based on two attributes that is on their methods of deployment or based on their level of interaction. Based on deployment, honeypots can be classified as (i) production honeypots and (ii) research honeypots, and based on the level of interaction classified as (i) high interaction honeypots (ii) low-interaction honeypots and (iii) pure honeypots. For our research, we have implemented a production and low-interaction honeypot. Production honeypots are easy to implement, capturing only limited information. Production honeypots are placed inside an organization's production network with other production servers to improve its overall security. Normally, production honeypots are low-interaction honeypots which simulate only the services frequently requested by at-tackers. Since they consume relatively few resources, multiple virtual machines can be hosted easily on one physical system itself, leading to shorter response times and requiring less code, thereby reducing the complexity of the system's security.

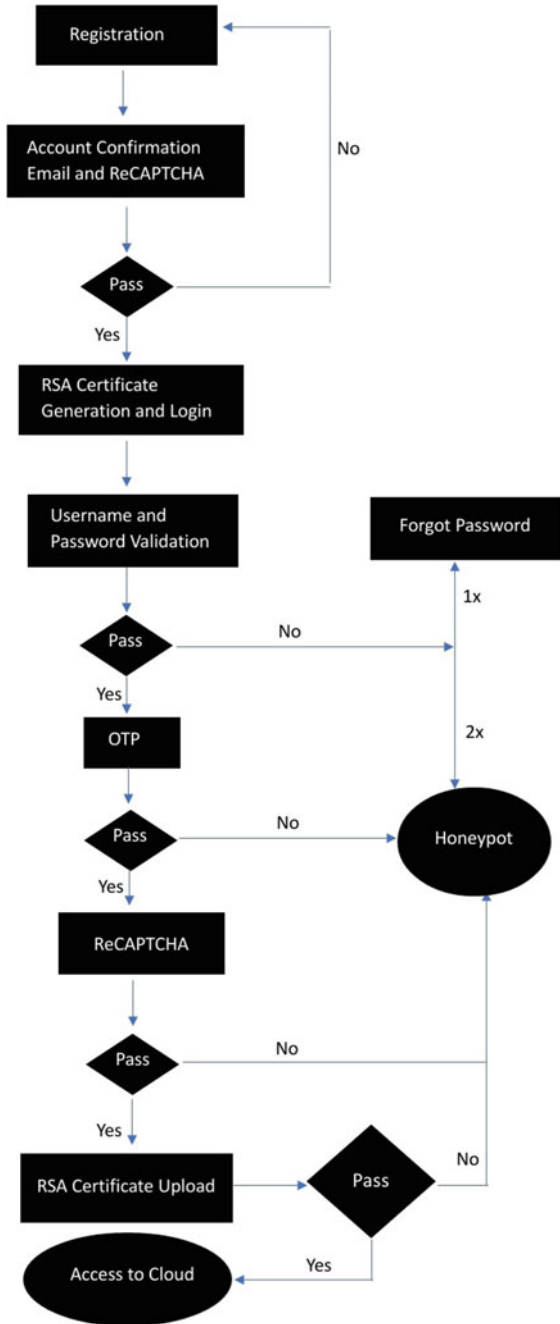
The honeypot's log file contains the IP address, time of intrusion and the duration for which the intruder accessed the contents of the honeypot. This can be modified to act as an intrusion detection system which will alert the cloud administrator in case of an intrusion. The IP address of the intruder could be used to track down the ISP, location, operation system used and other valuable information of the attacker. This setup allows us to ward off and track unwanted intruders.

6 Implementation

In our implementation, localhost on a local system is used to depict the working of the cloud. It provides the user with the resources similar to that of a cloud application hosted on a server and acts in accordance with the hardware and software specifications of a cloud environment. We have stored the static files of the application on the local server and files on Amazon S3 which fulfils the user's requirements like fast remote processing, dynamic resource allocation etc. and is therefore appropriate for demonstration purposes. Workflow of our methodology is depicted in Fig. 6.

We have created a Web application using Python, Django to serve as the interface of the cloud. We have implemented a private cloud environment. The functionality offered to users is registration, login, uploading files and downloading files, and users can also view and download files uploaded by other users.

Fig. 6 Workflow



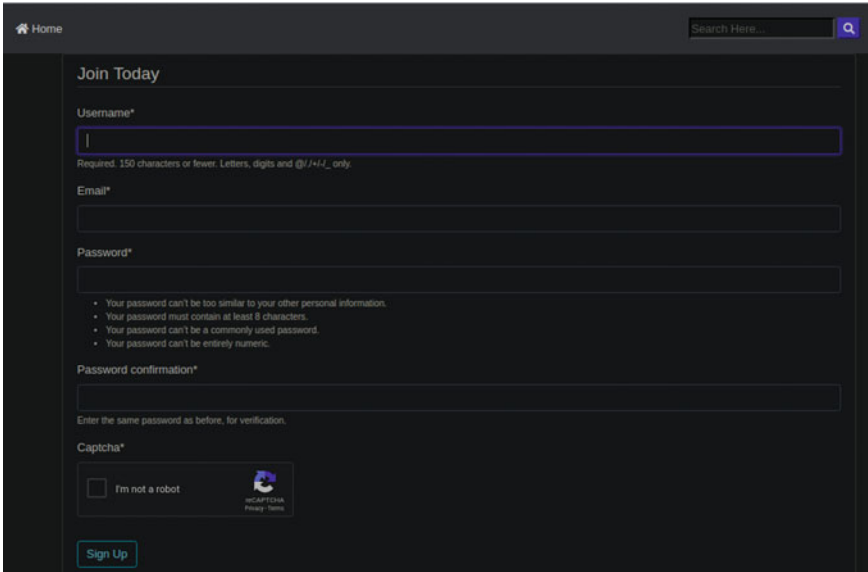


Fig. 7 Registration

First to use the cloud, a valid user approved by the organization using the private cloud has to register in the application.

For registration of new users, the user has to specify their names, email and choose a password which will be used to login later. We confirm that the email id provided is valid by sending an account confirmation email to the email is entered by the user on the registration page which contains a one-time passkey. We also have a ReCAPTCHA feature on the registration form to secure our website from bots that might be leveraged to perform a DDOS attack (Figs. 7 and 8).

On successful registration, an RSA certification file is created for the user which needs to be stored for future logins. After that, the user is redirected to the login page where they are required to login with their username and password first. If the password is typed wrong once, they are prompted the forgot password option, and if the user enters it wrong twice, the user is redirected to the honeypot directly (Fig. 9).

If the username and password are valid, the user is presented with the next layer of authentication. An OTP is sent to their email ID, which is to be filled on the authentication page (Fig. 10).

We again use a ReCAPTCHA feature to secure our website from bots, that might be leveraged to perform a DDOS attack which is the third layer (Fig. 11).

The fourth layer of authentication is the RSA certification file which was generated during registration. This RSA certification file is a unique identity for the user and should be stored in a secure manner (Fig. 12).

In case a user fails any of the steps involved in authentication, we devise a honeypot file system which looks identical to the cloud interface but with irrelevant content, that

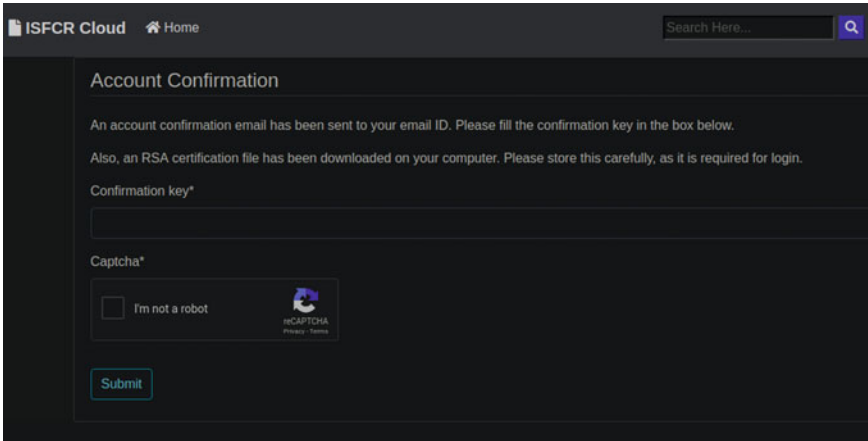


Fig. 8 Account confirmation

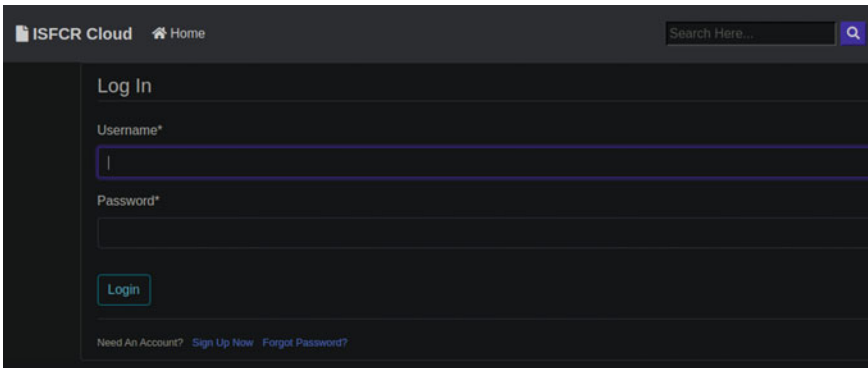


Fig. 9 Login

will be displayed to the intruder. Once the intruder is redirected to the honeypot, he will not be allowed to return to the main cloud interface. To implement the working of the honeypot, we have hosted the main cloud application on one terminal running on port 8000 and the honeypot on another terminal running on port 8080. The honeypot stores duplicate files which are hosted on a different S3 bucket will be displayed to the intruder. The honeypot offers limited functionality restricted only to viewing and downloading the garbage files. Other features such as upload, register and login are not available once directed into the honeypot. Figure 13 shows the honeypot home screen; the background has been set as such only to differentiate it for demonstration purposes, but in actual production, both the actual cloud home screen and honeypot home screen will look identical. The honeypot log file is used to find out the IP address and time of attack, which can be used to track down the intruder.

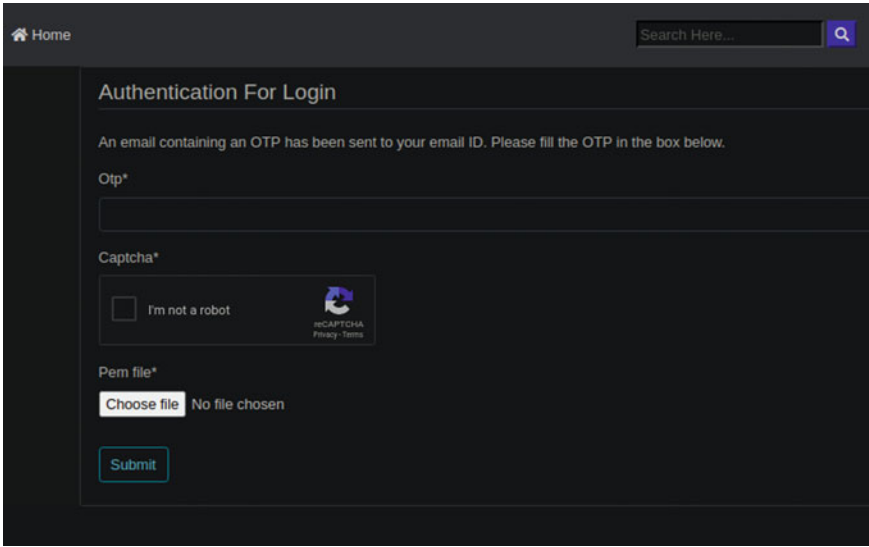


Fig. 10 Authentication

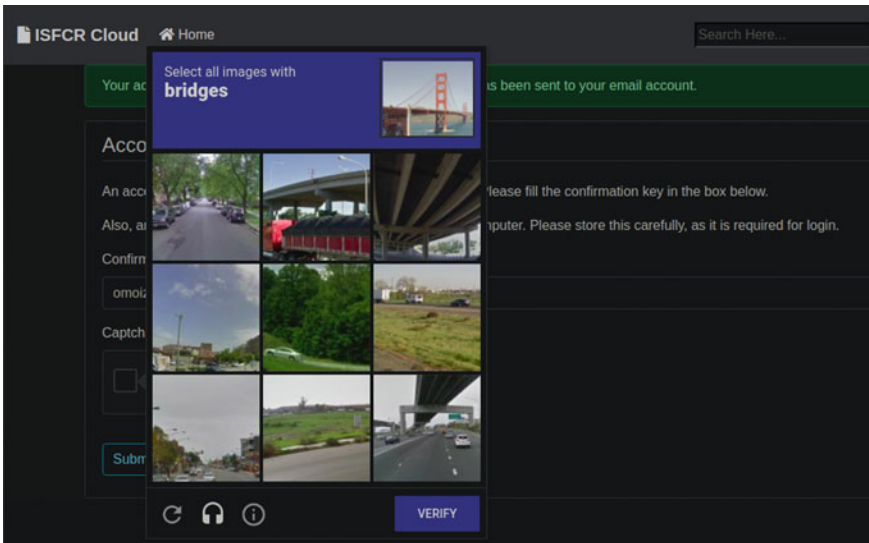


Fig. 11 ReCaptcha

Fig. 12 Pem/RSA Certification Key

```

user.pem
-----BEGIN CERTIFICATE-----
1  MIIFTCCA50CFGjphLbpcMMNuICUjL87ZnyKd2MA0GCSqGSIB3DQEBwCjUAMIGW
2  MQSwCQVDVQQGEWJTTJESMBAGA1UECAwJS2FybmF0YWhMREAYDVOQQHDLcYw5n
3  YWxvcmludDjAMBGNVBAoMBULTRkNSMRcwFYdVQQLDA5DbG91ZCBTZWNIcnleTEO
4  5MAWGA1UEAwFSVNGQ1IXJjAgBqkqhkiG9w0BCQEFW25vcnVwbHkuaXNMYzJAZz21h
5  6ahWuY29tMB4XDTIwMDYyNjA3MjkwN0XDTIxMDYyNjA3MjkwN0YyZWZyZXAzA3BGNV
6  8BAYTAkLOMRIeAYDVOQQIDALLYJuYXRha2E2eja0BgNVBACUJzhbmdhbG9yZTEO
7  8MAWGA1UEGwFSVNGQ1IXFzAV8BgNVBAsMDkNsb3V3IFNlYyY3YyYXRSMDQ4DAYDVOQQD
8  9DAVJU0ZDUjEmCQGCsqGSIB3DQEBQUAA4IC0wAwgIKAOICAQDnY20hsA0hJvZ/LZ/LUyMk
9  10ggTlMA0GCSqGSIB3DQEBQUAA4IC0wAwgIKAOICAQDnY20hsA0hJvZ/LZ/LUyMk
11 zG3z13SKeeUXnTymb17dLzmo+2ad1aKoR/VER0umCK08sY3usmoH/yCBTU0w+Ich
12 NbfRHLutZXsUANMv70BDJuzQ0lvcMMLCA0H1VKYU/p5FwIac008Ra0n2Fn9L5
13 tt2Zk8Bm7NCP32UxLZTz5ns1sLA35h/GwkSVC2eL27ADxh1Pvewa4Wj21Xz5Ml
14 bvnkk13/PR6NXZzB140QbfUwn6CBtEB2RkT5ho4mx4CZbqIsKqCFXarY1alNgrf
15 nVEPmsyYP9I/w8EX7L/hkpwD5HzHh3W1rukBLyAJzv5k3JvdzXL6HhywZw5u/3AET
16 ntYpZpWPpECSUSVQmEQzBoZJAJop539dfJtp17rdMCNRYJDC0oPuinQzCV0h5Qn7
17 laAatyVKL29ESIPRmdN70BS+PdnM6j5QLQLYy68dysGxXA6a/ndz4VY25zF19UWD
18 guIbjMNBwqdxEJtE5GjYUamb9a9umwJURam0Lknq4Jvz0WQzJw20jv0K0Bq40tNz
19 6rJ9TBr9Ho4c4aDuvqo7WkpyI8MV3JHG/CFw5K65NFHJLGT5ARyJh8BK7zdyM2ee
20 CLsEla+bfSytyc1dNuw404IhWjrdJj3X/Ln0wNRGFGj6ymkCd516gmBZCS57HK6JwZ
21 Qa0pxpHLCch+1Lpb4sL+DQIDAQABMA0GCSqGSIB3DQEBwCjUAMIGWGA1UEAzoIWI
22 ScorcNSznH0U/5NEwjge0/Y5SpkUzLVFVC2/V5eQ16l5wp38r+LSXyYe/QUH0DVY
23 +s04eCLyIEPX1yXBhyk+1/op3tkWDHsIDhAFS17hMVLub076Vq1InK0tpMK6H5a/
24 tJhcnYe8bnFzJYHE4SV+kTYcF5VJ2J7gs1oGfJhJGcxwf31bcX00F5nmBpB+Jjtm
25 wRGx8zgKFhXRa+2G25knL/r1rzfpp+L9BFr5S5QJr1ddo29T90WL9RTH5HFYE1tx
26 1RM7kEBw65pZMd0+qLk8t6mX5g9mYJ9uvnPDgrwY/I68cvDq2quyFGe00ZQuG
27 BLzn650kbX7PQJub52L36c7c01xLDHPRlpc2A+cty+21g1AYUTxGG00RTbMc0/
28 5Rl1K14K4/+206mCz/Rc4ZHTkFffTocTqPlf1pYb7LYnCFlohtBc+y/KZMPzdI3
29 t49tahu74k8VxTxHjvs57a0Me3xeJX9LKvd6YQp5bLJPTXI9b33yCh3c89Tx0Jh
30 XHNTENev+mICU1YAv0ueJxVQtXkKwnkabqccxCS5UAF82Q/a7o7nRkeruX8YRh8J
31 dMWL48SpIDCnWpJg+UtMcn0ncvGshq55t4tWld58G20kR1L20Z6zWRloyGeeX7G
32 5WjTmPUHdVYJApWl0PCKtz/F0UZsZL1aQ==
33 -----END CERTIFICATE-----

```

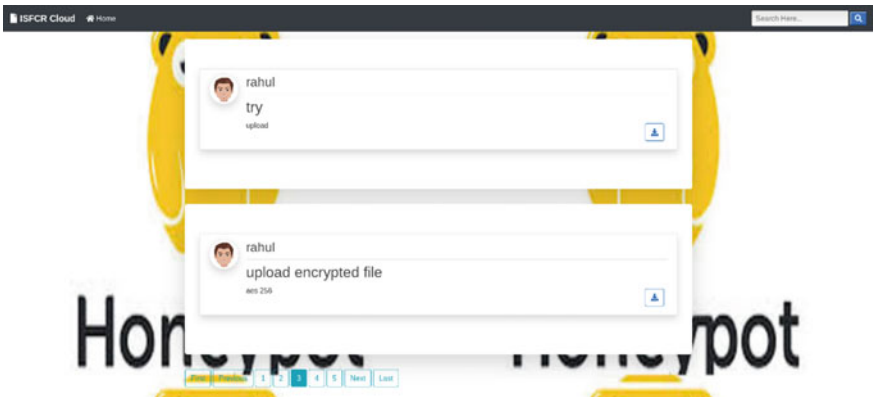


Fig. 13 HoneyPot home page

On successful login, the user is allowed to view and upload files. For file storage, we make use of an AWS S3 bucket with AES-256 encryption standard. Every time the files are to be uploaded or downloaded from the correct S3 bucket they are encrypted first with AES-256 encryption to prevent attacks like sniffing and spoofing and network-related attacks (Fig. 14).

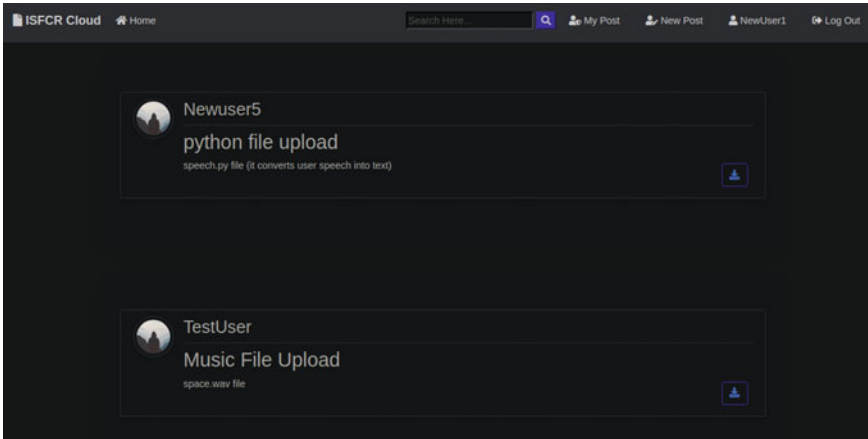


Fig. 14 Cloud home page

7 Conclusion and Future Scope

Security on the cloud is essential for businesses looking to store confidential information on cloud data storage. Our methodology attempts to implement maximum security precautions to access cloud data, using four-layer authentication with a password, OTP, RSA certification file and ReCAPTCHA. This ensures that only registered users can upload and download files on the cloud. We also encrypt the files using AES-256 encryption standard. AES is one of the most compact encryption techniques which consumes less storage space and has a short calculation time. AES-based cloud storage assures safe, flexible and privacy-based cloud user experience. Thus, we have proposed a model that implements maximum security measures for data storage on private cloud systems.

Future enhancements could include the concept of sessions and session timeout for login. Blockchain could be implemented to better store and secure files on the cloud. Vulnerability tests could be performed to identify potential pitfalls and handled adequately. In the future, hopefully it will be possible to achieve greater and even stronger security and lower the chances of security breaches compared to today.

References

1. Ms. Riddhi Mistry, Mr.Krunal Kantharia, Mr.Sandip Chauhan: A Collaborative Intrusion Detection System for Cloud Computing. JSRD- International Journal for Scientific Research Development— Vol. 1, Issue 4, 2013
2. Shouket Ahmad Kouchay: Data Protection in Cloud Computing-vulnerabilities, challenges and Solution. International Journal of Computer Trends and Technology (IJCTT) – Volume 34 Number 4 - April2016

3. Bih-Hwang Lee, Ervin Kusuma Dewi, Muhammad Farid Wajdi: Data Security in Cloud Computing Using AES Under HEROKU Cloud. The 27th Wireless and Optical Communications Conference (WOCC2018)
4. Kimaya Ambekar, R. Kamatchi: Performance Analysis of Various Encryption and Multilevel Encryption Techniques for Cloud Computing Security. International Journal of Computer Engineering in Research Trends Volume-5, Issue-4 ,2018
5. Boukari Souley and Hauwa Abubakar: A CAPTCHA – BASED IN-TRUSION DETECTION MODEL. International Journal of Software Engineering Applications (IJSEA), Vol.9, No.1, January 2018
6. Rahul Midha1, Simran Khajuria, Suman Samal, Aishwarya Mishra: Secure Data Protection. Cloud Computing International Research Journal of Engineering and Technology Volume: 06 Issue: 04 — Apr2019
7. Himanshu V. Taiwade, Bhagyashree T. Kaley, Aditya R. Raoot: Three Level Advanced Cloud Computing Security Mechanism. IJRECE VOL.7 ISSUE 1 (JANUARY- MARCH 2019)
8. S. Kumari, Princy, Reema, and S. Kumari: Security in Cloud Computing using AES DES. Int. J. Recent Innov. Trends Comput. Commun., vol. 5, no. 4, pp. 194–200, 2017
9. Sighom JRN, Zhang P, You L (2017) Security Enhancement for Data Migration in the Cloud. Secur. Enhanc. Data Migr. Cloud 9(23):1–13
10. Sonu Devi, Dr. Anuj Kumar Sharma: Understanding of Intrusion Detection System for Cloud Computing with Networking System. IJCSMC, Vol.9, Issue. 3, March 2020

Intrusion Resilience Analysis of Smart Meters



C. S. Sandhya, S. Nagasundari, and Prasad B. Honnavalli

1 Introduction

A smart meter is a device that records electrical energy consumption and sends this information to the supplier who monitors and also for billing. These instruments record energy per hour or more times per set and report fewer readings every day. This device enables two-way communication between the central system and the meter. Communication from the network to the meter can be wireless through any fixed wired connection such as a power line carrier. Various wireless communication options include the use of cellular communications, wireless ad hoc networks over Wi-Fi, low-power long-distance wireless, Wi-sun, ZigBee, wireless mesh networks, and Wi-Fi. The development of smart meter systems is gaining momentum worldwide, leading to similar deployments to realize the benefits.

1.1 Why Smart Meter?

A smart meter can identify behavior and patterns and track data usage. It has the ability to remotely connect or disconnect the consumer from the power grid. Interfaces with in-house smart devices which shall control the operation of the device during peak hours support the generation of on-site renewable energy and in turn export this energy to the grid using a net-metering agreement. It can alert the utility company immediately if there is a problem.

C. S. Sandhya · S. Nagasundari (✉) · P. B. Honnavalli
PES University, Bengaluru, India

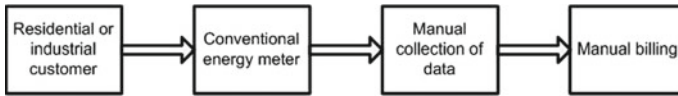


Fig. 1 Flow of conventional power meter [11]

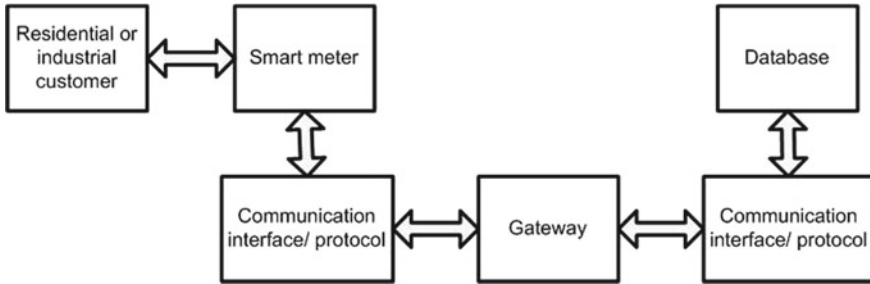


Fig. 2 Flow of smart meter [11]

1.2 Power Meters

In general, the electricity meters are traditional meters, wherein the energy consumption readings of individual households or industries are collected manually every month as shown in Fig. 1. Some disadvantages of traditional meter readings may be the occurrence of errors due to faulty readings that are forwarded by the meter reader. Monitoring energy consumption becomes difficult if meter readings are not accessible. A large time lag exists between recording meter reading, administration, delivery of accounts, and payment. The process of turning off power due to non-payment is difficult (Fig. 2).

1.3 Smart Energy Meter

The smart meter communicates with the supplier through a wireless communication medium. The device also communicates with in-house displays that provide consumers with details of energy consumption. It provides detailed feedback regarding energy use. It can adjust habits to reduce electricity bills. This reduces the number of massive blackouts and system-wide power failures. Vulnerability comes with the advantage of smart meters. The device being wireless for communication brings many concerns between consumer communities and service providers regarding information security.

For a system to be considered safe, it must meet four security requirements: confidentiality, integrity, availability, non-repudiation. Confidentiality is the prevention of unauthorized access, and securely communicate information such as power usage,

price information, and control orders that intrude consumer's privacy, revealing important information of utilities. In order to maintain integrity, prevent the critical information from being modified through sensing devices, electronic devices (e.g., smart meters), software, and control commands that can impede decision-making and corruption of data exchange in smart grids. Unlike the confidentiality situation, the integrity of the software must be kept critical as an adversary can control any device or electrical equipment through compromised software. Availability defines the prevention of an adversary from disallowing authorized personnel from using or controlling the system. Denial-of-service (DoS) attacks and distributed DoS (DDoS) attacks can delay block or corrupt information on the unavailability of power or exchange of information in a smart grid. In this case, the availability of control and price information is important because it can lead to revenue loss. Non-repudiation is an assurance that a person cannot deny the validity of something. Non-repudiation is a legal concept widely used in information security and refers to a service, which provides proof of origin and integrity of data. In other words, non-repudiation makes it very difficult to successfully reject who or where the authenticity and integrity of that message came from as well. Most security issues can undermine at least one of these security requirements. An attacker may be able to recreate the form of a valid meter data management system, and thus be able to gain access to confidential information, allowing the attacker to change control commands, denying access to the legitimate system, and cannot receive critical data to the system. The most important aspect regarding the safety of smart meters is network communication. This is because the smart meter must send data to the utility company and the data must not be manipulated during transmission to the smart grid.

In this paper, a smart meter testbed has been built, and each of the potential vulnerabilities has been analyzed systemically. Since the smart meter is deployed in large mass, its production cost is maintained low which directly influences its simplistic security measures. Furthermore, as it is a newly developing product, the availability of experimental testbeds for smart meters is comparatively less. Thus, there exists a need for low-cost testbeds before deploying the smart meters in the government pilot projects. These testbeds provide a suitable environment so that the various vulnerabilities associated with the smart meter could be explored. While cognitive radio is an emerging technique among wireless communication, Wi-Fi communication interface has been preferred for experiments mentioned in this paper.

A critical requirement of the smart meter is to enable accurate recovery of wireless transmissions at the central node or access point (AP). One of the impending challenges in this objective is the robustness of the data recovery in the presence of strong wideband interference which is caused due to easy access of the wireless data to any attacker, and inadequacy of existing physical layer security measures. Hence, power consumption dataset has been collected from IIT Bombay forum to analyze the usage of power in a household on a daily basis. In the process, random noise is injected and fluctuations in power consumed has been studied.

The key issues that are addressed in the paper are based on the following three use cases. The first one being measurement is about the smart meter's reading calculation regarding the usage of energy consumption. Corresponding threat to this use case

includes modification of parameters of the source code of the smart meter. The second component addressed is with respect to communication. Smart meter communicates usage information to AMI periodically. Corresponding threat to this use case includes flooding network lines with high frequency noise which in turn disconnects the network line. The third use case considered is remote access. Smart meter listens to remote commands from the AMI and act accordingly. Corresponding threat to this use case includes an MITM (man-in-the-middle) attack causing corrupted data to be transmitted or the smart meter taking remote instructions from an unauthorized entity.

The rest of this paper is organized as follows: Sect. 2 provides the background concepts and the related work. Section 3 narrates the design and working of smart meter testbed. The various potential threats on smart meter are discussed in Sect. 4. In Sect. 5, the study of attacks is described based on the results obtained out of experimentation. Finally, concluding remarks and future works have been addressed as a part of Sect. 6.

2 Background and Related Work

2.1 Types of High-Level Architecture

There are two types of communication architectures, point-to-multipoint and mesh architecture. With respect to point-to-multipoint architecture, neighborhood area network (NAN) is a collection of Wi-Fi hotspots and wireless local area network to enable users to connect to the Internet quickly. A NAN is generally installed by an individual to serve a family or a number of neighbors. WAN is a data network, usually used for connecting computers, that spreads across a wide geographical area. WAN can be used to connect cities, states, or countries. LAN is mainly a physical wire used to connect two devices. Data concentrator otherwise known as an access point is the nearest tower that the smart meter connects to. The access point has a transceiver which is used to send commands to smart meter from AMI and sends back requested data from smart meter to AMI.

In point-to-multipoint topology (star topology), each smart meter is connected to a data concentrator (router) making it a hub that in turn connects to the AMI supplier. Some of the pros are since each smart meter is independent of the other, an attack on smart meter shall have no effect on the other smart meters. And cons include single point of failure (Fig. 3).

In mesh topology, each smart meter is connected to every other smart meter. Only a certain number of smart meters communicate with data concentrators which may also be a mesh topology. Pros include no single point of failure and cons include smart meter being compromised; it becomes easy to gain access to all other smart meters (Fig. 4).

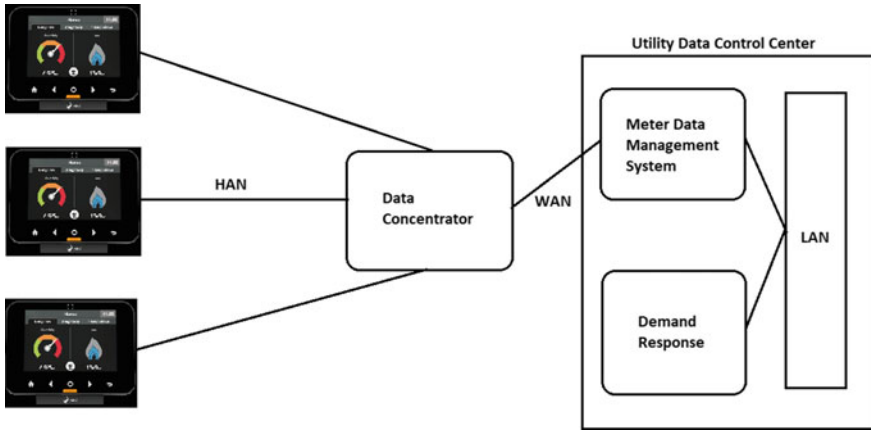


Fig. 3 Point-to-multipoint architecture

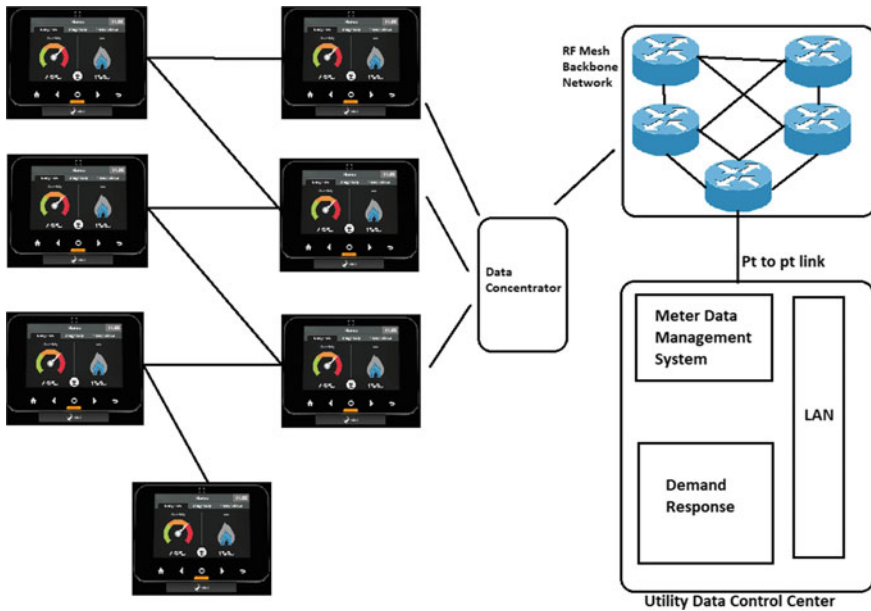


Fig. 4 Mesh architecture

3 Design of Smart Meter Testbed

Smart meter is a one-way communication device where it takes and responds to the commands from the AMI. Fig. 5 demonstrates the design of the smart meter testbed with the relevant components. The meter data management requests for the

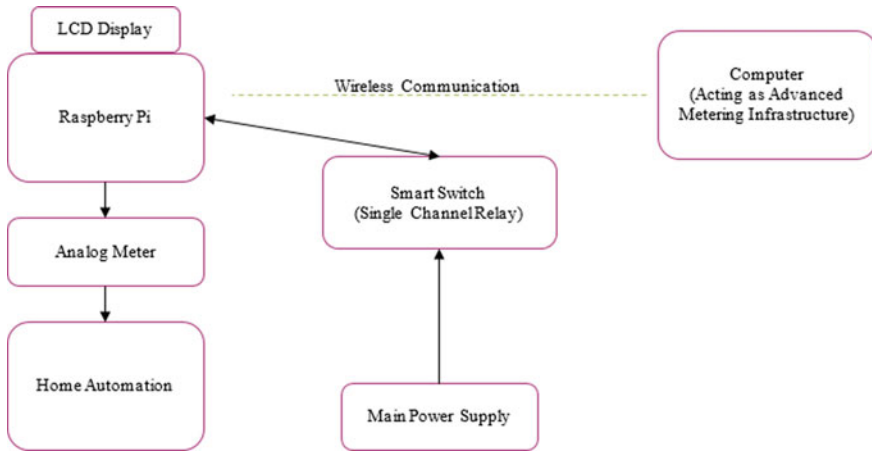


Fig. 5 System architecture of smart meter environment

metering data from time to time. This request signal is sent to the specific metering device through the nearest access point or data concentrator. The smart meter on receiving this request replies to the AMI with the latest meter reading which is sent back to the meter data management through data concentrator. Henceforth, this latest metering data is now forwarded to the pricing team which acknowledges the receipt of metering data. For the metering data received, the pricing team responds with the corresponding pricing which is then forwarded to the consumer.

3.1 Working of the Smart Meter Testbed

The testbed is created depicting the AMI and smart meter as per Fig. 6. The Raspberry Pi acts like the brain of the smart meter environment and wirelessly connects to the genuine laptop which acts like the advanced metering infrastructure providing

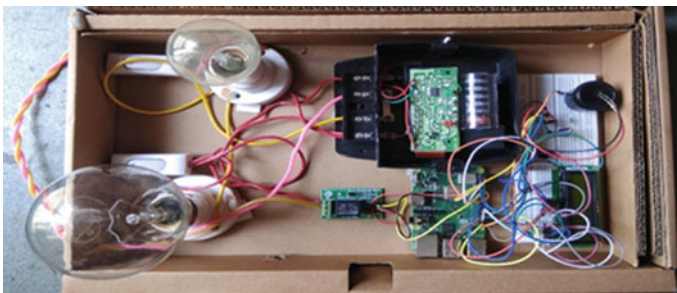


Fig. 6 Smart meter environment

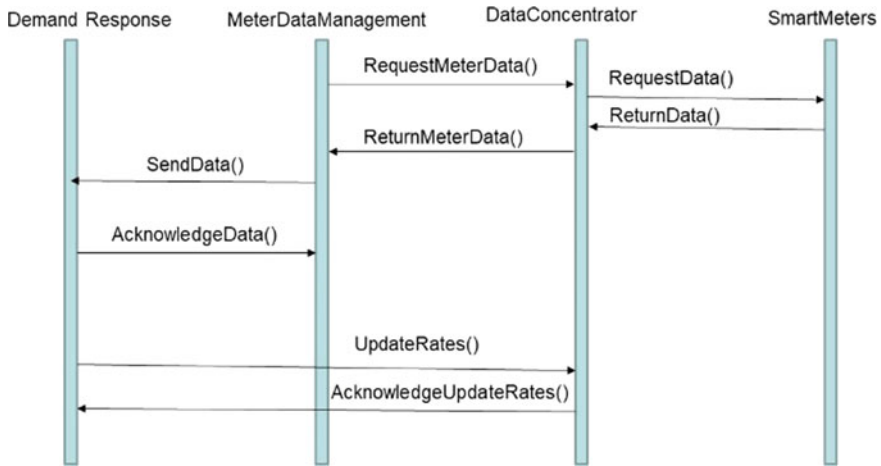


Fig. 7 Sequence flow of smart meter

commands to the smart meter. The single-channel relay acts like the heart of the smart meter testbed controlling the entire home automation system on the command of Raspberry Pi. The relay takes 240 V of power supply as input and converts it to 5 V DC output.

A 16X2 alpha-numeric LCD screen acts like the in-house display and provides information about the total cost with respect to the power consumed. Analog meter is connected to both the relay and the home automation system. Analog meter and the home automation get activated on the command of Raspberry Pi through relay. The LED light mounted on the analog meter blinks 3200 times which is equal to 1KWh. An LDR (light-dependent resistor) senses these blinks and sends the information to the Raspberry Pi which in turn processes the collected information to define the associated cost. Fig. 7 illustrates the sequence of actions that are initiated in the smart meter testbed.

4 Analysis of Potential Threats on Smart Meters

4.1 Component-Wise Threat Analysis

The smart meter components are depicted in Fig. 8. The first and foremost threat is with respect to the flash memory. Once on successfully gaining access to flash memory, one can retrieve secrets such as passwords, keys, and certificates. One can inject malicious code and also erase security logs. In case of analogic and application processor, post gaining access to analogic and application processor may happen so that the attacker can exploit the vulnerabilities of firmware. The attacker can easily

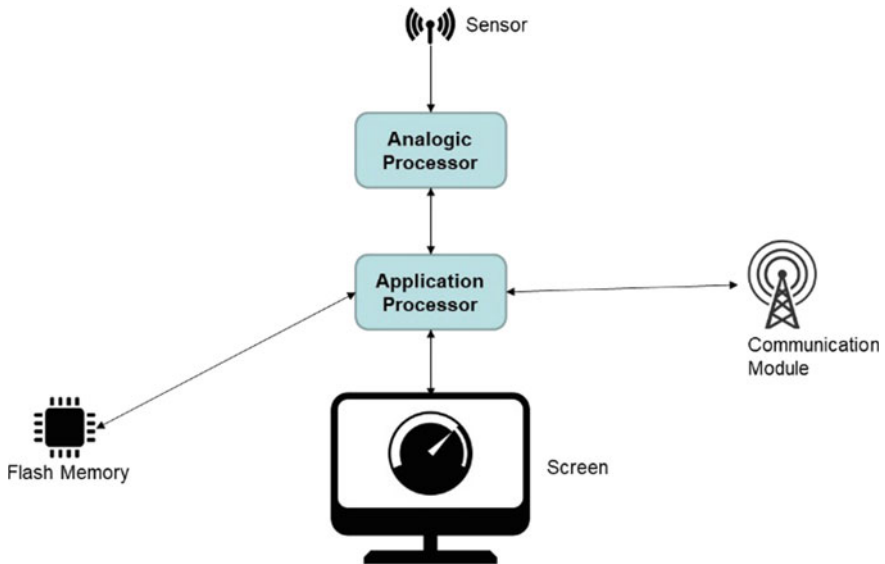


Fig. 8 Anatomy of smart meter

intercept networks if the port is open. One can easily gain access to the firmware and flash memory. Using the patched code, the attacker can modify chip logic causing the smart meter device to malfunction. The smart meter collecting system uses the radio system to communicate between the device and the supplier. It sends real-time data consumption readings and takes operational commands from data concentrator. Intercepting network communication between the device and supplier can result in disruption of power grid, denial of power usage. The attacker may also gain access to vital user information which might include personal information, payment information etc.

A home area network is a dedicated network for connecting devices in the house such as load control devices, displays, and smart appliances seamlessly into the overall smart metering system. Also, it contains software applications to control and monitor these networks. The HAN is a system within the smart grid, dedicated to demand response management (as mentioned in the sequence diagram Fig. 7), includes energy efficiency and demand response which are the key components in deploying of a smart grid. It facilitates communication between device and in-house display and intercepting can reveal critical user data. Optical interface provides physical access for service engineers for installation and maintenance of energy meters. Using physical probing, the logged information can be changed which shall again result in denial of power usage, theft etc.

4.2 Possible Potential Attacks on Smart Meters

There are various types of potential attacks that might affect the performance and authenticity of smart meters affecting the electricity suppliers as well as consumers.

4.2.1 DOS Attack

Flooding the target system with ICMP-Echo requests/Half-Syn packets to consume its buffer, thereby denying its service intended. The purpose of this attack is to launch a DOS attack on device and disconnect the device completely. As an effect, it can completely isolate the device from the network. Then, it may temporarily disconnect the system in order to gain connection with the victim. When the victim tries to reconnect, it may disconnect smart meters from the access points causing massive blackouts.

4.2.2 ARP Cache Poisoning Attack

Here, the attacker runs an arp spoof command to spoof itself as the legitimate router, thereby tricking the victim to forward packets to the attacker. The purpose of this attack is to launch a man-in-the-middle attack to intercept messages between sender and receiver. As an effect, it can pose to be a legitimate router and receive all the forwarded packets from the victim and then drop all the packets. It may successfully launch a MITM attack and intercept communication between victim meter and supplier.

4.2.3 Black Hole Attack

This is a type of active attack which sends packets to all the nodes connected in the same phase posing to be the shortest path to all the nodes and henceforth tricking the devices to forward the packets to it rather than to the access point and later dropping the packets. Its main purpose is to completely disconnect the smart meter device from the access point. By launching black hole attack, the attacker may successfully gain access to smart meters, and the device can be completely disconnected from the access point. This can cause a power cut to a specific residence. On mass launching, i.e., by launching this attack at once on various smart meters in a particular region that links a given access point, a massive blackout can be made possible.

4.2.4 Gray Hole Attack

It is a combination of active and passive attack that works as same as black hole attacks but forwards packets to the access points instead of the devices causing an MITM attack. Its purpose is to completely disconnect the smart meter device from the access point or passively either watch or intercept the movement of packets and manipulate the same. As an effect, the device can be completely cut-off from the access point. This can cause a power cut to a specific residence. But this is an active–passive attack as mentioned, and henceforth, the purpose of this attack is to mainly eavesdrop the network mostly in a passive manner.

4.2.5 Replay Attack

This is a type of MITM attack which intercepts the network, connecting both the sender and the receiver. Fig. 9 demonstrates the attacker posing to the sender to be a legitimate receiver and to receiver as the legitimate sender and can lead to stealing of confidential user data as well as modify the data before retransmission. Smart meters are highly vulnerable to replay attacks as there is no time log of packets sent and received. The purpose of this attack is to passively intercept the movement of packets and manipulate the same if needed. As an effect, the attacker can successfully get into the network when the smart meter and supplier are communicating. Posing as a legitimate supplier to the consumer redirects all the data to itself and pose itself as the smart meter to the supplier, pushing the data to the supplier. Using this access,

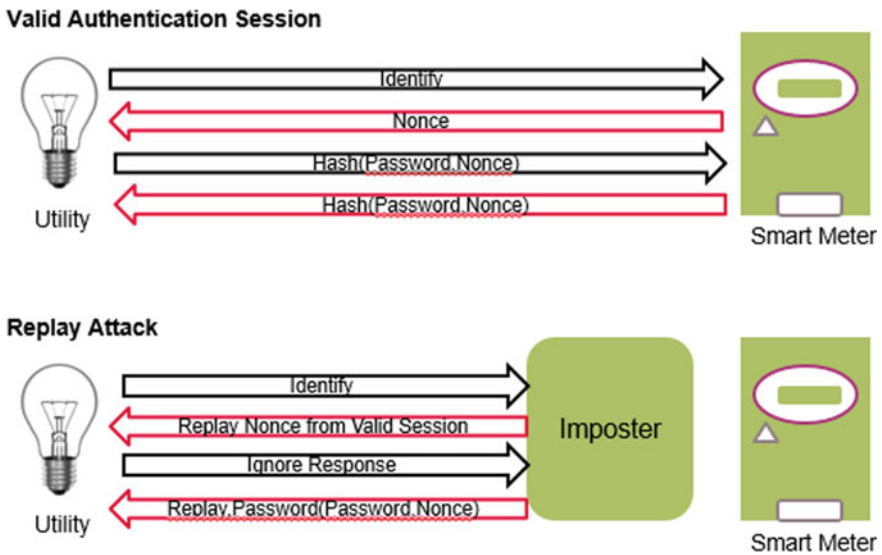


Fig. 9 Replay attack

the attacker can vary the reading and make the supplier quote a low price than the actual. The attacker can also vary the reading both ways showing the consumer and supplier different readings. The attacker can successfully get all the user personal details and also gain access to all the payment details.

4.2.6 Flood Repeater System Attack

The remote nodes listen to the beacons coming from the central node. The remote node does not hear any beacons when the central node is far off from remote node. It alerts the network about the same and request the nearest node to behave as a repeater between the node and the central node. This technique could be exploited by an attacker. On gaining access to multiple nodes and sending repeat requests, the central node can be tricked to think too many nodes in the network to be having poor connectivity which in turn grants permission for repeat request making the topology complex. The purpose of this attack is to flood the network or complicate the topology. As an effect, the attacker can pose to the supplier that multiple remote nodes have poor connectivity, forcing the central node to accept the repeater request on the other nodes and thus hijacked by the attacker.

5 Results and Discussion

The experimental setup considered four devices for creating the entire testbed. While one device (Windows 10 OS) acted as the advanced metering infrastructure, the other device (Kali Linux) acted as the attacker. Router was the common access point, and the constructed environment is shown in Fig. 5 to replicate the functionalities of smart meter. The purpose of this experiment was to study the behavior of smart meters by creating an environment similar to smart meter environment and test attacks that might turn out to be a potential vulnerability in case of smart meters.

Following are the discussions relating the various attacks in smart meter testbed. Considering the case of cracking access credentials, the main motive of testing this attack was to analyze the severity of the after effects of DOS attack. A smart meter usually communicates over ZigBee, but for testing purposes, Wi-Fi network communication medium is used. This attack is performed to demonstrate the DOS attack on smart meter which can completely disconnect the system causing a power blackout. When the attacker gain access to the network, man-in-the-middle attack is enabled. For this attack, a Kali Linux tool called Aircrack-ng is used. The de-authentication flood attack purpose is to cause a flooding or high frequency noise in the network to disconnect previously connected devices.

The graph in Fig. 10 represents the network traffic variation caused due to DOS attack. When the AMI is flooded with packets, it takes exponential amount of time to process the packet. As the incoming packet rate is much higher than the time taken by AMI to process each packet, it causes the device to disconnect from the network

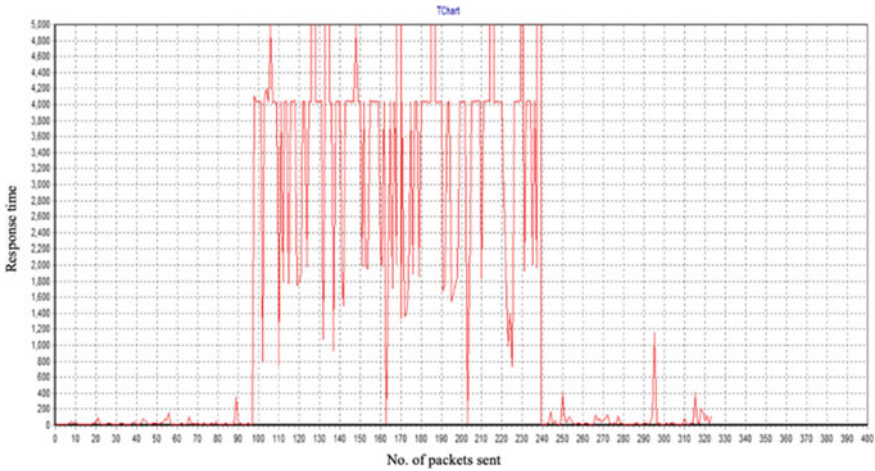


Fig. 10 Graph representing DOS attack

causing denial of service. The simultaneous access of smart meter is performed by two different devices (genuine and attacker device) to check the behavior of the system. Initially, the genuine device executes the smart meter source code, and the LCD screen acts like an in-house display, displaying the energy consumption and its cost at the user end. The attacker tries to execute the same code from Kali setup. This simultaneous execution causes one of the two scenarios—the LCD display either displays garbage values or displays refreshed values.

The ARP cache poisoning attack purpose is to showcase that on corrupting the ARP cache entries, and launching a DOS attack, the smart meter can be completely disconnected from its access point. Henceforth, when the smart meter device tries to re-establish connection, the attacker can pose himself as the nearest access point tricking the device to send the data to it. As a part of this attack, post a ping from attacker to victim. Device-1 (Victim’s System) lists the attacker’s MAC address and IP address that is communicated. On device-2 (Attacker’s System), attacker launches an arpspoof attack and re-routes the packet forwarding to itself from the router. While the arpspoof attack is in progress on the attacker’s system, the ARP cache table on device-1 shall show the attacker’s MAC address to be the router’s MAC address. Hence, the victim device forwards all its packets to the destination via attacker’s device. Once the ARP spoofing is successful, Kali device can restart the meter and the user gets to see the updated values while the genuine device (AMI) is unaware of this and continues to display the old values.

The parameters in the source code are altered to showcase the maximum benefit that the attacker can achieve through intrusion. In this experiment, attacker changes the parameters of the core program that takes blinks as input to calculate power consumption and also calculates relevant costs. In this attack, attacker tries to update the source code, and as a result, the values cause severe discrepancies.

As a part of smart meter power usage analysis, a Web application was developed which takes the malicious data and benign data as input and gives a comparison graph for a specified time span as depicted in Fig. 11. The green line signifies the malicious data readings, while the red line signifies the benign data readings.

Some of the suggested mitigation techniques include building a sniffer tool to monitor all the incoming traffic in the network to detect the flood (DOS flood, ARP flood) attack, incorporating a lock while executing the programs to maintain atomic executions on a device which avoids the simultaneous execution scenarios, and forcing a version control from the backend on source code that assist in tracking down the alterations to avoid discrepancies in values. Also, password protecting source code helps in avoiding any further alterations.

6 Conclusion and Future Work

As a part of this paper, the effects of various information-security vulnerabilities on the performance of smart meter device components were investigated. Cracking access credentials, flooding network with high frequency noise, simultaneous access on smart meter, gaining unauthorized access to source code and detailed analysis were performed on the dataset to plot graphs to detect suspicious activities. Results from virtual experiments depicted the lack of security functionalities in smart meter. The result analysis also revealed the requirement of intrusion detection, firewall-based packet filtering, or prevention mechanisms. The smart meter resiliency has to be enhanced by incorporating the security functions such as intrusion detection, prevention, encryption capabilities and network packet into the device. In the future, the mitigation techniques on smart meters will be devised using machine learning approach to formulate intrusion detection and prevention strategies.

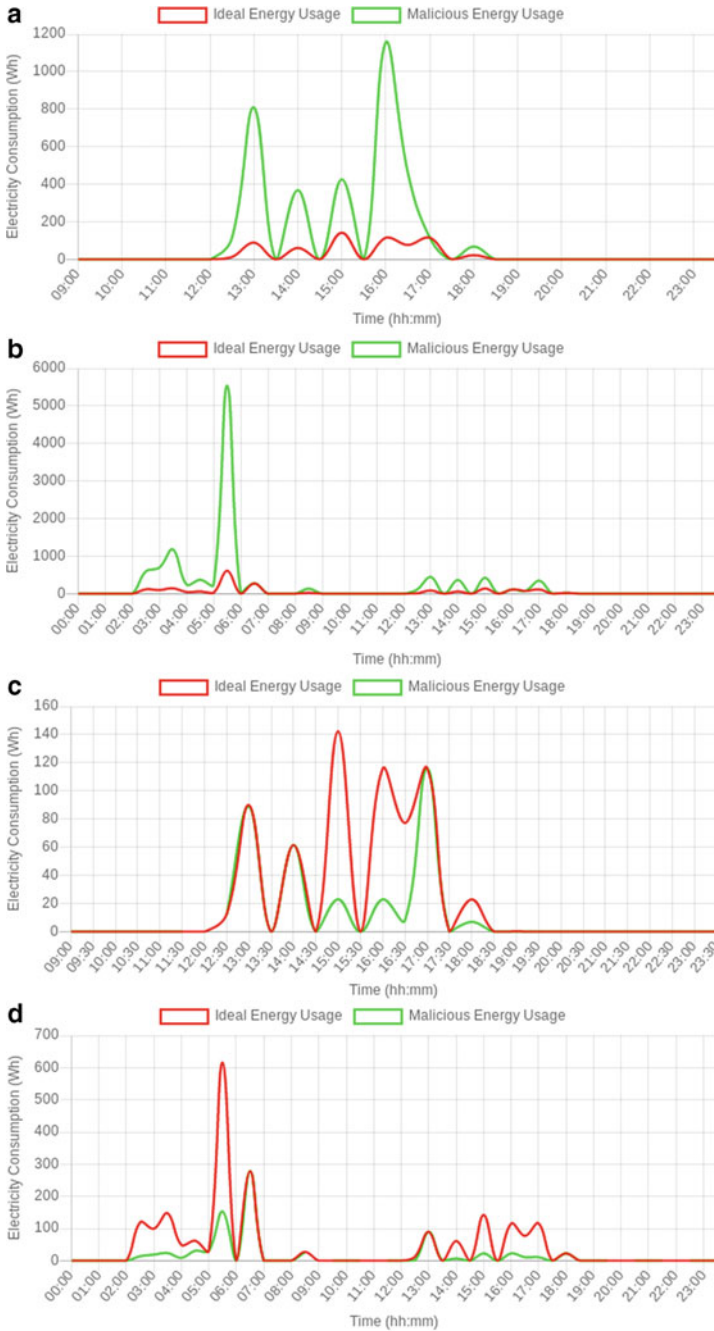


Fig. 11 **a** Comparison graph for malicious dataset-1 [20] and benign data for a time period between 9am and 11 pm, **b** Comparison graph for malicious dataset-1 [20] and benign data for a time period of 24 h, **c** Comparison graph for malicious dataset-2 [20] and benign data for a time period between 9am and 11 pm, **d** Comparison graph for malicious dataset-2 [20] and benign data for a time period of 24 h

References

1. Karimipour H, Dehghantanha A, Parizi RM, Choo KKR, Leung H (2019) A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. *IEEE Access* 7:80778–80788. <https://doi.org/10.1109/ACCESS.2019.2920326>
2. Prasad, G., Huo, Y., Lampe, L., & Leung, V. C. (2019, October). Machine learning based physical-layer intrusion detection and location for the smart grid. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (pp. 1–6). IEEE, pp. 1–6, doi: <https://doi.org/10.1109/SmartGridComm.2019.8909779>
3. Khattak A.M., Khanji S.I., Khan W.A. (2019) Smart Meter Security: Vulnerabilities, Threat Impacts, and Countermeasures. In: Lee S., Ismail R., Choo H. (eds) Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM) 2019. IMCOM 2019. Advances in Intelligent Systems and Computing, vol 935. Springer, Cham. https://doi.org/10.1007/978-3-030-19063-7_44
4. Swetha RBS, Meena KG (2015) Smart grid-a network based intrusion detection system. *International Journal of Computer Applications* 975:8887
5. Zabih, M., Tavasoli, M., & Yaghmaee, M. H. (2015, June). Analyzing the Effects of Various Attacks on Smart Meters. In *Accepted in International Conference and Exhibition on Electricity Distribution*
6. Halim F, Yusoff S, Rusli ME (2018) Cyber security issues in smart meter and their solutions. *International Journal of Computer Science and Network Security* 18(3):99–109
7. Shuaib, K., Trabelsi, Z., Abed-Hafez, M., Gaouda, A. M., & Alahmad, M. (2015, June). Resiliency of Smart Power Meters to Common Security Attacks. In *ANT/SEIT* (pp. 145–152)
8. Sankaranarayanan S (2017) Performance Analysis of Security Protocols in Smart Energy Meter System. *Int J Appl Eng Res* 12(19):8294–8315
9. Pandey, R. K., & Misra, M. (2016, December). Cyber security threats—Smart grid infrastructure. In *2016 National Power Systems Conference (NPSC)* (pp. 1–6). IEEE
10. Chockwanich, N., & Visoottiviseth, V. (2019, February). Intrusion Detection by Deep Learning with TensorFlow. In *2019 21st International Conference on Advanced Communication Technology (ICACT)* (pp. 654–659). IEEE
11. Oskan, S., Yildirim, E. N., Karatas, G., & Cuhaci, L. (2019, April). Intrusion detection systems with deep learning: A systematic mapping study. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1–4). IEEE
12. Papamartzivanos D, Mármol FG, Kambourakis G (2019) Introducing deep learning self-adaptive misuse network intrusion detection systems. *IEEE Access* 7:13546–13560
13. Kang, M. J., & Kang, J. W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. *PLoS one*, 11(6), e0155781
14. https://www.researchgate.net/figure/DDoS-attack-in-AMI-communication-network_fig1_268816940
15. https://www.researchgate.net/figure/Smart-meters-corrupted-ARP-cache_fig3_277727331
16. https://www.researchgate.net/figure/Corrupted-ARP-caches-for-performing-MiM-attack_fig4_277727331
17. https://www.pngkey.com/detail/u2t4a9r5r5q8y3r5_black-hole-attack/
18. https://www.researchgate.net/figure/The-replay-attack-discovered-in-the-studied-system-Because-the-two-messages-in-the_fig3_220816508
19. <https://www.tarlogic.com/en/blog/smart-meters-threats-and-attacks-to-prime-meters/>
20. <https://www.smartenergygb.org/en/about-smart-meters>
21. <https://sse.co.uk/smart-meters>
22. <http://seil.cse.iitb.ac.in/residential-dataset>

UCREDIT—Credit Scoring Using Social Media Platforms



M. Deeptha , G. S. Arpitha , S. D. Anjali , and D. B. Srinivas 

1 Introduction

For the vast majority, access to credit is a significant necessity for financial freedom and upward mobility. A positive Fair Isaac Company (FICO) assessment is imperative to get a home or vehicle, to endeavour into business, to seek employment or to attain different objectives. At present, however, people have almost no influence over how they are scored and even less capacity to challenge incorrect or biased evaluation of their credit. Conventional credit scoring software tools raise longstanding worries of accuracy and injustice. Billions of individuals need financial record and will get dismissed utilizing customary scoring strategies. Our study is motivated to broaden opportunities by using social network-based credit scoring and financing practices for a larger portion of the population and to benefit low-income individuals who would otherwise find it hard to obtain credit. We use alternative data sources including mobile phone, digital footprint, behavioural, and psychometric to assess the credit risk of anyone. We add minimal friction and maximum predictive power, providing customers more approvals and fewer defaults. An individual's social conduct and language can mirror the attributes of their conduct, which might be utilized as credit information. On the Web, the conduct and language of clients are frequently gotten from web-based life. Web-based life alludes to site and innovation that permit individuals to record, share, assess, examine, speak with each other, including long-range interpersonal communication locales, smaller scale blog, twitter, miniaturized scale letters, sites, discussions, etc. Presently, internet-based life has secured most parts of our life, through which we will know the freshest improvements of an individual.

M. Deeptha (✉) · G. S. Arpitha · S. D. Anjali · D. B. Srinivas
Nitte Meenakshi Institute of Technology, Bangalore, India

D. B. Srinivas
e-mail: srinivas.db@nmit.ac.in

Social information [1] is generally helpful for individuals with almost no financial record. An individual's social character, online notoriety, and expert contacts circle, which should turn into a component inside the appraisal of credit chance. The main contribution of this paper is to highlight social behaviour patterns into the typical credit scoring model, which are not considered in earlier exploration of credit scoring models. An individual's social character, online reputation and expert contacts circle, which should transform into a segment inside the examination of credit possibility. The principle contribution of this paper is to feature social behaviour patterns of into a credit scoring model, which are not considered in earlier exploration of credit scoring system.

We utilize the overall public information of borrowers on the platform to build the credit scoring model including the personal information, loan information, and social media footprints. The exploratory outcomes show that the credit scoring models can well recognize the default and the diligent clients, and it has higher forecast accuracy comparing with the model based on logistic regression and neural network [1].

Our proposed idea is to have an application to generate a UScore. Here, the user gives permission to analyse their social network and during this process, the user can also add additional information that may be applicable. Our algorithms could be used to generate a cumulative score based on a set of instructions/rules (standard scales) as per requirement based on the share held by each social site in the final calculation. The application can be deployed as a cloud service for easy provisioning and high availability.

Rest of the article is structured as follows. Section 2 describes related work. Section 3 describes the analytics process model. Section 4 describes results of various machine learning algorithms. In Sect. 5, algorithms performance with respect accuracy is presented. Finally, in Sect. 6, we conclude the paper.

1.1 Problem Statement

In today's financial environment, the concept of social network-based credit scoring system has gained appreciation. A credit scoring system is based on statistical model and this model distinguish "good" and "bad" loans using borrower data. Our idea is to have an application to generate a UScore here user:

1. User gives permission to analyse their social network.
2. User can add additional information that may be applicable.

Our mission is to provide one billion people access to powerful financial products at a lower cost, faster, and more conveniently. Our algorithms can be used to generate a cumulative score based on a set of instructions/rules (standard scales) as per requirement based on the share held by each social site in the final calculation. This would be the UScore desired. Our UCredit system can be deployed as a cloud service for easy provisioning and high availability, and it can be real and would possess the ability to manage and dispute using a mobile application.

1.2 Research Objectives

- To provide a better score so as to expand opportunities for larger portion of the population.
- To enable low-income individuals to access larger credit system who would otherwise find it hard to obtain credit.
- To enable growing and promising entrepreneurs and start-ups to obtain corporate loans.
- Provide an all-encompassing risk score and estimate defaults.

2 Related Work

From the literature review, we understand the various types of credit score-based systems followed by the financial institutions and professional banks for evaluating the customer credit credentials. The first category emphasis is on subsidizing achievement and default hazard and the subsequent category emphasis on a relationship social media information and default chance. In this paper, we consider investigating and implementing few ideas and related with second category. As a whole, each and every technique can have similar methods like finding social network ties, retrieving and saving the data from the central database, and predicting accuracy. Moreover, the vital differences between each and every methods are the type of methodology and implementation.

Authors in [1] target analysing the impact of using network-based measures with respect to client score precision and on tie development among clients. Credit scoring has been viewed as a basic subject and concentrated widely in the money field. Numerous computerized reasoning procedures have been utilized to settle credit scoring. Up until this point, credit scoring models dependent on information mining strategies has been applied to the fund field since data on clients' record as a consumer has been gathered and can be utilized to prepare the model. Credit scoring models can help chiefs of banks to settle on increasingly exact choices, and consequently viably control credit dangers. Accordingly, the setting up of a FICO rating model, which has incredible down to earth esteem and handy hugeness, has gotten one of the fundamental errands for banks [2]. Professionals and analysts have proposed numerous customary measurable techniques and man-made consciousness strategies for credit scoring.

Authors in [3, 4] say financial strength of a customer is more important than customer income, other loans, qualification etc. Social relationship and default risk of a customer is analysed in [5]. Riza et al. [6] use regression model to evaluate credit risk and measure loan performances. This regression model finds the credit-grade and debt-to-income ratio. Further, in [7, 8], authors used models like statistical models, linear or logistic regression model, profit analysis, linear discriminant analysis etc. But performance of these models is poor for nonlinear relationships.

In recent time’s machine learning and artificial intelligence algorithms are used to evaluate customer’s credit credentials and these algorithms perform better than statistical methods. Some of these are support vector machine (SVM) [9], artificial neural networks (ANN) [10], and random forest [11].

3 Architecture

Figure 1 shows the analytics process model. Model consists of various steps. Here, the step one is called as analytic hierarchy process (AHP). In analytic hierarchy model, we investigate the problem at different levels in detail and then express in a form suitable according to AHP requirements. The first step is to identify the business problem in which we try to answer the following questions.

- (a) What is our mission?
- (b) Who is our client?
- (c) What does the customer value?
- (d) What are our key results?
- (e) What’s our plan?

In step two, we identify the data sources. In our proposed work, the customers’ details and requirements act as the primary data sources. This data is amassed using the Kaggle data repository and surveying. Following this step is to sieve out the extraneous attributes and data pre-processing to convert raw data into clean data. The next step is training using machine learning algorithm; here we use data to improve our model’s ability to predict. Finally, we test our model against the data.

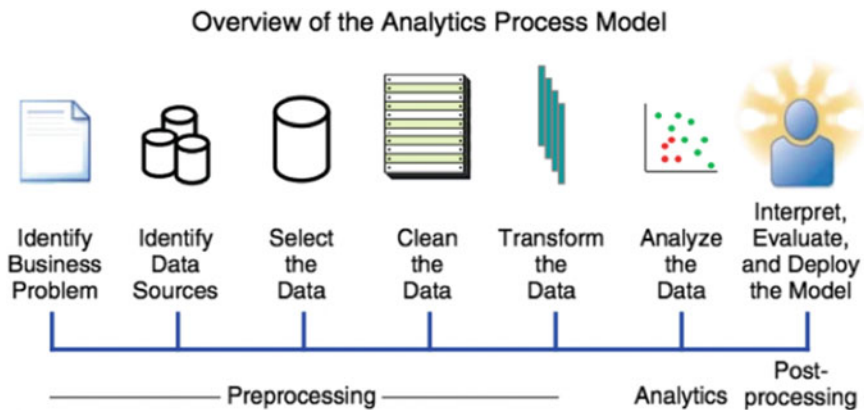


Fig. 1 Overview of the model

4 Results

Following algorithms are used for analysis. Here, dataset is divided into a training set and a test set. In a dataset, a training set is implemented to build up a model, while a test set is to validate the model built.

4.1 Logistic Regression

Logistic regression algorithm is a supervised classification algorithm. Here, the output target variable takes only discrete values for given set of features. It models the data using the sigmoid function

$$g(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

Figures 2 and 3 show logistic regression training set and logistic regression test set, respectively. Here, the green dots indicate the number of customers eligible for the loan designated by (1). Contrarily, the red dots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

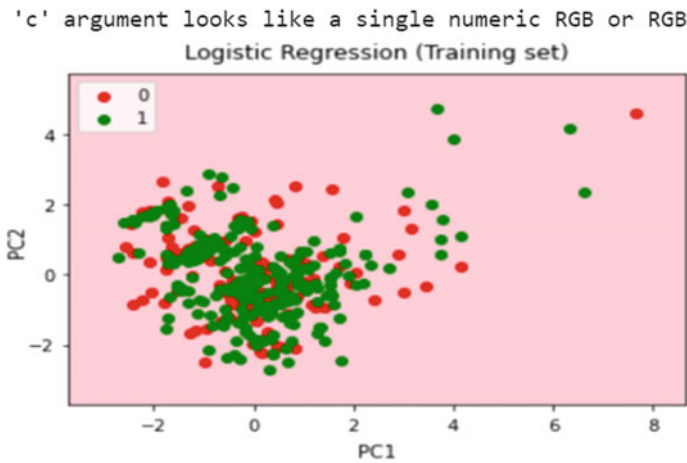


Fig. 2 Logistic regression training set

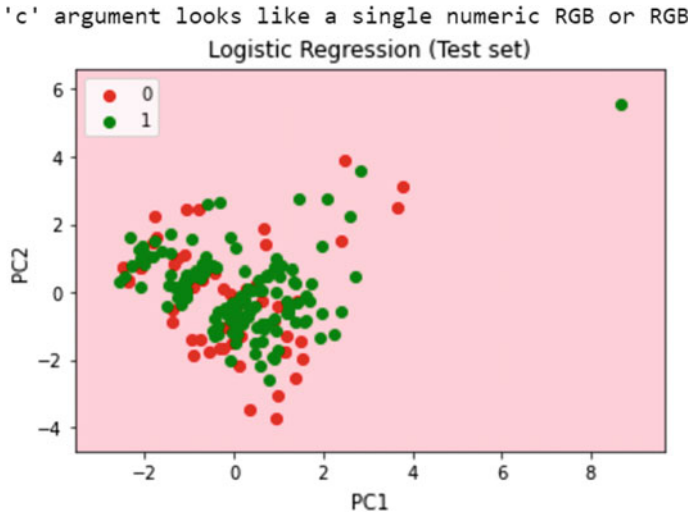


Fig. 3 Logistic regression test set

4.2 K-Nearest Neighbour

K-nearest neighbours is simple and commonly used machine learning classification algorithm. Figures 4 and 5 show KNN training set and KNN test set, respectively. The green dots indicate the number of customers eligible for the loan designated by (1). Contrarily, the red spots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

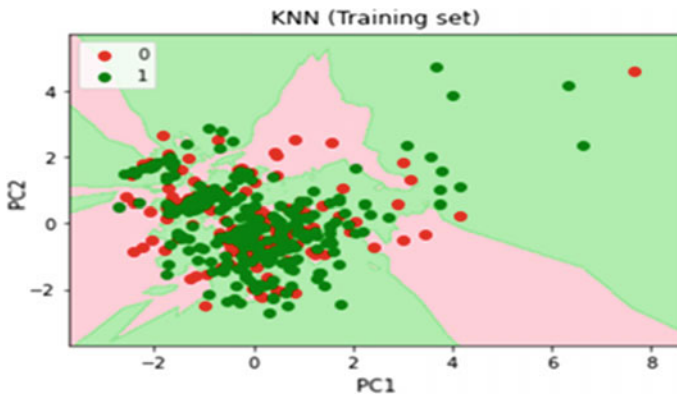


Fig. 4 KNN training set

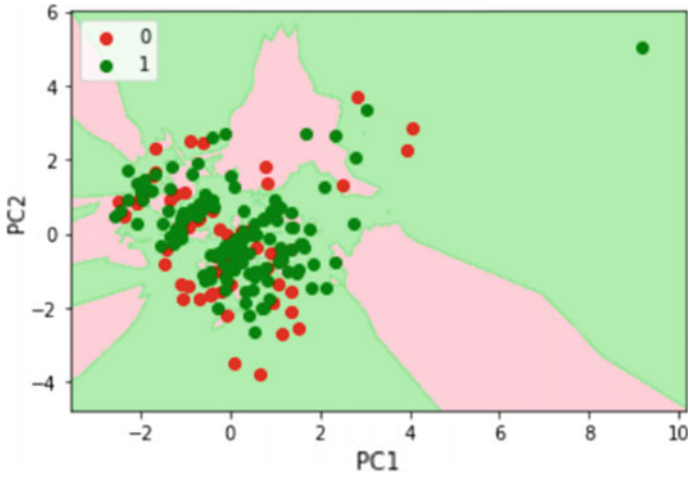


Fig. 5 KNN test set

4.3 SVM

A support vector machine (SVM) is another machine learning classifier; here given training data, algorithm outputs an optimal hyperplane which categorizes new examples. Figures 6 and 7 show SVM training set and SVM test set, respectively. The green dots indicate the number of customers eligible for the loan designated by (1).

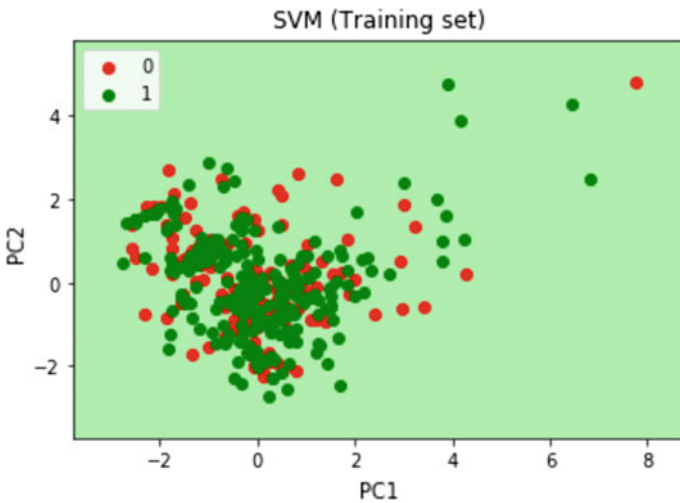


Fig. 6 SVM training Set

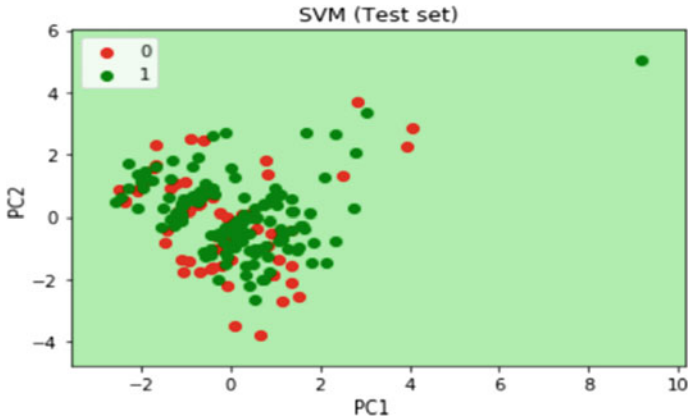


Fig. 7 SVM test set

Contrarily, the red spots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

4.4 Naïve Bayes

Naive Bayes classifier is based on Bayes' theorem. Here, given dataset is divided into two parts, namely feature matrix and the response vector. Feature matrix contains all the vector value of dependent features. Similarly, response vector contains the value of class variable for each row of feature matrix. Figures 8 and 9 show naïve Bayes training set and naïve Bayes test set, respectively. The green dots indicate the number of customers eligible for the loan designated by (1). Contrarily, the red spots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

4.5 Decision Tree

Figures 10 and 11 show decision tree training set and decision tree test set, respectively. The green dots indicate the number of customers eligible for the loan designated by (1). Contrarily, the red spots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

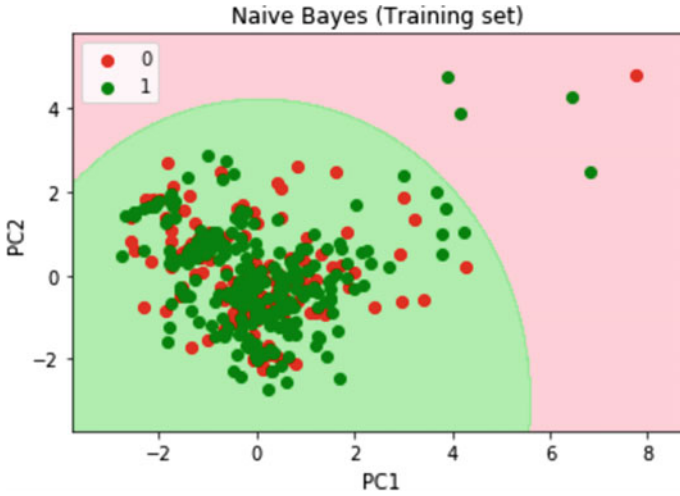


Fig. 8 Naïve Bayes training set

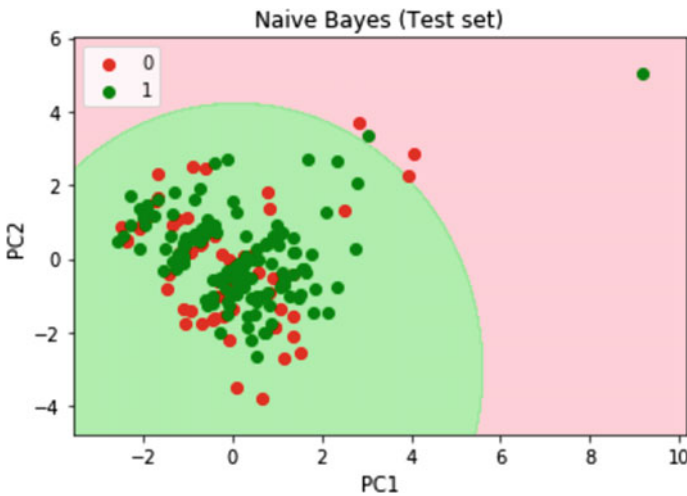


Fig. 9 Naïve Bayes test set

4.6 Random Forest

A random forest uses multiple decision trees to result. Figures 12 and 13 show random forest training set and random forest test set, respectively. The green dots indicate the number of customers eligible for the loan designated by (1). Contrarily, the red spots indicate the number of customers not qualified for the loans based on the attributes designated by (0).

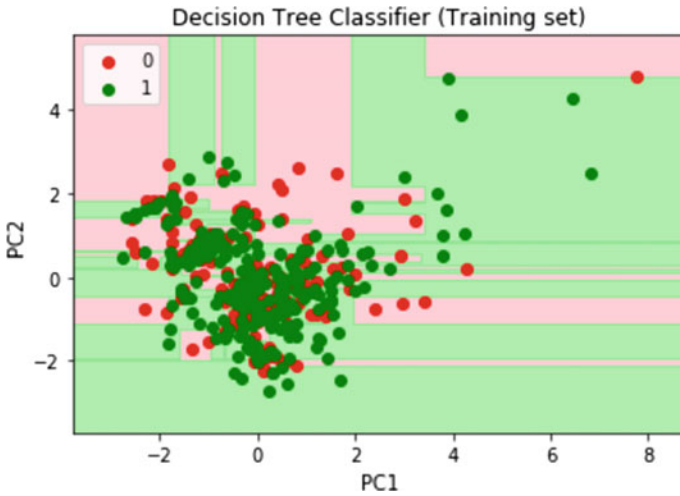


Fig. 10 Decision tree training set

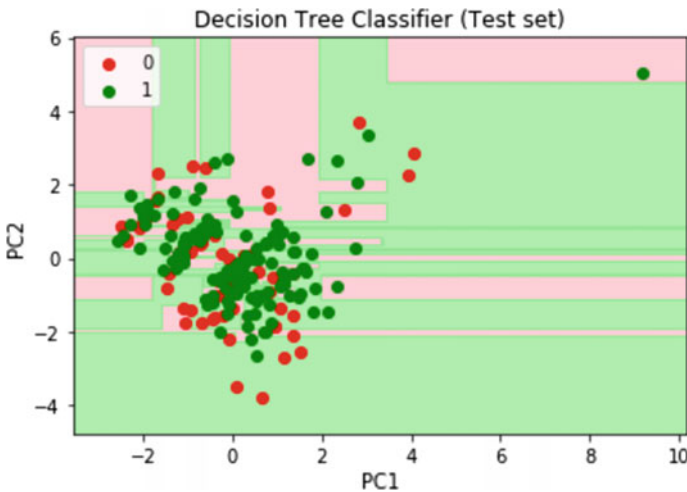


Fig. 11 Decision tree test set

5 Performance of the Algorithms

The performance of the above algorithms with respect to accuracy is listed in Table 1. Tabulated results in Table 1 show logistic regression, and SVM algorithms proved to be the most efficient and accurate algorithms with an accuracy of 70.732%.

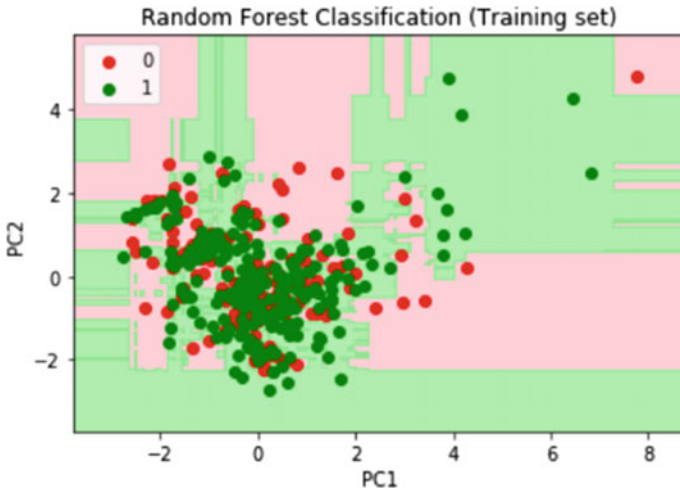


Fig. 12 Random forest training set

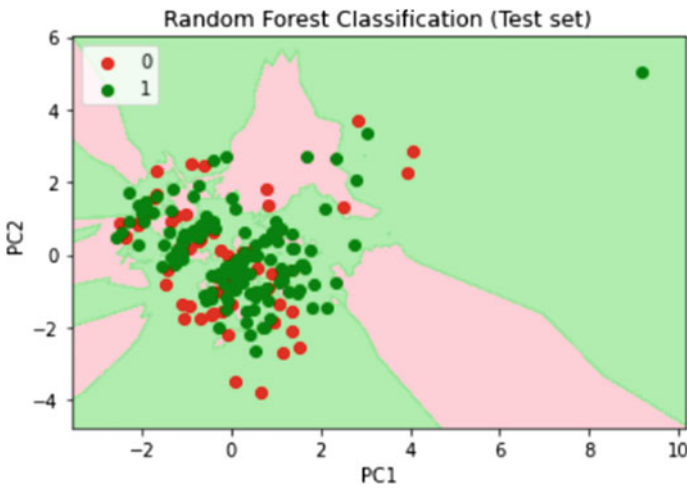


Fig. 13 Random forest test set

Table 1 Performance of algorithms

Algorithm	Accuracy
Logistic Regression	0.7073170731707317
SVM	0.7073170731707317
Naïve Bayes	0.7024390243902439
Decision Tree	0.5658536585365853
Random Forest	0.551219512195122
KNN	0.6731707317073171

6 Conclusion

In this paper, we have presented a credit scoring model using various online social networks. Here, we extract information about a customer on social networks such as twitter.com, linkedIn.com, and online payment platforms such as Phone Pay and Google Pay etc., and then we apply various machine learning algorithms to test the relationship between social network information and customer loan default. From this approach, nearly 70% of these customers may qualify mainstream loan products after factoring in our non-traditional measures. Our tool UCredit can be used for customers like immigrants and recently passed out college students. UCredit evidently suggests that social network information is more reliable and resourceful for faster and convenient credit scoring.

References

1. Wei Y, Yildirim P, Van den Bulte C, Dellarocas C (2016) Credit scoring with social network data. *Mark Sci* 35(2):234–258
2. Clayton J Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media, 2014
3. M. Herzenstein, R. Andrews, U. Dholakia, et al. (2008) The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities, Working Paper. Available at SSRN www.prosper.com (accessed 30 September 2014)
4. J. Qiu, Z. Lin, and B. Luo, Effects of borrower defined conditions in the online peer-to-peer lending market. E-life: web-enabled convergence of commerce, work, and social life, Lecture Notes in Business Information Processing, 2012, 108, 167–179
5. C. Everett, Group membership, relationship banking and loan default risk: the case of online social lending, SSRN Electronic Journal 03/2010. <http://www.researchgate.net/publication/228200235>
6. Emekter R, Tu Y, Jirasakuldech B, Lu M (2015) Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Appl Econ* 47(1):54–70
7. Hand DJ, Henley WE (1997) Statistical classification methods in consumer credit scoring: a review. *J R Stat Soc A Stat* 160:523–541
8. Baesens B, van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J (2003) Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *J Oper Res Soc* 54:627–635
9. Harris T (2015) Credit scoring using the clustered support vector machine. *Expert Syst Appl* 42:741–750
10. Zhao Z, Xu S, Kang BH, Kabir MMJ, Liu Y, Wasinger R (2015) Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Syst Appl* 42:3508–3516
11. Malekipirbazari M, Aksakalli V (2015) Risk assessment in social lending via random forests. *Expert Syst Appl* 42:4621–4631
12. Berger, Allen N., and W. Scott Frame. “Small business credit scoring and credit availability.” *Journal of small business management* 45.1 (2007): 5–22
13. Tan, Tianhui, and Tuan Q. Phan. “Social media-driven credit scoring: The predictive value of social structures.” Available at SSRN 3217885 (2018).
14. Yu, Xi, et al. “Data cleaning for personal credit scoring by utilizing social media data: An empirical study.” *IEEE Intelligent Systems* 35.2 (2020): 7–15
15. Kulkarni SV, Dhage SN (2019) Advanced credit score calculation using social media and machine learning. *Journal of Intelligent & Fuzzy Systems* 36(3):2373–2380

16. Huang C-L, Chen M-C, Wang C-J (2007) Credit scoring with a data mining approach based on support vector machines. *Expert Syst Appl* 33(4):847–856

Synthesis of Graphene Quantum Dots and Fabrication of Humidity Sensor



Aniruddh Holemadlu, G. Keerthana, Kunal B. Purohit, Habibuddin Shaik, and Veda Sandeep Nagaraja

1 Introduction

Humidity can be defined as the concentration of water vapour present in the air. It is an important characteristic of environment and affects the performance and working of industries and technologies. The continuous determination and monitoring of humidity in fields such as semiconductor manufacturing, soil moisture monitoring, food processing, packaging, civil engineering, air conditioning systems are necessary [1].

Humidity sensors were often ignored in high-volume application due to the scarcity of efficient and robust sensors. Thus, the key features a sensor must have to appeal to the market at large are: reliability, robustness, consistency, low-cost, simple structure, good repeatability and no hysteresis, durability and longevity, contaminant resistant and characterized to have a good behaviour towards humidity-temperatures cross influence, condensation and recovery [2].

Humidity sensors are made of different types of sensing materials, such as polymers [3], ceramics [4], metal oxide semiconductors (MOS) [5], carbon nanotubes and their composites [6]. Metal oxide semiconductor-based humidity sensors have an advantage over other types due to its easy and low-cost design, compact size and compatibility with modern devices. The most common material used for low-cost

A. Holemadlu · G. Keerthana · K. B. Purohit · V. S. Nagaraja
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Yelahanka, Bengaluru 560064, India

H. Shaik (✉)
Department of Physics, Nitte Meenakshi Institute of Technology, Yelahanka, Bengaluru 560064, India

H. Shaik · V. S. Nagaraja
Center for Nanomaterials and MEMS, Nitte Meenakshi Institute of Technology, Yelahanka, Bengaluru 560064, India

gas sensing is tin dioxide (SnO_2). This is due to its physiochemical properties that depend on the moisture around it. Tin dioxide being a stable n-type semiconductor has a band gap of 3.6 eV.

The SnO_2 sensors showed some advantages in contrast to the other types of available humidity sensor, but due to their high resistivity, they did not suddenly adjust the resistance values at higher RH, restricting their marketing and practical applications [7].

Carbon nanomaterials like carbon nanotubes and graphene have been shown to have potential applications in sensing field [8–12]. Graphene has drawn attention for its possible wide range of applications as an outstanding nanomaterial due to its large surface area of $2600 \text{ m}^2 \text{ g}^{-1}$, good chemical stability and outstanding electrical properties like low noise rates and a high carrier mobility. Investigations showed that graphene-based metal oxide nano-composite exhibited excellent improvement in gas sensing applications. A sensor developed by RGO/ SnO_2 hybrid composite showed high sensitivity, small response and recovery time over an extensive range of relative humidity levels [13].

Due to their one of a kind size-dependent electro-optical properties colloidal semiconductor quantum dots (QDs) have a plethora of powerful applications and as a result, they have been of significant research interest [14]. Graphene quantum dots (GQDs) showcase novel properties such as their behaviour as spin qubit with joint spin states and emission making it desirable for various sensing applications. Unlike typical QDs, GQDs are photo-stable, biocompatible, have increased surface grafting, and possess excellent electrical, mechanical and thermal properties. Features like this can tremendously assist to numerous high-tech applications including bioimaging [15], OLEDs [16], fuel cells [17], photovoltaic devices [18], composites [19] and biosensors [20].

Graphene quantum dots [21] synthesized by converting graphene nano-ribbons through a process of oxidative edge-roughening and cleavage have been reported to be an excellent humidity sensor with outstanding sensitivity [13] between 0 and 40% of the relative humidity.

Graphene Quantum Dots have drawn significant attention in the past few years. The fact that it is highly sensitive makes it an ideal material in the sensing field. The reason behind its high sensitivity is a phenomenon called Multiple Exiton Generation (MEG). MEG involves the generation of more than a few tens of exciton in the conduction band, from the absorption of a single photon [21–23].

In this paper, we report the synthesis of GQDs via carbonisation of citric acid [24] and the fabrication of humidity sensor using GQD as a sensing layer.

2 Materials and Facilities

Citric acid (anhydrous ($\text{C}_6\text{H}_8\text{O}_7$)) was purchased from Vasa Scientific Co., Bangalore. 0.1 M NaOH solution is used to adjust the pH of GQD solution. Pure aluminium

was used for the deposition of electrodes. The GQD was characterized via Raman Spectroscopy by Horiba Jobin Yvon T64000 and XRD by Panalytical x'pert2 powder.

2.1 Synthesis of GQD

Graphene Quantum Dots were synthesized via carbonization of citric acid monohydrate as reported by Chi et al. [24]. Figure 1 shows a Schematic representation of the synthesis method of graphene quantum dots by controlled carbonization of citric acid. Shortly, 2 g of citric acid (anhydrous ($C_6H_8O_7$)), was taken in a test tube and the test tube is subjected to heat through a heater belt, where the temperature rises to $200\text{ }^\circ\text{C}$ gradually. After 23 min, the test tube was removed and colour of the sample was observed to be pale yellow. These GQD samples were divided into 3 portions

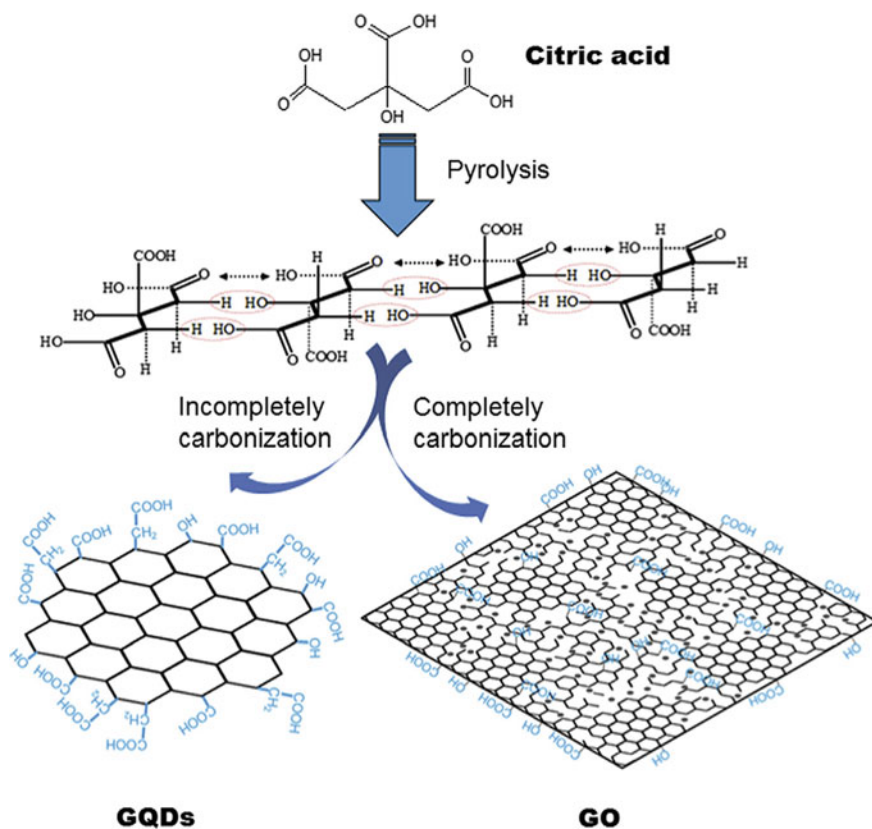


Fig. 1 Schematic of synthesis of GQD and GO via carbonization of citric acid [24]

Fig. 2 Schematic of electrodes of the sensor



and the pH of each portion was adjusted to 5.0 by adding 0.1 M NaOH in suitable quantity.

2.2 Fabrication of Interdigitated Electrodes

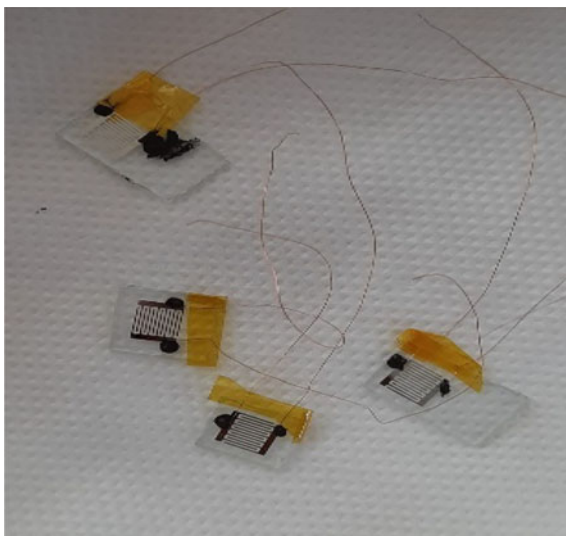
Interdigitated Aluminium electrodes with length and breadth of 0.8 cm each, thickness of 300 nm and spacing between each electrode equal to 0.5 mm, were deposited on the glass substrate by thermal evaporation. The schematic of interdigitated electrodes is shown in Fig. 2. Figure 3 shows the fabricated interdigitated electrodes with wire contacts taken. A small amount of graphene quantum dot solution was pipetted on the interdigitated electrodes to sense the humidity.

3 Results and Analysis

3.1 Material Characterization

Characterization refers to the specific and general method of probing and measuring the structure and properties of a substance. Raman spectroscopy and XRD were performed on the prepared GQD samples. In addition, Raman spectroscopy is a non-destructive process and involves easy sample preparation, and it can detect non-IR active vibrational modes. On the other hand, being also non-destructive, the XRD

Fig. 3 Humidity sensor with wire contacts



technique has marked sensitivity, reliability, simple and quick sample preparation, high resolution, and easy interpretation of the information that could be used for both qualitative and quantitative aspects of analysis.

3.1.1 Raman Spectroscopy

Raman spectroscopy determines the vibrational modes of molecules. Monochromatic light from LASER interacts with molecules vibrations or phonons resulting in a shift in laser energy. We have done Raman analysis for four different samples which were synthesized for different timings. It was observed that the results shown in Fig. 4, do not have characteristic peaks. This is due to the fluorescence spectrum of the precursor dominating the Raman peaks of GQD. Therefore, the results gave no conclusive evidence on the formation of GQD.

3.1.2 XRD Results

To study the structures of crystals and atomic spacing between them, X-ray diffraction is one of the most prominent techniques. It works on the principle of constructive monochromatic X-ray interference on a crystalline sample. The XRD pattern is unique for every crystalline material. The peak positions provide information that can be used to evaluate cell parameters. Each peak in the pattern has its own set of indices. Figure 5 shows the expected peak for Graphene oxide and GQD and Fig. 6 shows the obtained XRD pattern of a sample.

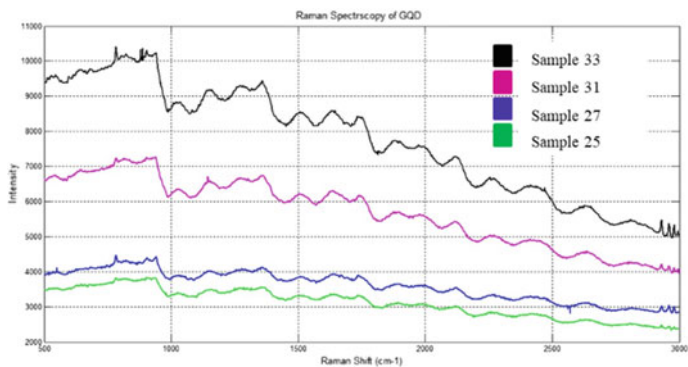


Fig. 4 Raman spectrum of samples

Fig. 5 Expected XRD pattern for GO and GQD

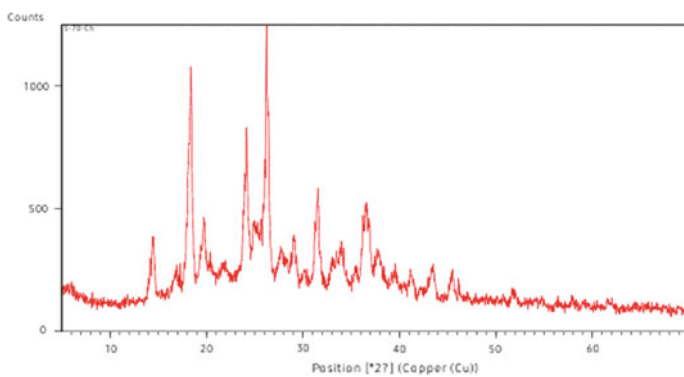
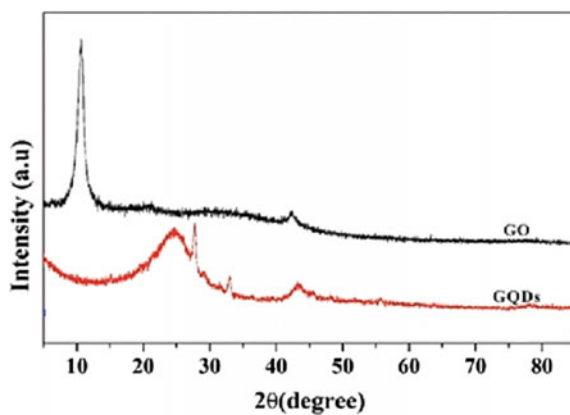


Fig. 6 XRD pattern obtained

3.2 Sensor Characteristics

Using the Electrochemical Analyser SP300 (Biologic), the voltage-current (V-I) characteristics of the sensor were analysed. The study was carried out on the sensor without and with GQD. A constant voltage of 1 V is applied between the electrodes and the current is measured. On variation of humidity, the sensor without GQD solution, there is a two orders change in the current as shown in Fig. 7, whereas there are six orders of change in current for the sensor with GQD. Also it is evident from Figs. 7 and 8 that the response time is also less for the sensor with GQD.

Fig. 7 Sensor response without GQD

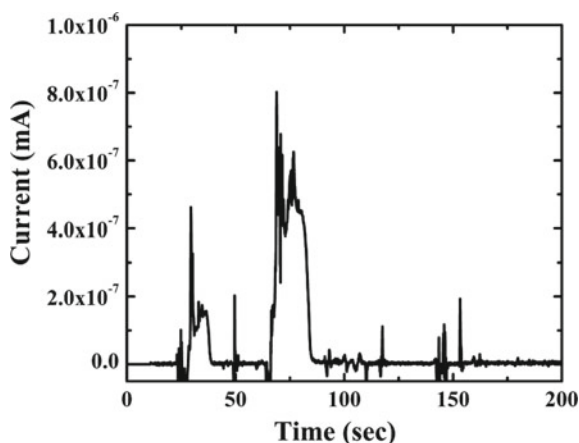
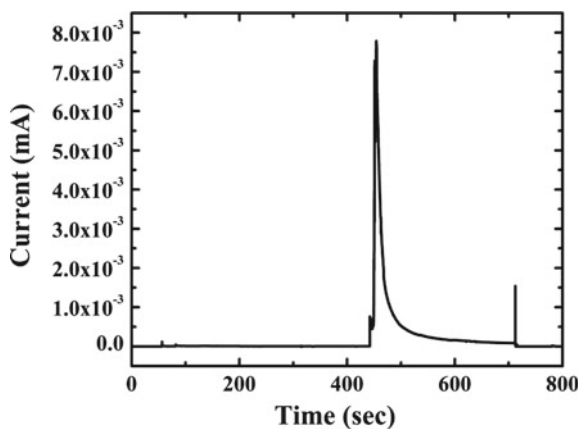


Fig. 8 Sensor response with GQD



4 Conclusion

In summary, a new version of humidity sensors based on Graphene Quantum Dots, via carbonisation of citric acid, is reported. It was found that Quantum Dots can be used for humidity sensing. Furthermore, GQD can be easily synthesized in laboratory level. The results obtained shows that the sensitivity is greater and response time is shorter with GQD solution. This was confirmed from the V-I characteristics of the sensor. For the sensor to be practically useful, other physical and chemical properties need to be characterized. As scope for future study, the sensors are to be tested under controlled conditions with variety of GQD samples along with variation in pH to validate its practical use. Electrodes can be deposited using different metals such as platinum or gold for comparing responsivity.

References

1. Alizadeha T, Shokri M (2015) A new humidity sensor based upon graphene quantum dots prepared via carbonization of citric acid. *Sens Actuators B Chem*
2. Yamazoe N, Shimizu Y (1986) Humidity sensors: Principles and applications. *Sens Actuators* 10
3. T. Fei, K. Jiang, S. Liua and T. Zha, "Humidity sensor based on a cross-linked porous polymer with unexpectedly good properties," *RSC Advances*, vol. 41, 2014
4. T. Nenov and Z. Nenov, "Multi-objective optimization of the parameters of TiO₂-based ceramic humidity sensors," *Elsevier Science*, 2012
5. S. Chatterjee, S. Chatterjee, A. Ray and A. Chakra, "Graphene-metal oxide nanohybrids for toxic gas sensor: A review," *Sensors and Actuators B: Chemical*, vol. 221, 2015
6. T. Fei, K. Jiang, F. Jiang, R. Mu and T. Zhang, "Humidity switching properties of sensors based on multiwalled carbon nanotubes/polyvinyl alcohol composite films," *Journal of Applied Polymer Science*, 2013
7. X. Song, Q. Qi, T. Zhang and C. Wang, "A humidity sensor based on KCl-doped SnO₂ nanofibers," *Sensors and Actuators B: Chemical*, vol. 138, 2009
8. T. Alizadeh and F. Rezaloo, "A new chemiresistor sensor based on a blend of carbonnanotube, nano-sized molecularly imprinted polymer and poly methylmethacrylate for the selective and sensitive determination of ethanol vapor," *Sensors and Actuators B*, vol. 176, 2013
9. T. Alizadeh and S. Mirzaghoolipur, "A Nafion-free non-enzymatic amperometricglucose sensor based on copper oxide nanoparticles-graphenenanocomposite," *Sensors Actuators B*, vol. 198, 2014
10. J. Robinson, F. Perkins, E. Snow, Z. Wei and P. Sheehan, "Reduced grapheneoxide molecular sensors," *Nano Letters*, vol. 8, 2008
11. R. Paul, S. Badhulika, N. Saucedo and A. Mulchandani, "Graphene nanomesh ashighly sensitive chemiresistor gas sensor," *Anal. Chem*, vol. 84, 2012
12. J. Han, B. Kim, J. Li and M. Meyyappan, "Carbon nanotube based humidity sensoron cellulose paper," *J. Phys. Chem. C*, vol. 116, 2012
13. D. Zhang, H. Chang, P. Li and R. Liu, "Fabrication and characterization of an ultrasensitive humidity sensor based on metal oxide/graphene hybrid nanocomposite," *Sensors and Actuators B: Chemical*, vol. 225, 2016
14. V. Ruiz, I. Fernández, P. Carrasco, G. Sevillano, H. Grande and J. Herrán, "Graphene quantum dots as a novel sensing material for low-cost resistive and fast-response humidity sensors," *Sensors and Actuators B: Chemical*, vol. 218, 2015

15. S. Zhu, J. Zhang, C. Qiao, S. Tang, Y. Li, W. Yuan, B. Li, L. Tian, F. Liu, R. Hu, H. Gao, H. Wei, H. Zhang, H. Sunb and B. Yang, "Strongly green-photoluminescent graphenequantum dots for bioimaging applications," *Chemical Communications*, no. 24, 2011
16. L. Tang, R. Ji, X. Cao, J. Lin, H. Jiang, X. Li, K. Teng, C. Luk, S. Zeng, J. Hao and S. Lau, "Deep Ultraviolet Photoluminescence of Water-Soluble Self-Passivated Graphene Quantum Dots," *ACS Nano*, vol. 6, 2012
17. Y. Li, Y. Zhao, H. Cheng, Y. Hu, G. Shi, L. Dai and L. Qu, "Nitrogen-Doped Graphene Quantum Dots with Oxygen-Rich Functional Groups," *Journal of the American Chemical Society*, 2012
18. X. Yan, X. Cui, B. Li and L. Li, "Large, Solution-Processable Graphene Quantum Dots as Light Absorbers for Photovoltaics," *Nano Letters*, 2010
19. W. Chen, G. Lv, W. Hu, D. Li, S. Chen and Z. Dai, "Synthesis and applications of graphene quantum dots: A review," *Nanotechnology Review*, 2018
20. H. Sun, L. Wu, W. Wei and X. Qu, "Recent advances in graphene quantum dots for sensing," *Materials Today*, vol. 16, no. 11, 2013
21. M. Li, R. Begum, J. Fu, Q. Xu, T. M. Koh, S. A. Veldhuis, M. Grätze, N. Mathews, S. Mhaisalkar and T. C. Sum, "Low threshold and efficient multiple exciton generation in halide perovskite nanocrystals," *Nature Communications*, vol. 9, 2018
22. Nozik AJ (2008) Multiple exciton generation in semiconductor quantum dots. *Chem Phys Lett* 457(1–3):3–11
23. M. C. Beard, "Multiple Exciton Generation in Semiconductor Quantum Dots," *Journal for Physical Chemistry Letters*, vol. 11, no. 2, 2011
24. Y. Dong, J. Shao, C. Chen, H. Li, R. Wang, Y. Chi, X. Lin and G. Chen, "Blue luminescent graphene quantum dots and graphene oxide prepared by tuning the carbonization degree of citric acid," *Carbon* 50, 2012
25. M. Bacon, S. Bradley and T. Nann, "Graphene Quantum Dots," *Particle and Particle System Characterisation*, 2013

Optimal Joint Scheduling and Cloud Offloading for Multi-component Applications



S. L. Shiva Darshan, Abdul Mueez, Aakash S. Shetty, B. A. Mohan, S. Ashok Kumar, and Roshan Fernandes

1 Introduction

Recent advancements in hardware and communication technologies have led hand-held mobile nodes to execute resource-intensive applications in them, though the size and type of applications may restrict the desired performance to be achieved, and/or the battery will drain faster compared with normal usage. One of the resource augmentation approaches is task offloading, which enables mobile nodes to push resource-intensive tasks to rich computing nodes at remote/cloud. This approach can throw some challenges like additional data communication cost, which in turn increases the task's remote completion time and/or energy consumption and may incur monetary cost for using a remote computing node. Thus, to determine whether task offloading is beneficial or not, a task scheduling process is needed. To decide this, the task scheduling process needs to continuously get updates on the status of currently available resources on the remote computing node.

This scheme also has some limitations like scaling of service. When multiple mobile device tasks need to be mapped on multiple remote computing nodes, there is repeated communication between these nodes, which can cause bottleneck and

S. L. Shiva Darshan (✉) · A. Mueez · A. S. Shetty · B. A. Mohan
Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India
e-mail: shiva.darshan@nmit.ac.in

B. A. Mohan
e-mail: mohan.ba@nmit.ac.in

S. Ashok Kumar
Capco Technologies Pvt. Ltd, Bangalore, India
e-mail: ashok.kumar@capco.com

R. Fernandes
NMAM Institute of Technology, Nitte, Karnataka, India
e-mail: roshan_nmamit@nitte.edu.in

communication overhead in the network. This bottleneck or overhead can cause delays for mobile devices that are waiting to perform task scheduling by getting an up-to-date status on the available resources from multiple remote cloud services. In such a situation, to minimize communication overhead, the use of a centralized broker node is proposed to perform resource monitoring tasks on behalf of all the mobile nodes, lowering communication overhead and delays due to queuing of queries for the current state of the available resources on the remote computing node.

2 Literature Survey

In the proposed work, a middleware system called broker is invoked to do the task of scheduling on behalf of a large number of mobile nodes. In [1], a mathematical model was proposed for a centralized task scheduling problem by considering various constraints. This model resulted in optimal solution for task scheduling problem by minimizing total energy consumption across all mobile devices. The proposed work emphasized mobile cloud computing, which is an evolving research field in cloud computing. In these kinds of environments, centralized broker architecture is utilized for resource augmentation for multi-component application on mobile devices.

The model proposed in [1] focused on energy optimization with centralized scheduling of tasks. Offloading decisions are now made on energy consumption, economic cost, or a combination of these two. Thus, an energy and monetary cost-aware mathematical task scheduler/broker architecture is proposed. The proposed model extends the model in [1] by considering minimization of the total monetary cost, as well as total energy consumption across multiple components of the application.

The work has been extended to allow mobile nodes to simultaneously offload multiple tasks to the cloud. Concerning task offloading, two different resource augmentation environments have been evaluated. In the first environment, computation resources are available from a local private cloud accessible through a Wi-Fi network. In the second environment, computation resources are available from public clouds accessible through the Internet. The advantage of the existing system is that it minimizes the total energy consumption when applied to a local private cloud. In this pre-existing model, user-defined delay tolerance is considered for every task, which puts a limit on the delay of the offloading tasks. In turn, the delay of a task defines the constraint for the minimum required data rate for a given input/output data size of the task.

In [2], optimal joint scheduling is used in most places to save resources. With the help of scheduling algorithms, it can be made sure that the resources used are kept in check and have made the best use of it. For example, electricity consumed by most home appliances continuously uses tight scheduling, whereas others consume it in a distributed fashion using flexible scheduling. With the use of an intelligent scheduling strategy, the overall electricity consumption of home appliances can be determined.

It can also determine the charging/discharging pattern of electric vehicles, which in turn helps to reduce the amount of imported electricity, capital, and operational costs.

In [3], the authors proposed optimal scheduling to reduce delay and power consumption for cognitive radio uplinks. They suggested many different scheduling algorithms to minimize the interference to the primary users, a decentralized scheduling algorithm that used on/off rate adaptation scheme, closed-form water-filling like power allocation policy to maximize the cognitive radio systems' per-user throughput, and an algorithm to increase the capacity region to a collision constraint on the primary users. The proposed algorithm aimed at optimizing the throughput for secondary users, while protecting primary users from interference.

In [4, 5], the main purpose was to create a home energy management scheduling solution with energy sources and thermally controllable appliances like washing machine, fridge, and geyser and optically controllable electrical equipment such as illumination. Their main focus with the energy management system was to reduce energy consumption by scheduling home appliances' usage.

The disadvantages of the previously existing systems are:

- (i) They mainly focused on energy optimization by including an economic element; and
- (ii) There was limitation to the number of tasks that could be offloaded because of the absence of a cloud.

The existing work relates to resource augmentation by considering offloading from a mobile device onto a cloud, though it may cause overhead on the network due to random allocations.

3 System Architecture

The proposed model of resource augmentation environments is to provide mobile nodes with computation resources. The factors considered while designing the system are:

- (i) Availability of resources (both local and in the cloud),
- (ii) Accessibility of resources over a Wi-Fi network,
- (iii) Cost of computational resources, and
- (iv) Task scheduling to meet optimization goals.

However, a resource augmentation environment is dynamic. The requirements for the needed resources may vary for a task with two things. One being the change in input data, and the other is the change in offloading goals like delay time and battery consumption. The available resources may also change at remote computational nodes due to the sharing of resources like CPU power, memory, and storage. Therefore, it is critical to dynamically decide on remote execution locations based on current requirements and availability of resources. The thing that poses a challenge is resource monitoring by a large number of mobile nodes. To overcome this

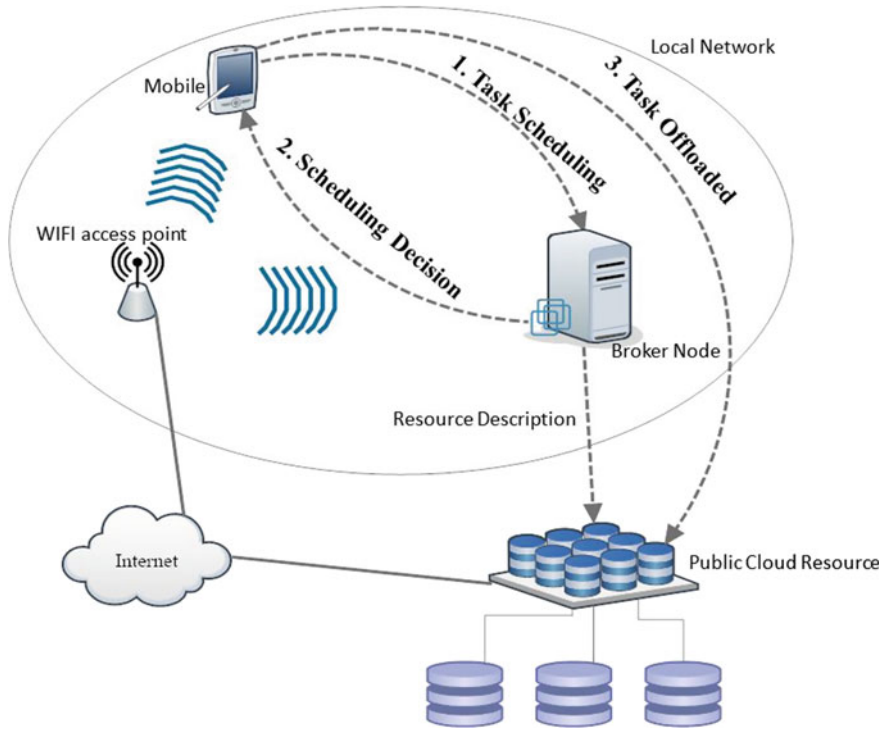


Fig. 1 System architecture for RAE using private clouds

challenge, a centralized node was used, which performed scheduling and monitoring of resources on behalf of all the mobile devices as shown in Fig. 1.

The advantages of the above-proposed system include:

Utilization of a centralized broker node architecture for resource augmentation. A mobile device might have many different offloading goals, which can be achieved using this system. The goals are:

- Battery power saving on mobile nodes,
- Minimizing the monetary cost of resources usage,
- Improving task execution time, and
- Achieving a combination of the above features.

The flow diagram in Fig. 2 depicts how the entire process takes place step-by-step. There are four important modules present in the system, which help in obtaining the objectives such as upload file module, download file module, task scheduling module, and cloud resource module.

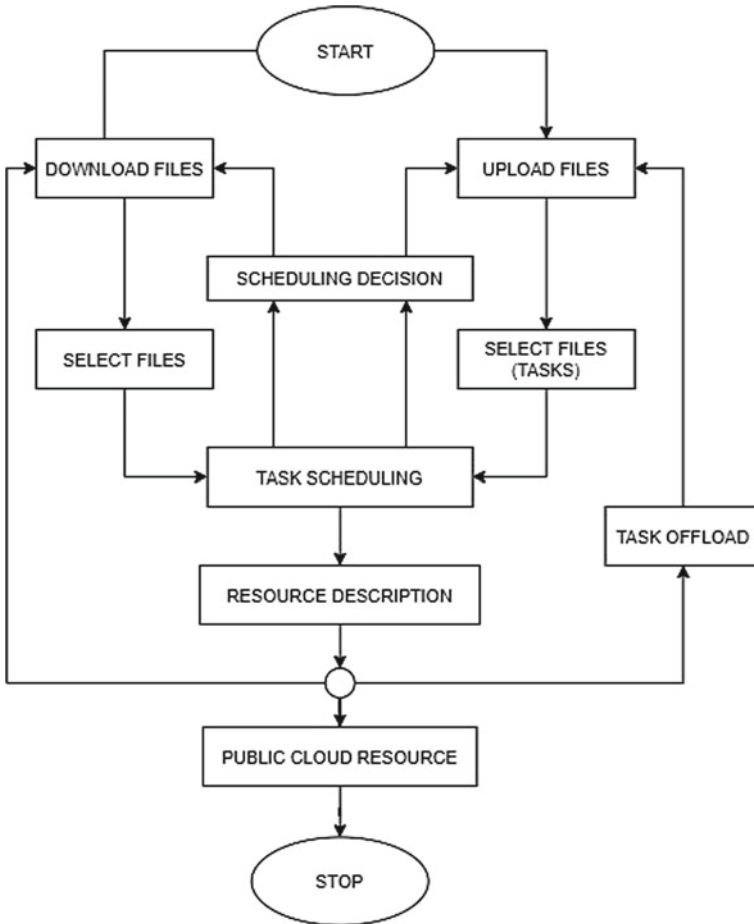


Fig. 2 Data flow diagram

4 Implementation Details

The proposed concept is implemented using Java and CloudSim. Java swing is used to design the user interface and CloudSim to simulate a private cloud in the local system. CloudSim is an open-source framework for modeling and simulating the cloud computing infrastructure and services. It allows the creating of multiple instances of compute node, storage, or networking resource.

There are three different types of users of the system such as:

- The mobile application user (client),
- Scheduler (middleware), and
- Cloud resource (server).

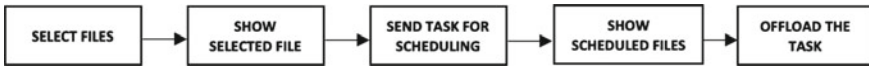


Fig. 3 Upload files module

All three users have a role to play in the proper functioning of the system and to get the desired result. Each contributes to achieve the optimized solution and goals of the proposed system.

The important modules of the system along with their functions are explained below.

Upload Files In this module, a set of files is selected from the mobile device. Here, each file is thought of as a task in the system. After selection, the file list is shown in the window. Then, the files (tasks) are sent for task scheduling to the broker node in the LAN. The scheduling order/decision is received from the broker node, and the files for offloading are shown. These files are then sent to the cloud resource for offloading, which is shown in Fig. 3.

Download Files This module is used to get files from the cloud resource to the account storage drive. A file can then be selected to be downloaded from the cloud resource. After selection, the file is sent to the broker node for scheduling. It is then displayed and downloaded to store in the mobile device, which is shown in Fig. 4.

Task Scheduling This is the most essential module in the system because it decides the number of resources to be allocated for the task. It analyzes the mobile device and the number of files (tasks). After analyzing, it comprehends the task size, sends time, and receive time-based on the task length. The module knows the available resources present in multiple cloud servers (public and private cloud). Finally, it allocates the cloud resources dynamically for the tasks, as depicted in Fig. 5.

Cloud Resource The cloud server is present in LAN, which can be communicated through a Wi-Fi access point. This module has a large number of cloud resources, and each cloud resource has a large number of virtual machines (VM) instances. This can also be scaled dynamically based on requirement and budget, and so ideally, it can be said that it has unlimited resources. It first gets the description of the resource

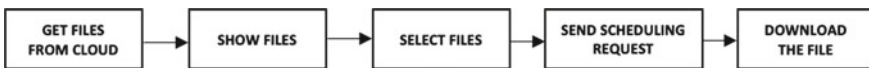


Fig. 4 Download files module



Fig. 5 Task scheduling module

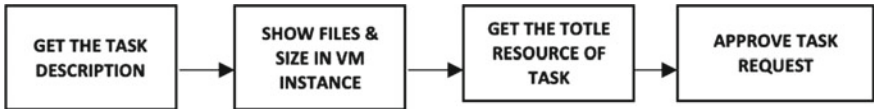


Fig. 6 Cloud resource module

from the scheduler present in the broker, and then checks for the availability of the resource. If the needed resources are available, it will accept the task request, as shown in Fig. 6.

5 Methodology

The mobile cloud ecosystem [6–8] consists of three distinct components as shown in Fig. 7, and the use case diagram is shown in Fig. 8.

The three aforementioned components of the mobile cloud ecosystem have an impact on the quality of experience (QoE) of the user. The QoE has four components like memory usage, energy consumption, monetary cost, and data transmission rate

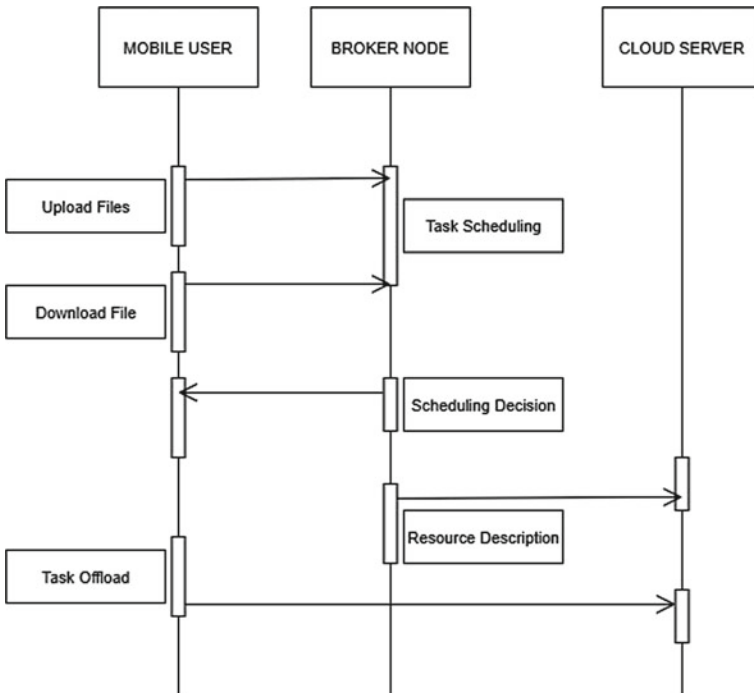


Fig. 7 Sequence diagram

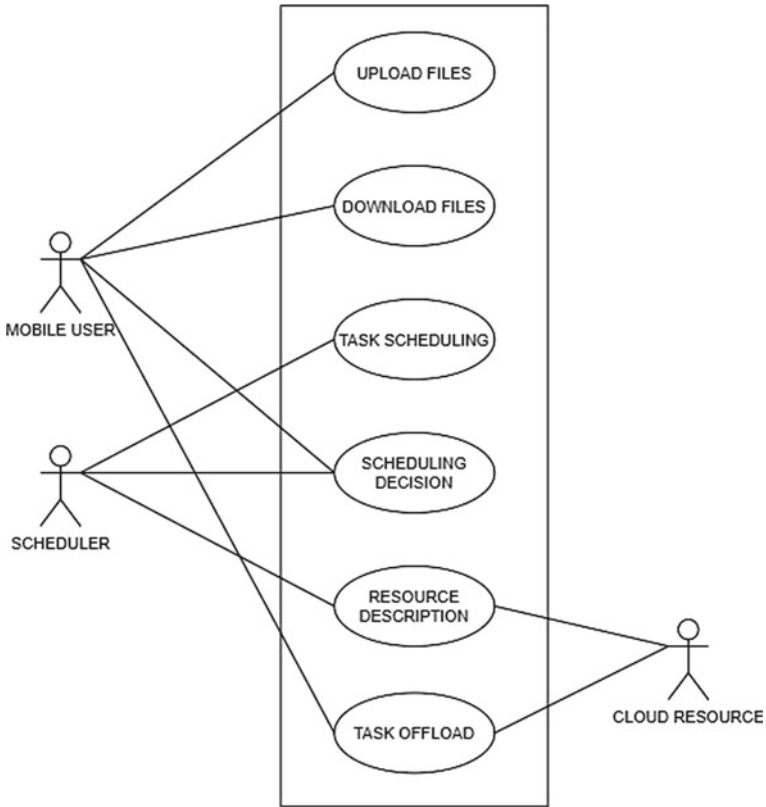


Fig. 8 Use case diagram

[9]. Using this ecosystem as the principle for architecture, the proposed application is implemented with three distinct components in its ecosystem and has unique roles to play. These components are:

- Mobile user—The user along with the application forms the mobile system and can upload, download, view scheduling decisions, and offload tasks.
- Scheduler—The scheduler along with the broker node is the network component. The scheduling is done here.
- Cloud Server—This component has all the resource descriptions, and offloading is done here [10–14].

6 Results

The proposed concept is implemented in Java and CloudSim. Figure 9 is a user interface that allows users to upload or download tasks to the cloud. Figure 10

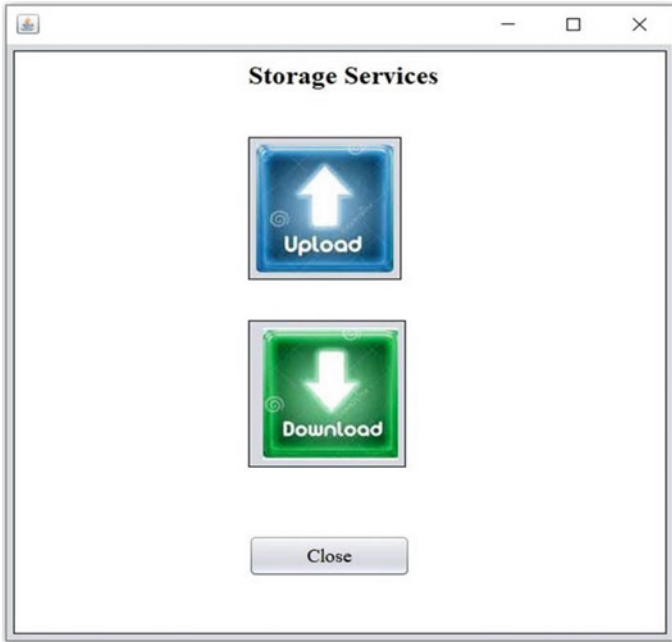


Fig. 9 Services selected

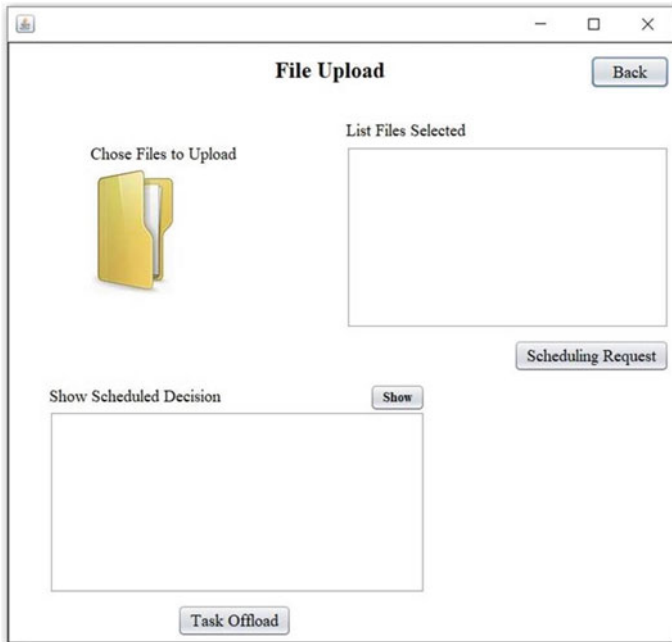


Fig. 10 File upload

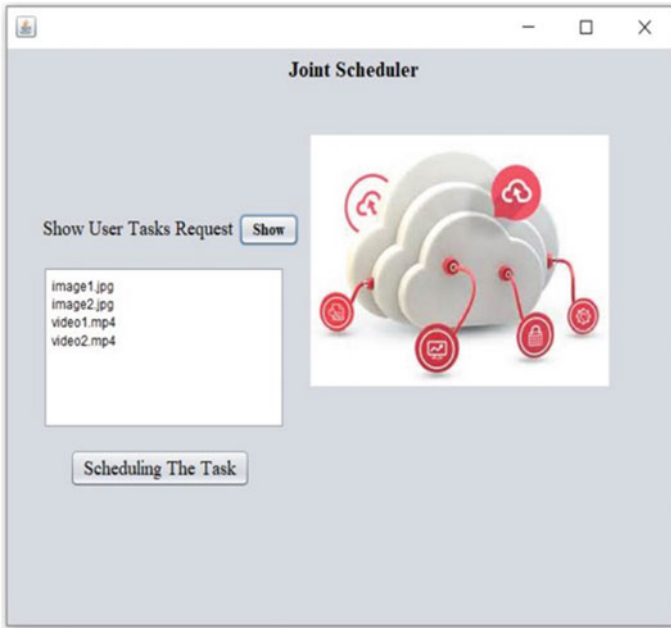


Fig. 11 Joint scheduler

shows the user interface of the user, where the user can list the tasks/files to be uploaded to the scheduler/broker. It also displays the scheduling decision done by the broker/scheduler.

Figures 11 and 12 are the user interface of the broker, which shows the uploaded tasks and resource requirements and the option to select the cloud service before executing on the cloud. Cloud resource availability is displayed in Fig. 13, where the user can choose the cloud to see the available resources in it. Once the task is offloaded to the cloud compute node, it is immediately updated in the user interface as shown in Fig. 14. Figure 15 shows the broker user interface, and Fig. 16 is a user interface from which the user can download the offloaded task after execution on the cloud.

7 Conclusion

The steps in implementing optimally scheduling and offloading have been done in five different modules. These modules are used in the process of scheduling the tasks. The files were considered as tasks in the system. The proposed project ensures that space is automatically made by sending the data to the cloud, which is being

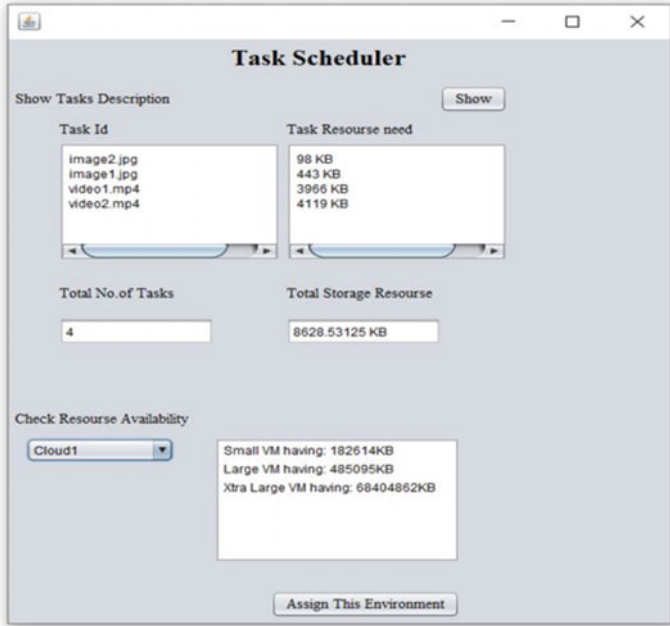


Fig. 12 Task scheduler

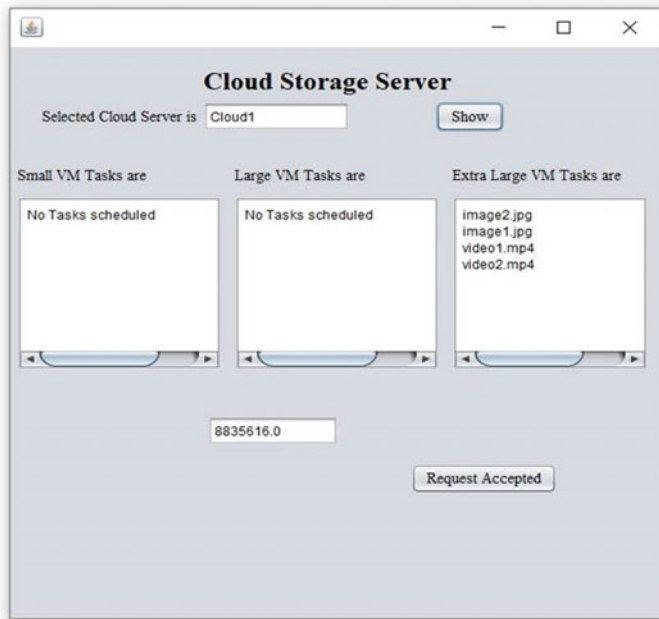


Fig. 13 Cloud storage server

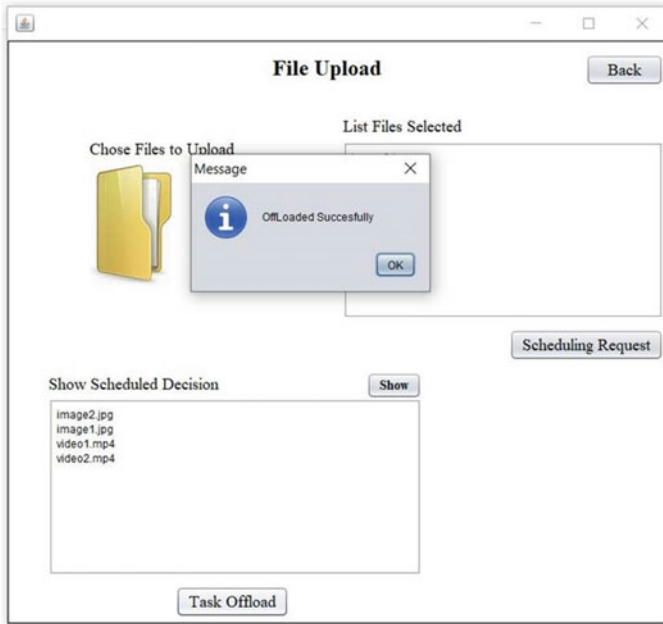


Fig. 14 Task offloading

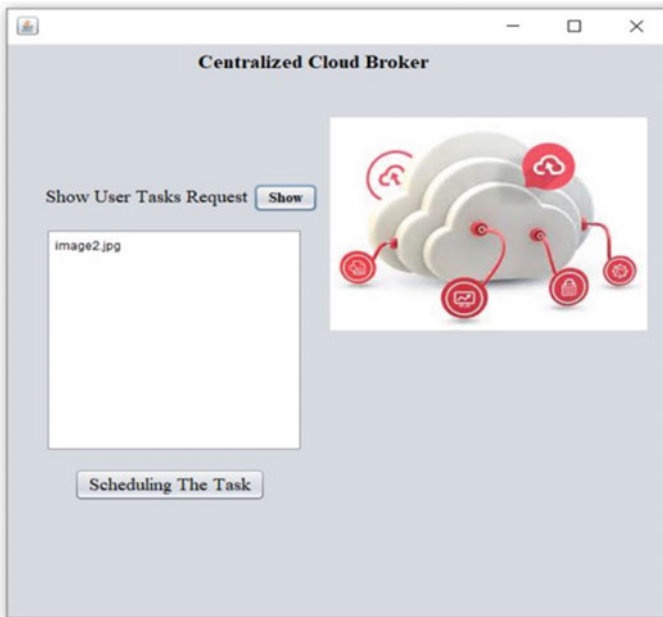


Fig. 15 Centralized cloud broker

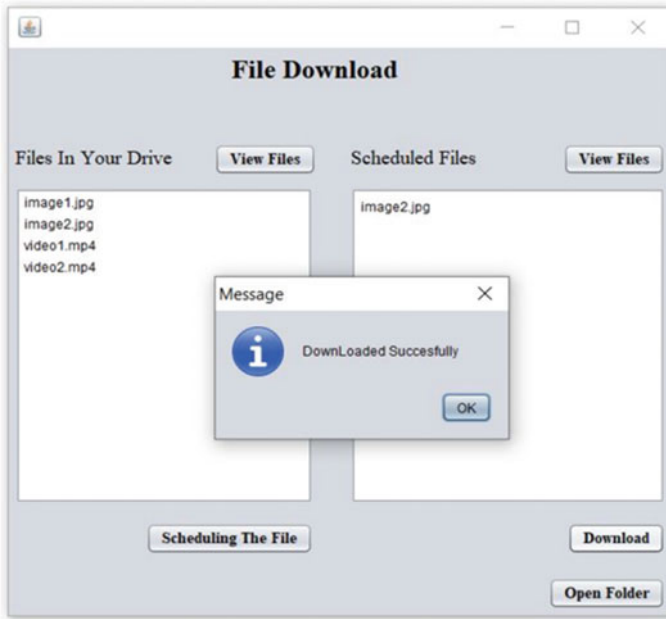


Fig. 16 File download to local storage

generated by the applications continuously. The main idea is to use the resources available in the cloud, by preferring private cloud services than public.

The proposed work provides a solution by moving the scheduling decision onto the broker and freeing the mobile nodes from scheduling and continuous resource monitoring of the cloud. The results show that using a broker provides optimal solutions for a task assignment problem and minimizing the total costs compared with a centralized scheduler without optimization. Cloud offloading is a solution to support computationally and storage-demanding applications on resource-constrained mobile devices.

All these along with the latest research and huge future scope makes mobile cloud computing strong in today's competitive market for applications requiring high performance. Various areas of the day-to-day life such as the education system, social network platforms, agricultural field, health industry, and government offices have high interest in this domain and have increased their research toward this end. In conclusion, the proposed model is not an end, but the first step in how to solve modern problems.

References

1. Mahmoodi SE, Uma RN, Subbalakshmi KP (2016) Optimal joint scheduling and cloud offloading for mobile applications. *IEEE Trans Cloud Comput.* <https://doi.org/10.1109/TCC.2016.2560808>
2. Tushar M, Assi C, Maier M et al (2014) Smart microgrids: optimal joint scheduling for electric vehicles and home appliances. *IEEE Trans Smart Grid* 5(1). <https://doi.org/10.1109/TSG.2013.2290894>
3. Ewaisha A, Tepedelenlioglu C (2016) Delay optimal joint Scheduling-and-Power-Control for cognitive radio uplinks. In: *IEEE global communications conference: cognitive radio and networks (Globecom'16—CRN)*. <https://doi.org/10.1109/GLOCOM.2016.7841726>
4. Shirazi E, Zakariazadeh A, Jadid S (2015) Optimal joint scheduling of electrical and thermal appliances in a smart home environment. *Energ Convers Manag* 106:181–193. <https://doi.org/10.1016/j.enconman.2015.09.017>
5. Dingwen Y, Lin Hsuan-Yin, Widmer J, Hollick M (2018) Optimal joint routing and scheduling in millimeter-wave cellular networks. 1205–1213. <https://doi.org/10.1109/INFOCOM.2018.8485929>
6. Upadhyay RD (2019) An SOA-based framework of computational offloading for mobile cloud computing. *Electronic theses and dissertations*, p 8185
7. Kumar K, Lu YH (2010) Cloud computing for mobile users: can offloading computation save energy? *Computer* 43. <https://doi.org/10.1109/MC.2010.98>
8. Ma X, Zhao Y, Zhang L, Wang H, Peng L (2013) When mobile terminals meet the cloud: computation offloading as the bridge. *IEEE Mag Netw* 27(5). <https://doi.org/10.1109/MNET.2013.6616112>
9. Vallina-Rodriguez N, Crowcroft J (2013) Energy management techniques in modern mobile handsets. *IEEE Commun Surv Tutor* 15(1). <https://doi.org/10.1109/SURV.2012.021312.00045>
10. Flores H, Hui P, Tarkoma S, Li Y, Srirama S, Buyya R (2015) Mobile code offloading: from concept to practice and beyond. *IEEE Commun Mag* 53(3). <https://doi.org/10.1109/MCOM.2015.7060486>
11. Balakrishnan P, Tham CK (2013) Energy-efficient mapping and scheduling of task interaction graphs for code offloading in mobile cloud computing. In: *IEEE/ACM international conference on utility and cloud computing (UCC)* <https://doi.org/10.1109/UCC.2013.23>
12. Nir M, Matrawy A, St-Hilaire M (2014) An energy optimizing scheduler for mobile cloud computing environments. In: *IEEE conference on computer communications workshops (INFOCOM workshops)*. <https://doi.org/10.1109/INFOCOMW.2014.6849266>
13. Ou S, Yang K, Zhang J (2007) An effective offloading middleware for pervasive services on mobile devices. *Pervasive Mob Comput* 3(4). <https://doi.org/10.1016/j.pmcj.2007.04.004>
14. Kavitha K (2018) Implementing joint scheduling approach in cloud computing for energy optimization. *Int J Pure Appl Math* 118(9). ISSN 1314-3395

Comparative Performance Analysis of Routing Topology for NoC Architecture



E. G. Satish and A. C. Ramachandra

1 Introduction

The network with Internet allows the computer to connect with various devices and communicate with different applications via any channel. There are different communication networks used for communication like local area network (LAN), metropolitan area network (MAN), and wide area network (WAN). Local area network (LAN) connects the network devices in smaller area like home and school. And, it is easily designed and computed. The communication channel used in LAN is twisted-pair cable and coaxial cables. LAN uses the bus, ring, and star topologies. Metropolitan area network (MAN) covers the larger area than LAN and lesser area than WAN. Wide area network (WAN) connects the larger area. A WAN could be a connection of LAN connecting to other LANs via telephone lines or radio waves.

Day by day, number of devices are connecting to network in more. For this, connecting cables required are more, and it increases the complexity. To reduce this, most of the VLSI systems are designed based on the system on chip (SoC). A system on chip is a single chip that integrates multiple functionalities. Most of the SoCs use the bus architecture. The speed and amount of logic built into a VLSI chip are increased day by day. This leads to the underperformance of system by using the bus topology. To overcome this problem, bus topology is replaced by network on chip which became a backbone for all the systems. A network on chip (NoC) is an on-chip interconnect technology used on SoC designs for efficiently interconnecting design blocks. In system on chip (SoC) to achieve the communication requirements which satisfies the properties like flexibility, scalability, and high bandwidth, the basic

E. G. Satish (✉) · A. C. Ramachandra
Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India
e-mail: satish.eg@nmit.ac.in

A. C. Ramachandra
e-mail: ramachandra.ac@nmit.ac.in

solution is implementation of network on chip (NoC). Figure 1 shows the generic NoC architecture. The generic network on chip architecture comprises with several processing elements (PE). PE can be audio, video cores, or memory elements, etc. Each processing nodes (PE) is connected to a router via a network interface. The routers are often connected in a predefined fabric. These processing elements can be placed in different topologies. There are many types of topologies like bus, ring, star, mesh, and torus. Most of the NoCs use the mesh and torus topologies due to their regularity and modularity. In Fig. 2, the memory core manager is mapped with the task mapping and scheduling unit with resource allocator which is connected to router for communication.

There are many types of topologies such as bus, ring, star, mesh, and bull. Most NoCs use mesh and bull topologies due to their regularity and modularity. Next, Fig. 4 shows the NoC router architecture. Reconfigurable calculation is necessary in such topologies so that the performance of all applications is acceptable. The set of cores that considers all applications will be assigned to routers in such a way that the established objective is met. The reconfigurable architecture is a computer architecture that combines some software flexibility with high hardware performance.

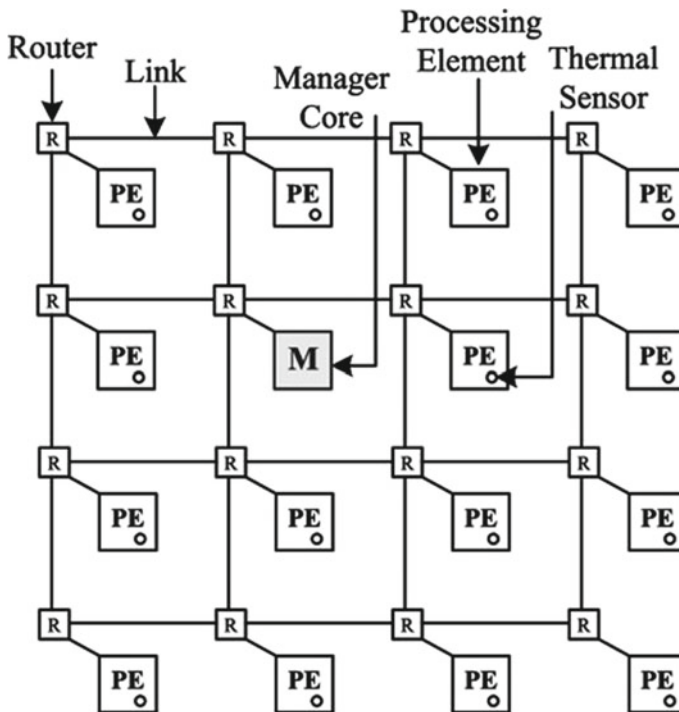
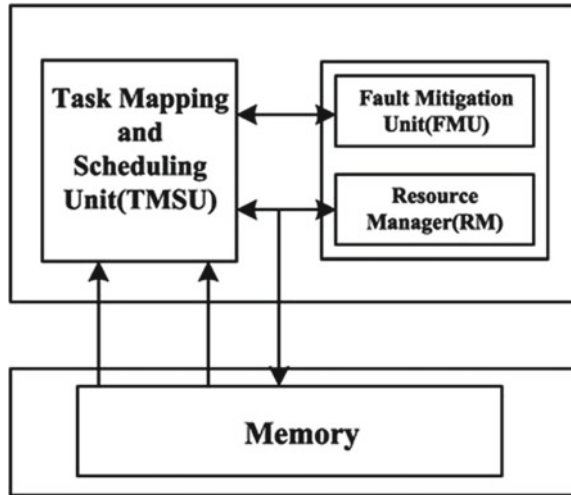


Fig. 1 NoC architecture with manager core and various processing elements (PE) linked with router

Fig. 2 Memory manager



The remaining of this paper is expressed as follows. In Sect. 2, a summary of NoC-related work is provided; in Sect. 3, the architecture of the NoC router is explained, and in Sect. 4, the NoC routing algorithm is explained. And in Sect. 5, the topology for NoC is explained. In Sect. 6, average latency is discussed and is concluded in Sect. 7.

2 Related Work

Madhubala et al. introduce a network on chip (NoC) design for interaction among the IP cores of multi-processor in a system on chip (SoC). The performance of multi-core systems is increased by using NoC architecture. The king torus topology is adopted to improve the system performance in parallel distributed applications by minimizing the processing time [1]. And, routing algorithms are used, and the router hops count is decreased from source and destination. So, the system performance is increased, and the speed is maximized by reducing the utilization of area. Madhubala et al. [2] introduced and explained the different routing algorithms which are based on mesh topology architecture. In 4×4 king mesh topology, the XY routing algorithm and weight-based path selection algorithms are implemented. The performance is better by adopting the weight-based path selection algorithm than the implementation of XY routing algorithm. The number of hops utilized in the network from the sender to receiver is reduced which lowers the latency delay between the source and destination. So, the weight-based path selection algorithm reduces the utilization of power and area. Upadhyay et al. [3] proposed for mesh of tree (MoT) topology-based network on chip (NoC) design with the multi-application mapping using the reconfigurable

Table 1 Comparisons of performance evaluation with routing mechanism or topology

S. No.	Routing mechanism or topology	Technology	Performance evaluation	Tools
1	Oblivious routing algorithm	Based on weight, hop-to-hop router transfer of packets	Reduced path selection and processing time	Xilinx 14.2
2	XY routing and weight-based path selection routing	Path selection technology in hop-to-hop router transfer	The hops count is decreased between source and destination	Xilinx 14.2
3	Mesh of tree topology	Two-phase particle swarm optimization	Communication cost is minimized	TGFF tool
4	Modified XY routing	On-demand buffer allocation	Latency decreases	NIGRAM 2.1
5	Mesh topology	Cluster-based deployment scheme	Reduced delay	Xilinx 14.2
6	Static and dynamic routing	Fault tolerance technique	Network performance increases 7	NIGRAM 2.1
7	Minimal routing	Wormhole switching technique	It gives the better clock frequency	Xilinx ISE 13.1
8	XY routing	Store and forward technique	Tours NoC architecture is designed	Xilinx ISE 13.1

architecture. The hops count also reduced in reaching the packets to its destination from the source in communication model.

The communication cost can be decreased in the reconfigurable architecture by implementing the two-phase particle swarm optimization PSO. In phase one [5], the global mapping can be achieved by mixing of various applications, and in two-phase, switching the cores making use of the multiplexers the reconfiguration can be done. The on-demand buffer allocation technology modifies the XY routing to minimize the latency [4]. The communication delay is reduced by adopting the cluster-based deployment scheme based on mesh topology, and routing mechanism's fault tolerance technique improves the network performance with making use of static and dynamic routing mechanism [9]. Table 1 shows the various routing comparison of performance of NoC by implementing the different routing mechanism.

3 NoC Router Architecture

In general, packets travel through routers to reach their destination. The following figure shows the architecture of the NoC router [15]. Based on the topology, the Noc

can be differentiated into two types regular topology and custom topologies. Regular topology is the best used in general purpose computing like torus, mesh or hypercube, and many, whereas the custom topologies are well-suited for the application-specific well-defined on-chip communication pattern. The network on chip flit format as shown in Fig. 3 with in a regular topology.

The router consists of several input ports, several output ports, a crossbar switch, registration, and control logic. For each router, the received packet is pushed and stored in the input buffer. Then, the channel arbitration and routing decision can be controlled and maintained by the control logic in the router. And then, the packet crosses through the crossbar and reaches the next available router. This process is repeated until the package reaches its destination. In each, the control logic of the routing unit is a finite state machine (FSM). Process the packet header to calculate an

Fig. 3 Network on chip flit format

Source id	Destination id	Pay load
-----------	----------------	----------

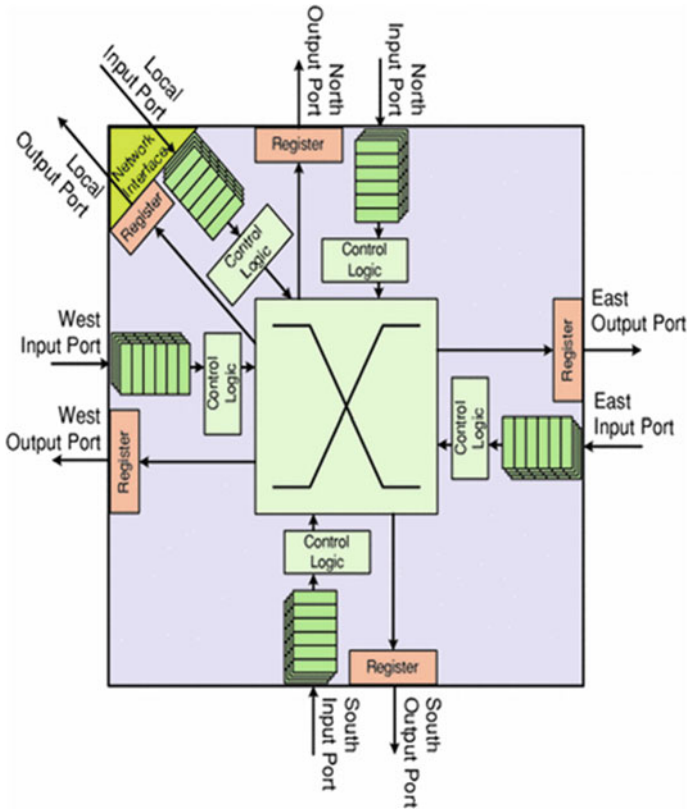


Fig. 4 NoC router architecture

appropriate output channel and generate requests for that output channel accordingly [12].

4 Routing Algorithms in NoC

The performance of a network on chip significantly depends on the routing algorithms implemented. Distributed routing algorithms are implemented in many existing NoC architectures to build the NoC platform. The flits are routed first in the X direction, until reaching the Y coordinate, and then in the Y direction, if another packet is using a network jump, the flit remains locked on the switch until the route is released. In a NoC architecture, the routing algorithm plays a vital role. The path between the source and destination of a packet, i.e., the packet routes between the hops defines the routing algorithm based on the implementation like lookup table and finite state machine. The main aim of the routing algorithm is to prevent the live lock, deadlock, and starvation conditions. For an interconnection networks, the major attributes of routing algorithm are adaptivity, connectivity, starvation and dead lock freedom, and fault tolerance. The combined routing algorithm with a topology in order to reduce the loss of packet during the reconfiguration stage. Fault-tolerant routing algorithm allows high reachability with increased scalability and autonomous in network size. Routing algorithm plays a vital role for efficient packet delivery in NoC design communication, mainly two varieties of routing algorithms, namely deterministic routing and adaptive routing. In deterministic routing, the route is fixed and determined prior to send the packet in the network, and adaptive routing algorithm selects the route depending on the status of network like network loads and queues.

Routing algorithm

Step 1: Read the incoming packet's destination address.

Step 2: Check the addresses of the current router with the incoming packets destination address.

If both address matches

then the packet is sent to the local core of the current router.

Else If the address of destination node matches to x (or y -axis) of the current router address

then the packet is sent to the neighboring router on the y -axis (or x -axis) towards the final destination.

Else

the stress values of current router’s neighbors are checked with the stress value of the destination then the packet is sent to the neighbor node which holds the minimum or low stress value.

5 Interconnect Topology for NoC

Topology is the placement of different nodes in the network and how the connectivity is achieved between the nodes. The torus interconnect topology [11] architecture achieves fault-tolerant and easy scalable interconnection networks for high-performance super computers. If the application runs on sub system or on a full system, it requires an add on hardware example switches to accommodate the network topology for any given job allocation scheme [7]. The number of portion switches are reduced by the designers to provide much flexibility mechanism for any given various job sizes. And, the below fig shows the tours NoC architecture. There is no single optimal topology for all the applications, depending on the specific requirements or application, and on various constraints, the suitable topology is selected from the various topology. Both the deterministic and adaptive routing algorithms are implemented based on the neighbor nodes load status. If the load congestion is more, then try to route the packets with low congested path. The low congested paths can be determined by using the 2-bits load value as per Table 2; the load can be known in the network topology. The load values are the information received from all four north, east, west, and south neighbors. Collectively, all four neighbor’s take responsibility to route the packets from source to destination.

6 Ideal Latency of NoC

The best topology plays a vital role to meet the characteristics of bandwidth and latency of any application. The goal is to achieve the maximum bandwidth with decreasing the latency delay. The throughput and latency depend on routing, flow control, design of the router, and topology. The latency is the time needed to traverse a packet from source to destination node via hops within communication network.

Table 2 2-bit quantized load in network topology

2-bit quantized load	Means
00	Ready to receive—free port
01 or 10	Threshold (buffer is half empty)
11	Buffer is full—busy port cannot receive anymore

Table 3 Average latency in ms for different topologies

Topology	Average latency in ms of 4 × 4 network	Average latency in ms of 8 × 8 network
Mesh	473.277	572.76
Full mesh	264.196	317.66
Torus	440.19	295.6
XX torus	433.661	281.84

$$\text{Average latency} = \frac{\sum_{i=1}^N \text{Packet latency}}{N}$$

N = the total number of packets arrives destination node, and packet latency refers to the latency of packet at the i th packet.

The ideal average latency T can be estimated by taking the average delay in routing for all participating nodes in the topology with no congestion, i.e., the topology with zero load. The average latency can be calculated with

$$T = H * dr + D/t + L/b$$

H = hop number average—from source to destination node.

dr = delay in routing.

D = distance from source node and destination node.

T = transmission speed.

L = packet length.

b = bandwidth

$$\text{Average number of packet Hops} = \frac{\sum_{i=1}^N \text{Packet Hop}}{N}$$

N = the total number of packets reached at destination node from the source.

packet hop = the number of routing hops essential to receive the packet to destination from its source.

The total number of packet hops divided by the total number of packets successfully reached to its destination. Table 3 shows the average latency of the various topologies which is simulated from the NS-2 tool [16].

7 Conclusion

As the torus interconnect topology is simplest topology and effective to design compared to any other interconnect topology, the NoC routing algorithm will make use of simple XY routing avoid deadlock in routing algorithm. So, the model is

required for implementation of 6D torus with high fault tolerance with XY routing by using the Xilinx ISE 13.1 simulator for simulation. In present technical era, high-performance computing systems are more popular and becoming much integrated and challenging to manage interconnections between the various participating nodes in the connected network. Torus interconnect topology architecture is going to become the fundamental base technology for all the future network on chip architecture.

References

1. Madhubala T, Karthika P, Sobana S (2017) A competent performance estimation of king torus topology. In: International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, pp 214–217
2. Madhubala T, Gayathri G, Karthikeyan I (2017) A competent performance analysis of king mesh topology. In: International conference on electrical, instrumentation and communication engineering (ICEICE), Karur, pp 1–5
3. Upadhyay M, Shah M, Bhanu PV, Soumya J, Cenkeramaddi LR (2019) Multi-application based network-on-chip design for mesh-of-tree topology using global mapping and reconfigurable architecture. In: International conference on VLSI design and international conference on embedded systems (VLSID), Delhi, NCR, India, pp 527–528
4. Lochana ASR, Arthi K (2017) Over looped 2D mesh topology for network on chip. In: 1st international conference on innovations in information and communication technology (ICICT), Chennai, India, pp 1–5
5. Phing NY, Warip MNM, Ehkan P, Zulkefli FW, Ahmad RB (2017) Topology design of extended torus and ring for low latency network-on-chip architecture. 15(2):869–876
6. Bahrebar P, Stroobandt D (2017) Adaptive and reconfigurable bubble routing technique for 2D Torus interconnection networks. In: International symposium on reconfigurable communication-centric systems-on-chip (ReCoSoC), Madrid, pp 1–8
7. Chemli B, Zitouni A (2017) Architecture and performances comparison of network on chip router for hierarchical mesh topology. In: 2017 international conference on engineering & MIS (ICEMIS), Monastir, pp 1–4
8. Kullu P, Tosun S (2019) MARM-GA: mapping applications to reconfigurable mesh using genetic algorithm. In: Euromicro conference on digital system design (DSD), Kallithea, Greece, pp 13–18
9. Kunthara RG, Neethu K, James RK, Sreeba SZ, Jose J (2019) DoLaR: double layer routing for Bufferless mesh network-on-chip. In: TENCON conference (TENCON), Kochi, India, pp 400–405
10. Khan S, Anjum S, Gulzari UA, Afzal MK, Umer T, Ishmanov F (2018) An efficient algorithm for mapping real time embedded applications on NoC architecture. IEEE Access 6:16324–16335
11. Priya S, Agarwal S, Kapoor HK (2018) Fault tolerance in network on chip using bypass path establishing packets. In: International conference on VLSI design and international conference on embedded systems (VLSID), Pune, pp 457–458
12. Pereira LMV, Melo DR, Zeferino CA, Bezerra EA (2018) Analysis of LEON3 systems integration for a network-on-chip. In: Latin-American test symposium (LATS), Sao Paulo, pp 1–3
13. Chen K, Wang T (2018) NN-Noxim: high-level cycle-accurate NoC-based neural networks simulator. In: International workshop on network on chip architectures (NoCArc), Fukuoka, pp 1–5
14. Upadhyay M, Shah M, Bhanu PV, Soumya J, Cenkeramaddi LR (2019) Multi-application based network-on-chip design for mesh-of-tree topology using global mapping and reconfigurable

- architecture. In: International conference on VLSI design and international conference on embedded systems (VLSID), Delhi, NCR, India, pp 527–528
15. Tsai W-C, Lan Y-C, Hu Y-H, Chen S-J (2012) Networks on chips: structure and design methodologies. Hindawi Publ Corp J Electr Comput Eng 2012. Article ID 509465
 16. Phing NY, Warip MNM, Ehkan P, Zulkefli FW, Ahmad RB (2017) Topology design of extended torus and ring for low latency network-on-chip architecture. TELKOMNIKA 15(2):869–876

LoRaWAN-Based Communication Protocol for Wearable Safety Devices in Mining Fields



Namratha Karanth , Deepak Choudhary , Jaideep Francis Reddy ,
and Ubay Athulla 

1 Introduction

This paper concentrates on administration, monitoring, and tracking health by processing the information collected from the wearable devices (also called as the end devices). The major issue faced in the mining industry is establishing proper communication between the miners and their supervisors. In the event of explosions and disasters, the existing communication framework is damaged. The re-establishment of connection is time-consuming. Therefore, a new framework is required which is robust, has high signal penetrating power, and ability to reach very long distances. Long-range (LoRa) communication helps in providing efficient solution in this case.

The effect of spreading factor (determining the number of chips in a symbol) and other physical layer (PHY) settings are discussed along with necessary simulations. The pictorial representation of Fresnel zone for each case is provided. The optimum position of placing the LoRa gateway device (which is very crucial) is also proposed with supporting calculations.

This paper comprises of three sections. Section 1 is the introduction and underlying theories concerning the problem; Sect. 2 gives the proposed network model, calculations, the implementation along with the simulation results supporting the same; Sect. 3 concludes the paper.

N. Karanth · D. Choudhary (✉) · J. F. Reddy
REVA University, Bengaluru, India

U. Athulla
Elmeasure India Private Limited, Bengaluru, India
e-mail: ubay@elmeasure.com

1.1 Introduction to LoRa

LoRa is a long-range, low-power, low-bitrate, wireless communication system which provides an infrastructural solution to the Internet of things. The ability of the module to adapt to different frequencies in order to optimize data rates, airtime, and energy consumption in the network is termed as adaptive data rate (ADR) which is the deciding factor in using LoRa as the protocol for our mining safety application. The entire model is implemented using star-of-stars topology [1, 2].

LoRa employs industrial, scientific, and medical (ISM) bands in the range of 169, 433, 868 MHz (in Europe and India) and 915 MHz. The end devices (also called as the motes) are connected via ISM bands to the gateway(s). The gateway(s) are connected to the network server via the standard IP network. LoRa defines the lower physical layer, and LoRaWAN defines the upper networking layers. LoRaWAN is a media access control (MAC) layer protocol that acts mainly as a network layer protocol. Data is received by multiple gateway(s) which forward the data package to the application servers via the network servers over an IP network [1].

LoRaWAN has three different classes of end device to address different application, and this paper highly suggests Class A bidirectional communication for mining sectors. Class A bidirectional model allows two-way communication that can schedule an uplink transmission based on their own needs. (In this case, we send pre-defined periodic heart rate of miners.) In these end devices, one slot uplink transmission is followed by two short downlinks. Class A devices have the least power utilization, but it offers less flexibility on downlink transmissions.

1.2 The Issue of Underground Mining Communications

Every day, one out of three miners die due to a lack of effective communication. When an accident occurs, the existing wired communication framework collapses, and it takes a lot of time and manpower to re-establish the communication network to start rescue operations. In the 2010 Copiapó mining accident [3], 33 miners were trapped for 64 days underground and at five kilometers from the entrance and it took them almost two weeks to establish a reliable communication network due to which the rescue operation was delayed.

Improper explosions lead to triggering massive widespread explosions which further collapse all the existing communication frameworks. Therefore, trapped survivors die from the poisoning of fumes like hydrogen sulfide and methane. While others die of dust explosions, usage of improper explosives, collapsing of mining stopes, the disintegration of huge boulders. These cases have proved that the existing wired communication system is unreliable and puts the miners' lives at stake. Hence, a wireless solution is required in a hazardous work environment for swift rescue operations.

1.3 Advantages of LoRa

In today's world, the mining sector lacks technological advancement; i.e., the communication framework used in most of the mining fields still depends on obsolete wired communication. LoRa provides a large area coverage with the installation of a few gateways at appropriate distances. LoRa having the ability to communicate via a multipath channel when there is no direct channel of communication between the two ends of the network model makes it the choice of communication protocol for wireless data transmission. The adaptive data rate LoRa provides further reinforces it as the *modus operandi* in the proposed scheme.

As wireless fidelity (Wi-Fi) and ZigBee are limited to their short range (up to 100 m) and high-power consumption, they are less preferred as wireless solutions in a mining environment more than 100 m below the earth's surface. Although ZigBee consumes lesser power than Wi-Fi it is still unmatched to the low-power attributes of LoRa. Wired communication protocols are sometimes rendered useless in catastrophes, whereas the LoRa gateways are usually placed on the surface far away from the area of impact. Unlike NB-IoT which is best suited for primarily static assets that have fixed locations, LoRa proves its advantage as it can be used with moving objects (end nodes). The discussed points make it clear for the need of a wireless communication protocol that triumphs in both range and minimum power consumption, and therefore, LoRa presents itself as a suitable solution [4].

LoRa sensors need to be serviced only once every 5–10 years. This helps the companies save a lot of money, which otherwise would be spent on human labor to perform the regular checks and setup. Although many IoT technologies are exclusive, the LoRa network management protocol allows the individual(s) and organizations to set up their own LoRa base stations provided the regulations over bands are followed. LoRaWAN uses unlicensed spectrum, and this results in 1% duty cycle in Europe which limits volume and density of traffic as well as the ability of base station to control the network and reduce the traffic. The exceptional battery life results in lower maintenance costs. Through gateways, it is possible to locate the miners where a mishap occurs without using GPS. With a wide range of merits, LoRa demonstrates the optimum choice for communication networks [5].

2 Proposed Network Model

The flowchart shown in Fig. 2 shows the end device (in our case the safety module) which initializes before the miner starts working and sends data of an individual's heart rate once every hour. As the safety module given to each worker has a unique ID which is identified at the gateway, the updated data is logged into the worker's profile at the webserver.

When the end node sends an uplink, it is acknowledged by the network server by sending a downlink that the message has been received. When there has been

no response from the worker, the network server times the window where there has been no response. If the time window exceeds 30 min, an emergency signal is sent as a text message to the authorities concerned and when the worker responds, the network server records the response and forwards it to the application server.

When the worker declares an emergency, a message is sent to the authorities so that swift and effective safety measures can be initiated (Fig. 3), which triggers the beep alarm embedded in the safety module to help sound navigate. Here, maintaining a track record of the worker's heart rate is of the essence to ensure a safer work environment for miners [6].

2.1 Implementation of LoRa in Major Mining Sectors

Accidents occurring due to natural calamities and other external factors make it difficult to rectify wired communication protocols placing many lives at stake. Considering a Fresnel zone from the gateway to the end node, the height of the gateway is calculated to maximize the uplink quality. Fresnel zone is an imaginary ellipsoid-shaped body around the direct line-of-sight path between the end node and the gateway. The proposed safety model is applicable to four major sectors in the mining industry, and the working of the model in each sector has been elaborated below:

Underground hard rock mining These mines are usually excavated 15–30 m below the surface. Considering the distance between the gateway and the end node to be 1 km, the distance at which the gateway must be placed above the ground is calculated (Table 2). Since LoRa communication can effectively function up to 300 m below the earth's surface, a suitable network can be established by placing gateways at appropriate distances. Uplink and downlink of data packets can be achieved at a higher bitrate. Medical personnel can quickly respond to dire situations and accidents upon the data received.

Underground coal mining As these mines are excavated much deeper than the hard rock mines (up to 90 m below the surface of the earth), the signal penetration and scalability of LoRa make it a befitting network protocol to establish a connection between both ends (Fig. 1). Assuming the appropriate distance between the gateway and the end node to be as 2 km, the height at which the gateway has to be placed above the ground is calculated (Table 2) and shown in Fig. 4

Surface Mining A single gateway can cover up to 3 km of surface area, thus allowing the coverage of a large area. This enables the LoRa network model to be cost-effective and completely wireless. According to the Fresnel zone, the optimum height at which the gateway has to be placed above the ground (r) is calculated (Table 2) (Fig. 5).

Tunneling The tunnels are usually narrow and deep (up to 100 m in depth). The temperature at the bottom is high, and RSSI is severely affected at such temperatures. Installation of gateways near the holes drilled for the purpose of ventilators can be a suitable approach for establishing communication (Fig. 6).

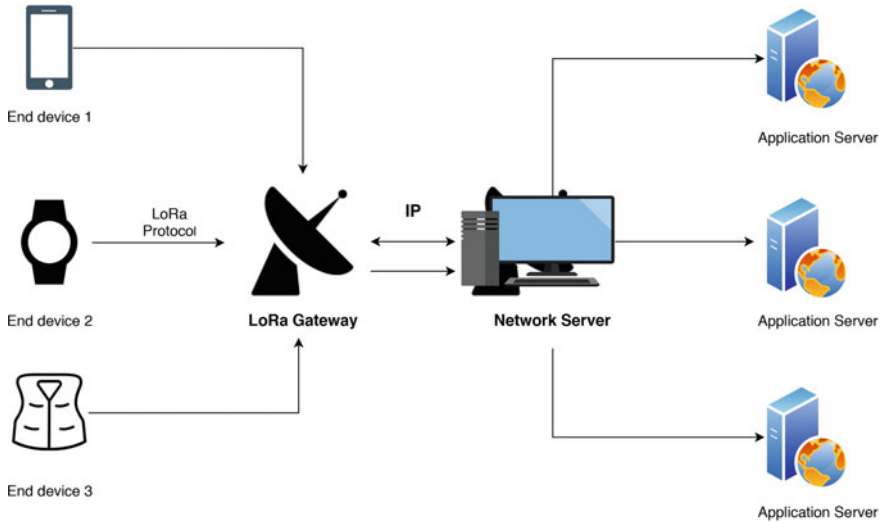


Fig. 1 Network model of LoRa mining safety applications

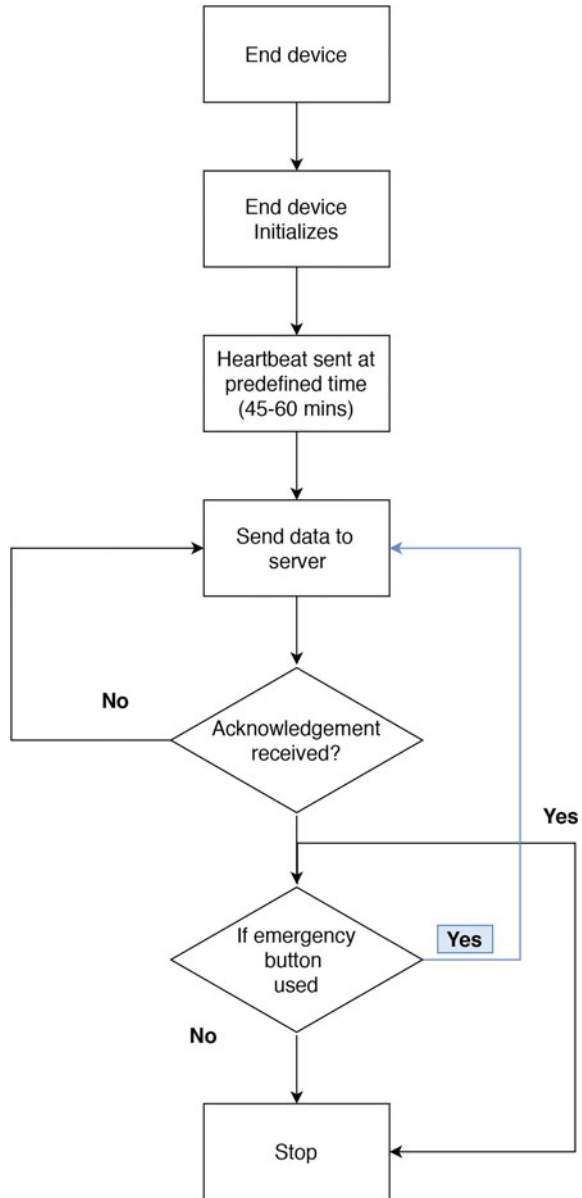
2.2 Physical Layer (PHY) Settings

When one fine-tunes physical layer settings (PHY), it gives the flexibility to acquire long-range communications with a low data rate which, in turn, impacts the network performance particularly the dependency of moderate reception in an underground environment. Impact of PHY settings on connectivity, RSSI, and sensitivity leads to better performance. The variation of PHY settings leads to two scenarios: The first one involves reducing the data rate in order to achieve a higher link quality and the latter involves having an intermediate link quality leading to some packet loss at a higher data rate. The maximum application payload depends on the selected data rate. For a node operating in harsh conditions, one should assume the worst data rate (spreading factor of 12), in which the node should not send more than about 51 bytes of data. At optimum conditions, the node can send about 222 bytes at a data rate of spreading factor = 7. In this specific application, the key objective to ensure that proper bidirectional communication is established even if data rate is compromised which leads us to the former scenario [7].

2.3 Simulation Settings and Inferences

Effect of different Spreading Factor (SF) LoRa (PHY) physical layer uses a chirp spread spectrum modulation that incorporates different SFs to tune the chirp modulation. To achieve a longer communication range, the following parameters are altered

Fig. 2 End device communication model



by reducing the end devices bandwidth (125 kHz for the below simulation): With higher airtime and sensitivity, the packet reception ratio improves along with the increment of bit redundancy and signal-to-noise ratio (SNR). Furthermore, increasing the spreading factor (7–12), increases the receiver’s sensitivity, longer transmission

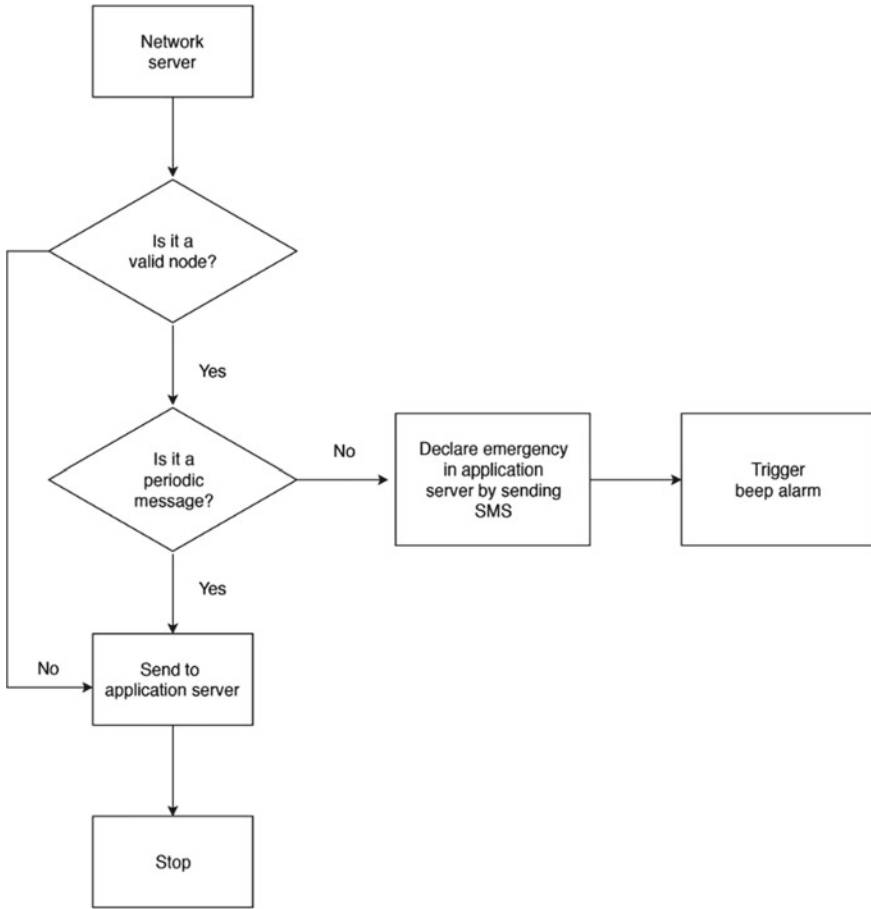


Fig. 3 Network server communication model

rate with comparatively low data rate, augmenting the communication range at its best efficiency [8, 9].

A trade-off between data rate (reception rate) and the spreading factor is the ultimatum of this paper, therefore the below simulation result [7]. To help achieve an effective signal penetration, installing numerous gateways in mining fields will ensure effective uplink and downlink, thereby improving the overall signal coverage. In case of catastrophes, one can find a relation between SFs and the RSSI when the emergency button has been pressed to find the approximate location of the survivor. Lower the SF, the closer the device is to the transmitter. Moving the gateway along the path of the mining field would yield correlations that could give us the approximate location of the survivors. This makes rescue operations faster (Fig. 7).

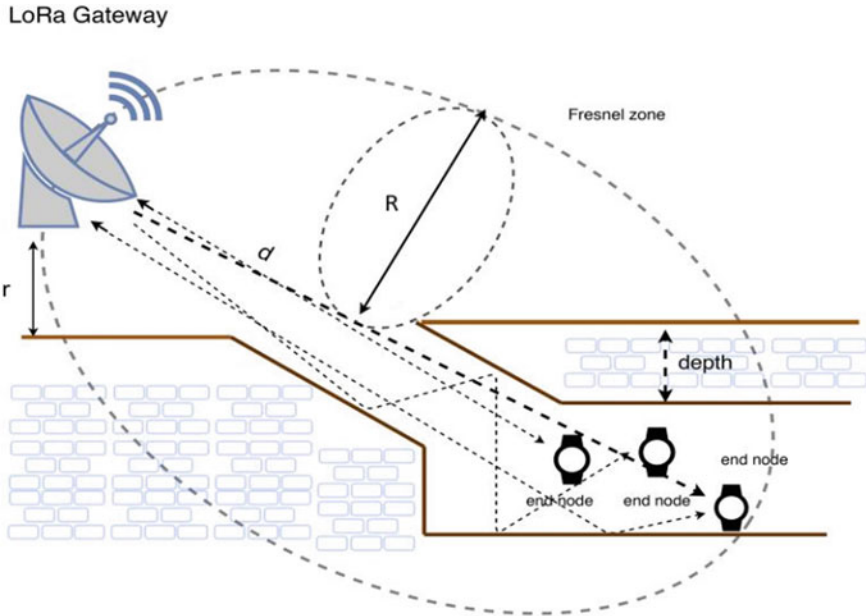


Fig. 4 Fresnel zone interpretation for underground hard rock/coal mining LoRa communication

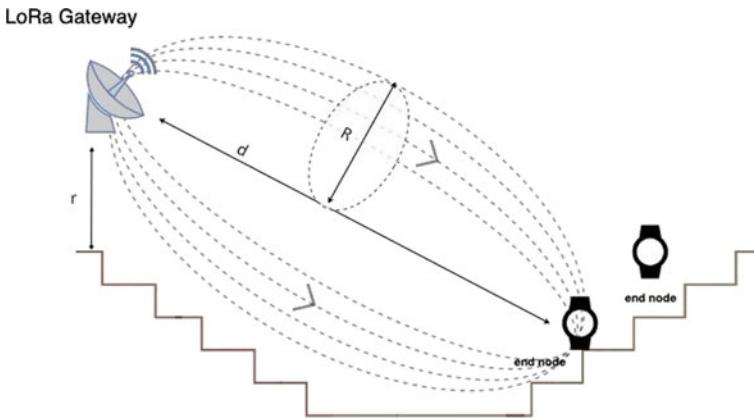


Fig. 5 Fresnel zone interpretation of surface mining LoRa communication

Symbol Rate Symbol rate (T_s) is defined as the time taken to transmit one LoRa symbol. Received LoRa symbols have been evaluated to assess the relationship between the total symbol length and time (Fig. 8).

Uplink messages use the LoRa radio packet explicit mode in which the LoRa physical header (PHDR) plus another header called PHDR-Cyclic Redundancy Check (PHDR_CRC) are included. For low data rate optimization, 125 KHz bandwidth,

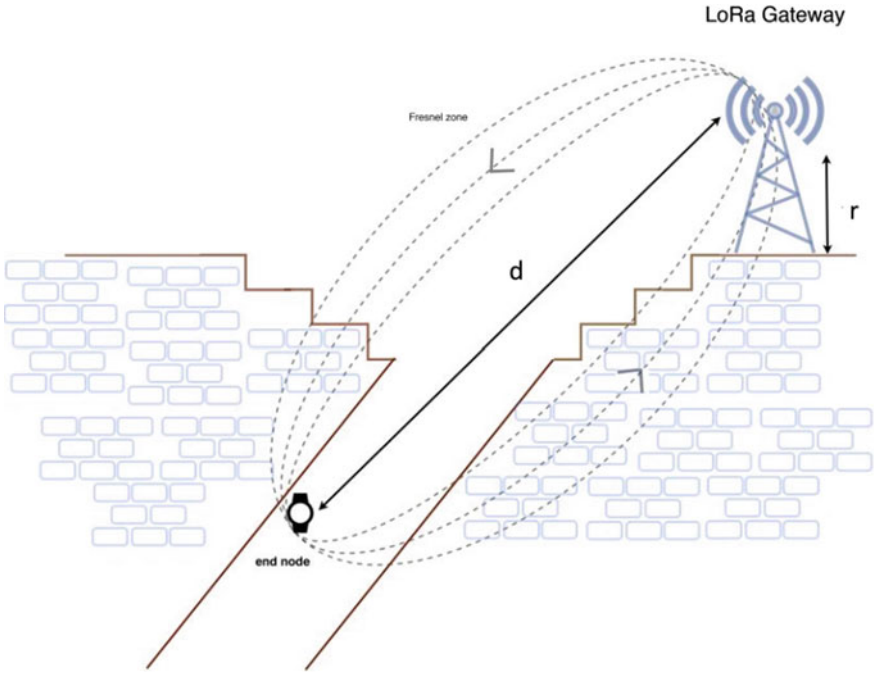


Fig. 6 Fresnel zone interpretation of tunneling LoRa communication

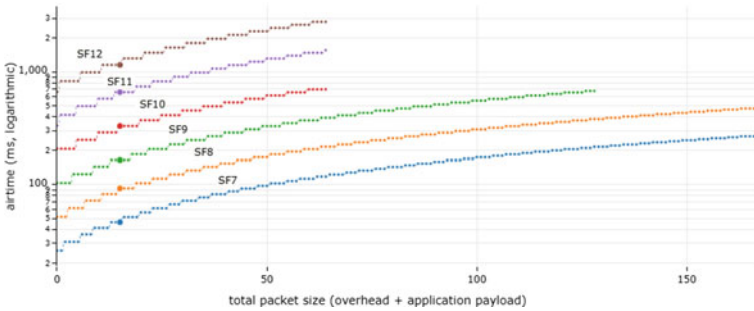
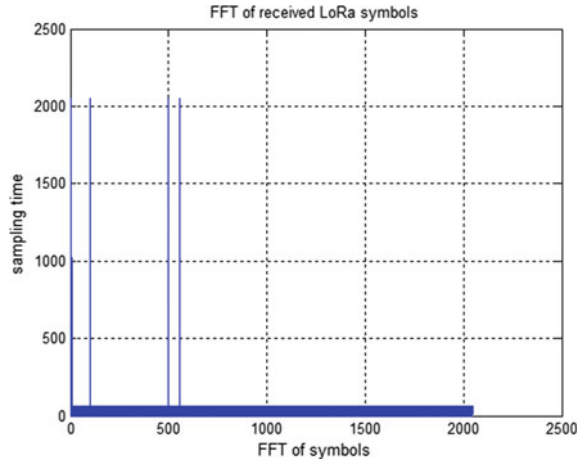


Fig. 7 Comparison of LoRa spreading factors (7–12)

and $SF = 11$ or 12 , adds a small overhead which increases robustness to reference the frequency variations over the timescale of the LoRa packet.

For the above simulation, with a sampling frequency of 125 KHz, the low-level header comprises a preamble of length 8 bytes, a physical header (PHDR) of 2 bytes, 4 bits of PHDR_CRC, a physical payload (PHY Payload) for this application as 15 bytes, and an error detection tail (CRC) of 2 bytes at spreading factor of 11, bandwidth = 125KHz is used to calculate the transmitted and received LoRa symbols. Also,

Fig. 8 FFT of the received LoRa symbols



here having a small payload, the probability of data transmission is very high given a situation with comparatively lower RSSI value [10, 11].

Effect of Received Signal Strength Indicator (RSSI) Received signal strength is determined mainly by two parameters, namely the transmission power and the distance between the transmitter and the receiver. LoRa received signal strength indicator (RSSI) can be defined as the received signal power in milli-watts and is expressed in dBm. The signal strength is considered stronger as its RSSI value approaches zero. Typically, RSSI values range between -40 and -120 dBm.

Receiver sensitivity is directly proportional to the spreading factor and inversely proportional to the bandwidth. From this, we deduce that an increase in the bandwidth increases the transmission rate. The symbol rate for a spreading factor is directly proportional to the bandwidth. The number of symbols required to transmit a payload is the major parameter for low data rate optimization, which is what this article aims for. Therefore, a bandwidth of 125KHz or lower and a spreading factor of 11–12 is optimal.

However, it is important to consider the obstacles in the path of data transmission which leads to reflections, refractions, and losses. It is impractical to have a clear path in the mining sector. Therefore, the loss of signal strength due to distance can be dealt with by installing a greater number of gateways by measuring the packet loss during transmission. Approximation in RSSI value and signal baud rate will be a breakthrough to help in the estimation of the last-known location of survivors and the location of the miners [12].

2.4 Mathematical Model

LoRa’s communication performance can be enhanced by varying the PHY settings such as bandwidth, spreading factor, receive signal strength indicator (RSSI), bitrate, coding rate, signal strength loss, symbol rate, height, and distance of the gateway(s) as summarized in the table below. The effect of each parameter has been discussed in PHY settings in detail. These mathematical models have been used in this research article.

This research article also suggests calculating the height of the earth’s curvature and the antenna polarization. If the distance between the gateway and the end device is greater than 5 km, we will have to consider the height of the earth’s curvature which will convert this into a flat earth scenario to calculate the height of the gateway. For a larger signal coverage, an omnidirectional antenna is used along with horizontal antenna polarization of both the gateway and the end nodes [13] (Table 1).

Table 1 Summary of LoRa’s configurable settings

Parameters	Values/equations
Bandwidth (BW)	125 kHz ... 500 kHz
Spreading factor (SF)	$2^{SF} = 2^7 \dots 2^{12}$ chips/symbol where SF = 7 ... 12
Coding rate (CR)	$CR = 4/(4 + N)$ where N is the number of symbols required to transmit payload
Height at which gateway is placed above the ground (r)	$r = 8.657 \times \sqrt{\frac{d}{f}}$ where d = distance from the end node to the gateway and f = frequency
Height of the earth’s curvature (h)	$h = 1000 \times \frac{d^2}{8 \times R}$ where R = Radius of the earth (which is 8504 km)
Symbol rate (T_s)	$T_s = \frac{2^{SF}}{BW}$

Table 2 Calculated height of the gateway for each mining sector

Mining sector	Height of the gateway (r) (m)
Underground coal/hard rock mining	9.29
Surface mining	6.57
Tunneling	13.14

2.5 Results

This paper has proposed the system design for tracking and monitoring people in underground mining locations by using the IoT-LoRa based platform. The system consists of LoRa end device (mote) as wearable devices, the LoRa gateway, and the local or cloud server. The data transmission between the LoRa mote and gateway is operated over the ISM frequency band. The Wi-Fi or other GSM networks are used for the connection between the gateway to the network server and the cloud. This device can help the concerned authorities to monitor miners as well as help them in catastrophes.

The extensive report on the simulation to make robust transmission settings is attained. The preference toward high scalability of LoRa module over Zigbee, NB-IoT, and Wi-Fi for deep mining application(s) has been discussed. The following parameters of the PHY settings were altered and simulated to make data transmission more reliable between the gateway and mote, decrease signal strength loss, and reduce the payload size: bandwidth = 125 kHz or lower, spreading factor = 10–12, preamble length = 8, sampling frequency = 125 kHz, and coding rate = 4/5. Tweaking the above parameters also results in enhanced network performance and lower power consumption.

3 Conclusion

The implementation of LoRa has been projected from the perspective of the end device side while also talking about the role of the network server. This paper provides a comprehensive study on the Fresnel zones for four types of mining scenarios to maximize link quality and RSSI. The optimum height at which the gateway is to be placed above the ground for superior link quality is tabulated. Furthermore, the paper provides a simulation result of comparison of different spreading factors and its effect on frequency–time, symbol rate. The entire design will be able to provide a cumulative health report of the miners and raise a red flag in case of any emergencies. The feasibility of the proposed device has been thoroughly studied. The future work of this device is to be implemented in real-time scenarios.

References

1. Ferré G, Giremus A (2018) LoRa physical layer principle and performance analysis. In: 2018 25th IEEE international conference on electronics, circuits and systems (ICECS), Bordeaux, pp 65–68. <https://doi.org/10.1109/ICECS.2018.8617880>
2. Hayati N, Suryanegara M (2017) The IoT LoRa system design for tracking and monitoring patients with mental disorder. In: 2017 IEEE international conference on communication, networks, and satellite (Comnetsat), Semarang, pp 135–139

3. Illiano C (n.d.) Rescue near for Chile miners trapped for two months. Retrieved 9/9/2020, from Reuters Alert Net <http://www.alertnet.org/thenews/newsdesk/N08212025.htm>
4. Georgiou O, Raza U (2017) Low power wide area network analysis: can LoRa scale? *IEEE Wirel Commun Lett* 6(2):162–165
5. Sinha R, Yiqiao W, Hwang S-H (2017) A survey on LPWA technology: LoRa and NB-IoT. *ICT Exp* 3. <https://doi.org/10.1016/j.ict.2017.03.004>
6. Petajajarvi J, Mikhaylov K, Yasmin R, Hämäläinen M, Iinatti J (2017) Evaluation of LoRa LPWAN technology for indoor remote health and wellbeing monitoring. *Int J Wirel Inf Netw.* <https://doi.org/10.1007/s10776-017-0341-8>
7. Cattani M, Boano CA, Römer K (2017) Evaluation of the reliability of LoRa long-range low-power wireless communication. *J Sens Actuator Netw* 6:7. <https://doi.org/10.3390/jsan6020007>
8. Augustin A, Yi J, Clausen TH, Townsley W (2016) A study of LoRa: long range and low power networks for the Internet of Things. *Sensors* 16:1466. <https://doi.org/10.3390/s16091466>
9. Waret A, Kaneko M, Guitton A, El Rachkidy N (2019) LoRa throughput analysis with imperfect spreading factor orthogonality. *IEEE Wirel Commun Lett* 8(2):408–411
10. Liu Y, Zhu H, Yu TTA, Tsang KF, Wu CK, Hung FH (2018) Packet loss analysis for LoRa-based heart monitoring system. In: *IECON 2018—44th annual conference of the IEEE industrial electronics society*, Washington, DC, pp 4668–4671
11. Ibáñez CL, Masnou BM, Ferré RV, Gomez C (2017). Modeling the energy performance of LoRaWAN. *Sensors* 17:2364. <https://doi.org/10.3390/s17102364>
12. Raza U, Kulkarni P, Sooriyabandara M (2017) Low power wide area networks: an overview. *IEEE Commun Surv Tutor* 19(2):855–873. <https://doi.org/10.1109/COMST.2017.2652320>
13. Dambal V, Mohadikar S, Kumbhar A, Guvenc I (2019) Improving LoRa signal coverage in urban and sub-urban environments with UAVs

Software Component Selection in CBSE Considering Cost, Reliability, and Delivery Delay Using PSO-integrated MVO and ALO



Anjali Banga and Pradeep Kumar Bhatia

1 Introduction

1.1 Component-based Software Systems

Research work has focused on cost estimation, reliability, and delivery delay for software component selection. Present work has detected the best cost, the best delivery delay, and the best reliability to select the best component. Here, three different mechanisms have been applied; first is traditional PSO, second is PSO-integrated MVO, and third one is PSO-integrated ALO. In Sect. 1, the component-based software system is discussed along with MVO and ALO optimized. Then, the related researches are explained in Sect. 2. Working of PSO, derivation of PSO-integrated MVO, and PSO-integrated ALO is explained in Sect. 3. The research methodology is representing the process from of work in Sect. 4, while the proposed work and results are represented in section Sect. 5. Here, the simulation to find the best value for cost, delivery delay, and reliability is performed using PSO, PSO-integrated MVO, and PSO-integrated ALO. The component dependency chart has been developed to select the best component considering cost, reliability, and delivery delay. Section 6 and Sect. 7 are presenting the conclusion and future scope, respectively.

CBS systems are the output results of the evolution of large-scale and complex systems. Rather than CBSSs, regular development mechanisms are less applicable in development of software. These mechanisms are suspected to have low productivity, high development cost, uncontrollable software quality, and high risk to move to new machinery. In component-based software systems, proficient components are used to essentially reduce the development cost and time to market. These components

A. Banga (✉) · P. K. Bhatia

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India

are very efficient to enhance the maintainability, reliability, and overall quality of software systems.

It is very difficult to check the CBSS class even when the compositional techniques are used in any specific systems. It is not easy when its components and structure are altered time to time.

In existing researches, two evolution processes are followed to describe the component-based software engineering (CBSE). First evolution process is the evolution of components for reuse. Second evaluation process is the evolution of component-based software systems (CBSSs) with reuse. In these systems, the additional components are integrated that can be expanded individually [1].

1.2 MVO (Multi-verse Optimizer)

Multi-verse optimizer (MVO) is considered as an innovative and effective nature-inspired optimization mechanism. It has been introduced by Mirjalili et al. The major inspiration behind this mechanism is depending on cosmological concepts. It is known as an innovative meta-heuristic optimization method that is known multi-verse optimizer. It is capable to solve OPF problem. MVO technique has been considered as a biological as well as sociological-inspired mechanism. This mechanism is depending on three concepts in cosmology. These concepts are white hole, black hole, and wormhole. The capabilities of MVO are search convergence rate rapidly. MVO makes use of roulette wheel selection. This mechanism supports in handling continuous and discrete optimization issues.

1.3 ALO (Antlion Optimizer)

The antlion optimizer is a recent meta-heuristic optimizer. In has been known as ALO also in its short form. This optimizer models communication of ants mathematically. This is an optimization algorithm which was proposed in order to provide solution related to optimization problems. This algorithm considers random walk of ants, building traps, entrapment of ants in traps. In addition to this, it also considers catching preys as well as re-building traps. This optimization algorithm is used for optimization of given issue. It is proposed in order to provide solution related to optimization problems. This algorithm is tested in foreign countries to get optimized modeling of solar cell and found very efficient. It is able to resolve power dispatch issues within power systems. The multi-layer neural networks are also required to be trained.

2 Literature Review

There are several researches related to soft computing techniques and software selection in software engineering. A review of these researches is provided here such as:

In 2015, Kumar et al. [1] discussed CBSE develops a quality software system. For this, they reused existing components. Here, a neuro-fuzzy model was discussed.

In 2010, Kwong et al. [2] wrote on a genetic algorithm (GA). It was discussed as it is efficient to solve the optimization model. It could be utilized to determine optimal selection of software components for CBSS development.

In 2017, Li et al. [3] considered component-based software development. Here, they proposed a reliable evaluation model. It was utilized to check component-based software systems. This research work focused to analyze influence of various components in software reliability. The results of this research work have shown that this model provides better accuracy than existing models.

In 2016, Verma et al. [4] developed autonomous tool automatic bad code detector. Their proposed model is sufficient to detect semantic in the source code.

In 2017, Bhardwaj et al. [5] did research on quality assurance. They used soft computing mechanism in component-based software. It was done to analyze different procedures and techniques. These are utilized to confirm quality of software in component-based software development.

In 2016, Gandhi [6] made assessment of component generality using fuzzy approach. This work was done to optimize software development cost. In this research work, they use fuzzy logic in order to estimate the reusability and software development cost.

In 2016, Diwaker et al. [7] in this work, ACO is used as methodology. They used this mechanism to find reusable components. Objective of research was to improve reliability and efficiency of system.

In 2015, Vodithala et al. [8] proposed a dynamic mechanism to get software components with help of genetic algorithm. Researcher proposed a dynamic approach in this work.

In 2015, Quadri et al. [9] wrote on SQA in component-based software. This research work consists of an investigation. Their research highlighted the way that is used to increase quality of a component-based software system.

In 2015, Kumar et al. [1] researched on neuro-fuzzy model. This research was made to predict and optimize quality as well as performance of CBSE.

In 2014, Singh, et al. [10] did estimation of software reusability. This research presented model depending on various factors.

In 2014, Tyagi et al. [11] provided neuro-fuzzy model to check reliability of CBSS. Adaptive neuro-fuzzy inference mechanism was proposed.

In 2011, Jha et al. [22] discussed fuzzy approach. This research focused on component selection. Author has introduced a framework which is helping developers to take decision whether to purchase or develop components.

3 Tools and Technology

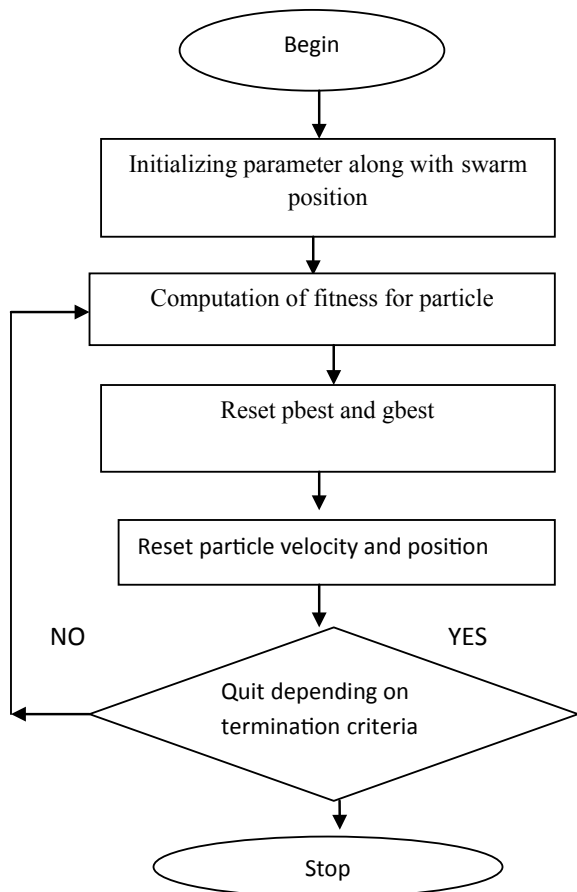
3.1 Particle Swarm Optimization [15]

Particle swarm optimization is a computation method which is very simple and used rapidly. It has been analyzed that this computation method is able to search the optimal solution efficiently (Fig. 1).

In computational science, PSO is defined as a method utilized to optimize issues. It works repetitively by putting the effort to make improvement in a candidate solution. This technique is able to resolve a problem with population related to candidate solutions. Here, the dubbed particles are moving around search space [15].

It is performing as per formula over position and velocity of particle. Every movement of particles is affected from the known local best position. On the other hand, it is guided to the best known positions within search space. Such search spaces

Fig. 1 Steps in PSO



are modified as per the better locations that are searched by different particles. It is supposed to make movement of the swarm for the best solutions. It makes little assumptions about the problem because PSO is a meta-heuristic. Therefore, PSO does not make sure that an optimal solution will be found in any time.

3.2 Deriving Equation for PSO Integrated ALO

The technical implementation of ALO has been shown below Here, $fobj$ is taken as objective function. Dim variable is showing count of variables. $Max_iteration$ is showing max count of generations. $SearchAgents_no$ is representing count of search agents. And, lb is equal to $[lb_1, lb_2, \dots, lb_n]$. Here, lb_n is presented as lower bound of n variable. Then, ub is equal to $[ub_1, ub_2, \dots, ub_n]$. Also, ubn is considered as upper bound of n variable.

Deriving equation for ALO

In 2015, Seyedali Mirjalili presented antlion optimizer. Such mechanism has been considered as nature-inspired mechanism. It is mimicking hunting mechanism of antlions. It is observed that there have been five major phases to hunt the prey in ALO which are ants make random walk, trap making, trapping ants in traps, catching of prey, trap reconstructing.

Generally, ALO is mimicking interaction in antlions and ants that are present in trap. Ants are allowed to move in stochastic manner in search space in order to search food. On other side, antlions hunt ants with the help of traps. The matrices

$$M_{Ant} = \begin{bmatrix} AN_{11} & AN_{12} & \dots & AN_{1d} \\ AN_{21} & AN_{22} & \dots & AN_{2d} \\ \dots & \dots & \dots & \dots \\ AN_{n1} & AN_{n2} & \dots & AN_{nd} \end{bmatrix} \tag{1}$$

and

$$M_{Antlion} = \begin{bmatrix} ANL_{11} & ANL_{12} & \dots & ANL_{1d} \\ ANL_{21} & ANL_{22} & \dots & ANL_{2d} \\ \dots & \dots & \dots & \dots \\ ANL_{n1} & ANL_{n2} & \dots & ANL_{nd} \end{bmatrix} \tag{2}$$

denote matrices to save location of n ants as well as n antlions.

Suppose, f is objective function at the time of optimization, the matrices

$$M_{OA} = \begin{bmatrix} f([AN_{11}, AN_{12}, \dots, AN_{1d}]) \\ f([AN_{21}, AN_{22}, \dots, AN_{2d}]) \\ \dots \\ f([AN_{n1}, AN_{n2}, \dots, AN_{nd}]) \end{bmatrix} \tag{3}$$

and

$$M_{OAL} = \begin{bmatrix} f([ANL_{11}, ANL_{12}, \dots, ANL_{1d}]) \\ f([ANL_{21}, ANL_{22}, \dots, ANL_{2d}]) \\ \dots \\ f([ANL_{n1}, ANL_{n2}, \dots, ANL_{nd}]) \end{bmatrix} \tag{4}$$

Complexity 3 is matrices to save fitness of n ants as well as n antlions. There is existence of six operators in case of ALO algorithm:

- (i) **Random Walks by Ants:** Ants update their location by random walk. They perform random walk during optimization at each step. Following are influencing random walk:

$$X(t) = [0, \text{cumsum}(2r(t_1) - 1), \text{cumsum}(2r(t_2) - 1), \dots, \text{cumsum}(2r(t_n) - 1)] \tag{5}$$

where $r(t)$ is a stochastic function explained below:

$$r(t) = \begin{cases} 1 & \text{if rand} > 0.5 \\ 0 & \text{if rand} \leq 0.5 \end{cases} \tag{6}$$

cumsum computes cumulative sum. Here, n is increasing number of iterations, t is showing present iteration, and rand is a random number produced with uniform distribution at interval of zero to one. Each search space is having a range of variable to allow random walks in search space. Ants are normalized with support of following equation min-max normalization before modifying location of ants:

$$X_i^t = \frac{(X_i^t - a_i)(d_i - c_i^t)}{(d_i^t - a_i)} + c_i \tag{7}$$

a_i has been min of random walk of i th variable and d_i has been max of random walk in i th variable.

c_i^t is min of i th variable at t th iteration.

d_i^t is indicating maximum of i th variable at t th iteration.

- (ii) **Trapping inside the pit of Antlion's**

Ants random walk is influenced by traps of antlions. Equations

$$c_i^t = \text{Antlion}_j^t + c^t \quad (8)$$

and

$$d_i^t = \text{Antlion}_j^t + d^t \quad (9)$$

are presenting that ants walk in random pattern in a hypersphere. This is defined by c and d vectors. These vectors are present around a chosen antlion. c^t is min of variables at t th iteration. d^t is indicating vector with max of variables at t th iteration. c_i^t is min of all variables in case of i th ant. d_i^t is max of variables in case of i th ant. Antlion_j is showing position of chosen j th antlion at t th iteration.

- (iii) Trap Making: Roulette wheel integrated in ALO has been used to find the fitter antlions. It is to trap the ants according to fitness value at the time of optimization.
- (iv) Sliding the Ants to the Antlion: As per fitness values, the antlions are capable to make the traps in proportion. The ants keep on moving on random basis. Once ants are in trap, antlions shoot sands outward center of pit. This behavior slides down trapped ant that is trying to escape, where radius of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 iteration values of w ants random walks hypersphere is decreased adaptively. The equations

$$c^t = \frac{c^t}{I} \quad (10)$$

and

$$d^t = \frac{d^t}{I} \quad (11)$$

shrink radius of updating ants positions and mimic sliding process of ant inside pits, where I is a ratio, c^t is minimum of all variables at t th iteration, and d^t indicates vector including maximum of all variables at t th iteration.

In (10) and (11), $I = 10^w(t/T)$, where T is maximum number of iterations and w is a constant defined based on current iteration t ($w = 2$ when $t > 0.1 T$, $w = 3$ when $t > 0.5 T$, $w = 4$ when $t > 0.75 T$, $w = 5$ when $t > 0.9 T$, and $w = 6$ when $t > 0.95 T$). Basically, constant w could adjust accuracy level of exploitation.

- (v) Prey capturing and Pit reconstruction: As ant goes to the bottom of pit, antlion catches it. Antlion is changing its location to new location of hunted ant to improve possibility to capture further prey:

$$\text{Antlion}_j^t = \text{Ant}_j^t \quad \text{if } f(\text{Ant}_j^t) > f(\text{Antlion}_j^t) \quad (12)$$

where Ant_{j}^t is representing location of chosen j antlion at t iteration. Ant_i^t represents the location of i ant at t loop.

- (vi) Elitism: It is considered as significant characteristic of evolutionary algorithms. It is to maintain the best solution found during optimization process at any stage. Here, the best antlion is known as elite. Each ant walks around a selected antlion randomly. It is performed by support of roulette wheel

$$Ant_j^t = \frac{R_A^t + R_E^t}{2} \tag{13}$$

where

R_A^t is walking randomly around antlion chosen by roulette wheel. Selection is performed at t th iteration. R_E^t is walking randomly around elite at t th loop. Ant_j^t is indicating the location of i th ant at t th loop.

Suppose, A is a function which is producing random initial solutions. B is changing population in beginning. It has been given by A . Here, C is returning true after condition is satisfied. ALO is described as three-tuple below based on operators:

$$ALO(A, B, C) \tag{14}$$

where functions A , B , and C have been defined as:

$$\emptyset \rightarrow A_{\{M_{Ant}, M_{OA}, M_{Ant}, M_{OAL}\}} \tag{15}$$

$$\{M_{Ant}, M_{Antlion}\} \rightarrow B_{\{M_{Ant}, M_{Antlion}\}} \tag{16}$$

$$\{M_{Ant}, M_{Antlion}\} \rightarrow C_{\{true, false\}} \tag{17}$$

M_{Ant} is matrix of position of ants and, $M_{Antlion}$ consists position of antlions. M_{OA} is containing corresponding fitness of ants, and M_{OAL} has fitness of antlions.

PSO integrated ALO Mechanism

In case of PSO, two updated methods known as velocity update and position update of particle are present. The modified velocity and location of each particle in PSO have been shown below:

$$v_i^{t+1} = wv_i^t + v_i^{t+1} + c_1 \text{rand}() (pbest^t - x_i^t) + c_2 \text{rand}() (gbest^t - x_i^t) \tag{18}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{19}$$

Here, v_i^t and x_i^t have been considered current velocity.

Position of i particle at t iteration.

Here, c_1 and c_2 have been considered as acceleration coefficients which are controlling influence of $pbest^t$ and $gbest$ in case of search process, respectively. $rand()$ is an arbitrary figure in $[0,1]$, $pbest^t$ becomes good position of particles at t th repetition, $gbest$ becomes good location in all particles at complete repetition, and w is inertia weight which is natural and steady less than 1.

The arrangement of (18) in PSO is used in the form of upgraded method of elitism engineer of ALO. Here, adopted method is that x_i^t , $pbest^t$, $gbest$, v_i^t , and v_i^{t+1} have been replaced using elite, R_A^t , R_E^t , $(R_A^t + R_E^t)/2$ in Eq. 13 and Ant_i^t , respectively. Thus, improved elitism operator of ALO is obtained as follows:

$$Ant_j^t = w \frac{R_A^t + R_E^t}{2} + c_1 rand()(R_A^t - elite) + c_2 rand()(R_A^t - elite) \quad (20)$$

Activities of ALO algorithm which incorporated PSO is listed below

Step 1: Start ants society and antlions in a random way. Determine whether the condition of ants and antlions is good. Identify finest antlions, and consider it in the form of the best (determined optimum).

Step 2: For each ant, identify a corresponding antlion by means of roulette wheel. Execute modification on c and d using equation ten and eleven. Form a walk and regulate it by means of equation five and seven. Switch the ant position with equation twenty.

Step 3: Identify physical condition of all ants and exchange antlion with its related ant if it becomes fitter 12. Modify elite if an antlion becomes better than elite.

Step 4: If conclusive criterion is convinced, return elite. If no, come back to Step 2.

Step 5: Stop.

3.3 Deriving Equation for PSO Integrated MVO

Phase 1: In phase one, equation of PSO is considered

Particle Swarm Optimization

Mechanism has been inspired by social expression of birds or fishes. The PSO consists of P_{best} and G_{best} . Position and velocity are updated over course of iteration from these mathematical equations:

$$v_{ij}^{t+1} = wv_{ij}^t + C_1 R_1(Pbest^t - X^t) + C_2 R_2(Gbest^t - X^t) \quad (1)$$

$$X^{t+1} = x^t + v^{2t+1} (i = 1, 2, \dots, NP) \text{ and } (J = 1, 2, \dots, NG) \quad (2)$$

where

$$W = w^{\max} - \frac{(w^{\max} - w^{\min}) * \text{iteration}}{\text{maxiteration}} \tag{3}$$

$$w^{\max} = 0.4$$

$w^{\min} = 0.9$. v_{ij}^t, v_{ij}^{t+1} has been considered velocity of “j” member of “i” particle in iteration number (t) as well as (t + 1). (Usually, $C_1 = C_2 = 2$), and r_1 and r_2 are random numbers (0, 1).

Phase 2: Multi-verse optimizer equation

Three such as are main algorithm of MVO gets encouragement from three concepts. These three concepts are black hole, white hole, and wormhole. All are developed in arithmetic form to examine, use, study, and local search, respectively. In the production of population, white hole becomes the essential part. Black holes becomes famous because of its powerful physical strength. The wormholes behave as time/space travel channels in which objects could move rapidly in universe. Steps involved in support of MVO population:

1. If level of inflation is higher, chances of white hole presence are also higher.
2. If level of inflation is higher, chances of black hole presence are lower.
3. Objects are transmitted by white holes when population has high level of inflation.
4. Objects are accepted by black holes when population has low level of inflation.

Arbitrary progress in direction of population whose physical condition is good is created by objects in every world. It is done by means of wormholes without taking into account expansion rate. Movement of objects takes place in the direction of world having low expansion rate from a world having greater expansion rate. It will surely improve the level of average inflation rates of complete cosmoses in the company of repetition. In all repetition, world are grouped depending upon their inflation rates. And, choose one from them by means of roulette wheel in the form of a white hole. The next steps are applied in order to achieve this procedure. Assume that

$$U = \begin{bmatrix} X_1^1 & X_1^2 & \dots & X_1^d \\ X_2^1 & X_2^2 & \dots & X_2^d \\ \dots & \dots & \dots & \dots \\ X_n^1 & X_n^2 & \dots & X_n^d \end{bmatrix} \tag{4}$$

In equation, d is showing the number of variables. The n is showing number of candidate solutions:

$$X_i^j = \begin{cases} X_k^j; & r_1 < NI(U_i) \\ X_i^j; & r_1 \geq NI(U_i) \end{cases} \tag{5}$$

X_i^j is representing j variable of i universe.
 U_i is representing i universe.
 $NI(U_i)$ has been considered as normalized inflation rate of i universe.
 r_1 is a random number from 0 to 1.
 X_k^j is showing j variable of k universe selected by a roulette wheel.

In order to deliver variations in case of universe and more possibility of growing inflation rate by wormholes, let the wormhole channels have been considered in universe as well as fittest universe made until now. Technique has been formulated as follow:

$$X_i^j = \begin{cases} X_j + \text{TDR} * ((ub_j - lb_j) * r_4 + lb_j); r_3 < 0.5 \\ X_j + \text{TDR} * ((ub_j - lb_j) * r_4 + lb_j); r_3 \geq 0.5 \\ X_i^j; r_2 \geq \text{WEP} \end{cases}; r_2 < \text{WEP} \tag{6}$$

X_j is showing j variable of fittest universe created until now.
 lb_j is indicating min limit of j parameter.
 ub_j is indicating max limit of j parameter.
 X_i^j is showing j parameter of i universe.
 r_2, r_3, r_4 are random numbers from 0 to 1.

It could be concluded by formulation that wormhole existence probability (WEP) and traveling distance rate (TDR) are chief coefficients. The formula for these coefficients is given by:

$$\text{WEP} = \min + l * \left(\frac{\max - \min}{L} \right) \tag{7}$$

where l shows present run, and L represents maximum run number/iteration.

$$\text{TDR} = 1 - \frac{l^{1/p}}{L^{1/p}} \tag{8}$$

where p shows the precision of operation in the company of repetition. In situation where p is higher, operation is speedy and more accurate. The difficulty of MVO algorithms based on no. of repetition, no. of population, roulette wheel method, and world arranging methods. The comprehensive computerized difficulty is as follows:

$$O(\text{MVO}) = O(1(O(\text{Quicksort})) + n * d * (O(\text{roulette}_{\text{wheel}}))) \tag{9}$$

$$O(\text{MVO}) = O(1(n^2 + n * d * \log n)) \tag{10}$$

n is showing number of universes
 l is showing maximum number of run/iterations

d is showing number of substances.

Phase 3: Deriving hybrid PSO-MVO equation

Set of hybrid PSO-MVO is integration of PSO and MVO. Hybrid PSO-MVO is merging the best strength of PSO and MVO toward targeted optimum solution. It is replacing PSO Pbest value to MVO universe value.

$$v_{ij}^{t+1} = wv_{ij}^t + C_1R_1(\text{Universes}^t - X^t) + C_2R_2(G\text{best}^t - X^t) \quad (11)$$

4 Research Methodology

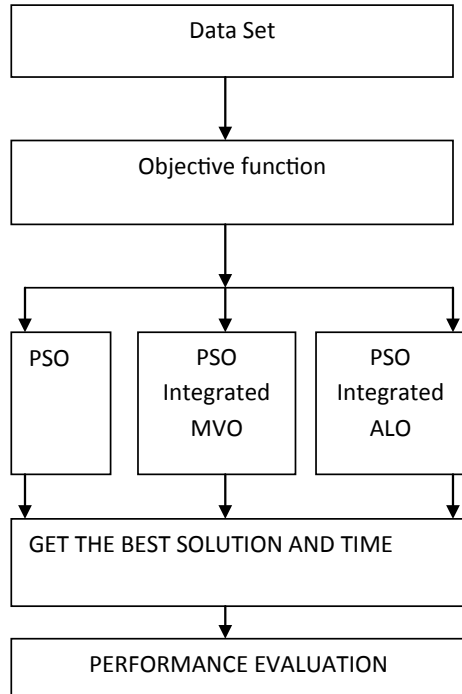
In this research, dataset representing version, modules, alternatives, cost, reliability, and delivery time is taken. This dataset has been taken from research [22] where fuzzy approach has been used for component selection among different versions of alternatives. This research had focused on fault-tolerant modular software system under recovery block scheme. Research is incorporating build-or-buy strategy. The steps included in proposed methodology are discussed below:

- Step 1: Get the dataset representing version, modules, alternatives, cost, reliability, and delivery time.
- Step 2: Develop an objective function in order to optimize the cost considering PSO, PSO-integrated MVO, and PSO-integrated ALO.
- Step 3: Get the best solution using above-mentioned techniques.
- Step 4: Compare the time taken in different techniques.
- Step 5: Stop (Fig. 2).

4.1 Dataset

Dataset representing version, modules, alternatives, cost, reliability, and delivery time has been considered during simulation. Total number of considered variable are 6, and population size is 27. The lower bound and upper bounds have been represented at the bottom of Table 1.

Fig. 2 Process flow



5 Result and Discussion

5.1 Proposed Work

Proposed work has focused on optimized model for software selection in component-based software engineering.

In proposed work, the simulation of cost optimization, efficiency, time consumption, and reliability optimization is made considering dataset of Table 1.

LB = [1 1 1 0 0.001 0]; and, % lower bound of variable

UB = [3 3 3 18 0.92 5]; and % upper bound of variable

$m = 6\%$ number of variable

$n = 27\%$ population size

maxiter = 500% maximum number of iteration.

a is the dataset, k is index, and x is input parameter for objective function. O is representing value returned by objective function.

Objective Function

Table 1 Dataset for COTS components

<i>m/n</i>	Version	Modules	Alternatives	Cost	Reliability	Delivery Time
1	1	1	1	0	0.001	0
2	1	1	2	0	0.001	0
3	1	1	3	0	0.001	0
4	1	2	1	0	0.001	0
5	1	2	2	0	0.001	0
6	1	2	3	0	0.001	0
7	1	3	1	0	0.001	0
8	1	3	2	0	0.001	0
9	1	3	3	0	0.001	0
10	2	1	1	14	0.9	3
11	2	1	2	12.5	0.86	4
12	2	1	3	17	0.9	2
13	2	2	1	13	0.87	4
14	2	2	2	11	0.91	5
15	2	2	3	18	0.89	2
16	2	3	1	13	0.86	4
17	2	3	2	16	0.85	3
18	2	3	3	16	0.89	3
19	3	1	1	11	0.88	4
20	3	1	2	18	0.92	2
21	3	1	3	15	0.88	3
22	3	2	1	17.5	0.86	2
23	3	2	2	12	0.89	4
24	3	2	3	15	0.86	3
25	3	3	1	14	0.88	3
26	3	3	2	18	0.9	2
27	3	3	3	17	0.87	2

$$(bs) = \sum_{k=0}^n (x - a(k))^6$$

$$O = \left(\frac{1}{500} \right) + 1 / \sum_{k=0}^n 1/b_s$$

Matlab script for objective function

```
aS = [0 0 0 0 0 0 0 0 14 12.5 17 13 11 18 13 16 16 11 18 15 17.5 12 15 14 18 17];
```

```

for j = 1:27
    bS(j) = sum((x' - aS(:,j)).^6);
end
o = (1/500 + sum(1./([1:27] + bS))).^(-1);

```

5.2 Cost Optimization

Simulation result of cost estimation in case of PSO

```

Iteration# 1 Swarm.GBEST.O = 2.1321
Iteration# 2 Swarm.GBEST.O = 2.1317
Iteration# 3 Swarm.GBEST.O = 2.1293
Iteration# 4 Swarm.GBEST.O = 2.1293
Iteration# 5 Swarm.GBEST.O = 2.1293
Iteration# 6 Swarm.GBEST.O = 2.1293
Iteration# 7 Swarm.GBEST.O = 2.1293
Iteration# 8 Swarm.GBEST.O = 2.1293
Iteration# 9 Swarm.GBEST.O = 2.1293
Iteration# 10 to 500 Swarm.GBEST.O = 2.1293
The best solution found
ans =

    13.5126

The best objective value
ans =

    2.1293

Elapsed time is 6.656739 s (Fig. 3).

```

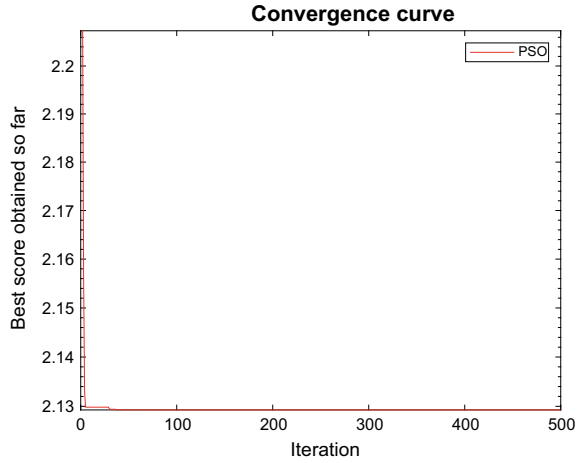
Simulation result of cost estimation in case of ALO integrated PSO

```

At iteration 50, the elite fitness is 0.35324
At iteration 100, the elite fitness is 0.35324
At iteration 150, the elite fitness is 0.35324
At iteration 200, the elite fitness is 0.35324
At iteration 250, the elite fitness is 0.35324
At iteration 300, the elite fitness is 0.35324
At iteration 350, the elite fitness is 0.35324
At iteration 400, the elite fitness is 0.35324
At iteration 450, the elite fitness is 0.35324
At iteration 500, the elite fitness is 0.35324
The best solution obtained by ALO is 0.043477

```

Fig. 3 PSO simulation



The best optimal value of the objective function found by PSO integrated ALO is 0.35324
Elapsed time is 2.286426 s (Fig. 4).

Simulation result of cost estimation in case of PSO Integrated MVO

- At iteration 50, the best universes fitness is 2.1549
- At iteration 100, the best universes fitness is 2.1435
- At iteration 150, the best universes fitness is 2.1295
- At iteration 200, the best universes fitness is 2.1293
- At iteration 250, the best universes fitness is 2.1293

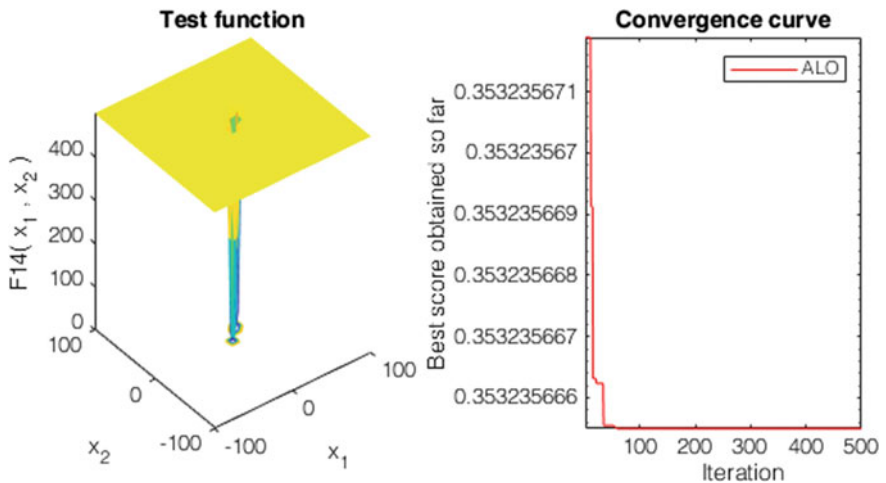


Fig. 4 PSO-integrated ALO simulation

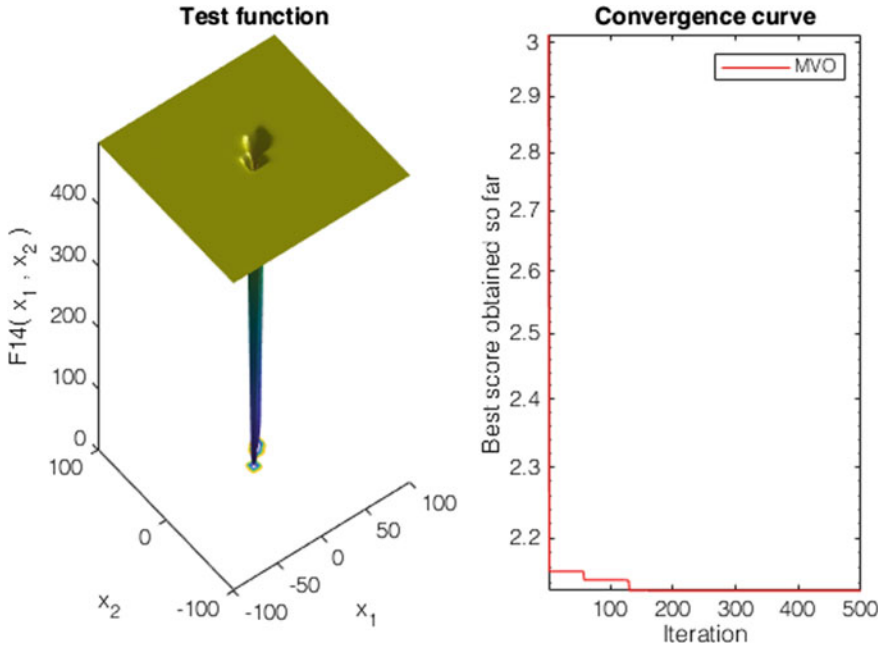


Fig. 5 PSO-integrated MVO simulation

At iteration 300, the best universes fitness is 2.1293
 At iteration 350, the best universes fitness is 2.1293
 At iteration 400, the best universes fitness is 2.1293
 At iteration 450, the best universes fitness is 2.1293
 At iteration 500, the best universes fitness is 2.1293
 The best solution obtained by MVO is 13.5135
 The best optimal value of the objective function found by PSO integrated MVO is 2.1293
 Elapsed time is 1.981937 s (Fig. 5).

5.3 Efficiency

Efficiency in case of cost estimation has been calculated by taking inverse of time taken in all cases

$$\text{Efficiency} = 1/T$$

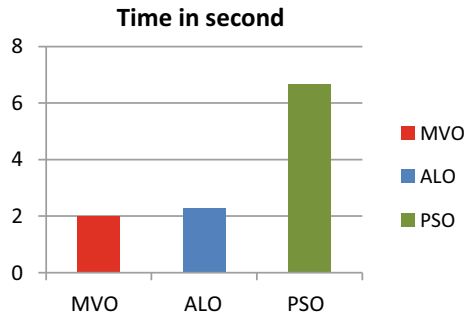
Time taken by MVO, ALO, and PSO is shown in following Table 2.

Considering above table, following chart has been plotted. The time taken in case of MVO is less than other (Fig. 6).

Table 2 Time consumption during cost estimation

Technique	Time in second
PSO-integrated MVO	1.981937
PSO-integrated ALO	2.286426
PSO	6.656739

Fig. 6 Comparison of time consumption during cost estimation



Efficiency of three techniques has been represented in Table 3. The efficiency is found using following equation.

$$\text{Efficiency} = 1/\text{Time.}$$

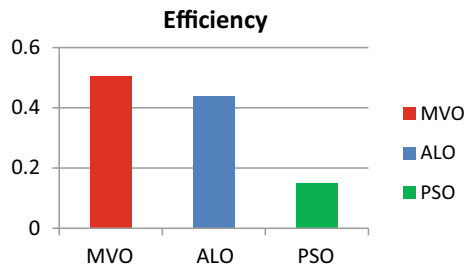
Considering above table, chart representing efficiency is shown below

See Fig. 7.

Table 3 Efficiency table

Technique	Efficiency
PSO-integrated MVO	0.504556906
PSO-integrated ALO	0.437363816
PSO-integrated PSO	0.150223706

Fig. 7 Comparison of efficiency



5.4 Reliability

During reliability optimization, lower bound, upper bound, number of variable, population size, and maximum number of iterations are considered to find efficiency, time consumption considering six different objective functions.

Lower bound of variable $LB = 0.001$

Upper bound of variable $UB = 0.92$

Number of variable $m = 6$

Population size $n = 27$

Dataset = [0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.9 0.86 0.9 0.87 0.91 0.89 0.86 0.85 0.89 0.88 0.92 0.88 0.86 0.89 0.86 0.88 0.9 0.87]

maxiter = 500%maximum number of iteration

a is the dataset, k is index, and x is input parameter for objective function. O is representing value returned by six objective functions.

Objective function 1:

$$(bs) = \sum_{k=0}^n (x - a(k))^2$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

Objective function 2:

$$(bs) = \sum_{k=0}^n (x - a(k))^3$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

Objective function 3:

$$(bs) = \sum_{k=0}^n (x - a(k))^4$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

Objective function 4:

$$(bs) = \sum_{k=0}^n (x - a(k))^5$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

Objective function 5:

$$(bs) = \sum_{k=0}^n (x - a(k))^6$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

Objective function 6:

$$(bs) = \sum_{k=0}^n (x - a(k))^7$$

$$O = \left(\frac{1}{500}\right) + 1/\sum_{k=0}^n 1/b_s$$

MATLAB script for objective function

For ll = 2:1:7

aS = [0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.9 0.86 0.9 0.87
0.91 0.89 0.86 0.85 0.89 0.88 0.92 0.88 0.86 0.89 0.86 0.88 0.9 0.87];

for j = 1:7.

 bS(j) = sum((x' - aS(:,j)).^ll);

end

o = (1/500 + sum(1./([1:7] + bS))).^(-1);

end

After simulation in PSO, the best solution has been calculated considering above dataset and six objective functions. Following table is representing best solution and time taken in case of PSO (Table 4).

After simulation in PSO-integrated ALO, the best solution has been calculated considering above dataset and six objective functions. Following table is representing the best solution and time taken in case of PSO-integrated ALO (Table 5).

Table 4 Solution chart for reliability using PSO

Objective function	PSO best solution	Time taken
1 (ll = 2)	0.2601	6.921177
2 (ll = 3)	0.2536	5.380807
3 (ll = 4)	0.2579	6.713173
4 (ll = 5)	0.2543	5.671892
5 (ll = 6)	0.2571	6.736498
6 (ll = 7)	0.2549	5.604441

Table 5 Solution chart for reliability using PSO-integrated ALO

Objective function	PSO-integrated ALO best solution	Time taken
1 (ll = 2)	0.26011	2.274776
2 (ll = 3)	0.2536	2.368447
3 (ll = 4)	0.25794	2.440933
4 (ll = 5)	0.25433	2.502913
5 (ll = 6)	0.25709	2.502815
6 (ll = 7)	0.25489	2.497580

After simulation in PSO-integrated MVO, the best solution has been calculated considering above dataset and six objective functions. Following table is representing best solution and time taken in case of MVO-integrated ALO (Table 6).

Following table is representing the time consumption in case of PSO, PSO-integrated ALO, and PSO-integrated MVO (Table 7).

Considering above table, following chart has been plotted (Fig. 8 and Table 8).

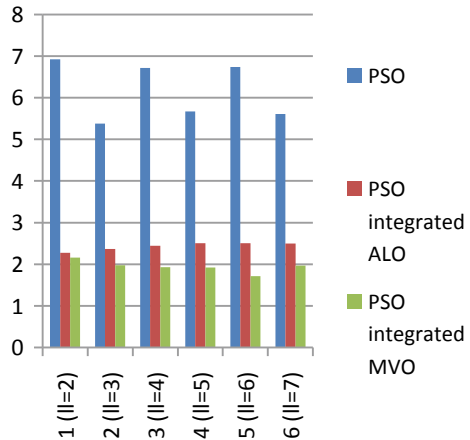
Table 6 Solution chart for reliability using PSO-integrated MVO best solution

Objective function	PSO-integrated MVO best solution	Time taken
1 (ll = 2)	0.26011	2.156834
2 (ll = 3)	0.2536	1.973367
3 (ll = 4)	0.25794	1.931486
4 (ll = 5)	0.25433	1.925430
5 (ll = 6)	0.25709	1.716243
6 (ll = 7)	0.25489	1.965474

Table 7 Comparison table for reliability

Objective function	PSO	PSO-integrated ALO	PSO-integrated MVO
1 (II = 2)	6.921177	2.274776	2.156834
2 (II = 3)	5.380807	2.368447	1.973367
3 (II = 4)	6.713173	2.440933	1.931486
4 (II = 5)	5.671892	2.502913	1.92543
5 (II = 6)	6.736498	2.502815	1.716243
6 (II = 7)	5.604441	2.49758	1.965474

Fig. 8 Comparison chart for reliability



5.5 Component Dependency

In order to find the best component, the best solution of cost, reliability, and delivery delay is considered. As the best solution for cost and reliability has been calculated, the pending **delivery delay** would be calculated using PSO-integrated MVO approach. The best solution for delivery delay is as follows:

- At iteration 50, the best universes fitness is 0.3366
- At iteration 100, the best universes fitness is 0.33123
- At iteration 150, the best universes fitness is 0.33049
- At iteration 200, the best universes fitness is 0.33016
- At iteration 250, the best universes fitness is 0.33016
- At iteration 300, the best universes fitness is 0.33016
- At iteration 350, the best universes fitness is 0.33016
- At iteration 400, the best universes fitness is 0.33014
- At iteration 450, the best universes fitness is 0.33014
- At iteration 500, the best universes fitness is 0.33014
- The best solution obtained by MVO is 0.52484

Table 8 Component dependency chart

m/n	V e r	Md l	Al t	Cos t	dCost	Rel	dRel	iRel	DD	dDD	Average
1	1	1	1	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
2	1	1	2	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
3	1	1	3	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
4	1	2	1	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
5	1	2	2	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
6	1	2	3	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
7	1	3	1	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
8	1	3	2	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
9	1	3	3	0	13.5135	0.001	0.2561	0.7439	0	0.33014	4.862513
10	2	1	1	14	0.4865	0.9	0.6429	0.3571	3	2.66986	1.171153
11	2	1	2	12.5	1.0135	0.86	0.6029	0.3971	4	3.66986	1.693487
12	2	1	3	17	3.4865	0.9	0.6429	0.3571	2	1.66986	1.83782
13	2	2	1	13	0.5135	0.87	0.6129	0.3871	4	3.66986	1.523487
14	2	2	2	11	2.5135	0.91	0.6529	0.3471	5	4.66986	2.510153
15	2	2	3	18	4.4865	0.89	0.6329	0.3671	2	1.66986	2.174487
16	2	3	1	13	0.5135	0.86	0.6029	0.3971	4	3.66986	1.52682
17	2	3	2	16	2.4865	0.85	0.5929	0.4071	3	2.66986	1.854487
18	2	3	3	16	2.4865	0.89	0.6329	0.3671	3	2.66986	1.841153
19	3	1	1	11	2.5135	0.88	0.6229	0.3771	4	3.66986	2.18682
20	3	1	2	18	4.4865	0.92	0.6629	0.3371	2	1.66986	2.164487
21	3	1	3	15	1.4865	0.88	0.6229	0.3771	3	2.66986	1.511153
22	3	2	1	17.5	3.9865	0.86	0.6029	0.3971	2	1.66986	2.01782
23	3	2	2	12	1.5135	0.89	0.6329	0.3671	4	3.66986	1.850153
24	3	2	3	15	1.4865	0.86	0.6029	0.3971	3	2.66986	1.51782
25	3	3	1	14	0.4865	0.88	0.6229	0.3771	3	2.66986	1.17782
26	3	3	2	18	4.4865	0.9	0.6429	0.3571	2	1.66986	2.171153
27	3	3	3	17	3.4865	0.87	0.6129	0.3871	2	1.66986	1.84782
Minimum											1.171153

The best optimal value of the objective function found by PSO-integrated MVO is 0.33014

Elapsed time is 1.860964 s

Now the best solution for cost is (bCost) = 13.5135

The best delivery delay is (bDD) = 0.33014

The best reliability is (bRel) = 0.2571

n is number of modules = 27.

Algorithm to select the best module depending on cost, delay, and reliability

Step 1 set $i = 1$ and repeat Steps 2, 3, 4, 5, 6, 7 until $i \leq n$

Step 2 get delta cost dcost

$dcost(i) = \text{abs}(\text{cost}(i) - \text{bcost})$

Step 3 get delta reliability dRel
 $dRel(i) = \text{abs}(\text{Rel}(i) - bRel)$
Step 4 get inverse of reliability iRel
 $iRel(i) = 1 - dRel(i)$
Step 5 get delta delivery delay dDD
 $dDD(i) = \text{abs}(DD(i) - bDD)$
Step 6 calculate the average AVG
 $AVG(i) = (dCost(i) + iRel(i) + dDD(i))/3$
Step 7 $i = i + 1$
Step 8 get minimum average
 $\text{min} = \text{MIN}(AVG(i))$
Step 9 set $i = 1$ and repeat steps 10, 11 unit $i \leq n$
Step 10 if($\text{min} = \text{avg}(i)$)
 solution = i.
Step 11 $i = i + 1$.
Step 12 stop

In above table, the minimum average is 1.171153. Thus, module 10 is considered as solution considering cost, delivery delay, and reliability.

6 Conclusion

From above experiments, it is concluded that the PSO-integrated MVO and PSO-integrated ALO take less time in finding the best solution as compared to traditional PSO. The simulation to get the best solution in case of cost and reliability has been made in research. In case of reliability simulation, six objective functions have been used to compare the solution and time consumption. Result also represents that if comparison of PSO-integrated MVO and PSO-integrated ALO is made, then PSO-integrated MVO is found more efficient. The best module has been chosen considering the best cost, delivery delay, and reliability.

7 Future Scope

The research could be further extended considering another parameter of software component selection. These parameters might be operational profile, and portability would be considered. Many soft computing approaches from artificial neural network could improve the further performance. And, these may help in resolving several issues.

References

1. Kumar G et al (2015) Neuro-fuzzy model to estimate and optimize quality and performance of component based software engineering. *ACM SIGSOFT Softw Eng Notes* 40(2):1–6
2. Kwong CK et al (2010) Optimization of software components selection for component-based software system development. *Comput Ind Eng* 58(4):618–624
3. Li K et al (2017) Reliability evaluation model of component-based software based on complex network theory. *Qual Reliab Eng Int* 33(3):543–550
4. Verma A et al (2016) Inconsistency detection in software component source code using ant colony optimization and neural network algorithm. *Indian J Sci Technol* 9(40)
5. Bhardwaj O et al (2018) Quality assurance through soft computing techniques in component based software. In: *Proceedings of 2017 international conference on smart technology smart nation, SmartTechCon 2017*, pp 277–282
6. Gandhi P (2016) Assessment of component generality using fuzzy approach to optimize software development cost. *7(2):647–654*
7. Diwaker C et al (2016) Assessment of ant colony using component based software engineering metrics. *Indian J Sci Technol* 9(44)
8. Vodithala S et al (2015) A dynamic approach for retrieval of software components using genetic algorithm. *Proceedings of IEEE international conference on software engineering and service science, ICSESS, November 2015*, pp 406–410
9. Quadri AT et al (2015) Software quality assurance in component based software development—a survey analysis. *Int J Comput Commun Syst Eng* 2(02):305–315
10. Singh C et al (2014) Estimation of software reusability for component based system using soft computing techniques. In: *Proceedings of 5th international conferences confluence 2014 next generation information technology summit. ICM*, pp 788–794
11. Tyagi K et al (2014) An adaptive neuro fuzzy model for estimating the reliability of component-based software systems. *Appl Comput Inform* 10(1–2):38–51
12. Ghasemi N et al (2013) Software reuse metrics for component qualification in reusable verification environment. *Int J Res Comput Appl Rob* 3(4):168–176. ISSN 2320-7345
13. Roopa YM (2013) Particle swarm optimization approach for component based software architecture. *3(12):557–561*
14. Tang JF et al (2011) An optimization model for software component selection under multiple applications development. *Eur J Oper Res* 212(2):301–311
15. Saed AAA et al (2011) Applying particle swarm optimization to software performance prediction an introduction to the approach. In: *2011 5th Malaysian conference software engineering, MySEC 2011*, pp 207–212
16. Palviainen M et al (2011) The reliability estimation, prediction and measuring of component-based software. *J Syst Softw* 84(6):1054–1070
17. Xvlqj H et al (2003) Departament de Ciències Matemàtiques i Informàtica
18. Diwaker C et al (2019) A new model for predicting component-based software reliability using soft computing. *IEEE Access* 7:147191–147203
19. Tomar P et al (2018) Prediction of quality using ANN based on Teaching-Learning Optimization in component-based software systems. *Softw Pract Exp* 48(4):896–910
20. Diwaker C et al (2018) Prediction of software reliability using bio inspired soft computing techniques. *J Med Syst* 42(5)
21. Mu L et al (2018) A multi-objective optimization model of component selection in enterprise information system integration. *Comput Ind Eng* 115:278–289
22. Jha PC et al (2011) A fuzzy approach for component selection amongst different versions of alternatives for a fault tolerant modular software system under recovery block scheme incorporating build-or-buy strategy. *Am J Oper Res* 1:249–258. <https://doi.org/10.4236/ajor.2011.14029>. Published Online December 2011 <http://www.SciRP.org/journal/ajor>

Analysis of Hybrid Power Generation System on the Highway and Tracking Through IoT



Sujata Shivashimpiger , Purshottam Kamat , and S. Anand 

1 Introduction

Energy is the most basic requirement to move the world. It is the backbone factor for the socio-economic development of any nation. The practices of everyday requirement of energy resources are also the factor for direct or indirect effects on the environmental consequences. Keeping in the mind about the disastrous consequences which are happening all over the world to living creatures, renewable energy sources are an alternative option to bring stability in the nature and power consumption for mankind. Out of various renewable energy sources, the most feasible one is the energy from the sun and the wind, which are free and easily available at any corner of the world. According to the studies, solar [1] has a high potential to bestow the demand for energy through an active or passive way. This clean and abundant source of energy will not be sufficient to give continuous power supply through the PV panel due to unpredictable weather conditions and inflexibility period of solar insolation. To get maximum output power from solar irradiation, an automatic solar tracking system has been also introduced. But still, it is not sufficient to get sustained power supply for long periods, so the integration of wind turbines will help out to fill the gap of continuous generation, and hence, the model is turned out to be a hybrid power generating source. The wind is another clean energy source. This is also very dynamic in nature. The wind energy varies according to the geographical regions. In the metropolitan cities where big structures of building block the free flow of wind energy, it is not possible to stall the tall structure of horizontal axis wind turbines

S. Shivashimpiger (✉) · P. Kamat · S. Anand
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: sujata.s@nmit.ac.in

S. Anand
e-mail: Anand.s@nmit.ac.in

which is a costly, skillful worker requires for maintenance and work in drag and lift force.

In the city highway is the only place where a huge number of moving vehicle creates wind turbulence which can be harnessed to generate electricity. Highway meridian is the perfect place to install small scale operation of power generation through hybrid technology. Here, the vertical axis wind turbine will be very cost-effective, efficient, less maintenance, and safe to operate. VAWT works on the principle of drag force which acts parallel to the direction of wind flow [2] that will be getting from the moving vehicles while the aerodynamic loss of wind on the highway. The solar panel can be easily superimposed upon VAWT and collectively will act as the source of standalone power systems that will be suitable in urban environments.

In this paper, the analysis of a designed model of cup type of VAWT is done in simulation software called ANSYS workbench. The testing is done in the software upon the static state of the designed model by applying various forces and pressure. After successful testing of the proposed model in simulation software under various conditions, the prototype fabrication is done. A standalone hybrid energy system consisting of wind and PV is proposed with the battery for energy storage. The voltage generating through different sources individually gets monitored through voltage sensors. The NODEMCU ESP32 microcontroller allows exchanging the data through WiFi or via Bluetooth to the personal system. These are the very secured and reliable techniques to monitor and manage the energy over an intra-network through an open source of IoT platform called Thingspeak.

2 Literature Survey

VAWT will not be sufficient to generate power, so the integration of solar panels will aid to supply power continuously on the highway lighting system as being a hybrid power generation source. VAWT is specially designed to capture wind from all the directions at low wind velocity which is created by moving vehicles on the highway [2]. By increasing the number of the blade at lower wind speed, the maximum rotor power coefficient can be obtained [3]. The VAWT is very suitable for a small scale energy system which will occupy less space in the populated urban area [4]. The computational analysis shows that the force acting on the designed turbine blade is capable to withstand without deforming under extensive pressure. Although aerodynamic turbulence coming from the vehicle on the highway is very less, however, under such condition VAWT is able to utilize to produce a good efficiency power output up to 6.1035 W for given parameters, theoretical efficiency calculated comes out to be 50.69% [5, 9]. The designed blades are light but strong and simple in structure. It can capture wind from any direction and can withstand any weather condition. The main advantage of this VAWT is that its cost is limited and it can easily be affordable for commercial power generation. Considering the all-weather point of view the material used which are noncorrosive [6, 8]. The starting-up wind speed of the semi-circle Cup-Rotor vertical axis wind turbine is about 1.6 m/s,

the rotational velocity and the tip-speed ratio increases with wind speed from 1.6 to 12 m/s [7]. A hybrid power system having VAWT, solar panel, and integration of IoT controlling system will be cost-effective and help to reduce power requirements in roadside applications for power generation [10]. Monitoring through IoT helps in regular maintenance by transferring data over a network which will sort out defects in the system by conveniently [11].

3 Design of the Proposed System

The expected highway hybrid power generation system consists of the following types of equipment:

PV Panel Photovoltaic (PV) technology, use to convert photons from solar energy into electricity. Polycrystalline type solar of 12 V, 10 W having specification is installed in this hybrid system. The top-mounted solar panel consists of semiconductor material to release electrons. To store the maximum amount of charge into battery, the tracking mechanism will help to detect the direction of the solar radiation through the help of light detecting resistor (LDR) and DC motor.

C-Type VAWT To minimize the dependency upon solar energy, the vertical axis wind turbine together adding to make the system hybrid. This VAWT is very suitable for low-speed wind which generates from vehicles and can easily capture wind from any directions. This turbine is connected to DC dynamo having 12 V, 1 Amp specification which is connected through a 1:7 ratio of gear mechanism to enhance power generation. The selected turbine has a low noise level, high efficiency, long lifespan, and maintenance-free.

DC TO DC Converter A DC-to-DC is a power electronics converter that converts DC supply from one voltage level to another. DC output steps up voltage while stepping down current from its input supply to its output load. To reduce voltage ripple, filters are made of capacitors containing at least two semiconductors: a diode and a transistor. This will allow storing a common voltage from two different power generating sources.

Battery A 12 V, 7 A lead–acid battery is an electrochemical energy storage device used for this hybrid power system. It helps to store voltage generated from both solar and DC dynamo passing through the DC-DC converter. This will help to preserve charge in the form of electrons for lateral power requirements. This battery is chosen because of the longevity, low maintenance, low-cost, and high recycling rate.

WIFI-ESP32 ESP32 is a type of microcontroller module which has WiFi and Bluetooth enabling function. It helps in the monitoring of the system by relaying the electrical data wirelessly into the personal operating devices. This two-way data transmitter microcontroller (Fig. 1) is interfaced with various sensing devices which

Fig. 1 NodeMCU ESP32

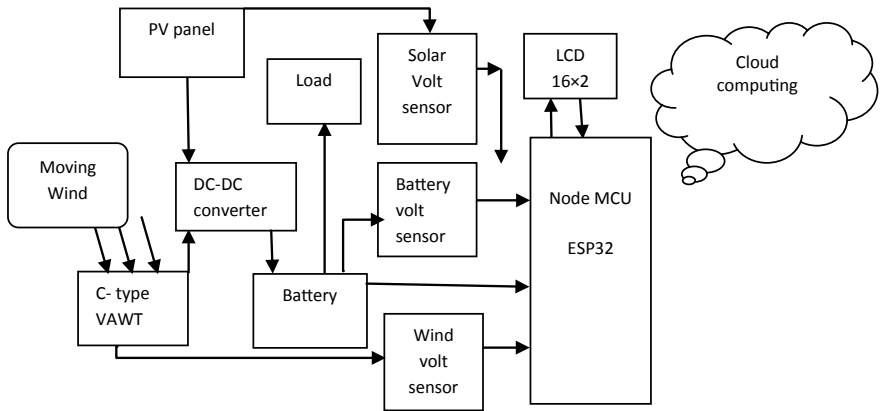


Fig. 2 Designed structure of Hybrid System

will sense the electrical parameters and help to monitor the load automatically from the base station (Fig. 2).

4 Methodology

The highway power generating hybrid renewable energy system comprises of solar photovoltaic panel and vertical axis wind turbine (VAWT). This C-type VAWT is designed in mechanical modeling software called CATIA. The designed 3-D model is simulated in CFD ANSYS software. The stress and strain values are analyzed separately at different values of wind velocity, force, and pressure. The deformation, stress, and strain effects can be seen thoroughly on the static condition of the blade in the CFD ANSYS simulation software. Thereafter fabricated model is being designed

by taking corresponding calculation and safety analysis values from the software. For the practical analysis, the hybrid power generation is tested under the solar radiation and normal wind velocity from 1.56 to 4.2 m/s condition. DC to DC converter helps to store a common voltage from two different fluctuating voltage generating sources like solar photovoltaic panel and DC dynamo integrated VAWT. A lead-acid battery having 12 V, 4.7 A which is the only source of recharging and discharging component, is connected with all other voltage generating sources and with power-consuming equipment. ESP32 is a WiFi-based microcontroller that helps to monitor and track the data through the protected website over the internet into the personal system. This ESP32 module helps to measure the voltage parameter separately on the solar, wind, and battery system through the connection of voltage sensors. This value will display into visualization form by MATLAB data simulation open IoT platform called Thinkspeak in a personal system.

5 Proposed Designed Model

5.1 3D Designed Model in CATIA

The most important power generation component of this hybrid system is VAWT which aids to generate power when sunshine is not available and reduces the total dependency of solar radiation. VAWT is designed in the CATIA modeling software tool. The 2-dimensional model is designed by determining the suitable measurement for the available condition like the highway side power generating system. Then, designed model is converted into 3-dimensional types to figure out the practical analysis of the model (Fig. 3).

Fig. 3 Isometric view of 3D C-Type VAWT



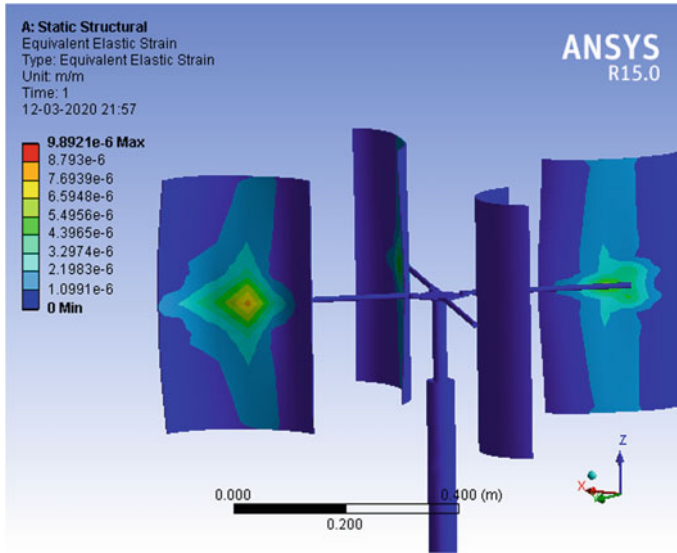


Fig. 4 Strain at 3.4 m/s wind velocity

6 CFD Analysis

The study of computational fluid dynamic behavior is done in the ANSYS workbench simulation software. The 3-D designed model is transferred for studying the practical approach by applying the physical analysis to determine the different parameters like stress, strain, and deformation under various conditions. ANSYS workbench is the software tool platform to get the virtual analysis on the model. The model has been tested under different wind velocities and pressures conditions for better endurance and efficient final prototype (Figs. 4 and 5).

Figures 6 and 7 show CFD Analysis on the simulation software. The effect of stress and strain distribution on the designed model of wind blade at different wind velocities can be observed (Fig. 8).

7 IoT System

Internet of Things is the process of exchanging data wirelessly through designed personal systems. The hybrid highway power generation system is easily monitored by such technology from the far base station. The working of the IoT system is shown in Fig. 10: in which it consists of ESP32 WiFi module microcontrollers, internet providers' source like WiFi router, and the controller is connected through a secured website via cloud computing platform (Fig. 9).

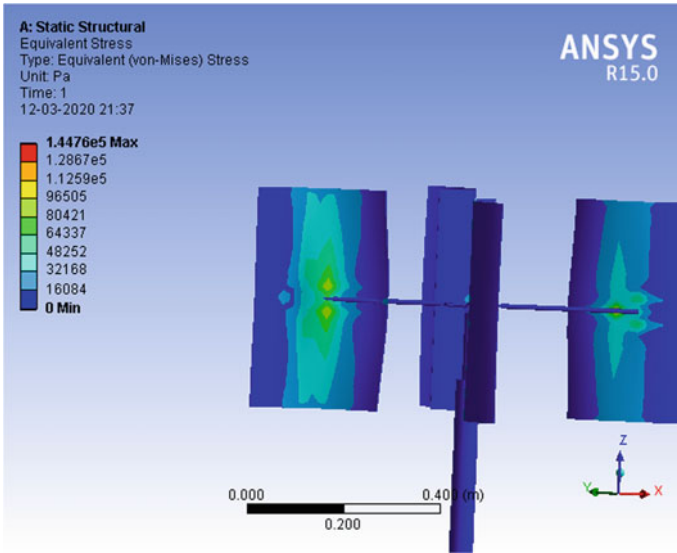


Fig. 5 Stress at 3.4 m/s wind velocity

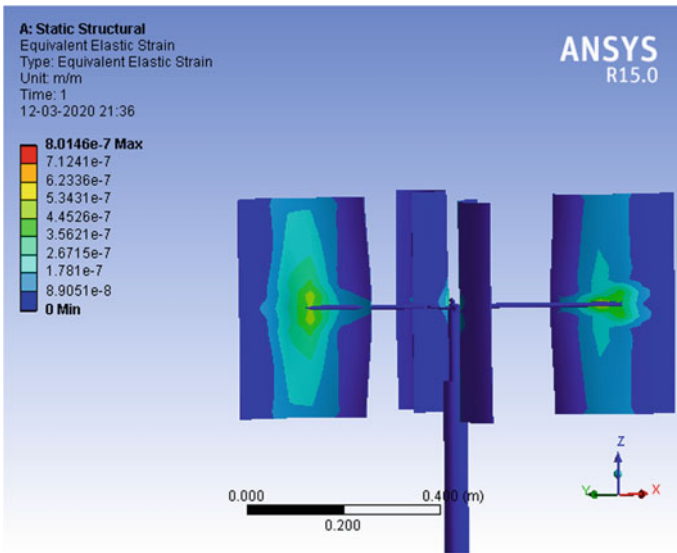


Fig. 6 Strain at 12 m/s wind velocity

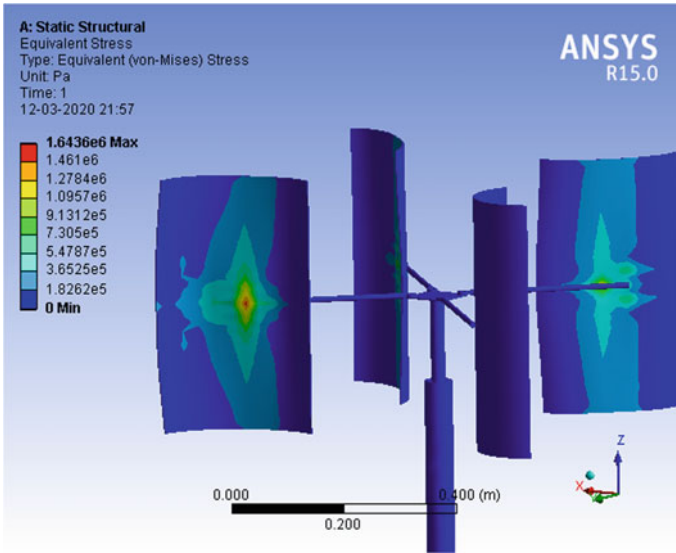


Fig. 7 Stress at 12 m/s wind velocity

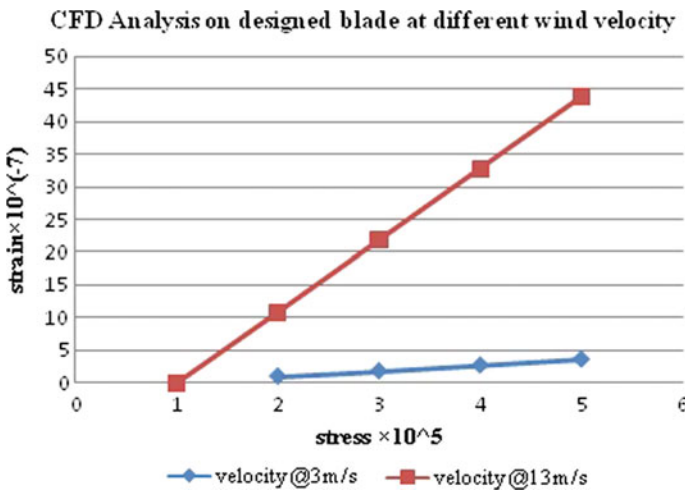


Fig. 8 CFD Analysis of the designed blade at different wind velocity

ESP32 microcontroller senses the voltage through the voltage sensor which has generated by the battery, wind turbine, and solar panel. This helps to exchange the electrical data wirelessly over the unique internet protocol. The data is monitored on the opened IoT platform source called Thingspeak. This is bidirectional controlling system software that enables us to read and write the monitoring process. It helps to eradicate manual operation. Thingspeak is a web protected platform, which has

Fig. 9 Prototype of IoT monitoring system

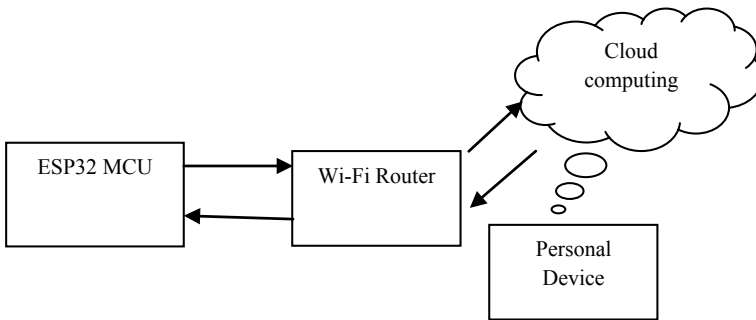


Fig. 10 Block diagram of the IoT working system

made it convenient to set parameters according to requirement. This system is having MATLAB analytical programming features that analyze data and processes to represent in the visualization method. The final output from IoT through integrated ESP32 and Thingspeak platform can be seen in Charts 1, 2 and 3.

8 Result and Calculation

Height of the blades (h) = 0.457 m

Radius of the rotor (R) = 0.422 m

Swept area,

$$S = 2 \times R \times h$$
$$S = 2 \times 0.422 \times 0.457$$

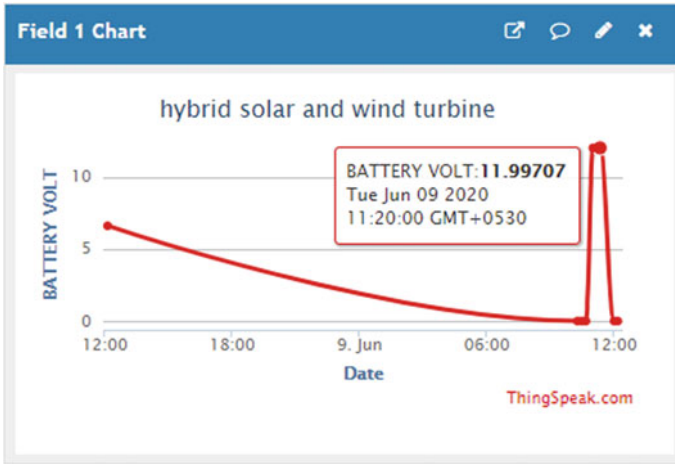


Chart 1 Battery Volt of the hybrid system

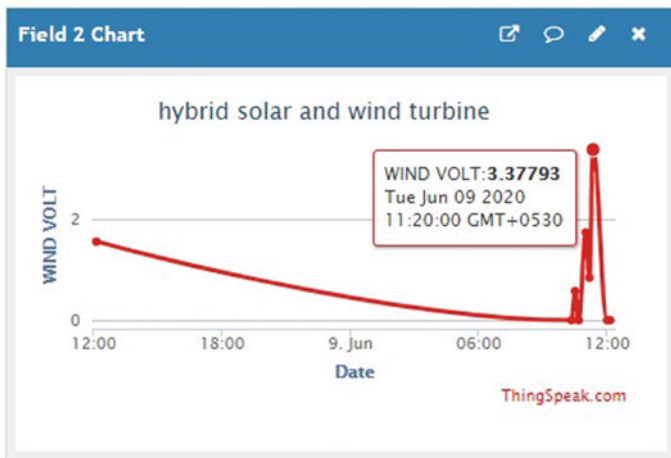


Chart 2 Wind volt of the hybrid system

$$\therefore S = 0.38 \text{ m}^2 \tag{1}$$

where S denotes the Swept area (m^2).

Torque produced by turbine blade,

$$\begin{aligned} \text{Torque } (\tau) &= 0.5 \times \rho \times S \times R \times v^2 \\ \tau &= 0.5 \times 1.225 \times 0.38 \times 0.422 \times 1.35^2 \\ \therefore \tau &= 0.179 \text{ N m} \end{aligned} \tag{2}$$

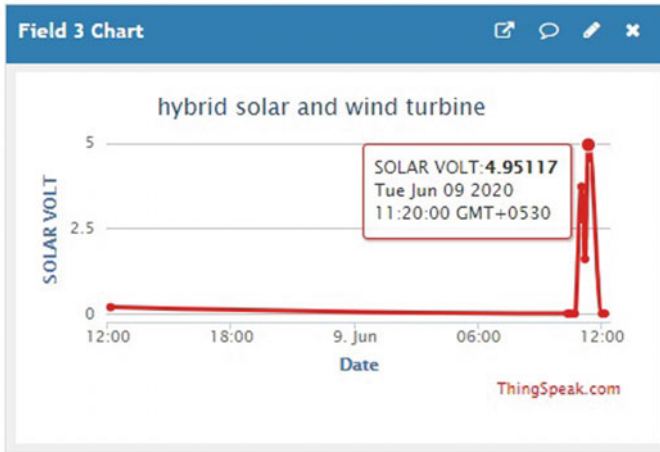


Chart 3 Solar volt of the hybrid system

where ρ denotes air density (kg/m^3) at STP, v denotes wind velocity at 1.35 m/s.

Output Power produced by turbine,

$$\begin{aligned} \text{Power}(P) &= 0.5 \times \rho \times S \times v^3 \\ P &= 0.5 \times 1.225 \times 0.38 \times 1.35^3 \\ \therefore P &= 0.5726 \text{ W} \end{aligned} \tag{3}$$

Angular velocity ω of turbine,

$$\begin{aligned} \omega &= \frac{\text{Power}}{\text{Torque}} \text{ rad/s} \\ \omega &= \frac{0.572}{0.179} \\ \therefore \omega &= 3.19 \text{ rad/s} \end{aligned} \tag{4}$$

Tip-Speed Ratio (λ) of blade,

$$\begin{aligned} \lambda &= \frac{\omega R}{v} \\ \lambda &= \frac{3.19 \times 0.422}{1.35} \\ \therefore \lambda &= 1.00039 \end{aligned} \tag{5}$$

A huge number of moving vehicles on city highway creates wind turbulence that brings low wind velocity but is capable enough to generate electricity. The torque

Table 1 Wind turbine parameters calculation

Wind turbine parameters	Wind velocity		
	1.35 m/s	2.4 m/s	3.6 m/s
Torque (N m)	0.1816	0.5742	1.292
Power (W)	0.5812	3.2657	11.022
ω (rad/s)	3.200	5.687	8.5310
TSR (λ)	1	1	1

produce by the aerodynamic loss of wind on the highway can be seen in the given table. As wind velocity increases the power generation through the turbine also increases simultaneously (Table 1).

9 Final Fabricated Model and Power Generation

After the final designed calculation, the prototype was tested under the varying pressure conditions in the simulation CFD Ansys software. The fabricated model is developed which is shown in Fig. 11.

The designed fabricated model is tested under varying wind velocity. The output power generated from the model can be seen in the given Fig. 12.

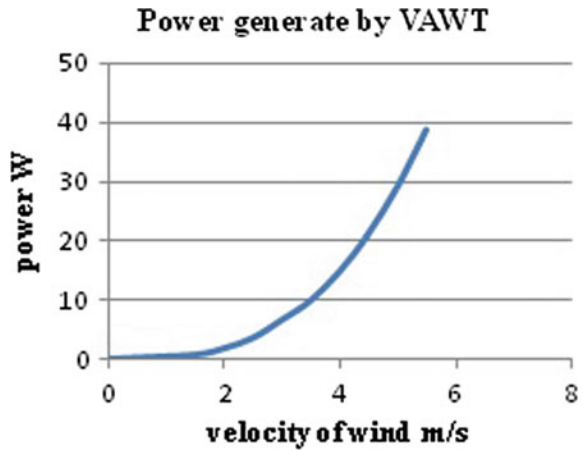
10 Conclusion

This paper has introduced the performance endurance of the designed fabricated model of hybrid renewable energy along with the management of the characteristic electrical parameter through wireless technology called IoT. The highway-based small scale hybrid renewable energy system help to serve the power solution for street light and standalone power source for the rural area. The implementation of a flexible wireless controlling system makes it easier to access the output of the system on devices such as smartphones and personal computers. The management of energy utilization and storage system help to reduce the problem of electricity inadequacy by controlling over secured web-enabled systems at a far distance. Such a hybrid system not only helps to avoid intermittent power generation through reducing solar dependency but also recapturing available wind by VAWT beside the busy highway, and it is very cost-effective, low maintenance, self-drive, and easily monitored and installed.

Fig. 11 Final model of Hybrid model



Fig. 12 Power generated by VAWT



References

1. Future of solar photovoltaic. <https://www.irena.org/remap>
2. Pandey A, Devi R (2017) Study and development of hybrid wind turbine for highway side application. *IJAREEIE* (2017)
3. Saat AF, Rosly N (2019) Aerodynamic analysis of vertical axis wind turbine. *J Aviation Aerospace Technol* 1(1):1–6
4. Li Z, Han R, Gao P, Wang C (2019) Analysis and implementation of a drag-type vertical-axis wind turbine for small distributed wind energy systems
5. Punugu VMN, Kabra TS (2018) Design and analysis of adaptable flexed-cup vertical axis wind turbine. *IJEDR* 6
6. Malge P, Ganesha T (2017) Study and analysis of Savonius vertical axis wind turbine with neodymium permanent magnet rotor. *Int Res J Eng Technol (IRJET)* 4
7. Hu Y, Wang T, Jin H, Cao X, Zhang C (2017) Experimental study on aerodynamic characteristics of vertical-axis wind turbine. *Int J Smart Grid Clean Energy* 2017. 10.12720
8. Fedak W, Anweiler S, Gancarski W, Ulbrich R (2017) Determination of the number of vertical axis wind turbine blades based on power spectrum. In: *EEMS* 2017. 10.1051
9. Menaka R, Mohan K, Muthu Vijay P, Ranjith I, Ragul D (2018) Power generation by hybrid VAWT system for highway applications. *Int J Adv Res Dev*
10. Siddiqui A, Gajbhiye M, Sawarkar P, Thakare R, Samrit V, Thote V (2020) A review paper on power generation with vertical axis wind turbine. *Int J Grid Distrib Comput* 13(1)
11. Srivastava P, Bajaj M, Rana AK (2018) IOT based controlling of hybrid energy system using ESP8266. *IEEE*
12. Castillo J (2011) Small-scale vertical axis wind turbine design
13. Kulkarni SA, Birajdar MR (2016) Vertical axis wind turbine for highway application. *Imperial J Interdisc Res (IJIR)* 2(10)
14. Das JBK, Murthy PLS (2014) Design of machine elements-1
15. Das JBK, Murthy PLS (2014) Design of machine elements-2

A Software Defect Avoidance Technique for E-commerce Website Development



R. Sudarshan, S. K. Srivatsa, K. N. Chaithra, and K. N. Mohan Kumar

1 Introduction

According to Laudon et al. [1], e-commerce is the use of the Internet, the Web, and applications to transact business. More formally, digitally enabled commercial transactions between organizations and individuals are referred as e-commerce. Digitally empowered dealings include all transactions facilitated by electronic technology. These transactions happen through the Internet, the Web, or mobile applications.

1.1 Present Status of e-commerce Growth in India

In India, steadily the tendency of e-commerce is rising. In support of this, there are numerous software business organizations asserting to build e-commerce websites.

The e-commerce trade is estimated to form the largest part of Indian economy with a worth of around USD 100 billion by 2020 as shown in Fig. 1.

R. Sudarshan
Department of Computer Science and Engineering, VELS University, Chennai 600117, India

S. K. Srivatsa
Department of Electronics Engineering, Anna University, Chennai 600044, India

K. N. Chaithra (✉)
Department of Electronics and Communication, Nitte Meenakshi Institute of Technology,
Bengaluru 560064, India
e-mail: chaithra.kn@nmit.ac.in

K. N. Mohan Kumar
Department of Computer Science and Engineering, MS Ramaiah University of Applied Sciences,
Bengaluru 560058, India
e-mail: mohan.cs.et@msruas.ac.in

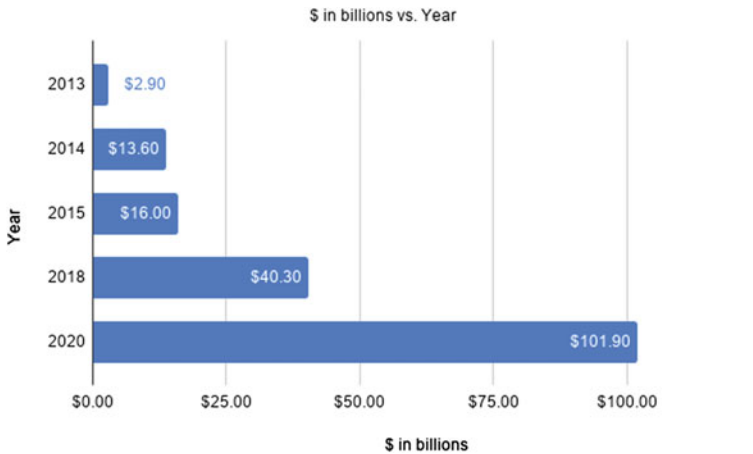


Fig. 1 E-commerce size (\$ billions) in India

Advancement of technology led to inventions such as digital payments and local logistics. Large-scale consumer engagements and digital marketing have empowered the e-commerce industry to grow quickly [2].

1.2 An Inquest on Software Defects

A defect is “An imperfection or deficiency in a work product as a result of which it does not meet its condition. This results in a correction or replacement of that work product.” as per Institute of Electrical and Electronics Engineers (IEEE) Standard 1044–2009 [3].

The cost of discovering and fixing imperfections are the major single overhead in the software development life cycle. Cost per defect penalizes quality. Most methods of testing are less than 35% in defect removal efficiency or eliminate roughly one bug out of three. Around 7% of bug fixes introduce new bugs. Nearly, 6% of test cases have defects of their own [4]. Defects are infused into the artifacts of the product at different phases [5].

Not all defects in the software development are attributed to coding. They may be also due to:

- misinterpretation of requirements
- incorrect study and recording of requirements
- partial requirements gathering
- non-identification of non-functional requirements
- erroneous design.

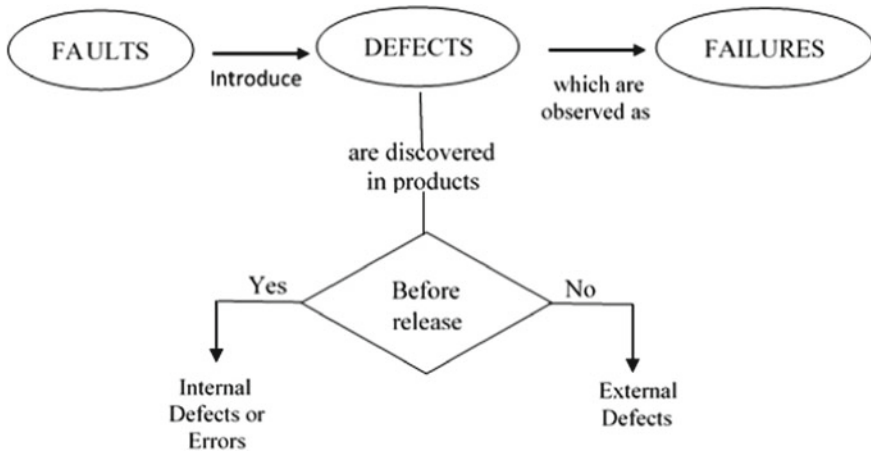


Fig. 2 External and internal defects

Defect elimination is one of the main overheads in any software project, and it significantly disturbs delivery timelines. Effective defect elimination leads to decrease in the delivery schedule and thereby leads to a better quality product [6]. Figure 2 indicates the assumptions made in this work regarding the defects and failures [7, 8].

The e-commerce project development consists of numerous highly coupled components and a numerous external integrations. Every component or module is built according to the phases of RUP, namely inception, elaboration, construction, and transition. Profound ambiguity exists at the ending phases of lifecycle as to which defect originated in which phase of the SDLC. There is a necessity to establish a procedure to pin point a defect to its origin and eliminate it at the stage itself. An attempt is made through this work to address this gap.

2 An Architectural Perceptive of E-commerce Sites

This study was conducted on five sequentially built e-commerce products by adopting ATG framework. The defect data collected from one project was utilized in the succeeding projects. This facilitated to improve upon the weak process areas. The size of each project was around 90 Executable Kilo Lines of Code (EKLOC) [9, 10].

2.1 E-commerce Application Architecture

An overview of application architecture based on ATG e-commerce platform is presented in this section. The objective is to capture significant architectural decisions, providing guidelines for detailed design and implementation. This information is crucial for planning to develop an unambiguous, workable technical solution so that defects could be avoided at an early stage.

Under application architecture, each subsystem identified represents a group of *logically related business functionality*, as shown in Fig. 3. The grouping is based

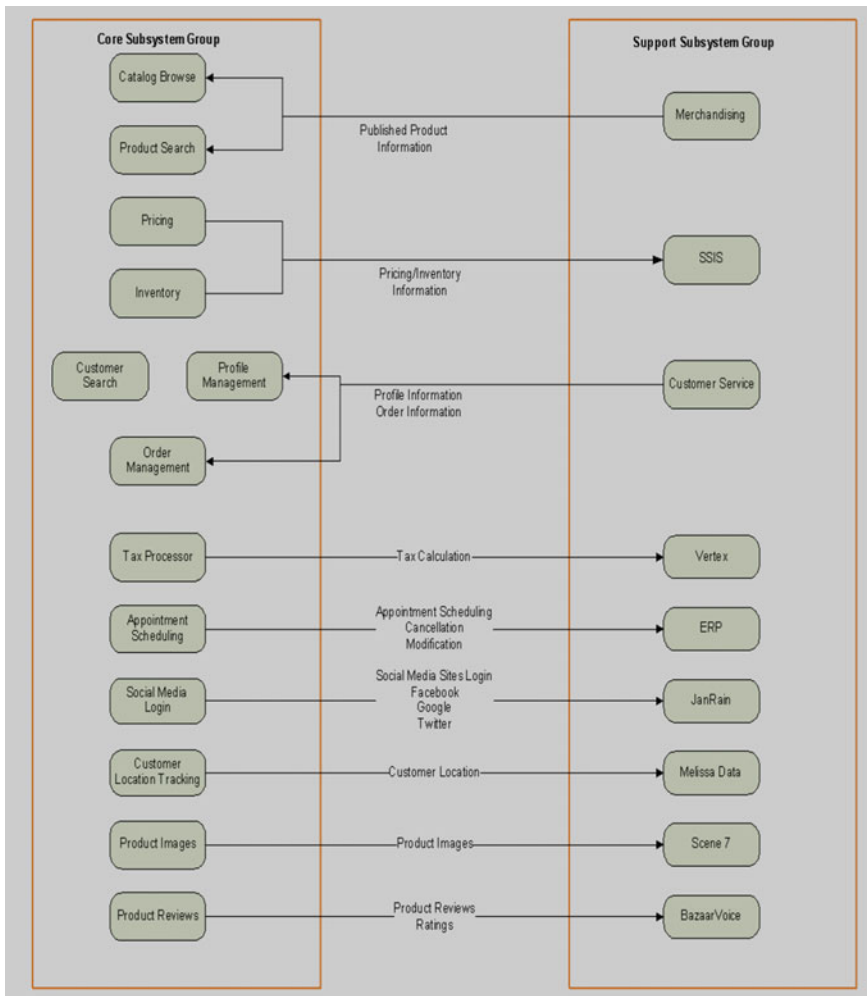


Fig. 3 Application architecture of ATG e-commerce website

on the architectural principle of high cohesiveness and low coupling between the subsystems. Subsystems are divided into core and support groups. Subsystems under the core group are responsible for all tasks related to the online store and contain all the online store-related business functionality. Subsystems under the support group are backend systems, responsible for supporting the online store in terms of, handling data publishing, reporting, etc. [7, 9].

2.2 Core Group Subsystem

The subsystems under the core group are further divided into tiers as per the technical architecture. The subsystems in both core and support groups and the information exchanged between them are shown in Fig. 4. This section details the subsystems identified under the core group for the application.

Browse

The browse component group defines components used to support the online store browsing experience. This component allows user to browse various category pages throughout the site. It is supported by the personalization, inventory, catalog, and pricing repository components. It uses these components to generate the home page, category landing pages, and product detail pages. Navigation and bread crumbs are also generated using these components. The search component helps the browsing component provide search capability anytime during the browsing experience.

Catalog

This repository component stores all the catalog content including products, stock keeping units (SKU) and categories. Catalog content is pushed from the *Merchandising* system to this component during deployments. It utilizes the services of inventory system for displaying and managing inventory levels. *Pricing* subsystem's services are leveraged to display prices of products and SKUs at various locations while displaying the catalog.

Product Reviews

This component is responsible for integrating with *Bazaar voice* to pull up the product reviews and customer rating.

Personalization

This backend component drives the dynamic promotional content on the browsing pages. It uses a combination of *Slots*, *Scenarios* and *Targeters* to give the site a more personalized feel. All promotional content including advertisements and special offers are managed by this component.

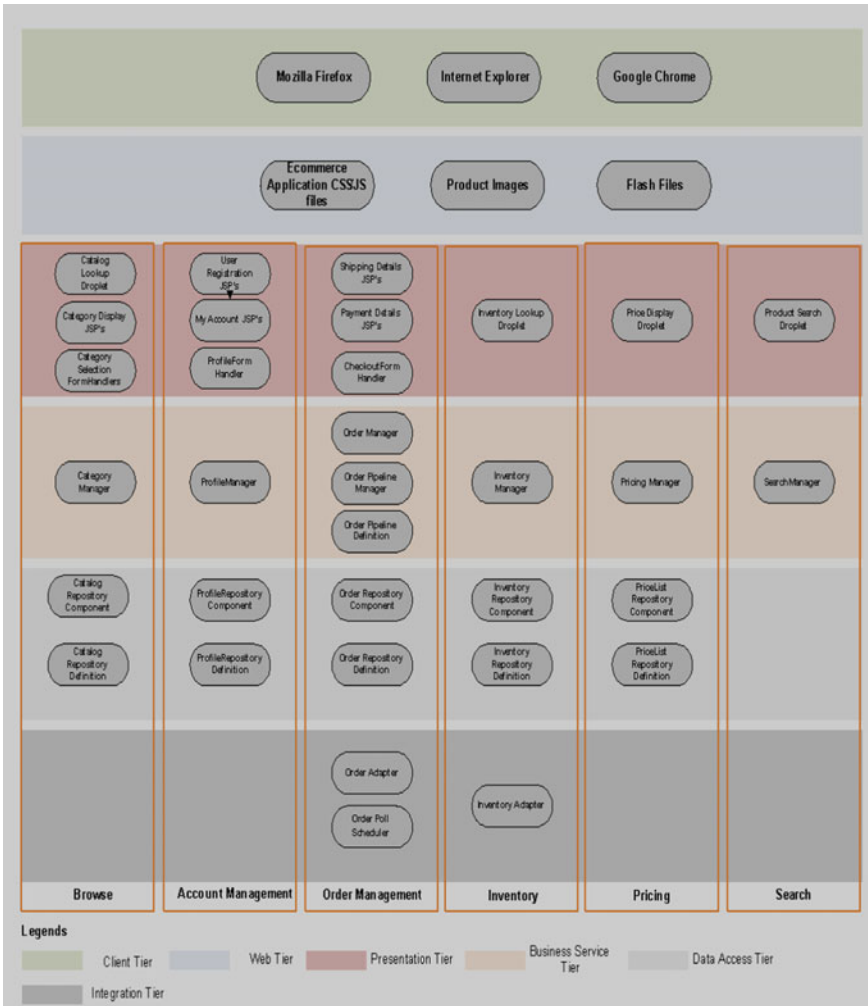


Fig. 4 Core group subsystem

Inventory

This subsystem is responsible for supporting inventory functionality for the site. This provides services to check stock, pre-order and back order, inventory levels at a SKU level. This also manages altering stock level as part of checkout and refund process. The inventory stock levels are updated by an external feed multiple times a day. The feed updates the inventory on the live site.

Following is the list of inventory status which is displayed on the site based on the status of the stock:

- *Estimated Availability:* in <n> number of days

- *In Stock*
- *Back Order*
- *Special Order*.

Pricing

This subsystem is responsible for supporting pricing functionality for the site. This provides services to fetch pricelist and manages prices for the products and SKUs. The prices are pushed to this component from the *Merchandising* system during a deployment. There could be different pricing options like *Special Pricing*, *Regular Pricing*, *Closeout Pricing*, etc.

Search

This component is responsible for supporting the search functionality for the site. It is used to run keyword searches, sorting search results, and results filtering. It sources the front-end search result pages with facets or refinements for search results filtering. Some of the search facets which are used include:

- Brand
- Price
- Performance category
- Size in different units
- Color
- Recently bought/searched.

Account Management

The account management component group defines components used to manage a user's account.

My Account

This component is responsible for allowing a user to update account specific information such as password, email address, and address book. It uses the profile component to get and store this information. The component also manages payment-related information and communication preferences. The user will also have the option to set the preferred vendor for a particular product.

Registration

This component allows new users to register for an account on the site. It collects all required information to create a new Web account. It also includes the opt-in preferences for newsletters, updates, and promotions. All these information are stored in the profile component.

Social Media Login

This component allows following social media users to login to the e-commerce application with their user Id. These social media users are not required to register on

the application. They can use login with *Facebook/Twitter/Google* button to provide their login and password and that will allow the user to login to the application as a regular user.

Following is the list of social media sites which support social media login.

- Facebook
- Twitter
- Google Apps

Profile

This repository component manages all the user-related information. This includes all information to support the *My Account*, *Order History* and *Checkout* components.

Order Management

The order management component group defines components used to manage the *shopping cart* and *checkout* processes.

Checkout

This component is used to manage the checkout process. It uses the *Profile*, *Order* and *Promotion* components. The *Profile* is used to give users the ability to login during the checkout process. All the shipping and billing information are stored in the order component. The *Promotion* component is used to apply associated discounts during the checkout process. During the last step of checkout process, the inventory is validated. After a successful validation, the entire order information is sent to the fulfillment component.

Shopping Cart

This component is used to manage the *shopping cart* functionality. It uses the *Order* component to store the items and pricing totals. The *Promotion* component is used to apply discounts during the pricing process. This component is integrated with the *Tax Calculator* component to calculate the tax and shipping costs for an order.

Tax Calculation

This component is integrated with *Shopping Cart* component to calculate the tax for a particular order. The tax is calculated either at the order level or at the line item level which needs to be finalized as part of functional requirement.

Order History

This component allows users to check the status of their orders. It provides the order details, status of the orders like, whether it is shipped, canceled etc. It uses the *Order* component to get this information.

Order

This repository component stores all the order-related information. It supports the checkout, shopping cart and track order components.

Promotion

This repository component is used to store the promotions, coupons, and reward codes. Promotions and coupons are created using this component to apply global or redeemable discounts.

Merchandizing

The merchandising component group defines components used for publishing content to the application and Web servers.

Publishing

This component is responsible for the management of all site contents and catalog contents. It allows admin users to create projects to edit content objects and deploys them to the production site. It also allows admin users to update user information.

Catalog and Content Feed

This component is responsible for importing catalog data from third-party vendors to the publishing system. The import process creates a new project and translates the data into publishing assets. Admin users can modify the assets before deployment.

Customer Service

The customer service component defines components used to support customer service tasks. The customer service component inherits extensions made to the catalog, pricing, order and promotion components.

Agent

This component is responsible for allowing customer service users to create, modify, and delete orders before it is submitted. The customer service user can also view a user's order history including the details and tracking numbers. It uses the shopping cart, track order, checkout, and order history components to perform these tasks.

3 Functionality Module Dependency Matrix for E-commerce Portal

All the working units of an e-commerce product (as detailed in segment 2.2) are first examined for the "Dependence" on each other, to comprehend the coherent order of artifacts (such as use-cases, architecture, test plan) that are created. For instance, inventory module functionality relies on catalog module, fulfillment module relies on checkout, browse and shop module depends on pricing and promotion, and so on. This functionality-dependency is shown in detail in Tables 1a, b.

Each intersection of the cell was given one of the appropriate values (depending on the functionality) from the following list:

Table 1 Functionality module dependency matrix

Modules	Internationalization	Membership and account management	Catalog	Search	Browse and shop	Pricing and promotions	Inventory
Internationalization		This (Row) uses to some features of the Column	No dependency	No dependency	No dependency	No dependency	No dependency
Membership and account management	Finish (Column) to start (Row)		No dependency	This (Row) uses to some features of the Column	No dependency	No dependency	No dependency
Catalog	Finish (Column) to start (Row)	No dependency		No dependency	No dependency	No dependency	Start (Column) to start (Row)
Search	Finish (Column) to start (Row)	Start (Column) to start (Row)	Finish (Column) to start (Row)		This (Row) uses to some features of the Column	Finish (Column) to start (Row)	No dependency
Browse and shop	Finish (Column) to start (Row)	This (Row) uses to some features of the Column	Finish (Column) to start (Row)	Finish (Column) to start (Row)		This (Row) uses to some features of the Column	This (Row) uses to some features of the Column
Pricing and promotions	Finish (Column) to start (Row)	This (Row) uses to some features of the Column	Finish (Column) to start (Row)	No dependency	No dependency		No dependency
Inventory	No dependency	No dependency	Finish (Column) to start (Row)	No dependency	No dependency	No dependency	
Modules	Shopping cart	Checkout	Fulfillment	Content management	Merchandising	Customer service	Reporting analysis
Shopping cart		No dependency	No dependency	No dependency	No dependency	Start (Column) to start (Row)	No dependency

(continued)

Table 1 (continued)

Modules	Shopping cart	Checkout	Fulfillment	Content management	Merchandising	Customer service	Reporting analysis
Checkout	Finish (Column) to start (Row)		No dependency	No dependency	No dependency	Start (Column) to start (Row)	No dependency
Fulfillment	Finish (Column) to start (Row)	Finish (Column) to start (Row)		No dependency	No dependency	Start (Column) to start (Row)	No dependency
Content management	No dependency	No dependency	No dependency		No dependency	Start (Column) to start (Row)	No dependency
Merchandising	No dependency	No dependency	No dependency	No dependency		Start (Column) to start (Row)	No dependency
Customer service	Finish (Column) to start (Row)	Finish (Column) to start (Row)	Finish (Column) to start (Row)	No dependency	No dependency		No dependency
Reporting analysis	This (Row) uses to some features of the Column	This (Row) uses to some features of the Column	No dependency	No dependency	No dependency	Start (Column) to start (Row)	
Feeds to data warehouse	Start (Column) to start (Row)	Start (Column) to start (Row)	Start (Column) to start (Row)	No dependency	No dependency	No dependency	No dependency

- Start (Column) to Start (Row)
- This (Row) uses some features of the above (Column)
- Finish (Column) to Start (Row)
- No dependency.

These tables could be used as a ready reference at the inception phase so that a better project plan could be drawn out for resource allocation. A meaningful, just-in-time resource allocation leads to cost saving by employing only necessary resources.

When every dependency is defined for all the working unites, then the developmental actions like construction of use-cases, architecture, test plan, and test cases are sequenced on the basis of logical flow. By this technique, numerous unidentified scenario of flow that may lead to defects could be discovered and corrected.

4 Conclusion

In e-commerce product development, a vital characteristic is the execution of various interfaces of the product subsystems to guarantee compatibility between them. These interfaces are not restricted only to user interfaces. They include information and data exchange among components of the product, like external data sources, middleware, and other outside sources. Interface handling has to be considered meticulously during the execution of the project.

A module functionality dependency matrix that is developed *before* the commencement for the project development as discussed in Sect. 3. This matrix helped in:

- Bringing in greater understanding of the sequence in which the activities have to be performed by each team member.
- Suggested the timing of the artifact that has to be *signed-off* by the relevant stake holders.
- The relevant artifacts were kept ready to be used by the subsequent stages in development without any anomalies.
- Bringing down the count of both, internal and external defects (as shown in Fig. 2).

References

1. Laudon KC, Guercio TV (2017) E-commerce—business, technology, society, 13th ed. Persons
2. Indian Brand Equity Foundation (2018, December) The rise and rise of e-commerce in India. Retrieved August 2019, from ibef.com
3. IEEE Standard Classification for Software Anomalies (2010). <https://doi.org/10.1109/IEEESTD.2010.5399061>

4. Jones C (2010) *Software engineering best practices—lessons from successful projects in top companies*, 1st ed. McGraw-Hill
5. Woody C, Ellison R, Nichols W (2014) Predicting software assurance using quality and reliability measures. In: Technical note, CMU/SEI-2014-TN-026, December 2014, pp 4–12
6. Kan SH (2004) *Metrics and models in software quality engineering* 2nd ed. Pearson Education
7. Benjamin AK, Carla EB, Hilmi O, Vijaykumar TN (2005) Detection and prevention of stack buffer overflow attacks. *Commun Assoc Comput Mach ACM* 48(11):50–56
8. Reinertsen D (2009) *The principles of product development flow: second generation lean product development*, celeritas
9. Meyer AN et al (2014) Software developers' perceptions of productivity. In: *Proceedings of the 22nd ACM SIGSOFT international symposium on foundations of software engineering (FSE 14)*, pp 19–29
10. Cunningham W (1992) The WyCash portfolio management system. In: *Proceedings of 1992 conference object-oriented programming systems languages and applications (OOPSLA 92)*. Available <http://c2.com/doc/oopsla92.html>

A Novel Methodology to Restructure Legacy Application onto Micro-Service-Based Architecture System



T. R. Vinay  and Ajeet A. Chikkamannur

1 Introduction

Designing and developing software application has undergone a sea change right from its inception to the present day. In its early stages, software engineers designed and developed stand-alone applications (desktop applications), where the application used to run on a single system. These systems were not interacting with any other applications or with the outside world. After the advent of www (internet), the need for interaction between applications was the main requirement. So engineers came up with new software development architectures to support web-based applications.

A legacy system, in the context of computing, refers to outdated computer systems, programming languages or application software that are still in use. Legacy modernization [10], or software modernization, refers to the conversion, rewriting or porting of a legacy system to a modern computer programming language, software libraries (improvement in Software Technology), protocols (Rapidly changing business rules), or hardware platform (improvement in Hardware Technology). Legacy transformation aims to retain and extend the value of the legacy investment through migration to new platforms.

On the other hand, hardware technologies were also improved rapidly. According to Moore's Law, we were able to insert more chips on a single die and more dies on a Mother Board. Initially, we had a single-core system, later multi-core system became very common. These multi-cores were used to build large servers.

T. R. Vinay (✉)

Nitte Meenakshi Institute of Technology, Bengaluru, India

A. A. Chikkamannur

R L Jalapa Institute of Technology, Bengaluru, India

T. R. Vinay · A. A. Chikkamannur

Visveshwaraya Technological University, Belgavi, India

Also because of the availability of very high-speed internet, cloud technology is becoming very popular.

Because of rapid improvements in hardware technology and the advent of cloud computing, there is no limit on hardware resources availability. Now the challenge to software engineers is to design and develop software applications that are independent, massively parallel modular applications that can serve a large number of requests (Scalable). To achieve these requirements, new architectures which support the above characteristics were developed.

2 The Problem

The legacy systems which were designed and built in the previous decade were traditionally single-threaded applications serving single request at a time. There was a need to migrate these business processing or data processing systems to modern hardware systems so that they can run as multi-threaded parallel applications thereby increasing its performance. The main challenge [13] to migrate these systems is to ensure that the migrated system satisfies all the business requirements, and it ensures completeness and correctness. Traditionally legacy system was built using procedural programming techniques that did not support parallelism. To understand this system, there was a shortage of skilled engineers where they needed to understand this legacy system and have knowledge of modern systems.

This can be achieved by simply ignoring the legacy system and build a new system, understanding all the business requirements by following the software development life cycle. This method leads to an increase in budget costs, takes time to develop and not fit in industrial schedules. In the end, we cannot assure it completely satisfies all the business requirements and leads to software failure [13]. Thus, we need to develop a novel methodology to migrate these legacy systems to modern systems which ensures correctness and completeness and well within timing and budgetary constraints.

The second problem is traditional software development process. In initial stages, a single team used to develop and manage large monolithic applications. As business requirements grow and change dynamically, more teams were roped in to develop large and complex monolithic applications. Many independent development teams are working in the same time simultaneously on the same code. This leads to less reliable applications, poor performance and might crash as well. The entire scenario is shown in Fig. 1.

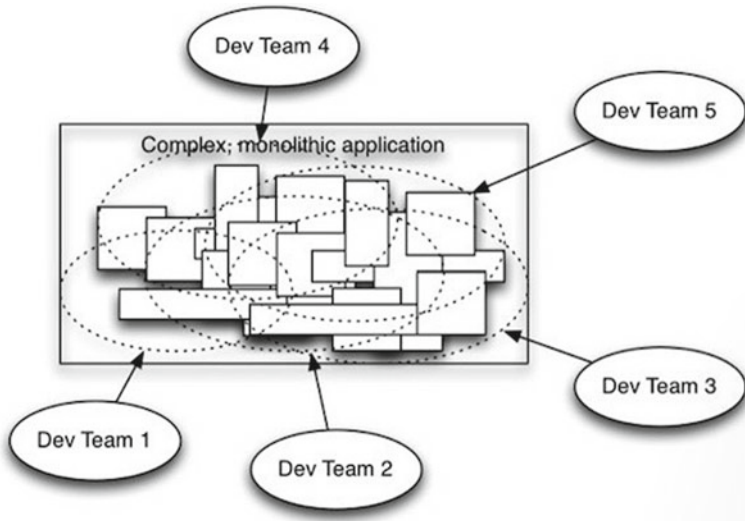


Fig. 1 Large complex monolithic application managed by many teams

3 Past Work

This section focuses on strategies employed for migrating software or system by different researchers under some predefined conditions. It also reviews the various technologies and makes observations on each of them.

4 The Proposed Methodology

This work proposes a novel semi-automatic methodology for restructuring legacy systems in the same relative abstraction level (Implementation level) [11, 12] to modern computing hardware platform such as multi-core systems or to the cloud. Here we apply the reverse engineering technique thereby ensuring that all business requirements are taken and fulfilled (completeness). Since the transition is done at the implementation level, it can also ensure correctness.

The input to our methodology is software requirement specification (SRS) [8, 9] of the legacy system and its complete program. The target architecture is micro-service-based application architecture system, where each service is a combination of well-defined user interface (UI) and its business logic (functionality). The services are small, independent, and loosely coupled. This leads to enterprises to become more agile and move toward DevOps practices and continuous testing. Below is in Fig. 2, showing micro-service-based application architecture.

Table 1 Literature review observation

S. No.	Author/title	Observation
1	Miller [1], “Reverse engineering Strategies for Software Migration”	<ul style="list-style-type: none"> • Identifying and minimizing risk in migration project • Strategy: Three categories of tools • Analysis Tool: to extract software artifacts such as flow graphs, slices and pointer analysis • Program understanding environments: uses parsers, software requirement specification (SRS) • Integrated forward and reverse engineering strategies • Modern system architecture (Target System) is not considered
2	Salopek [2], “Migration of Legacy Test Program to modern programming environment”	<ul style="list-style-type: none"> • Test programs are only considered for migration from legacy programming language into Newer programming language. Ex: ATLAS, BASIC to JAVA • New features of modern languages such as threads are not taken into account while migrating • Non-functional requirement of system such as security, performance is not considered • Newer hardware, new architecture machines are not taken into account while migrating
3	Zou et al. [3, 4], “Incorporating Quality Requirements in Software Migration Process”	<ul style="list-style-type: none"> • The authors take into account specific non-functional requirements such as maintainability and reusability of the new migrated system. It should confirm to this quality requirement • The process relies heavily on people having domain knowledge, documentation (SRS), and customer interviews • Here each task is identified and migrated • Specifically migration is applied on procedure oriented paradigm into object oriented paradigm • Migration process is depicted as state transition system, where the system changes states as being migrated • The bottleneck of this methodology is, it is dependent on documentation (design flaws) and domain knowledge of people involved in migration • Also parallelism available in object oriented technology is not considered

(continued)

Table 1 (continued)

S. No.	Author/title	Observation
4	Forite and Hug [5], “FASSM: Fast and Accessible Software Migration Method”	<ul style="list-style-type: none"> • The authors propose a technique based on model driven engineering and to capture knowledge as transformation rules • Here the development team defines the structure and border of the functionalities by consensus • The target language may offer different options to implement for a given problem • This technique is accessible to any team with knowledge in the field of modeling (UML)
5	Tepe [6], “ARNO Project: Challenges and Experiences in a Large-scale Industrial Software Migration Project”	<ul style="list-style-type: none"> • The paper presents the process involved in migrating a large software project, the tools used, migrating the data and re-implementing the job control. Also testing played a big role and consumed significant part of resources • Here first the existing code was renovated, i.e., identifying the dead code, unreferenced data and removing it. Here they are cleaning the code. Comment lines are retained • Converting procedure oriented system into object oriented system • Identifying the constraints and mitigating it was challenging • Customers/User involvement during testing was very crucial in each phase of migration from understating the existing task to migrating into new paradigm and testing it to satisfy customer requirements
6	Menychtas et al. [7] “ARTIST Methodology and Framework: A approach for the migration of legacy software”	<ul style="list-style-type: none"> • The authors emphasis on first conducting Migration feasibility assessment before taking up migration project. If the expenditure is very high, projects that never end and migration risk and failures • Next study on spplication discovery and understanding is carried out. For each task, model is generated and migration takes place. For each task, testing, verification, and certification are generated to gain confidence of customers and users

From Fig. 2, it is seen that each micro-service is managed by small development team, and there is no overlapping over each other. So, it is easier to maintain and modify according to change in business requirements.

The high-level design diagram in Fig. 3 shows the input to our methodology and how it is transforming the input to our desired small, independent micro-services which can be run in parallel on multi-core systems or on the cloud.

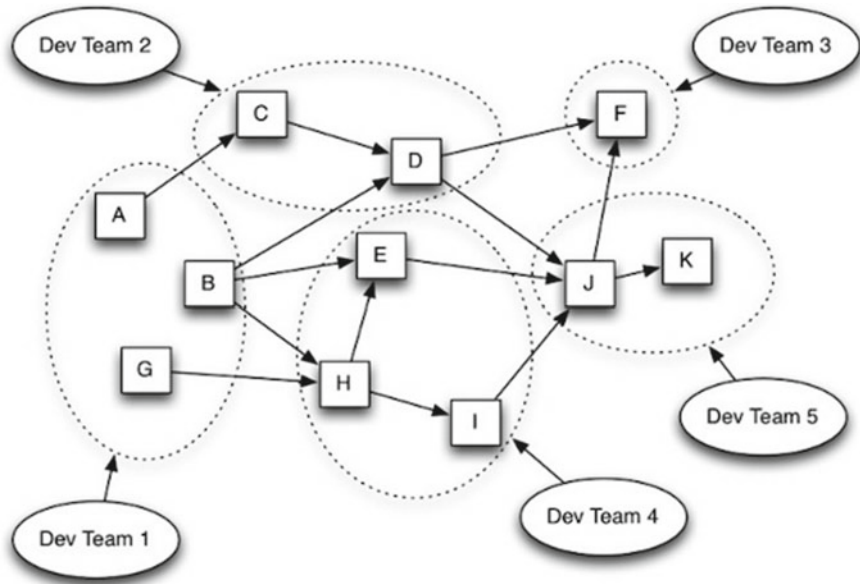


Fig. 2 Micro-service-based application architecture

4.1 Steps of the Methodology

1. From SRS document, extract functional requirements (FR) by using noun-phrased approach.
2. Similarly from the legacy program code, extract user interface (UI) and its corresponding functionality.
3. To ensure correctness and completeness, mapping between FR extracted from SRS to UI from legacy program is applied as shown in Fig. 3. The mapping results are shown in a Table format known as FR-UI hyper graph as shown in Table 2.
4. There may be three different outcomes of this mapping process
 - i. **1:1 mapping** (For every FR there is one UI). This is ideal case.
 - ii. **1:M mapping** (for every FR, there may be Many corresponding UI to achieve it).
 - iii. **M:1 mapping** (for many FR, there is only one UI which will satisfy all the requirements).
5. For any legacy application, after extracting FR and UI; if there is 1:1 mapping then, migrating it into micro-service-based architecture is straight forward.
6. If our legacy application has 1:M or M:1 mapping between FR and UI. There will be data dependency and control dependency issues.

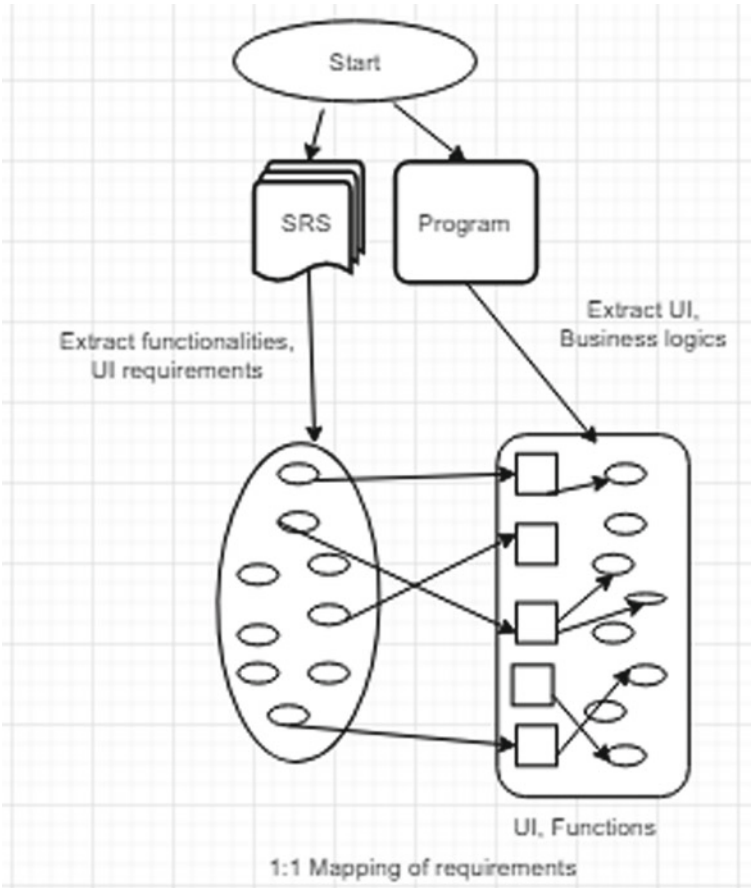


Fig. 3 Extracting requirements from SRS and program of a legacy system

Table 2 Functional requirement—user interface hyper graph

FR					
UI	UI-1	UI-2	...	UI-N	Remarks
FR-1		✓			1:1 Mapping
FR-2					
...					
FR-N	✓		✓	✓	1: M or M:1 Mapping

7. To resolve the issued defined in step 6, replicating the code and data has to be taken up, i.e., we need to ensure that for every FR, their exists only one UI. The mapping has to be brought down to 1:1. This step leads to redundancy of code. But the trade-off between achieving parallelism and having redundancy, the former will score over as the target system is micro-service-based architecture, and it wants each service to be a small and independent entity.
8. For every FR-UI, create a new service (combination of UI and corresponding functionality) in Micro-service architecture system.
9. Now all these service can be run in parallel and independently in the new system.
10. END.

Thus, by using SRS and legacy program and extracting functional requirements, user interface and corresponding code. The legacy application can be transformed into modern architecture using the above semi-automated technique and in the same abstraction level.

5 Case Study and Results

Considering a sample legacy bench mark application, we extracted its functional requirement from SRS and user interface and its corresponding code from the above application.

The sample functionality of this legacy program is as follows:

1. Begin
 2. Matrix[N][N] = {Random Values}, Here N is the size of the matrix which will have 10,000,20,000,40,000, 60,000,80,000 and 100,000 in each run.
 3. Determinant = FindDeterminant (Matrix).
 4. Sum = FindSumDiagonalelements (Matrix)
 5. Inverse[][1000] = FindInverse (Matrix)
 6. End
- [Example program for finding values of a Matrix]

5.1 Program Analysis

5.1.1 The Worst-Case Execution Time

The above program is implemented in legacy programming language, and for each functional requirement, a separate C-Function was implemented. The above program will run as a single-threaded program performing all the matrix operations

Table 3 Execution time of legacy program and its restructured parallel program

Data size	Legacy program (C-Program) in ms	Multi-threaded parallel program (using Java) in ms
10,000	419	366
20,000	1467	982
40,000	7493	5185
60,000	12,269	9629
80,000	25,516	15,871
100,000	39,293	26,026

in sequence in the order of implementation. The recorded worst-case execution time with different data sizes is shown in Table 3.

Now the above legacy program was transformed into Java program creating parallel threads for each of the functionality in the same abstraction level in accordance with micro-service architecture. This program will execute all the threads in parallel on a multi-core system. The recorded worst-case execution time is also shown in Table 3.

It is evident that multi-threaded parallel program will execute faster compared with single-threaded legacy program.

5.1.2 Total Memory Consumption

Memory consumption is the amount of memory used by a particular program under execution (process). It is expressed as a percentage. Memory available = 2 GB.

The legacy C program consumed less space, whereas Java program with multiple threads consumed more memory space. Since JVM transforms java code into byte code, it consumes more memory space and also space is consumed for creating, maintaining multiple parallel threads as shown in Table 4.

Table 4 Total memory consumption of the two programs in percentage

Data size	Legacy program (C Program) in %	Multi-threaded parallel program (using Java) in %
10,000	0.00083	0.00560
20,000	0.00087	0.00830
40,000	0.00093	0.02250
60,000	0.00098	0.02790
80,000	0.00124	0.03250
100,000	0.00156	0.03860

6 Conclusion

The objective to take this research work is to develop a migration and management tool to migrate the legacy systems into new hardware platforms and evaluate the performance of this new system. Migration is a continuous process in software industry because of rapid improvement/change in Hardware Technology, Software Paradigm viz. Computer Languages, DBMS Technologies, Operating System functionality and Software Engineering Process, Rapid, and Dynamic change in Business Operations. Due to emergency of new powerful, efficient modern computer languages, migrating to newer languages is necessary because it will be easier to maintain and to update them as business rules changes according to the organization needs and customer needs. To migrate from procedure-oriented paradigm to object oriented paradigm is also very much in need as OO methodology has proven itself to be powerful and shows correctness and completeness than the other paradigm. Also with micro-service architecture and multi-core systems, programs can be designed and executed by creating parallel and independent threads.

References

1. Miller HA Reverse engineering strategies for software migration. ACM Trans ICSE
2. Salopek PR (2000) Migration of legacy test program to modern programming environment. IEEE Trans
3. Zou Y Incorporating quality requirements in software migration process. In: 11th annual STEP-04
4. Kontogiannis K et al (1998) Code migration through transformations: an experience report. In: IBM CASCON 1998
5. Forite L, Hug C (2014) FASSM: fast and accessible software migration method. IEEE Trans Universite Paris
6. Teppe W (2009) ARNO project: challenges and experiences in a large-scale industrial software migration project. IEEE Trans
7. Menychtas A et al (2014) ARTIST methodology and framework: a approach for the migration of legacy software. IEEE Trans
8. Handigund SM reverse engineering of legacy COBOL systems. Doctoral dissertation, IIT, Bombay
9. Handigund SM, Arunakumari BN, Chikkamannur A (2018) Automated methodology to streamline business information flow embedded in SRS. In: Sa P, Bakshi S, Hatzilygeroudis I, Sahoo M (eds) Recent findings in intelligent computing techniques. Advances in intelligent systems and computing, vol 709. Springer, Singapore. https://doi.org/10.1007/978-981-10-8633-5_33
10. Jalote P An integrated approach to software engineering, 3rd ed. Narosa Publishing House
11. Chikkamannur A, Handigund SM (2009) An ameliorated methodology to design normalized relations. In: 2009 IEEE/ACS international conference on computer systems and applications, Rabat, pp 861–864. <https://doi.org/10.1109/AICCSA.2009.5069431>
12. Chikkamannur AA, Handigund SM An ameliorated methodology for ranking the tuple. Int J Comput Technol (IJCT) 14(4):5616–5620

13. Vinay TR, Chikkamannur AA (2016) A methodology for migration of software from single-core to multi-core machine. In: 2016 international conference on computation system and information technology for sustainable solutions (CSITSS), Bangalore, pp 367–369. <https://doi.org/10.1109/CSITSS.2016.7779388>

A Survey of Static and Dynamic Metrics Tools for Object Oriented Environment



Manju  and Pradeep Kumar Bhatia 

1 Introduction

Metrics are supposed to obtain more accurate estimations of project milestones and developing a software system that has minimal faults. Now most popular approach to measuring the quality of software is object-oriented software metrics. It is of mainly two types, i.e., static metrics and dynamic metrics. The static software metrics are obtained from static analysis of the software in software development life cycle, whereas dynamic software metrics are computed based on the data collected during execution of software. So, static metrics can be extracted in early phase of software development life cycle for better estimation of software quality, whereas dynamic metrics are extracted during and after the coding phase. There is number of tools available like open-source, free, commercial for static metrics but very few exist for dynamic metrics [1]. In this paper, 18 static metrics and 3 dynamic metrics tools are compared mainly on 4 characteristics. Firstly, type of languages supported by tool, type of tool whether it is free, open-source or commercial, format of output supported by tool and lastly, number of metrics supported by the tool. There is number of tools available for static metrics calculation mainly in three categories: free, commercial and open-source [2]. Authors studied all major tools and include only those tools that are updated and easy to install. Exclusion of¹ tools are done on installation basis, outdated and incompatible with the operating system. On this basis, authors include only 18 tools for static metrics evaluation. Dynamic metrics tools measures run time behavior of software. Authors found that there is number of tools available for dynamic analysis of software but our main focus is only on metric

¹ Excluded Static Metrics Tools: jpeek, cloc, ohcount, Metrics, cyvis, javaNCSS, SLOC-Count, CodeCount, LOCC, BCML, JMetric, ES2, Semml, source code metrics, refactorIT, SonarJ, JCSC, SAME, PMD, CPD, Xradar, checkstyle, QALap, JLint, Classycle, Squal, Soar Plugins.

Manju (✉) · P. K. Bhatia
Guru Jambheshwar University of Science and Technology, Hisar, India

evaluation tools [1]. Therefore, only metrics evaluation tools are included in study and other metric analysis tools are excluded² like memory leak, heap allocation, buffer overflow, etc. The main purpose of this comparative study is to provide availability of object-oriented static and dynamic metrics tools and various characteristics of these tools. On the basis of this study, authors tried to find out the answer to the following questions:

Q1: Whether all the object-oriented features are evaluated by static and dynamic metrics tools?

Q2: Is there enough tools available to calculate dynamic metrics?

Q3: Is there enough languages supported by existing tools?

The rest of the paper is organized as follows: Sect. 2 provides the detailed literature review of object-oriented static and dynamic metrics tools. Section 3 explains various static and dynamic metrics tools with their characteristics. Section 4 discusses the key observations drawn from the current study, and lastly, the paper presents conclusion of the current study in Sect. 5.

2 Related Work

Several object-oriented metrics tools have been proposed and developed over the past several years. The tools differ in variety of attributes, features, language supported, metrics supported, etc. This literature study has been done to analyze the object-oriented static and dynamic metrics tools. Summary of the literature review is shown in table in decreasing order of year from 2019 to 1997. From Table 1, it is observed that most of the studies made by different researchers were based on static metrics evaluation tools and very few researchers worked or analyzed dynamic metrics tools.

3 Software Metrics Tools

3.1 Static Metrics Tools

Static metrics tools are used to analyze software in the early phases of software development life cycle by measuring static metrics given by various researchers. In this study, authors mainly focus on 18 static metrics tools and explained them with the help of 4 main characteristics. Firstly, type of software (freely available, commercial, open-source), language supported (java, C++, etc.), output format (xml, html etc.), and lastly, number of metrics supported by various tools. All the commercial tools

² **Excluded Dynamic Metrics Tools:** DMA, SSS, Daikon, Valgrind, Rational Purify, Parasoft insure++, Pin, Javana, DIDUCE, DJProf, Racer, JInsight.

Table 1 Summary of literature review on static and dynamic metrics tools

Author's name	Work done	Conclusion	Described tool
Babur et al. [3]	Comparison of various static and dynamic object oriented tool with a new hybrid tool AndroPyTool	AndroPyTool is a hybrid tool and best for android applications	Hybrid (Static and dynamic)
Schnoor et al. [4]	Comparison of 3 measurement approaches named dynamic weighted, dynamic unweighted and static coupling metrics	There is a significant correlation between static and dynamic weighted metrics	Static and dynamic
Gupta et al. [5]	Comparison of 8 static metrics tools on 14 metrics	Metrics value are tool dependent	Static
Fregnan et al. [6]	Survey on vast quantity of coupling metrics and their relation with tools	There are not enough tools available to calculate coupling metrics	Static and dynamic
Mshelia et al. [7]	Comparison of 10 tools and explain these tools with the help of 15 characteristics	Metrics and language supported by these tools is very limited	Static
Reddy [8]	Introduction of new tool named software metrics tool designed in c# language worked as web application	Tool is good to calculate statistics as well as important metrics	Static
Kayarvizhy [2]	Review of object oriented metrics tools	Need of more tools to support extensibility and flexibility in the object oriented environment	Static
Gosain et al. [1]	Survey of dynamic metrics analysis techniques and tools	Only few dynamic metrics evaluation tools are available	Static and dynamic
Sarvari [9]	Explanation of efficient and easy collection of dynamic metrics with MapReduce program	73% increasing performance metrics evaluation with this approach	Dynamic

(continued)

Table 1 (continued)

Author's name	Work done	Conclusion	Described tool
Gitika et al. [10]	Survey on various dynamic coupling metrics	Most of the metrics are object level coupling metrics and mapped to maintainability quality factor	Dynamic
Dogra [11]	Comparison of 4 open source software with JHawk tool	Some metrics are highly correlated as compared to others in predicting faults	Static
Tomas et al. [12]	Comparison of 17 static metrics tools	Only 3 tools support internal quality models	Static
Tahir et al. [13]	Survey of dynamic metrics and their mapping to software quality	Most of the research is focused on coupling, cohesion and maintainability metrics	Dynamic
Baker et al. [14]	Comparison of 4 static metrics tool on various object oriented features	Metrics values are tool dependent	Static
Rani et al. [15]	Description of SDmetrics tool for UML language	This tool trace error in design phase before execution and save time and efforts	Static
Novak et al. [16]	Comparison of 5 tools for 5 metrics based on c# language	Various tools give different value for various metrics on same source code	Static
Kocagumeli et al. [17]	Introduction of new tool named Prest to calculate static metrics	Metrics calculated by this tool is 28 and 5 languages are supported by this tool	Static
Singh et al. [18]	Introduction of new tool named DynaMetric	Only tool to evaluate static and dynamic metrics	Static and dynamic
Lincke et al. [19]	Comparison of 10 object oriented tools	Different tools give different metrics value for same source code	Static
Dufour et al. [20]	Introduction of a new tool named *J for dynamic analysis of java program	Very efficient tool	Dynamic
Mayo et al. [21]	Introduction of interface and dynamic metrics	Static and dynamic both metrics should be calculated for better measurement of quality	Static and dynamic

included in the current study are available on trial basis. Therefore, we can say that tools with their availability, flexibility and regularly updated are only included in our study. 18 static metrics tools with their 4 characteristics are shown in Table 2. SDMetrics and JHawk tool support highest number of metrics and both tools come under the category of commercial tools but are available on trial basis. The most common metrics set calculated by almost all the tools are CK metrics suites and then MOOD metrics suite.

3.2 *Dynamic Metrics Tools*

Dynamic state of program is the realistic view rather than static state. It gives a more precise view in terms of actual behavior of the program. Dynamic metrics can handle better run time programming features like dynamic binding, polymorphism, dynamic allocations, etc. Dynamic metrics tools are used to calculate metrics during run time of a program. Authors mainly found 3 dynamic metrics tools. Firstly, DynaMetrics [18] was proposed by Singh and Singh for C++ and Java languages. It supports 13 static and 15 dynamic metrics. Dynamic metrics evaluated by this tool are mainly coupling and complexity metrics. Secondly, JDissect [23] was proposed by Arisholm et al. mainly for Java language and mainly focus on coupling metrics. Lastly, *j [20] was proposed by Bruno Dufour et al. only for java language. Main focus of this tool is on polymorphism-based metrics. Dynamic metrics tools with their 4 main characteristics (tool type, language supported by tool, output format of tool and numbers of metrics supported by tool) are described with reference in Table 3.

4 Key Observations

The authors conducted a systematic study on existing static and dynamic object-oriented metrics tools. It has been found that most of the static metrics tools are freely available and tools under commercial category are also available on trial basis, whereas mostly dynamic metric evaluation tools are offline that results in lack of validation of dynamic metrics. The study also identified that only limited numbers of dynamic metrics tools are available as compared to static metrics tools. Following are the key points drawn from the study.

14 static metrics tools support java language as input to the tool and after java, C++ and C# are the most preferred languages as inputs to the various existing tools. Many other languages are also supported by static metrics tools like UML, HTML, VB, .NET, etc., whereas dynamic metrics tools support very limited number of languages as input as shown in Fig. 1.

- Authors found that number of languages supported by static metrics tools are 9, and dynamic metrics tools are only 2 as shown in Fig. 2 which indicates that there

Table 2 Static metrics tools

Tool name	Tool type	Language supported	Output	Metrics
CCCC ³	Open-source	C, C++, Java	.xml, .cc	8
CodeAnalyser ⁴	Free	C++, C, Java, html, ass	.html, .txt, .csv	11
Sourcemonitor ⁵	Free	C++, C, C#, VB.NET, Java, Delphi, VB6	.smp, .smproj, .xml, .csv	2
JDepend ⁶	Open-source	Java	.xml, .dot	7
LOCMetrics ⁷	Free	C#, C++, Java, SQL	.cpp, .cc, .h, .inl, .cs, .java, .html, .jpg	12
NDepend ⁸	Commercial	VB, C#	.xml	82
Dependency finder ⁹	Open-source	Java	.xml	>33
CKJM ¹⁰	Open-source	Java	.xml, .xslt	8
SDMetrics ¹¹	Commercial	UML	.xml, .html	128
RSM ¹²	Commercial	C++, C#, Java	.xml, .html, .txt, .csv	>16
JHawk ¹³	Commercial	Java	.csv	102
QMOOD++ [22]	Free	C++	.xml	>30
Analyst4j ¹⁴	Free	Java	.xml	>25
Eclipse Metrics Plugin 1.3.6 ¹⁵	Free	Java	.xml	23
Eclipse Metrics Plugin 3.4 ¹⁶	Free	Java	.xml	25
Understand ¹⁷	Commercial	C, C++, Java, Jovial, Pascal, ADA, .NET, Python	.html, .csv	>50
VizzAnalyser ¹⁸	Free	Java	.xml	>10
CodeMR ¹⁹	Free	C++, Java, Scala	.html	27

³ <https://sourceforge.net/projects/cccc/>

⁴ <http://www.codeanalyzer.teel.ws/>

⁵ <http://www.campwoodsw.com/sourcemonitor.html>.

⁶ <http://www.testingtoolsguide.net/tools/jdepend/>

⁷ <http://www.locmetrics.com/>

⁸ <https://tomassetti.me/ndepend/>

⁹ <http://depfind.sourceforge.net/>

¹⁰ <https://www.spinellis.gr/sw/ckjm/doc/indexw.html>

¹¹ <https://www.sdmetrics.com/download/SDMetricsManual.pdf>

¹² <https://msquaredtechnologies.com/>

¹³ <http://www.virtualmachinery.com/jhawkprod.htm>

¹⁴ <https://www.theserverside.com/discussions/thread/44693.html>

¹⁵ <http://metrics.sourceforge.net/>

¹⁶ <https://marketplace.eclipse.org/content/eclipse-metrics>

¹⁷ <https://scitools.com/category/release/>

¹⁸ https://www.researchgate.net/publication/241753972_VizzAnalyser-_A_Software_ComprehensionFramework

¹⁹ <https://www.codemr.co.uk/downloads/>

Table 3 Dynamic metrics evaluation tools

Tool name	Tool type	Language supported	Output	Metrics supported	References
DynaMetrics	Offline	C++, Java	.xml	Static-13 Dynamic-15	[18]
JDissect	Offline	Java	.txt	52	[23]
*J	Online	Java	.xml	21	[20]

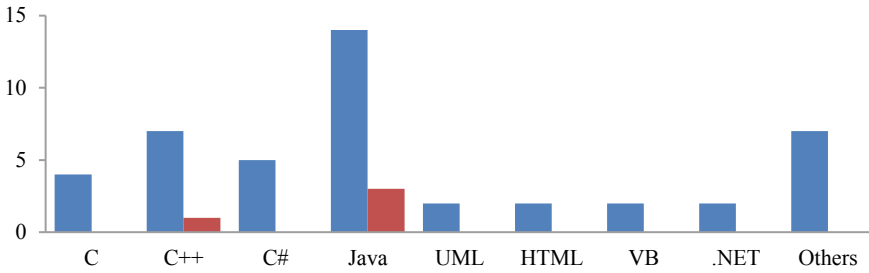
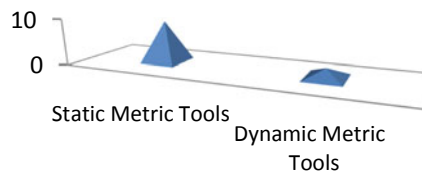


Fig. 1 Comparison of languages supported by static and dynamic metrics tools

Fig. 2 Comparison of number of languages supported by static and dynamic metric tools



is lot of work is to be done in the field of dynamic metrics tools so that they can support more number of languages as input.

- Authors observed from the study that many researchers worked on coupling and complexity object oriented feature and designed respective metrics to evaluate these features with the help of existing tools, whereas there are not enough tools available to measure run time features of software like dynamic cohesion, run time polymorphism, dynamic binding, etc., as shown in Fig. 3.
- From the study, authors found that static metrics tools measure more number of metrics as compared to dynamic metrics tools. Maximum number of metrics are evaluated by static metrics tools like SDMetric [32] and JDepend [27], whereas in case of dynamic metrics tools, JDissect evaluates the maximum number of metrics as shown in Fig. 4. Metrics covered by dynamic metrics tools are mainly coupling metrics. From the study, authors found that most of the metrics mapped to maintainability and reusability quality features leaving the other necessary features like fault proneness, reliability, testability, etc., to measure quality of a software that revealed that there is lots of work required in the field of mapping of metrics to quality factors and their validation too.

Fig. 3 Object-Oriented features supported by static and dynamic metrics tools

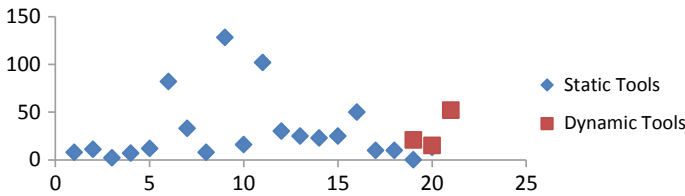
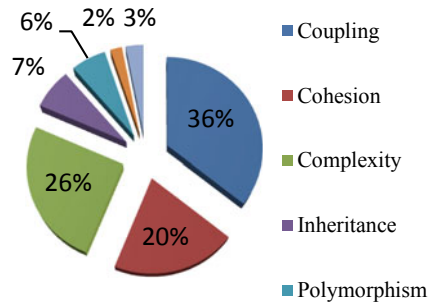


Fig. 4 Number of metrics supported by static and dynamic metric tools

5 Analysis Results

Authors analyzed 18 static and 3 dynamic tools for metric evaluation. Based on this study, authors concluded that there are various tools available for static metrics calculation but only few for dynamic metrics. Further, metrics and language supported by dynamic metrics tools are very limited. From this study, author tried to answer Q1, Q2 and Q3 as followed:

Q1: Whether all the object-oriented features are evaluated by static and dynamic metrics tools?

Figure 3. Not exactly, all the object-oriented features are not explored by these tools. There exist limited number of tools to support features like dynamic binding, run time polymorphism and inheritance.

Q2: Is there enough tools available to calculate dynamic metrics?

Table 3: No, there is not enough tools available to calculate dynamic metrics. There is a number of tools available for dynamic analysis like Daikon (Invariant Detection), Valgrind (Memory Leak), Pin (Cache Modeling), DJProf (Heap Allocation), Racer (Heap Allocation), Caffeine (Object LifeTime Analysis), JInsight (Deadlock Detection) but very limited number of tools available for metrics computation.

Q3: Is there enough languages supported by existing tools?

Figure 2. Authors observed that static metrics tools support 9 languages as input, whereas dynamic metrics tools only support 2 languages and the most supported language is java (Fig. 1) in both the cases as input that indicates more work is required to design dynamic metrics tools so that these tools help the real world and captures the run time features of any software that is programmed in any object-oriented language.

6 Conclusion

This paper compared the existing static and dynamic metrics evaluation tools with the help of 4 characteristics that is type of tool, language supported by tool as input, output format of tool and number of metrics supported by tool. From the survey of existing literature, some observational results have been revealed. Firstly, there is very limited number of dynamic metrics evaluation tools exist as compared to static metrics tools. Secondly, number of metrics and language supported by dynamic metric tools is very less when compared with static metrics tools. Lastly, existing dynamic metrics tools mainly evaluate coupling object-oriented feature and very little emphasis is done on features like dynamic cohesion, run time polymorphism, dynamic binding, etc. Therefore, it is concluded from the survey that more work is required in the field of dynamic metrics tools so that they are available online or free to calculate static and dynamic metrics and statistics can also be applied on metrics evaluated by tools so that quality measurement of software gets easier.

References

1. Gosain A, Sharma G (2014) A survey of dynamic program analysis techniques and tools. In: 3rd international conference on frontiers of intelligent computing: theory and applications (FICTA) 2014. *Advances in intelligent systems and computing* 327, vol 1. Springer, Cham, pp 113–122. https://doi.org/10.1007/978-3-319-11933-5_13
2. Kayarvizhy (2016) Systematic review of object oriented metric tools. *Int J Comput Appl* 135(2):8–13
3. Malik B, Khalid J, Arif H, Sadiqa A, Tanveer A, Mumtaz A, Afzal Z, Azhar S (2019) Comparing hybrid tool for static and dynamic object-oriented metrics. *Int J Adv Comput Sci Appl* 10(5):525–530. <https://doi.org/10.14569/IJACSA.2019.0100568>
4. Schnoor H, Hasselbring W (2019) Comparing static and dynamic weighted software coupling metrics. In: *information and software technologies—25th international conference (ICIST 2019)*, vol 1078, pp 285–298. https://doi.org/10.1007/978-3-030-30275-7_22
5. Gupta EM, Mehta RK, Rai EM (2019) A consolidated and comparative analysis of software metrics tools for systems performance evaluation: a survey. *IJARCCCE* 8(4):160–167. <https://doi.org/10.17148/IJARCCCE.2019.8427>
6. Fregnan E, Baum T, Palomba F, Bacchelli A (2018) A survey on software coupling relations and tools. *Inf Softw Technol* 107:159–178. <https://doi.org/10.1016/j.infsof.2018.11.008>

7. Mshelia YU, Apeh ST, Edoghogho O (2017) A comparative assessment of software metrics tools. In: International conference on computing networking and informatics (ICCNI). IEEE, Lagos, pp 1–9. <https://doi.org/10.1109/ICCNI.2017.8123809>
8. Reddy VR (2016) Software Metrics tool. Master of science dissertation, North Dakota State University Of Agriculture and Applied Science, North Dakota
9. Sarvari S, Singh P, Sikka G (2015) Efficient and scalable collection of dynamic metrics using MapReduce. In: 2015 Asia-Pacific software engineering conference (APSEC), pp 127–134. <https://doi.org/10.1109/APSEC.2015.21>
10. Geetika R, Singh P (2014) Dynamic coupling metrics for object oriented software systems: a survey. ACM SIGSOFT Softw Eng Notes 39:1–8. <https://doi.org/10.1145/2579281.2579296>
11. Dogra S (2013) Metrics evolution using open source software. M. Tech thesis, School of Mathematics and Computer Applications, Thapar University
12. Tomas P, Escalona MJ, Mejias M (2013) Open source tools for measuring the Internal Quality of Java software products: a survey. Comput Standards Interfaces 36(1):244–255
13. Tahir A, MacDonell SG (2012) A systematic mapping study on dynamic software quality metrics. In: 28th IEEE international conference on software maintenance, Riva del Garda, Italy. IEEE Computer Society Press, pp 326–335
14. Bakar NSAA, Boughton CV (2012) Validation of measurement tools to extract metrics from open source projects. In: IEEE conference on open systems, Kuala Lumpur, pp 1–6. <https://doi.org/10.1109/ICOS.2012.6417648>
15. Rani T, Sanyal M, Garg S (2012) Measuring software design class metrics: a tool approach. IJERT 1(7):1–7
16. Novak J, Rakić G (2010) Comparison of software metrics tools for: net. In: 13th international multiconference information society-IS, vol A, pp 231–234
17. Kocaguneli E, Tosun A, Bener AB, Turhan B, Caglayan B (2009) Prest: an intelligent software metrics extraction, analysis and defect prediction tool. In: 21st international conference on software engineering and knowledge engineering (SEKE'2009), pp 637–642
18. Singh P, Singh H (2008) DynaMetrics: a runtime metric-based analysis tool for object-oriented software systems. ACM SIGSOFT Softw Eng Notes 33:1–6
19. Lincke R, Lundberg J, Löwe W (2008) Comparing software metrics tools. In: International symposium on software testing and analysis. ACM, pp 131–142
20. Dufour B, Hendren L, Verbrugge C (2003) A tool for dynamic analysis of Java programs. In: Proceedings of 18th annual ACM SIGPLAN conference on object-oriented programming, systems, languages, and applications. ACM Press, pp 306–307
21. Mayo KA, Wake SA, Henry SM (1990) Static and dynamic software quality metric tools. Department of Computer Science, Technical report, Virginia Tech, Blacksburg
22. Bansiya J, Davis C (1997) Using QMOOD++ for object oriented metrics. Dr. Dobb's J
23. Arisholm E, Briand LC, Foyen A (2004) Dynamic coupling measurement for object-oriented software. IEEE Trans Soft Eng 30(8):491–506. <https://doi.org/10.1109/TSE.2004.41>

Hierarchical Block Chain-Based Authentication Management Scheme for IoT Devices



M. Revanesh and V. Sridhar

1 Introduction

Identity sensor-based Internet of things (IoT) represents a set of devices that are capable of detecting, gathering and processing information related to a certain phenomenon. These sets of integrated devices play a major role in empowering the thoughts of smart homes, precision agriculture and reliable healthcare systems. The advancement in the technology accompanied with vast potential of these devices has profoundly impacted changes in the human life in many more ways than imagined. However, integration of such devices on large scale has also presented the shortcomings of the technology by exposing some of the major system-building challenges faced by these types of networks, major of them being security issues in scalable environment, thereby hampering the widespread adoption of the technology.

In order to ensure widespread adaptation of the technology, it is necessary to address the security issues pertaining to reliability and security of data communicated through these devices along with service availability and authentication, preferably in a decentralized way because centralized approaches largely rely on trusted third-party certificates, which pushes the entire network operation to the edge of single point failure in case if centralized authorities are compromised [1, 2].

Block chain, a decentralized distributed ledger technology, concurs with the distributed attributes of IoT, which guarantees an environment to conduct trusted transactions without a third party, in which every task and request is recorded on

M. Revanesh (✉)

Department of Electronics and Communication Engineering, PES College of Engineering, Mandya, Karnataka, India

V. Sridhar

Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India

the chain of blocks along with a digital signature of the owner node for public verification. The ledger is generated and maintained in a decentralized fashion by all participants in the system, which proves to be a promising platform if adopted in the right manner to provide a solution for problems of the IoT. However, as the integration of block chain for IoT security is still in its trial stage, there are as yet numerous issues in the current block chain-based strategies that are expected to be addressed before its widespread adoption in IoT network [3, 4].

Most of the current solutions designed for IoT security using block chain can be classified into two main streams [5]: (1) analyzing the architecture of IoT and signifying how they better fit with the characteristic structure of block chain and (2) using block chain for solving the authentication issue of IoT devices in large-scale network. But, these security solutions still do not guarantee resilience against single point network failure problem and fail to assure service availability and authentication collectively. So in order to overcome these problems, a block chain-based authentication management scheme is designed for IoT devices with the following objectives:

- An IoT network is designed by considering the basic structure of wireless sensor network where every node in IoT is classified as base stations, cluster heads and data collection nodes based on the capabilities of nodes.
- Black widow optimization-based clustering technique is adopted in the proposed network along with hierarchical block chain model to ensure authentication in a scalable IoT network.

The rest of the paper is structured as follows: Sect. 2 reviews some of the existing related works in the field of IoT security, Sect. 3 describes the new proposed methodology that can be adopted in IoT devices, Sect. 4 compares the results of proposed methodology with few existing standard ones in the field, and finally, Sect. 5 summates the entire work.

2 Related Work

Ouaddah et al. [6] designed a block chain-based control framework for accessing IoT devices called “fair access” which is a decentralized pseudonymous authorization framework which ensures privacy and provides access to users to have control over the data and eases the task of integrating smart objects into the current Internet. The framework balances the advantages offered by block chain technology to provide a stronger and transparent control on the network. But, this solution when adapted to real-time IoT devices requires large memory and power which makes it not suitable for battery-operated IoT devices.

A block chain authentication and trust module (BATM) [4] is proposed for IoT devices to address authentication and trust-level management issues by addressing

two important constraints of sensor network such as adaptability (evolution and mutability) and the node restrained resource (computational complexity, energy consumption and memory usage). The BATM with its decisive human-like knowledge-based trust model demonstrates how to use immutable services of block chain to provide solutions to large set of problems in the field of distributed WSN. But, the generation and storage of distributed block chain on a resource-constrained WSN environment in a global scenario could be a challenging task.

A block chain-based decentralized Docker trust (DDT) was designed by Xu et al. [7] which greatly reduces the risk of DoS and provides signature verification services for Docker images. Saad et al. [8] have presented a POSTER: Detering DDoS attacks on block chain-based crypto currencies through mem pool optimization for the use of distributed denial of the service (DDoS) causes the massive transaction of the crypto currencies memory, the size and the transaction of the optimization as mem pool optimization that includes designs as fee based and age based. The problem of this paper is low accuracy because of the DDoS attacks.

Haseeb et al. [9] have designed a malicious activity prevention framework for secure routing in WSN-based mobile IoT, which provide optimum secured data transmission solution for IoT devices in a cost-effective manner. The framework improvises the data security and delivery ratio between IoT devices under the presence of malicious entities by adopting lightweight cryptosystems. But the problem with the proposed paper is its limitations in addressing multi-hop communications which is an essential requirement in sensor-based IoT devices.

Ren et al. [10] have presented incentive mechanism of data storage based on block chain for WSN used to construct the block chain, and storing the node with it rewards for controlling the data access. The new data blocks reduce the power computation to the work proof mechanism. The node which is closest to the existing data block helps in storage of data by collecting and hashing only different sub blocks, thus helps in reducing a lot of hashing power. The problem of this paper is deploying the perceived object and collects the information of coverage area in cooperative manner.

Gorodnichev [11] has presented a next-generation industrial block chain-based WSN used as a network of the protocol in the IoT for the sensor nodes that allows the network data to transferring and storing. The problem of this paper is solving the computing and firmware of the devices in the configuration.

Liu et al. [11] have presented an optimization technique for industrial Internet of things (IIoT) systems, which was designed to address the effectiveness issue of the IIoT information, which was done by integrating IoT with block chain for empower information putting away/handling/partaking in a protected furthermore, effective way. This work also describes a novel profound fortification learning-based execution streamlining structure for block chain-empowered IIoT frameworks to achieve the following objectives: (1) giving a procedure for assessing the framework from the parts of versatility, decentralization, dormancy and security; (2) improving the versatility of the fundamental block chain without influencing the framework's decentralization, inertness and security; (3) planning a modulable block chain for

IIoT frameworks, where the square makers, accord calculation, square size and square stretch can be chosen/balanced utilizing the DRL procedure.

3 Proposed Method

The IoT network is framed by combining large sets of sensor integrated nodes which work with a common principle of collecting data by working in a cooperative manner. In our proposed model, the nodes in the network are classified as base station, cluster heads and sensing nodes based on their individual capability. Base station in each network operates like a node manager by collecting and processing the data from the sub-networks (either through cluster head or end users), and cluster head node acts like an intermediate base station which helps in easy processing and collection of data from the end nodes (sensing nodes). The end nodes represent the set of IoT devices such as sensor integrates processors, cameras, etc., which helps in collecting the data. The clustering is accomplished by adopting black widow optimization algorithm, and an authentication scheme based on block chain is deployed based on hierarchy of data transfer.

General public block chain network in which each node joins an unauthenticated network and tries to build a decentralized trust network through consensus algorithm requires repetitive authentication which might become a bottle neck in limited resource IoT devices, whereas private block chain helps in establishing trusted network easily within a set of peer in a closed network but does not ensure scalability requirement of IoT devices; hence in order to address the issue of both scalability and authentication, a hybrid block chain [5] model which uses a combination of public and local block chain is used in the proposed network. Figure 1 describes the Block

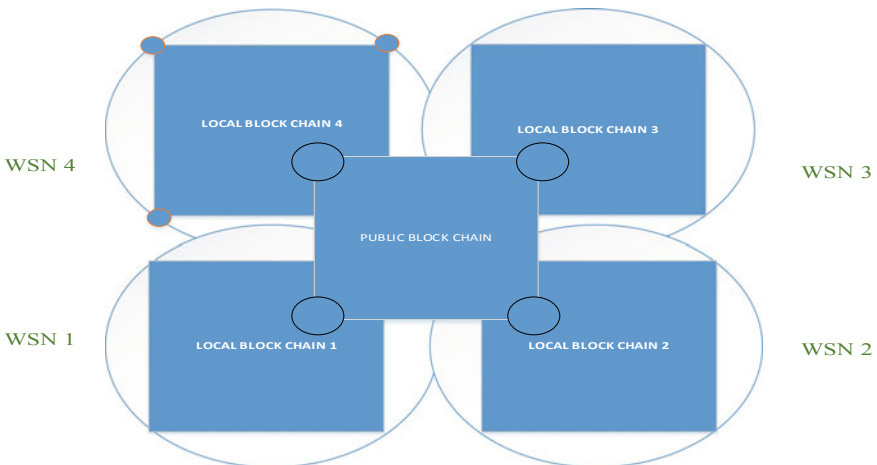


Fig. 1 Basic diagram of block chain

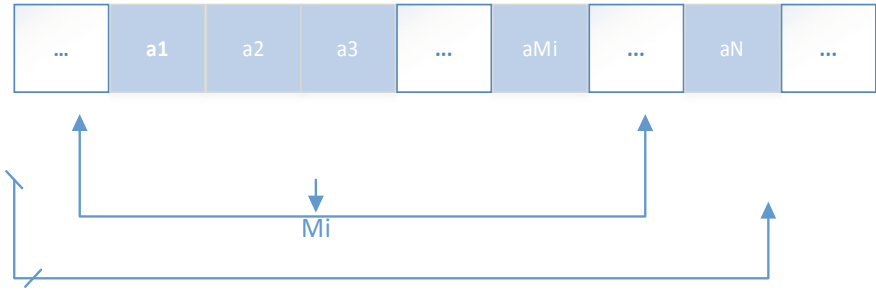


Fig. 2 Block selection process

diagram of adopted hybrid block chain.

The peer in the network are joined with the Internet of things in the devices that are blocked in the generation that are treated as the operation of the $a_1, a_2, a_3 \dots a_{M1} \dots a_N$. Figure 2 shows the blocks with the clouds in the constant process to save the local storage.

Black widow optimization algorithm

Black widow optimization (BWO) is an optimization algorithm inspired by unique mating behavior of black widow spiders. One of the unique stages involved in the BWO called as cannibalism helps to omit species with inappropriate fitness from the circle of competition, which also leads to an early convergence [12]. A brief overview of BWO is described in Fig. 3.

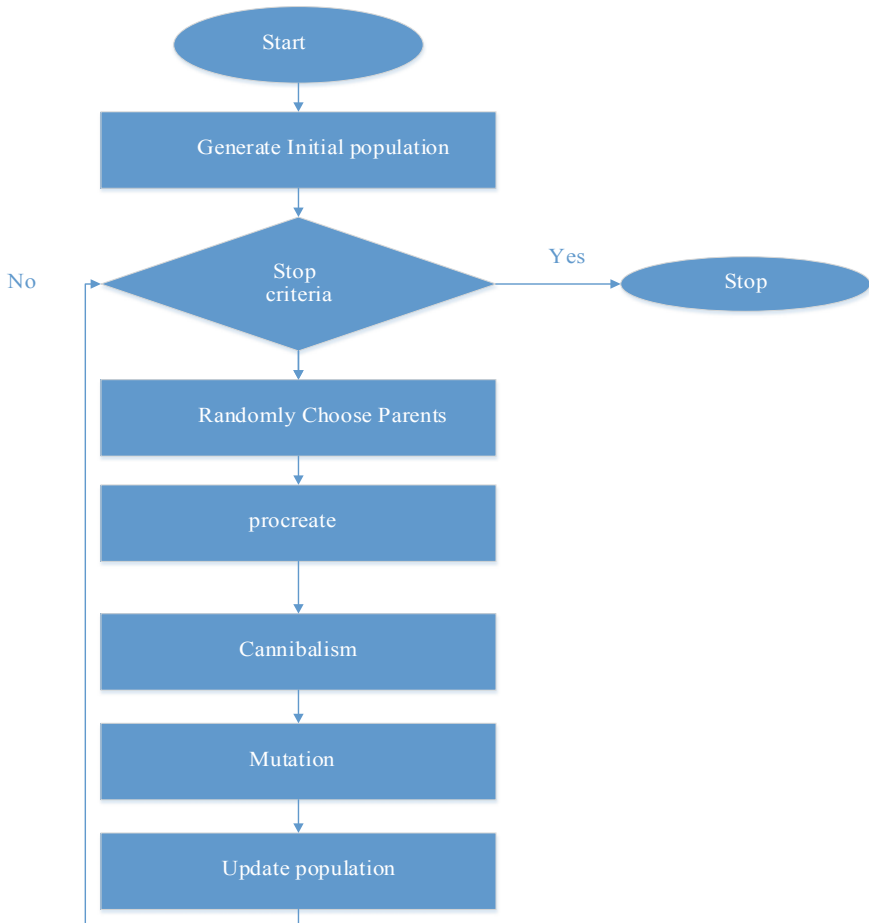


Fig. 3 Basic block of black widow optimization

<i>Pseudo Code of Black Widow Optimization algorithm</i>
Input: Maximum number of iteration, Rate of procreating, rate of cannibalism, rate of mutation
Output: near-optimal solution for the objective function
<pre> // initialization 1. The initial population of black widow spiders Each pop is a D-dimensional array of chromosomes for a D-dimensional problem // Loop until the terminal condition 1. Based on procreating rate , calculation the number of reproduction " nr" ; 2. Select the best nr solutions in pop and save them in pop1 ; // Procreating and cannibalism 3. For i=1 to nr do 4. Randomly select two solutions as parents from pop1 ; 5. Generate D children using equation1 ; 6. Destroy father ; 7. Based on the cannibalism rate, destroy some of the children (new achieved solutions) ; 8. Save the remain solutions into pop2 ; 9. End for // Mutation 10. Based on the mutation rate, calculate the number of mutation children "nm" ; 11. For i=1 to nm do 12. Select a solution from pop1 ; 13. Mutate randomly one chromosome of the solution and generate a new solution ; 14. Save the new one into pop3 ; 15. End for // Updating 16. Update pop = pop2+pop3 ; 17. Returning the best solution ; 18. Return the best solution from pop ; </pre>

4 Results and Discussion

The proposed scheme is simulated and evaluated by using the NS 3 software with the following parameters listed in Table 1.

Figure 4 shows the message size handled. The limitations for the experimental results were clearly expressed in the exponential decay.

The figure shows that the message sizes with different nodes for the message of the authentication message of the analysis of reply attack the signature in the analyzing message of cluster heads which is needed.

Figure 5 demonstrates the performance of the throughput for the previous models [13, 14] and the proposed model, which clearly shows the proposed model outperforms other models by more than 35%.

Table 1 List of parameters

Parameters		Specifications
Network settings	Sensor nodes	100
	Block chain gateway	1
	Simulation area	1000 × 1000 m ²
	Transmission range in meter	100
	Sensor node initial energy	1000 J
	Block chain gateway location	(X = 500, Y = 500)
	Data transmission energy	2 × 10 ⁻⁸ J/bit
	Idle status energy	10 E ₀ J/bit
	Convergence energy	5 E ₀ J/bit
	Interpacket interval	10 ms
	Threshold of the distances between nodes	20 d ₀ /m
	Control packet	32-bit
	Data packet speed	1 bit/s
	Packet size	1024 bytes
	Medium access control protocol	IEEE 802.11
	Packet flow speed	4 packets/second
	Data rate	88 Mbps
	Data flow type	UDP/CBR
	Number of simulation rounds	2000
	Number of retransmission	Min.-5 Max.-7
Queue size	80 kb	
Block chain	Block header size	80 bytes
	Transaction counter	1-9 bytes
	Number of transactions	Variable
	Hashing technique	SHA-256
	RRoof type	PoC
Transaction fees	No	
Simulation time		100 s

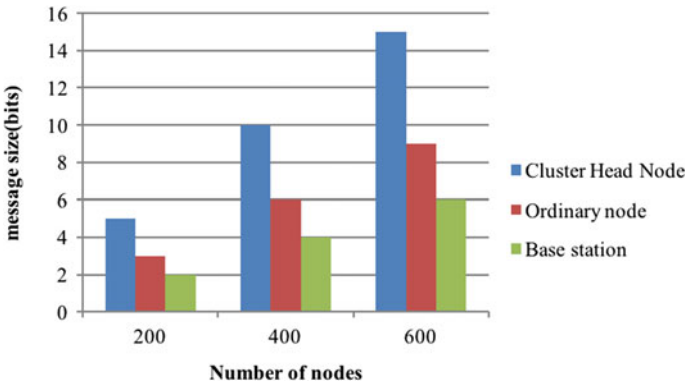


Fig. 4 Comparison of number of bits of messages by different nodes

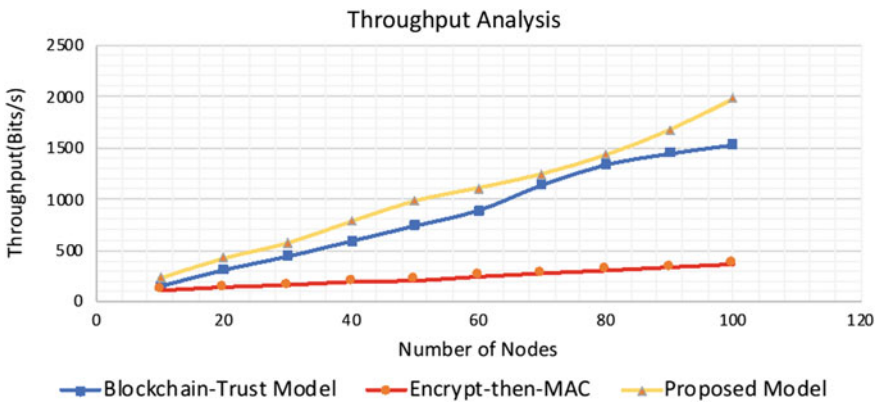


Fig. 5 Comparison of throughput

Figure 6 shows the comparison of packet loss rate as a function of number of nodes, and it clearly shows a considerable reduction in packet loss as compared to other standard models.

5 Conclusion

This paper proposes a multi-WSN-based model for IoT devices which adopts black widow optimization for clustering and a hybrid hierarchical block chain model for secure transfer of data from end devices to base station (server). The comparison of throughput and packet loss rate clearly depicts the improvement in performance of the network when tested for a set of 100–300 nodes, and adaptation of hybrid block chain ensures optimum security required by the network by optimally using

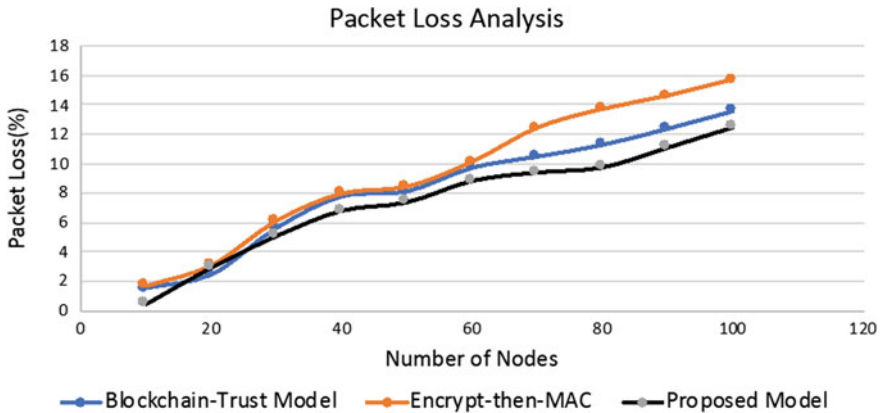


Fig. 6 Comparison of packet loss versus number of nodes

the scarcely available resources in the devices. In future work, the proposed network will be evaluated for other parameters such as improving the scalability of the network without compromising security and performance.

References

1. Jain SK, Garg K (2009) A hybrid model of defense techniques against base station Jamming attack in wireless sensor networks. In: First international conference on computational intelligence, communication systems and networks, Indore, pp 102–107. <https://doi.org/10.1109/CICSYN.2009.57>
2. Chen J, Zhang H, Du X, Fang B, Liu Y, Yu H (2012) Base station location protection in wireless sensor networks: attacks and defense. In: IEEE International Conference on Communications (ICC), Ottawa, ON, pp 554–559. <https://doi.org/10.1109/ICC.2012.6364186>
3. Christidis K, Devetsikiotis M Blockchains and smart contracts for the Internet of Things. *IEEE Access* 4:1–1. <https://doi.org/10.1109/ACCESS.2016.2566339>
4. Moinet A, Darties B, Baril J (2017) Blockchain based trust and authentication for decentralized sensor networks. CoRR abs/1706.01730
5. Cui Z et al (2020) A hybrid blockchain-based identity authentication scheme for multi-WSN. *IEEE Trans Serv Comput* 13(2):241–251
6. Ouaddah A, Abou Elkalam A, Ait Ouahman A (2016) FairAccess: a new blockchain-based access control framework for the Internet of Things. *Secur Commun Netw* 9(18):5943–5964. <https://doi.org/10.1002/sec.1748>
7. Xu M et al (2020) On cloud storage optimization of blockchain with a clustering-based genetic algorithm. *IEEE Internet Things J*
8. Saad M, Thai MT, Mohaisen A (2018) POSTER: deterring ddos attacks on blockchain-based cryptocurrencies through mempool optimization. In: Proceedings of the 2018 on Asia conference on computer and communications security
9. Haseeb K et al (2019) Intrusion prevention framework for secure routing in WSN-based mobile Internet of Things. *IEEE Access* 7:185496–185505
10. Ren Y et al (2018) Incentive mechanism of data storage based on blockchain for wireless sensor networks. *Mobile Inf Syst* 2018

11. Liu M et al (2019) Performance optimization for blockchain-enabled industrial Internet of Things (IIoT) systems: a deep reinforcement learning approach. *IEEE Trans Industr Inf* 15(6):3559–3570
12. Hayyolalam V, Pourhaji Kazem AA (2020) Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems. *Eng Appl Artif Intell* 87:103249. <https://doi.org/10.1016/j.engappai.2019.103249>
13. She W, Liu Q, Tian Z, Chen J, Wang B, Liu W Blockchain trust model for malicious node detection in wireless sensor networks. *IEEE Access* 7. <https://doi.org/10.1109/access.2019.2902811>
14. Lara-Nino CA, Diaz-Perez A, Morales-Sandoval M (2018) Energy and area costs of lightweight cryptographic algorithms for authenticated encryption in WSN. *Secur Commun Netw* 2018:1–14. <https://doi.org/10.1155/2018/5087065>

Segmentation of Skin Lesion Using Adaptive Contours Color Method



Ramya Srikanteswara  and A. C. Ramachandra 

1 Introduction

Melanoma is considered as the fastest growing cancer all around the globe. Several statistical reports suggest that skin cancer has highest mortality rate midst all cancer types. Reports suggest that there is a significant enhancement in the cases of skin cancer patients than compared to any other cancer type [1]. According to US statistical reports, every year 5 million cases are identified with skin cancer in United States of America (USA) [2]. The reason of skin cancer is unrestrained abnormal progress of cells. The most common, lethal and destructive form of skin cancer is Melanoma which has highest mortality rate all across the world. For instance, 9000 people lose their life from Melanoma, from 91,000 cases are diagnosed every year in USA [3]. Similarly, Melanoma causes 2000 deaths from 14,000 cases and 22,000 deaths from 100,000 cases diagnosed every year in Australia [4] and Europe [5]. Moreover, most concerning factor is that the growth in number of melanoma patients is very rapid in recent time. For instance, melanoma cases have risen around 225% in last 30 years [1990–2019] in USA [3, 6]. The reports suggest that the survival rate in melanoma cases is reduced from 99% to approximately 14% [7] when diagnosed in later stages. This paper discusses the problem and its solution in the following sections which are as follows. In Sect. 2, background of melanoma and various existing methods are discussed. In Sect. 3, the proposed model for segmentation technique is discussed. In Sect. 4, experimental results are presented and Sect. 5 gives a comparative analysis, and in Sect. 6 conclusion is discussed.

R. Srikanteswara (✉) · A. C. Ramachandra (✉)
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: ramya.srikanteswara@nmit.ac.in

A. C. Ramachandra
e-mail: ramachandra.ac@nmit.ac.in

2 Related Background

The human skin contains three types of tissues such as hypodermis, epidermis and dermis. Epidermis consists of melanocytes which are type of a skin surface cells. The abnormal growth of this melanocytes is the reason for deadliest Melanoma cancer disease [8]. The affected region by Melanoma remains same as skin color in initial stages. However, skin color becomes pink, red, purple, blue and ultimately becomes black or dark brown due to melanocytes in later stages [9]. Melanoma is most concerning disease in terms of diagnosis due to its metastasis characteristics which have potential to spread.

However, Melanoma can be cured completely, if detected in early stages of disease. According to several research reports, the survival rate of Melanoma patient becomes 96% in case of detection of Melanoma disease in primary stage [10]. Due to this reason, detection of Melanoma disease in primary stage becomes a crucial priority. However, identification of melanoma in early stages is a challenging process even for skin experts and specialists due to high visual resemblances between malign and benign skin lesions. Detection of Melanoma from naked eye is very complex and challenging procedure. Therefore, various imaging techniques are presented by several researchers. However, Dermoscope imaging technique is one of the most proficient and competent technique for the identification of Melanoma in initial stages. The accuracy of Dermoscopic high resolution and non-invasive imaging technique is quite high which utilizes light magnifying equipment and immersion fluid for enhancing visualization of skin surface [11]. Various Reports states that diagnosis results are improved by 50% for malignant cases in Melanoma using Dermoscopic imaging technique [12]. Skin experts can deeply study the structures of skin surface using Dermoscopy and reduction in surface reflectance provides additional advantage in Melanoma diagnosis. However, understanding of Dermoscope imaging technique is a challenging process even for medical experts and requires a proficient training to study them. Therefore, a proper and effective analysis of Dermoscopic imaging technique is very essential for the proper diagnosis of Melanoma disease.

Study of Dermoscope imaging technique consists of three key aspects such as Dermoscope image segmentation, feature extraction of Dermoscope images and classification of these images. All three stages are very crucial in analysis of Dermoscopy Segmentation is a primary phase for the assessment of Dermoscopic images which is utilized for detecting lesion border in Dermoscopic images. This technique helps to confine grazes effectively. Although segmentation process requires moderate knowledge to perform lesion detection operation but variety of lesion shapes, structures, color and sizes makes this process quite complicated. Moreover, intrinsic factors such as hairs, blood vessels make segmentation process very complex [13]. Therefore, development of a proficient and robust segmentation technique becomes a challenging task. Several researchers have shown interest in developing a proficient segmentation technique for Melanoma diagnosis.

A comprehensive analysis [14] of Dermoscope Imaging Technique is presented in which segmentation, feature extraction and classification of Dermoscope images

is explained in brief. Melanoma classification technique is presented in one of the techniques for the proficient diagnosis of digital images. Here, Gaussian filter [15] is utilized for eliminating noise. DERMIS dataset is utilized for performance evaluation. Skin lesion segmentation technique is introduced in one of the method [16] based on adversarial training to extract discriminative features. Here, artifacts are reduced using DenseNet sampling methods. Performance is evaluated on ISIC dataset. In a process described [17], an automated Melanoma detection technique is introduced using Deep-learning methods and this method is used for effective classification of lesion pixels. ISIC dataset is utilized for performance comparison. A statistical analysis is performed on skin lesion detection, their challenges and solutions in one of the technique [18]. Various techniques and their drawbacks are discussed. However, there are several challenges arising while segmenting Dermoscopic images such as skin and non-skin pixel overlap in presence of complex background, color diversity in Melanoma affected region, existence of intrinsic factors in affected area, high overhead, ineffective feature extraction, etc. Therefore, a research on a proficient lesion segmentation technique is very essential to detect Melanoma in early stages.

3 The Proposed Method

3.1 Introduction to ACM

A proficient segmentation method constructed using Adaptive Contour Model (ACM) for the identification of Melanoma disease at primary stages is required. Besides, Gaussian samples are adopted to manage heterogeneous entities. The mean and variance values of Gaussian samples can differ. The contour features of Dermoscopic image lesion can be exploited with the help of Gaussian structures. The contour features are extracted to form a smoother boundary of a lesion using Adaptive Contour Model. The color pixels in demographic images may defer due to varied contour feature extraction for different classes. Pre-processing methods are utilized along with ADM for the elimination of noise present in Dermoscopic images. The proposed Adaptive Contour Model (ACM)-based lesion segmentation technique is evaluated upon PH2 dataset. The performance of proposed Adaptive Contour Model is highly effective and can be applied in diagnosis of Melanoma.

Numerous skin cancer cases have been detected in recent years all across the world because of universally varying temperature. Melanoma is a harmful skin disease whose mortality rate is 1.62% [19]. Australia and USA are the worst hit countries by Melanoma. A survey carried out by World Health Organization (WHO) states that every year 13 million people comes in contact with deadly Melanoma disease [20]. Identification of Melanoma disease at primary phase is extremely essential for the reduction of high mortality rate and diagnosis cost. Therefore, Adaptive Contour

Model is introduced for the lesion segmentation of Dermoscopic images in preliminary phase. This section discusses about mathematical modeling of proposed Adaptive Contour Model (ACM) for the segmentation process of Dermoscopic images and methods adopted to extract contour features. Moreover, Pre-processing methods are utilized for the elimination of noise present in Dermoscopic images. The following section discusses about mathematical representation of Adaptive Contour Model (ACM) for effective segmentation process.

3.2 Mathematical Model

This section discusses mathematical representation of proposed ACM model for image segmentation and contour feature extraction. The effective segmentation of Dermoscopic images helps to identify Melanoma in primary phase. In proposed Adaptive Contour Model after removal of noise with the help of pre-processing block, input image is masked with the ground truth for efficient segmentation of proposed ACM. The steps in ACM model can be represented diagrammatically in the form of a flow chart as given in Fig. 1.

Consider representation of an image area, denoted as δ and dermoscope input image can be expressed as $K(m) : \delta \rightarrow X$ and noise exists in the dermoscope images can be represented as $j(m) : \delta \rightarrow X$, unspecified bias region in the image can be indicated as $Z(m) : \delta \rightarrow X$ and finally, reconstructed actual signal can be represented as $L(m) : \delta \rightarrow X$ then color variation in dermoscopic image pixels can be represented as,

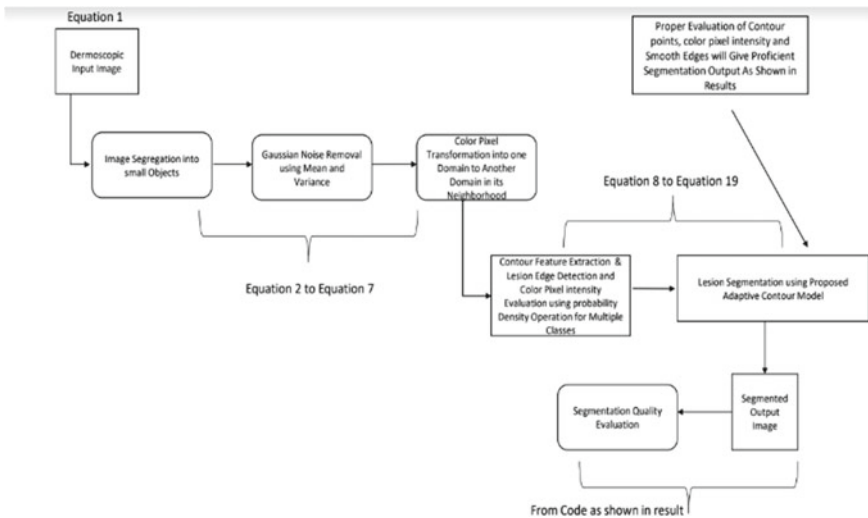


Fig. 1 Flowchart representing segmentation process through ACM model

$$K(m) = Z(m).L(m) + j(m) \quad (1)$$

Let R entities be present in dermoscopic image domain δ then image domain of a th entity can be represented as δ_a . Suppose that the reconstructed actual signal $L(m)$ is represented as a piecewise constant for every entity domain, i.e., $L(m) = b_a$ for $m \in \delta_a$. Here, b_a can be expressed as a constant. The bias region $Z(m)$ is always considered to remain smooth in an image domain δ .

The noise j exist in the Dermoscopic images is always considered as Gaussian dispersed sample error in which variance and mean can be represented as v_a^2 and ω_a , respectively, where mean is considered as zero. Therefore, variance v_a^2 and mean ω_a are the two parameters which can evaluate color pixels of dermoscopic image. Thus, Gaussian dispersed samples can define contour characteristics of image pixels.

However, effective and proficient lesion segmentation of an image cannot be extracted using single Gaussian dispersed sample. Therefore, many Gaussian dispersed samples are required to extract contour characteristics of diverse color pixels for every entity domain. The Gaussian dispersed samples for their respective entity domain δ_a can be expressed as,

$$d(K(n)|\Theta_a) = \frac{1}{((2\pi)^{1/2}).v_a)} e^{-(K(n)-\omega_a(m))^2.(2v_a^2)^{-1}} \quad (2)$$

Here, v_a can be represented as standard deviation and diverse contour mean can be denoted as $\omega_a(m)$. The bias region $Z(m)$ can be considered as zero due its slow variation. Thus, it can be considered that,

$$\omega_a(m) \cong Z(m).b_a \quad (3)$$

where Θ_a representation can be denoted as $\Theta_a = \{b_a, v_a, Z\}$. For every spot m in the image domain δ , their neighboring central area can be represented as

$$P_m = \{n || n - m || \leq d\}. \quad (4)$$

where Eq. (4) represents small neighborhood. Maximum neighborhood considered is equal to the radius of the entity. Here, d can be expressed as the radius of neighboring central area P_m . R , non-overlapping entities are present in the image domain δ_a of a th entities then the complete set of image domain δ can be represented by the following Eq. (5),

$$\delta = \cup_{a=1,\dots,R} \delta_a \quad (5)$$

where the intersection of δ_a and δ_c is represented by ψ where a is not equal to c . Here, a and c are two entities.

The actual color domain $Y(H)$ of an image can be transmitted into any other domain $X(H)$ with the help of mapping as,

$$H : K(m|\Theta_a) \rightarrow \mathbb{R}(m|\Theta_a) \quad (6)$$

Here, can be $\mathbb{R}(m|\Theta_a)$ further factorized as,

$$\mathbb{R}(m|\Theta_a) = 1/i_a(m) \sum_{n \in \delta_a \cap P_m} K(n|\Theta_a) \quad (7)$$

Let, dispersion of color pixels in an image can be denoted as m and dispersion of pixel color is an independent process and $i_a(m) = \delta_a \cap P_m$. Then, its relative probability density operation remains Gaussian for all $\mathbb{R}(m|\Theta_a) \in X(H)$ and can be represented as,

$$\mathbb{R}(m|\Theta_a) = R\left(\omega_a, \frac{v_a^2}{i_a(m)}\right) \quad (8)$$

Let, product of all probability density operation still remains Gaussian for all $R(n|\Theta_a)$,

$$K(n|\Theta_a) \cong K(m|\Theta_a) \quad (9)$$

Here, the intersection of δ_a and P_m is belongs to all n , i.e., $\delta_a \cap P_m \in \forall n$. Then, the Gaussian color pixel even function can be represented as

$$\prod_{n \in \delta_a \cap P_m} d(K(n|\Theta_a)) = d(K(m|\Theta_a))^{i_a(m)} \quad (10)$$

where Eq. 8 is directly proportional to $R(\omega_a, v_a^2 \cdot (i_a(m))^{-1})$. Thus,

$$d(\mathbb{R}(m|\Theta_a)) = \prod_{n \in \delta_a \cap P_m} d(K(n|\Theta_a)) \quad (11)$$

Suppose that,

$$Y = \{\mathbb{R}(m|\delta_a), m \in \delta, a = 1, 2, 3 \dots, R\} \quad (12)$$

where Y is a variable representing a th entity. Then, probability density operation of the a th entity representation can be described as,

$$d(Y|\Theta_a) = \prod_{a \in \delta} d(\mathbb{R}(m|\Theta_a)) \quad (13)$$

Then, the reconstructed combined probability density operation can be performed using ACM for segmentation of images as,

$$d(Y|\phi) = \prod_{a=1}^R d(Y|\Theta_a) \quad (14)$$

Which is further decomposed in Eqs. (15) and (16) and can be represented as follows,

$$d(Y|\phi) = \prod_{a=1}^R \prod_{m \in \delta} d(\mathbb{R}(m|\Theta_a)) \quad (15)$$

$$d(Y|\phi) = \prod_{m \in \delta} t(\mathbb{R}(m|\phi)) \quad (16)$$

where $\phi = \{\Theta_a, a = 1, 2, \dots, R\}$. and $\prod_{a=1}^R d = t$ then,

$$t(\mathbb{R}(m|\phi)) = \prod_{a=1}^R d(\mathbb{R}(m|\Theta_a)) \quad (17)$$

Which is further factorized as follows,

$$t(\mathbb{R}(m|\phi)) = \prod_{a=1}^R \prod_{n \in \delta_a \cap P_m} d(K(n|\Theta_a)) \quad (18)$$

Form Eqs. (10) and (18), the covariate of Gaussian dispersed samples can be extracted as follows,

$$t(\mathbb{R}(m|\phi)) = \prod_{a=1}^R d(\mathbb{R}(m|\Theta_a)) \propto R(\omega, \rho) \quad (19)$$

where $t(\mathbb{R}(m|\phi))$ is directly proportional to $R(\omega, \rho)$ and ω can be represented as $\rho \sum_{a=1}^R i_a(m) \cdot \frac{\omega_a}{(v_a^2)}$ and ρ can be expressed as $\frac{1}{\sum_{a=1}^R i_a(m) \cdot (v_a^2)^{-1}}$.

where the reconstructed combined probability density operation represented by Eq. (16) provides the composition of image pixels with multiple contour class intensities. The actual color pixels are transmitted to another domain which provides the information about same class adjacent pixels of a dermoscopic image from Eq. (7). Therefore, its classification provides noise free output result and an even border can be achieved while segmenting the Dermoscopic images.

4 Results

The analysis of performance for the lesion segmentation of Dermoscopic images using Adaptive Contour Model (ACM) is discussed in this section, to detect Melanoma skin cancer disease in primary stage. Melanoma detection at primary phase is a very challenging process due to the presence of intrinsic factors such as hair, blood and due to high visual resemblances between malign and benign skin lesions. Numerous varieties of lesion shapes, structures, color and sizes make lesion segmentation process even more challenging. Detection of Melanoma is challenging for Experts and medical specialists as well from naked eye. However, primary stage Melanoma detection can be found efficiently. Thus, a proficient segmentation technique based on Adaptive Contour Model (ACM) for the identification of Melanoma disease at primary stages is presented here.

The proposed technique segregates vital entities from the background of an dermoscopic image and reconstructs actual color information from its neighboring pixels of same class so that an even border can be reconstructed. The contour features are extracted to form a smoother boundary of a lesion using Adaptive Contour Model. The performance of proposed Adaptive Counter Model is evaluated based on various factors such as efficient counter feature extraction, noise removal and quality of segmented image. The performance of proposed Adaptive Counter Model is evaluated using PH 2 data set [21]. The PH2 dataset contains 200 dermoscopic diagrams for testing. The dermoscopic image resolution of this dataset is of size 768×560 . Out of 200 Dermoscopic images present in these PH2 dataset 40 images are of Melanoma type, 80 atypical nevus type and 80 only benign nevi type. The proposed Adaptive Contour Model-based lesion segmentation technique is simulated over windows system. Segmented images and Features are extracted using MATLAB. The comparative analysis of proposed Adaptive Contour model with various state-of-art-technique is presented in the following paragraph.

5 Comparative Analysis

The performance comparison of projected Adaptive Contour Model with various state-of-art segmentation techniques based on the factors like efficient counter feature extraction, noise removal and quality of segmented image is presented here. Pre-processing methods are utilized for the removal of noise from the Dermoscopic images as well as enhances the efficiency of Adaptive Contour Model. The contour features of image lesion can be exploited with the help of Gaussian structures. The contour features are extracted to form a smoother boundary of a lesion.

The proposed Adaptive Contour Model is compared with mFCN – PI [22], Peng et al. [23], DCLPSI [24], FrCN [25], iMSCGnet [26] in terms of Accuracy (AC), Jaccard Index (JA) and Dice Coefficient (DI). This index is utilized for evaluating

Table 1 Average performance evaluation metrics (%)

Algorithm	JA	DI	AC
mFCN-PI	83.99	90.66	94.24
PENG	85.00	90.00	93.00
DCLPSI	85.90	92.10	95.30
Frcn	84.79	91.77	95.08
iMSCGnet	88.21	93.36	95.71
ACM	88.95	93.97	96.77

efficiency of lesion segmentation process. The simulation results of segmentation are very precise and proficient. Here, Table 1 shows the evaluation of proposed.

Adaptive Contour Model with various state of the art segmentation technique in terms of AC, JA and DI. Here, Accuracy (AC) for the segmentation process is achieved quite high as 96.77 compared to other existing techniques. Likewise, Jaccard Index (JA) and Dice Coefficient (DI) factors are obtained using proposed Adaptive Contour Model as 88.95 and 93.97, respectively.

The qualitative analysis of dermoscopic image present in PH2 dataset using Adaptive Contour Model is demonstrated in Fig. 2 and is very proficient and effective. Here, Fig. 2a reveals the input dermoscopy images, Fig. 2b demonstrates the ground truth and Fig. 2c demonstrates the projected even border segmented images.

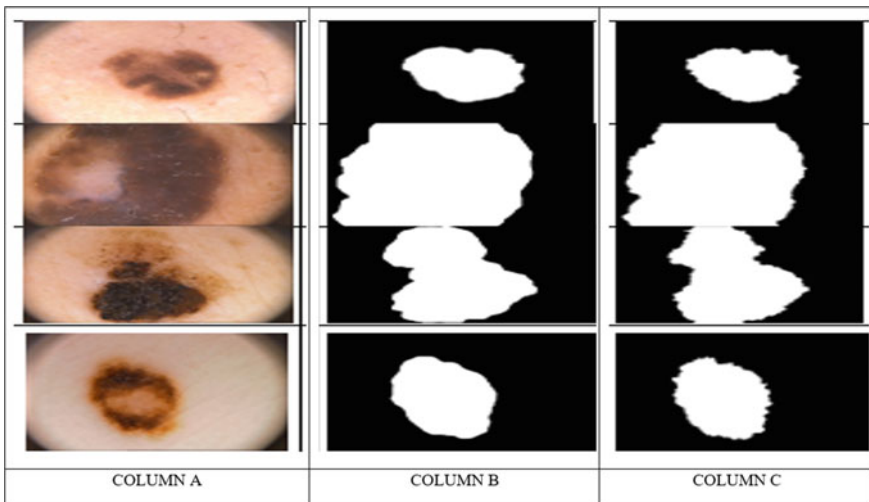


Fig. 2 Segmentation of PH2 dermoscope images column a original images b ground truth c segmentation using the proposed Adaptive Contour Model (ACM)

6 Conclusion

The segmentation of Dermoscopic images to detect Melanoma in primary stage is quite challenging and complex process due to presence of hairs, blood and variety of lesion shapes, structures, color and sizes. Therefore, in this paper, a proficient lesion segmentation technique based on Adaptive Contour Model (ACM) for the identification of Melanoma disease at primary stages is introduced. The contour features are extracted to form a smoother boundary of a lesion using Adaptive Contour Model. The proposed technique segregates vital entities from the background of a dermoscopic image and reconstructs actual color information from its neighboring pixels of same class. Pre-processing methods are utilized along with ADM for the elimination of noise present in Dermoscopic images. A comprehensive mathematical representation of proposed lesion segmentation technique is presented to evaluate contour feature based on probability density operation. The performance of proposed Adaptive Counter Model is evaluated using PH2 dataset. The evaluation of projected Adaptive Contour Model with various state of the art segmentation technique in terms of Accuracy (AC), Jaccard Index (JA) and Dice Coefficient (DI). Here, Accuracy (AC) of lesion segmentation process is quite high than any other state-of-art-techniques which is 96.77. Similarly, factors like Jaccard Index (JA) and Dice Coefficient (DI) are evaluated which are 88.95 and 93.97, respectively. In the future work, efficient feature extraction and classification of lesion segmentation process will be discussed based on proposed Adaptive Contour Model (ACM).

References

1. Kong B, Sun S, Wang X, Song Q, Zhang S (2018) Invasive cancer detection utilizing compressed convolutional neural network and transfer learning. Lecture notes in computer science. Springer, Cham
2. American Cancer Society (2018) Cancer facts and figures 2018. [Online]. Available: <https://www.cancer.org/research/cancer-factsstatistics/all-cancer-facts-figures/cancer-facts-figures-2018.html>
3. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics. CA: Cancer J Clin 68(1):7–30
4. Australian Government (2019) Melanoma of the skin statistics. [Online]. Available: <https://melanoma.canceraustralia.gov.au/statistics>
5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J Clin 68(6):394–424
6. Silverberg E, Boring CC, Squires TS (1990) Cancer statistics. CA: Cancer J Clin 40(1):9–26
7. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM (2015) Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the U.S. population, 2012. JAMA Dermatol 151(10):1081–1086
8. Feng J, Isern NG, Burton SD, Hu JZ (2013) Studies of secondary melanoma on C57BL/6J mouse liver using 1H NMR metabolomics. Metabolites 3:1011–1035. [CrossRef] [PubMed]
9. Kasmi R, Mokrani K (2016) Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Process 10(6):448–455
10. Siegel R, Miller K, Jemal A (2018) Cancer statistics. CA Cancer J Clin 68:7–30. [CrossRef] [PubMed]

11. Pellacani G, Seidenari S (2002) Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy. *Clin Dermatol* 20:222–227. [CrossRef]
12. Ali A-RA, Deserno TM (2012) A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. In: *Medical imaging 2012: image perception, observer performance, and technology assessment*. International Society for Optics and Photonics, p 8318, Bellingham, WA, USA
13. Abbas Q, Celebi ME, Garcia IF (2011) Hair removal methods: a comparative study for dermoscopy images. *Biomed Signal Process Control* 6(4):395–404
14. Celebi ME, Codella N, Halpern A (2019) Dermoscopy image analysis: overview and future directions. *IEEE J Biomed Health Inform* 1–1 <https://doi.org/10.1109/JBHI.2019.2895803>, 2019.
15. Khan MQ et al (2019) Classification of melanoma and nevus in digital images for diagnosis of skin cancer. *IEEE Access* 7:90132–90144. <https://doi.org/10.1109/ACCESS.2019.2926837>
16. Wei Z, Song H, Chen L, Li Q, Han G (2019) Attention-based dense unet network with adversarial training for skin lesion segmentation. *IEEE Access* 7:136616–136629. <https://doi.org/10.1109/ACCESS.2019.2940794>
17. Berkay M et al (2019) Deep learning based melanoma detection from dermoscopic images. In: *Scientific meeting on electrical-electronics and biomedical engineering and computer science (EBBT)*, Istanbul, Turkey, pp 1–4. <https://doi.org/10.1109/EBBT.2019.8741934>
18. Chyad MA, Alsattar HA, Zaidan BB, Zaidan AA, Al Shafeey GA (2019) The landscape of research on skin detectors: coherent taxonomy, open challenges, motivations, recommendations and statistical analysis, future directions. *IEEE Access* 7:106536–106575. <https://doi.org/10.1109/ACCESS.2019.2924989>
19. Tarver T, American Cancer Society (2012) Cancer facts and figures 2014. *J Consum Health Internet*. 16:366–367. [CrossRef]
20. Skin cancer, Ultraviolet radiation and the INTERSUN Programme, World Health Organization (WHO)
21. Mendonça T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J (2013) PH2—a dermoscopic image database for research and benchmarking. In: *35th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, pp 5437–5440. <https://doi.org/10.1109/EMBC.2013.6610779>
22. Bi L, Kim J, Ahn E, Kumar A, Fulham M, Feng D (2017) Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Trans Biomed Eng* 64(9):2065–2074
23. Peng Y, Wang N, Wang Y, Wang M (2019) Segmentation of dermoscopy image using adversarial networks. *Multimedia Tools Appl* 78(8):10965–10981
24. Bi L, Kim J, Ahn E, Kumar A, Feng D, Fulham M (2019) Stepwise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern Recognit* 85:78–89
25. Al-masni MA, Al-antari MA, Choi M-T, Han S-M, Kim T-S (2018) Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput Methods Programs Biomed* 162:221–231
26. Tang Y et al (2020) iMSCGnet: iterative multi-scale context-guided segmentation of skin lesion in dermoscopic images. *IEEE Access* 8:39700–39712. <https://doi.org/10.1109/ACCESS.2020.2974512>

Assessing the Cognitive Levels of the Students Based on Their Learning Styles



K. S. Mohan and Y. P. Gowramma

1 Introduction

1.1 Research Problem

Cognitive thinking of a person always plays major role in solving any real-time problems. Cognitive thinking skills or levels [1] vary from person to person, and it also depends on the cognitive load that a person can feel while solving any problem. Different problems need different cognitive skills. This is same for the students who are in the learning domain. And, on the other side, everyone have different learning style [2, 3] and every learning style is associated with respective cognitive levels. This study is to assess the cognitive levels based on the learning styles.

1.2 Why This Study?

In spite of many different educational reforms in last, many decades teachers are still struggling to improve the academic performance of the students. This work is going to help the teaching fraternity to understand the learning requirement of the students by mapping the learning style with the cognitive levels of the students. This work is aiming to achieve the mapping between cognitive levels and learning style with more accuracy.

K. S. Mohan (✉)
Government Polytechnic, Joida, Karnataka, India

Y. P. Gowramma
Kalpataru Institute of Technology, Tiptur, Karnataka, India

1.3 About the Proposed Work

The objective of the work is to assess the cognitive levels of the students based on their learning styles. This work concentrates on only visual learners. The major aim is to use physiological sensing to collect the data by generating video stimuli by varying the input complexities. The cognitive abilities of the learners change according to the complexities of the stimuli. But, this process of using physiological sensors like electrocardiogram (ECG), electroencephalogram (EEG), Galvanic skin response (GSR), and eye trackers is being restricted because of COVID-19 pandemic.

Instead, we have used a different approach in our study. We have collected data using online questionnaire in the form of quiz. For this, we have selected two best quality and useful videos. Based on the content of the videos, assessment was done, and data have been collected for the assessment of cognitive levels.

2 Literature Survey

2.1 Related Work

There is always a close relationship between how the students learn and to their cognitive levels. In today's digital era, the students prefer more to learn through videos materials available in different platforms like LMS and other social networks. There are many works done to personalize the learning materials based on LMS and social networks. Some of the existed works are listed as below.

Uma et al. [4] presented a study which focuses on analyzing the thinking skills of students present in an assessment using revised Bloom's taxonomy standards. Different categories of courses have been considered for the analysis of complexity level present in an assessment and know the quality of the system. The study proposed a weight-based data mining approach to classify the Bloom's categories and the thinking levels associated with that. The thinking levels such as low, medium, and high are decided based on contribution of the cognitive dimension parameters. This study helps the organization to check the quality of assessment paper and to decide whether to accept or not. The cognitive domain parameters used in the assessment reveal whether the order of thinking as low, medium, or high. In case of low or medium, one can improve by giving weightage to higher order parameters.

Afifa Yasmeen et al. [5] the purpose of the study is to assess the level of cognitive learning with the help of question paper analysis. The data collected were descriptively analyzed and presented through tables and graph for easy and quick interpretation of the results. The results showed that out of six courses, only one course reached the highest level of cognitive learning. The study tried to measure the achievement of cognitive knowledge through levels of cognitive domain proposed by Bloom's taxonomy. The study suggests that the teachers need to pay more attention to their teaching and assessing practices and make an effort to achieve and assess cognitive

learning on all six levels. The work can be enhanced by considering the Bloom's taxonomy for psychomotor and affective domain.

According to Ananthu Kuttattu et al. [6], the learning style can be referred to as the way a student prefers to acquire, process, and retain information. The prominent learning style classification model is the VAK model. According to this model, visual, kinesthetic, and auditory are the three major kinds of learning styles. Many researches have shown that people prefer more than one way of learning; hence, categorizing a person to just one of the above types as done in traditional methods is not accurate. A method to identify our learning styles more accurately is required. Machine learning can be applied in this field to achieve our aim in the most efficient way. Once we have accurate information about learning styles, we can use it to suggest career options. This research aims to predict the learning style combinations of students and suggests field of study using algorithms like k-means, SVM, and decision tree.

2.2 Limitations Addressed

According to [4], the author's aim was to classify the cognitive skills based on the assessment using Bloom's taxonomy. But, they have not addressed the learning style. In this work, learning styles of the student play a vital role in analyzing the cognitive skills.

In [5], the author assessed the cognitive learning levels of the students by analyzing the question papers. In this study, they have also not mentioned what learning styles the students have before assessment. It is better to know the learning style before assessment so that it gives clarity on using better learning materials. Since this study maps learning style with cognitive levels, we can use this as model while preparing teaching/learning materials so that students can advantage of it.

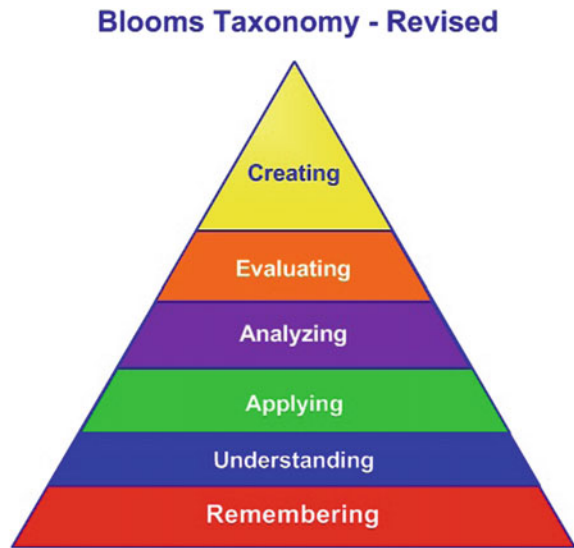
Physiological sensing was not used to assess cognitive levels based on learning styles. Bloom's taxonomy was used only for the academic assessment purpose where attainment of cognitive levels is targeted but not for mapping between learning style and cognitive levels. So, our work is going to address all these limitations.

3 Research Methodology

3.1 Bloom's Taxonomy for Cognitive Domain of Learning

The Bloom's taxonomy is to understand and assess the thinking levels (cognitive levels) of the students during their learning process. It is one among the best methods which has been adopted in education system while framing the academic curriculum. There are six different cognitive levels in Bloom's taxonomy [7]. The revised Bloom's taxonomy is shown as in Fig. 1.

Fig. 1 Revised Bloom's taxonomy



The levels from bottom to top are arranged based on the level of thinking. The bottom most level indicates the lowest order of thinking, and top level indicates the higher order thinking. Each level from bottom to top increases the order of thinking, and top to bottom decreased the order of thinking. The first two levels from bottom (remembering and understanding) specify lower order thinking, the next two (applying and analyzing) specify medium order thinking, and last two (evaluating and creating) specify higher order thinking skills. All the cognitive levels are equally important for the students in order to mold their professional carrier.

3.2 Learning Styles

The learning style is the description of the attitudes and behavior which determines an individual's preferred way of learning. There are several systems of learning styles. VAK model is the one which is more frequently used learning style inventory. As per the model, most of us prefer to learn in one of three ways, visual, auditory, or kinesthetic.

Visual: A visual—dominant learner absorbs and retains information better when it is presented in, for example, pictures, diagrams, and charts.

Auditory: An auditory—dominant learner prefers listening to what is being presented. He or she responds the best to voices, for example, in a lecture or group discussion.

Kinesthetic: A kinesthetic—dominant learner prefers a physical experience. He or she likes a “hands on” approach and responds well to being able to touch or feel an object or learning prop.

A variation of the VAK model developed by Neil D. Fleming is VARK or visual, auditory, reading/writing, and kinesthetic.

Reading/Writing: The learner belongs to this style uses repetitions of words or writing. The person who prefers this style remembers or organizes things best in his/her mind by taking down notes.

3.3 Learning Style Identification

Because of the unprecedented situation occurred due to the COVID-19 pandemic, the whole process has been carried out online. A simple questionnaire in the form of Google form was prepared to know the preferred learning style of the students. For this, a total of 17 students from fourth and sixth semester from computer science branch of Government Polytechnic, Joida, Karnataka, were selected as the subjects.

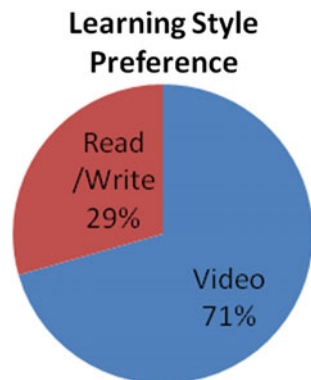
Their responses were recorded in an auto-generated Excel sheet using online questionnaire. Among the 17 students who were the subjects, 12 members preferred to learn through video, and only 5 students preferred the method of read/write process (Table 1, Fig. 2).

Since majority of the students preferred to learn through video, in this study, the focus is on analyzing the cognitive levels of the video learners. Hence, videos are used as the learning material.

Table 1 Preferred learning styles

Preferred learning styles		
Total students	Video	Read/write
17	12	5

Fig. 2 Learning style preference



3.4 Learning Material Selection Process

Since our focus is on using videos as learning materials, the important task was to prepare or to select the suitable videos. We have selected two videos which are significantly different in their difficulty level, and we took the opinion n from five subject material experts (SMEs) who have vast experience in teaching and industry experience.

The feedback obtained from SMEs is categorized as follows.

1. Do you feel the content is useful for the students?
Response: 100% useful.
2. How engaging the videos for the students?
Response: 50% said it is efficient, and 50% said it is at average level.
3. Which cognitive levels can be addressed?
Response: understanding, apply, analyze, and evaluating levels can be addressed.
4. How significantly different the two videos are in terms of difficulty?
Response: 4 out of 6 subject experts felt both the videos are significantly different in their difficulty levels.

Based on the opinions obtained as above from the SMEs, the two videos were selected and decided to address understanding, applying, analyzing, and evaluating cognitive levels.

3.5 Steps to Identify the Cognitive Levels

Since most of the students preferred to learn through video materials available in LMS, YouTube, and other sources, we have used sample videos with good quality and useful content on data structures course as learning material and then we did the assessment based on that learning material. The assessment has been prepared using verbs available in rubrics based on Bloom's taxonomy for different cognitive levels. Based on the correct answers recorded for various assessment questions, we were able to know which cognitive thinking levels are present or not in those students.

3.6 Assessment Based on Learning Material

The process is carried out using online questionnaire. The assessment questions are prepared based on the content in the videos. The verbs used in the questions are taken from the rubrics based on Bloom's Taxonomy. Since the assessment carried out online, timer had been set in order to make sure that the responses are recorded in time, and there should not be any misuse of data from other sources by the students.

The questions were prepared in such a way that the responses from the students are directly related to their cognitive thinking levels. The assessment questions are purely new, and no question is directly or indirectly linked to any other sources.

4 Experiment and Results

Earlier, we have decided to conduct the experiment and to collect the data in a closed experiment room. But due to current unprecedented situation due to COVID-19 pandemic, we then decided to collect the data online. Everything has been done through online questionnaire. The questions are prepared keeping mind that they must address the cognitive levels based on Bloom’s taxonomy (except creating level). Experiment was conducted like a quiz with suitable assessment points.

As we mentioned already, two videos which are significantly different in their difficult level were selected, and questions were based on the content in the videos. Before conducting the quiz, we have taken the feedback from the students who preferred to learn through videos. Based on this, the learners are classified as follows (Table 2).

4.1 Experimental Setup

Quiz 1: Based on First Video

The learning material in this video is about basic of stack data structure for the computer science students. The content in video was able to address all cognitive levels except creating level.

There were five questions, and points are assigned based on required thinking capacity of the students (Table 3).

Table 2 Classification of learners

S. No	Description of the type of learner
1	Fresh learners who feel the learning material is easy
2	Fresh learners who feel the learning material is average
3	Fresh learners who feel the learning material is difficult
4	Concrete learners who feel the learning material is easy
5	Concrete learners who feel the learning material is average
6	Concrete learners who feel the learning material is difficult

Table 3 Assessing cognitive levels

Cognitive level	No. of questions	Points per question	Total
Understanding	2	2	4
Applying	1	4	4
Analyzing	1	4	4
Evaluating	1	6	6
Total			18

Table 4 Assessing cognitive levels

Cognitive level	No. of questions	Points per question	Total
Understanding	1	2	2
Applying	1	4	4
Analyzing	1	6	6
Evaluating	1	6	6
Total			18

Quiz 2: Based on Second Video

The learning material in this video is introduction to linked list data structure for the computer science students. The content in video was able to address all cognitive levels except creating level.

There were four questions, and points are assigned based on required thinking capacity of the students (Table 4).

4.2 Results and Discussions

After the assessment is over, students were made to see their scores so that they can assess themselves, and this would definitely made them to think about their answers.

Based on the answers, it was easy for us to assess which cognitive levels they would map strongly and to which they would map weakly. Based on this, teachers can modify the teaching methodology to improve the cognitive thinking skills of the students and also can personalize the learning materials in order make sure that students must map strongly to all the cognitive levels.

The data analysis is as follows (Tables 5 and 6)

The assessment of cognitive levels mapping has been done as follows.

For the types of learners, refer Table 2.

Table 7 shows the mapping percentage of cognitive level of the students based on the first assessment quiz. Row 3 and Row 6 show that there were no learners to that category for that video learning material.

Table 5 Data analysis of video learners—first video

S. No	Name	Sem	Learning style preference	Feedback on stack video	Cognitive level score						
					About difficult level	Understanding/remember	Apply	Analyze	Evaluate	Total	
1	Aftab	6	Video	No	Less	4	4	4	4	6	18
2	Karthik	6	Video	No	Less	4	0	4	4	0	8
3	Ganesh	6	Video	No	Average	2	0	4	4	6	14
4	Jujepin	4	Video	No	Less	4	4	4	4	6	12
5	Akshata	6	Video	Yes	Average	2	4	4	4	0	10
6	Anusha	4	Video	Yes	Average	4	0	4	4	0	8
7	Priyanka	4	Video	No	Less	2	4	4	4	6	16
8	Roshani	6	Video	Yes	More	4	4	0	0	0	8
9	Jennifer	6	Video	Yes	Less	4	4	4	4	6	18
10	Pranjali	6	Video	No	Average	2	0	4	4	0	6
11	Jestin	6	Video	No	Average	4	4	4	4	6	18
12	Vinayak	6	Video	Yes	Average	2	4	4	4	6	16

Table 6 Data analysis of video learners—second video

S. No	Name	Sem	Feedback on linked list video		About difficult level	Cognitive level score					
			Familiar or not			understanding/remember	Apply	Analyze	Evaluate	Total	
1	Aftab	6	No			2	4	6	6	18	
2	Karthik	6	Yes	Average		2	4	6	6	18	
3	Ganesh	6	Yes	less		2	0	0	6	8	
4	Jujepin	4	No	average		2	0	0	6	8	
5	Alshata	6	Yes	less		2	4	6	6	18	
6	Anusha	4	Yes	average		2	4	0	6	12	
7	Priyanka	4	No	average		2	4	0	6	12	
8	Roshani	6	Yes	less		2	0	6	6	14	
9	Jennifer	6	Yes	more		2	0	6	0	8	
10	Pranjali	6	Yes	less		2	4	0	6	12	
11	Jestin	6	No	average		0	0	0	6	6	
12	Vinayak	6	No	average		2	4	0	6	12	
		6	Yes	average		2	0	6	6	14	

Table 7 Data analysis 1

Type of learner	Understanding %	Applying %	Analyzing %	Evaluating %
1	87.5	50	100	75
2	65	44	100	66
3	–	–	–	–
4	75	100	100	100
5	82.5	65	65	0
6	–	–	–	–

Table 8 Data analysis 2

Type of learner	Understanding %	Applying %	Analyzing %	Evaluating %
1	100	50	100	100
2	65	65	33	66
3	–	–	–	–
4	100	50	0	100
5	100	50	25	100
6	100	50	12.5	100

Table 9 Final result

Type of learner	Understanding %	Applying %	Analyzing %
Fresh	79.35	52.25	83.25
Concrete	89.35	66.25	47.5

Table 8 shows the mapping percentage of cognitive level of the students based on the second assessment quiz. Row 3 shows that there were no learners to that category for that video learning material. Row 6 will be neglected for the final assessment because of less no. of learners in that category (only one learner in our study). The above two tables are simplified by taking averages between them by neglecting Rows 3 and 6, and the final result is as follows. Before this, for the simplification purpose, first find the average of Row 1 and Row 2 and average of Rows 4 and 5 separately in both the tables. The final result is obtained as below (Table 9).

5 Limitations and Conclusion

Cognitive levels of learning domain are important for the students. There are various methods to assess and analyze the cognitive levels. The focus of this study is to map the learning style of the students to their cognitive levels. In our study, 70% of the

students selected prefer to learn through video material. Hence, we have selected only visual learners for this study. Initially, our main objective was to conduct the experiment in close room. Even our main aim of the work was to use physiological sensors (EEG, ECG, GSR, eye trackers, etc.) to collect the data using different stimuli. But due to COVID-19 emergency situation, we were unable to go in that direction, and it was also not possible to take data from more subjects (students). So, we have modified our work of experiment without changing the objective.

We have taken video materials and with the help of feedback from subject material experts (SMEs) used them in this experiment to conduct online assessment. Based on the feedback and answers from the students, we were successful in drawing the conclusion that fresh learners who felt that video materials were less and average and difficult will map strongly to understanding (79.35%), analyzing (83.25), and evaluating (76.55%) cognitive levels and weak map to applying level with 52.25% of score. The concrete learners will map strongly to understanding (89.35%) and evaluating (75%) cognitive levels and moderately map applying level 66.25% and weakly map to analyzing level with 50% score in the online assessments.

In the further our work, this work is going to be extended with more subjects so that online data collection will be replaced with physiological sensing. And also, other type of learners (audio/hands on/read-write) will be addressed. The work can also be enhanced to map the learning style to create cognitive level.

References

1. Mayilvaganan M, Kalpanadevi D (2017) Recent trends in human computer interface to analysis the cognitive skill of students based on user interface. In: International conference on advanced computing and communication systems (ICACCS-2017), Coimbatore, India
2. Ariyaratne MKA, Marikar FMMT (2019) Identification of the best teaching practice by VAK model in the computer degree programme. In: International conference on advancements in computing (ICAC), Malabe, Sri Lanka
3. Rashid NA, Taib MN, Lias S, Sulaiman N (2010) Classification of learning style based on Kolb's learning style inventory and EEG using cluster analysis approach. In: 2nd International congress on engineering education, Kuala Lumpur, Malaysia
4. Uma D, Thenmozhi S, Hansda R (2017) Analysis on cognitive thinking of an assessment system using revised Bloom's taxonomy. In: 5th IEEE international conference on MOOCs, innovation and technology in education. <https://doi.org/10.1109/MITE.2017.00033>
5. Yasmeen A, Yasmin M, Saleem MS (2019) Cognitive learning in outcome-based education: a case study of bachelor of science in electrical engineering. In: International conference on innovative computing (ICIC), 978-1-7281-4682-9/19/\$31.00 ©2019 IEEE
6. Kuttattu AS, Gokul GS, Prasad H, Murali J, Nair LS (2019) Analysing the learning style of an individual and suggesting field of study using machine learning techniques. 978-1-7281-1261-9/19/\$31.00 ©2019 IEEE
7. Patil SK A comparative study of question bank classification based on revised bloom's taxonomy using SVM and K-NN

A Comprehensive Review on Deep Learning-Based Community Detection in Networks



Naveed Ul Islam and R. Sunitha

1 Introduction

Real-world systems often exist as networks corresponding to entities and the relations that exist within the system. Physical, biological, information, and social networks such as the Internet, ecosystems, electric grids, and telephony are typical examples of networks which effectively model the features, structure, and dynamics of such systems. Due to this mapping of real world, the analysis of networks, as surrogates, is crucial toward understanding the structure and dynamics of such complex systems. This has motivated an enormous interest in the fields of network science and social network analysis which deals with the study of organization, functions, and relationships that exist within networks at various scales and levels of granularity. While synthetic and real-world networks arise in a multitude of diverse fields, they share various concepts and properties. One of the important network properties is the existence of its *community structure* [1, 2]—which allows for uncovering the organization of a network by discovering its structural or functional units. A community generally refers to a group of entities that correlate more to each other as compared to the rest of a given network. Intuitively, communities in large-scale data domains refer to such data points as related by some objective or subjective criterion so as to form a cluster (closely knit group of points) distinct from the rest of the dataset. Traditionally, a community is defined in the context of a graph representation as a group of nodes having a denser edge volume between them as compared with the rest of the graph [3]. This property provides insights into the organization of the network at a mesoscopic level beyond entity-pair relationships and, therefore, helps discern the overall network structure in terms of communities.

N. Ul. Islam (✉) · R. Sunitha
Department of Computer Science, School of Engineering and Technology, Pondicherry University, Pondicherry, India

Community formation is a result of structural or functional similarities of nodes in a network [4]. Therefore, discovering communities contributes to knowledge extraction in relation to the structure of the networks about the utility and interactions of entities within a community. For instance, communities on the web would mean a set of web pages that share the same content or topic [5]; in biochemical classification, a community would refer to the identification of functional groups in a metabolic or interaction network [6, 7]. Similarly, in citation analysis, communities can be papers belonging to the same discipline, theme, or author [8]; these can be food webs in ecological niches [9]; a group of friends or users sharing same interest on a social network, for example, groups on Facebook. Similarly in neuroscience, brain networks and neuron interactions can be analyzed for understanding the functional structure of the brain [10] and so on. Community detection has a wide range of applications in molecular biology and computational network biology [11–13], and scientometrics [14] among other areas, including image processing and econometrics, as is reported broadly in [2].

Though the graph representation seems intuitive and a mere collection of edges and nodes, a complex system could easily scale to millions of nodes and edges. This makes community detection a very challenging task and an open and active research problem. In fact, the problem is shown to be *NP-hard* [15] and has not yet been resolved to a satisfactory level. Community detection, also known as *graph clustering* [16] in network science literature, involves different approaches that vary in the way a community is defined. Community detection is, therefore, an ill-defined problem in literature since the definition of a community is subjective or (otherwise) not known in advance. However, there are two broad approaches toward the definition of a community [17]

- *Structural Approach*: a community based on some structural property like topology or link structure of the nodes or network; such as edge density.
- *Semantic Approach*: a community based on some similarity metric between individual nodes to group the most similar ones; typical examples of similarity are a shared interest in a topic or common network attributes such as metadata or link content.

Traditional approaches to detect or extract such communities involve the use of hand-engineered statistical models, probability theory [18, 19], and linear algebra [20]. However, more recently, the use of deep machine learning [21] has been proposed for the task of community detection, owing to the power of deep structures to learn highly approximate representations of the data [22]. More specifically, graph variants of deep neural networks such as graph autoencoders [23] and graph convolutional networks [24] have been employed recently to directly learn representations of the graph structure or the underlying data manifold. Specific algorithmic procedures are embedded or applied onto this learned representation to detect communities. In the subsequent sections, we discuss the motivation for using deep learning for community detection followed by detailing the state-of-the-art deep learning frameworks applied to the task. The survey concludes with a discussion on results of the study along with recommendations for the future work.

2 Rationale

Community detection has been an active research area in SNA and machine learning since past two decades. Particularly, since the seminal work of Girvan and Newman [25], community detection has proved very useful in understanding the structure, organization, dynamics, and evolution of networks [26, 27]. Earliest approaches required hand-engineered representation models to extract and manipulate the graph structural properties. For example, with social network analytic methods, the aim is to find a hard partition of nodes into disjoint groups referred as communities [28]. Further on, with probabilistic methods, a softer partition is sought where a measure of statistical similarity is introduced to cluster nodes with higher probabilities of such similarity [3] into a community. Since the notion of similarity is subjective and as such quantifying, the similarity measure is difficult for a large network that has rich node and link features or contains structure only. Therefore, approaches that can capture both semantic and structural information are proposed under what is called *graph embedding* [29]. The general approach is to embed a high-dimensional network space (geodesic) into a low-dimensional latent space (Euclidean) and then apply a general clustering algorithm to the latent embedding. Non-negative matrix factorization (NMF) [30], spectral clustering (SC) [31], and similar kernel methods are typical examples of this approach. The main drawback with this approach is the computational cost involved when dealing with networks at a scale and the complexity involved in the process such as the time-consumption when designing features and their inflexibility to adapt during the learning process [32].

A more recent approach that overcomes the complexity of previous approaches in terms of scalability and computational power is to directly learn representations that embed structural or feature information of a network. This is based on the intuitive property of deep neural network architectures and frameworks that perform a mapping from a given high-dimensional data space to a low-dimensional representation space. The general idea behind using GNNs is to encode facts and latent beliefs about a network as *probabilistic* graphical models [33]. This can be realized by the assumption that the nodes are sampled from a particular probabilistic distribution, and the learning algorithm validates the truth of this assumption by learning the actual distribution. The algorithm can be applied for individual nodes, sub-networks, or whole networks. For example, *random walks* [34] are used to find node embeddings that can further be replaced by a machine learning algorithm to cover a large graph space. The goal is to populate some measure of node similarity over the course of the walk through the graph structure. For spanning whole networks, the difference is that an entire network is mapped to a low-dimensional representation such that similar nodes in the original graph are embedded close together in the representation manifold. The task of community detection is, therefore, formulated as a learning problem, i.e., to learn a particular kind of clustering or graph partitioning with minimum information loss. Typically, a variational autoencoder (AE) [35] or a stack of autoencoders is employed for learning representations of graph data through reconstruction of the manifold in a latent space. This neural network performs a mapping in a way that

pairwise similarity in original space and the latent space is preserved. Evaluation and results of autoencoders on various datasets [36] show computational efficiency, scalability features, and effective cluster quality [37]. Furthermore, AEs provide for including certain a priori information about the data when modeling the probability distributions for underlying data and can also be fine-tuned for particular downstream applications.

3 Related Work

One of the classic and widely cited works of the first decade (2000–2010) since the inception of the problem of community detection is by Fortunato et al. [2] introducing the area of community detection and detailing at large the traditional social network approaches and methods. A more recent work [3] by the same authors provides an elementary detail into the broad research area and a user guide to approach the problem of community detection. The survey by [38] details widely the developments in the area for the last decade (2010–2019). The authors provide taxonomy of traditional and generative approaches (statistical and probabilistic models) for static and temporal networks from both a structural and semantic standpoints. Specific surveys on graph embedding and graph representation learning (GRL) approaches are detailed in [29, 32] where the authors detail *shallow* and *deep* embedding approaches to learn graph features. But, the subsequent downstream tasks are restricted to node classification, clustering, and recommendation in these works. To the best of our knowledge so far, no survey has been reported recently detailing the application of deep learning approaches specific to community detection.

This survey is motivated by the previous wide-range surveys aforementioned and aims to bridge the gap in the literature from the perspective of deep learning methods that have recently been adapted for the task of community detection. The process of collecting literature was broadly focused on machine learning approaches to the problem. However, the search specifically was tuned by two key parameters: (1) works that exploit and detail deep learning architectures, algorithms, and methods and (2) good-quality articles published within the last decade. This set of recent papers was collected using a phrased search from Google, Google Scholar, IEEE digital library, ResearchGate network, SpringerLink, Elsevier's Science Direct and ACM Digital Library, and other reputed conference proceedings such as AAAI, ICDM, and NIPS. Phrases "community detection in social networks using deep learning," "deep neural networks for community detection," "autoencoder-based community detection" and refined keywords such as "sparse autoencoders" and "deep learning" among others were used over the course of search. The papers collected, thus, were graded by their year of publication and subject further to quality assessment. The criterion was a detailed study of the relevance, efficiency, and improvement over previous approaches and competing works to recommend superior, better-quality algorithms. A schematic representation for the survey's methodology is illustrated in Fig. 1.

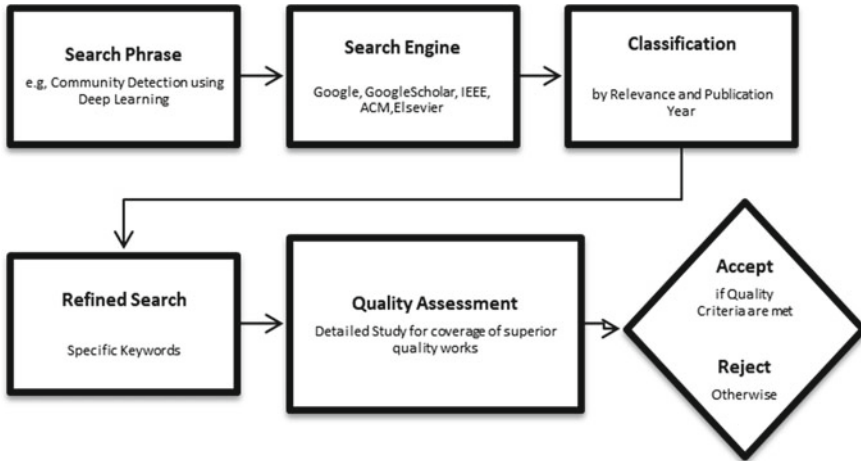


Fig. 1 Paper collection process

4 State-of-the-art

The challenges posed by large graph sizes of real-world networks and the recent development of GNNs, only a limited number of methods have been proposed so far reporting the application of deep learning to the problem of community detection. Therefore, we discuss each one of these approaches separately and provide a summary in Table 1. A discussion on results of the study is presented in the next section.

Graph encoder (GE) [39]—one of the first works employing deep learning to the task of community detection—posits that optimizing the objective function of the autoencoder is similar to finding a solution for spectral clustering. Straightforward implementation of spectral clustering using eigenvalue decomposition is computationally expensive, of cubic order to the number of nodes and super-quadratic for the fastest implement. Further, to ensure a sparse representation so as to improve the efficiency of data when dealing with large-scale data, spectral clustering is not a straightforward method to incorporate or guarantee sparsity without destroying its spectral nature. Autoencoders provide a highly efficient alternative having a complexity of near linear order to the number of nodes usually leads to local optimum due to back propagation, allows for sparsity constraints by introducing a L1 regularization term to the original objective function, and can be stacked into multiple layers of sparse encoders to achieve features and benefits of deep structures.

The authors propose a deep neural network of stacked autoencoders which takes as input the normalized graph similarity matrix. A greedy layer-wise training is employed to learn the best nonlinear representations that satisfy desired sparsity properties and closely approximate the input matrix through reconstruction. On the final learned embedding, *k*-means is run on the sparse encoding output by the last

Table 1 Summary of the performance and advantages of deep learning-based community detection methods on real world and benchmark networks evaluated with standard metrics

Year	Method	Deep architecture	Dataset (s)	Evaluation metric (s)	Advantage (s)
2014	GE [39]	Sparse autoencoder	Wine, 20-newsgroup, database of interacting proteins, and BioGrid	Normalized mutual information (NMI)	Deeper layers yield better communities and outperforms <i>k</i> -means and spectral clustering
2020	VGAECD [33, 40]	Variational graph autoencoder	GN and LFR benchmarks Karate, PolBlogs Cora, PubMed and	Ground truth, NMI, modularity (Q), conductance, and triangle participation ratio	Outperforms baseline methods on ground truth. Performs better than Node2Vec and DeepWalk on synthetic and benchmark graphs
2017	MGAE [36]	Graph convolutional network and stacked autoencoder	Cora, Citeseer, and Wiki	Accuracy (ACC), normalized mutual information (NMI), <i>F</i> -score, precision, recall, average entropy (AE), and adjusted rand index (ARI)	Content and structure-augmented autoencoder. Scalability to a larger graph size with high efficiency and outperforms baseline methods using either or both content and structure including VGAECD
2016	DNR [16]	Stacked autoencoder	Karate, Dolphins, football, Polbooks, PolBlogs, Cora, friendship 6–7, and GN and LFR benchmarks	NMI	Incorporates nonlinear network features Outperforms all competing methods
2018	DIR [41]	Deep feedforward autoencoders	WebKB, Citeseer, Cora, UAI2010, and PubMed	NMI, F-score, ACC, Jaccard similarity, and generalized NMI	Learns multi-level representations. Combines modularity and normalized-cut model in a nonlinear fashion. Outperforms competing methods except with PubMed dataset for NMI

(continued)

Table 1 (continued)

Year	Method	Deep architecture	Dataset (s)	Evaluation metric (s)	Advantage (s)
2018	UWMNE UWMNE-LE [42]	Deep autoencoder	Cornell, Texas, Washington, Wisconsin, Citeseer, UAI 2010, and PubMed	ACC and NMI	Uniform representation for topological and semantic information. Preserves local structures and provides parameter free data integration. Performs best on 6 out of 7 networks in terms of accuracy and 5 out of 7 in NMI. Improves upon the best baseline by 19% in terms of NMI

layer to obtain the clustering results, i.e., communities. The method is evaluated on various real-world graphs and is shown to outperform baseline algorithms including spectral clustering. The stacked architecture boosts cluster quality in terms of accuracy when going from shallow to deep layers. The authors claim the work to be first employing deep learning to community detection by employing stacked autoencoders, and the evaluation results imply that sparsity on graph embedding improves clustering results. The results further show that *Graph Encoder* beats k-means when run on original normalized similarity graph implying the efficiency of deep structures to learn better graph representations.

Variational graph autoencoder (VGAE) [23] framework, based on *variational autoencoder* (VAE) [35], is proposed for automated graph learning on graph-structured data which makes use of latent variables for learning latent representations. The autoencoder model is implemented using a GCN encoder and an inner product decoder for the task of link prediction on several citation networks in the original paper. VGAE projects convolutions onto a univariate Gaussian space and has been considered for graph classification and link prediction only.

Variational graph autoEncoder for community detection (VGAECD) [33] extends the VGAE framework for the task of community detection by introducing a mixture of Gaussians making the model desirable for capturing higher order nonlinear patterns forming community structures. VGAECD employs Bayesian approach to include clustering capabilities in the VGAE for the community detection task. This eliminates the need for a predefined community structure definition as community assignment of nodes is based on the criterion best reducing the loss function. Further, the VAE features provide a lower bound which guarantees convergence to a local minimum, scales well to the size of the network and does not require additional priors

which is, for example, the case with stochastic block model approximation. In practice, the restriction of a univariate Gaussian distribution in the original VGAE model restricts all nodes to be sampled from a single clustering space. This means that dissimilar nodes tend to stay away from the centroid (Gaussian mean). By relaxing the latent space to be a mixture of Gaussians, the mean should be a better representative of each community such that the more similar nodes should stay closer to its represented mean. The objective function is defined as a KL-divergence optimization problem to minimize the difference between assumed prior and the posterior probabilities. This is implied through a maximization of the lower bound (ELBO) with the objective that this difference is minimized. The parameters are learnt using stochastic gradient descent in an inference model using a 2-layer GCN model to arrive at optimal values so that the variational posterior approximates the true posterior as best for the graph data sample. The sampling is done using a Monte-Carlo stochastic gradient variational Bayes estimator where the mean and variance of the multivariate Gaussian distribution are obtained by convoluting through the GCN layers. The evaluation of VGAECD is done on both synthetic (GN and LFR benchmarks) and empirical real-world citation networks with ground truth (Karate, PolBlogs, Cora, and PubMed). The method is compared against seven state-of-the-art methods including primarily the degree-corrected SBM, spectral clustering, DeepWalk, Node2Vec, and VGAE and is shown to have a superior performance almost consistently in terms of widely used community detection evaluation metrics as accuracy (ACC), normalized mutual information (NMI), convergence, and modularity maximization.

In a follow-up paper [40], the authors demonstrate that VAE and its variants are biased toward the reconstruction loss minimization over a clustering loss. Thus, a poor performance as training is prolonged over time. A dual optimization approach is proposed to increase the learning speed and better practices for linearizing the encoder. The results of these modifications show better results in that the learnable parameters are reduced in number and the network converges faster alongside a faster training speed. The performance gain, however, in terms of community detection is marginal.

Marginalized graph autoencoder for graph clustering (MGAE) [36]—a unified framework—which integrates both content and structural information using stacked autoencoders for nonlinear representation to extract graph clusters as communities. The algorithm takes the graph content and structure as input and learns a content-augmented deep representation of nodes which is refined and clustered using spectral clustering. The autoencoder architecture is an extension of the GCN to the task of unsupervised graph clustering. The important difference here is the idea of employing a marginalization process which effectively yields an effective solution to the task of community detection. Using marginalized autoencoders, the process allows for the interplay between content and structure, adds random noise for scaling the training to a large dataset, and provides for an efficient derivation of a closed form solution to the autoencoder. Further, the stacked architecture provides deep representation learning with each graph autoencoder trained sequentially along with corrupted information and optimization to minimize the reconstruction loss

between encoded and clean content void of random noise. Furthermore, the representation, thus, yielded is fed into a spectral clustering algorithm for the final clustering results. Summarily, the algorithm employs a spectral graph convolution for the task of community detection, which is compared against baseline methods exploiting structure or content information only such as k-means, BigClam, Graph Encoder, and DeepWalk, or methods using both structure and content such as circles, relational-topic model (RTM), robust multiview spectral clustering (RMSC), and VGAE on three benchmark datasets—Cora, Citeseer, and Wiki. The evaluation results show the framework yielding better than the baseline methods in terms of ACC, NMI, F-score, precision, recall, average entropy (AE), and adjusted rand index (ARI) metrics.

Deep nonlinear reconstruction (DNR) [16] algorithm employs a stacked autoencoder inspired by the singular value decomposition-based modularity maximization, to find a nonlinear embedding that reconstructs the modularity matrix. As an improvement, a semi-supervised DNR algorithm which incorporates pairwise constraints is proposed. This offers a variation in the autoencoder stack configuration, the number of layers, and a feedforward from the output of a previous autoencoder to the next up in the stack. The method is shown to perform better on Girvan-Newman (GN) and Lancichinetti-Fortunato-Radicchi (LFR) benchmarks and other real-world graphs in comparison with certain traditional baseline algorithms.

Deep integration representation (DIR) [41]—motivated by the similarity between autoencoders and spectral clustering in terms of matrix reconstruction, the algorithm based on nonlinear approach of deep autoencoders combines topological and node content information of networks for community discovery. DIR effectively integrates network modularity and normalized cut as two community detection models or objectives in a deep joint reconstruction framework. The algorithm takes as input the adjacency and similarity matrices and computes modularity and Markov matrices, respectively, using the node contents through the degree matrix. A spectral matrix is set with modularity and Markov matrices and the rows of which are fed to a single hidden-layer autoencoder as the input. The parameters of the autoencoder are optimized to find the latent representation. The representation is further fine-tuned and clustered using k-means to find the communities. The algorithm is evaluated on various citation datasets with various layer configurations and is shown to have consistently better results in terms of NMI, F-score, Jaccard similarity, and generalized NMI when compared to the baseline approaches considering topology or node contents only or both.

Unified weight-free multi-component network embedding (UWMNE) [42]—an effective representation learning framework for community detection using a deep autoencoder for network reconstruction. This approach effectively enhances the network embedding approaches utilizing both topological and semantic information as features through a graph representation of both kinds of information. Further, the representation is enhanced using local neighborhood structure, i.e., pairwise node relationships, and the topological and semantic graphs are learnt using a deep autoencoder to propose an integrated network representation. The significant difference is that the previous class of network embedding methods employs a feature vector representation for capturing both content and structure, while in this approach,

the intuition is that graph representation provides a better representation scheme for relational information and describes the geometry of latent manifolds underlying the semantic information. The network topologies are maintained in graph form, and semantic information is represented using neighborhood graphs. This information is integrated in a joint reconstruction framework of neural networks.

A variant model of UWMNE with local enhancement (**UWMNE-LE**) is proposed in the same work where additional information about link strength and semantic similarity is included to reinforce a local structure consistency for better downstream learning tasks. This optimized model is applied for three network applications, viz. node classification, user recommendation, and community detection on seven datasets and is compared with nine state-of-the-art methods employing topological embedding or both topological and semantic embedding. This deep autoencoder model performed the best on six of the seven networks in terms of accuracy AC and best on five out of the seven networks in terms of NMI.

5 Discussion

Community detection is a challenging task in the domain of social network analysis and machine learning. Over the years, many approaches, methods, and solutions have been proposed to solve the problem for a world of ever-increasing data. Classic approaches based on structural or semantic information either do not scale computationally or do not yield neat communities. Furthermore, there is no agreed-upon definition of a community in the existing literature, and therefore, the methods vary greatly in data representation, underlying mathematical models and implementation details. This paper provided a brief overview of the existing approaches to introduce the broad research area. However, proceeding toward larger data domains, with high frequency of arrival and wide variety of kind, an immediate and urgent task is to develop approaches, methods, and algorithms to handle the data for subjective data clustering and downstream tasks based on such analyses. The emergence of deep machine learning architectures as neural networks provides for immense power and flexibility to deal with such data. The application of deep machine learning to the task of community detection is a recent development in the research community and has not been detailed in the literature so far, to the best of our knowledge. This work detailed the type of learning architectures and expanded on the subsequent methods and algorithms proposed so far.

6 Conclusion and Future Scope

A community is defined in a given network as a group of nodes having a denser edge volume between them as compared with the rest of the network and helps in analyzing the functional and structural organization of a network. Community detection has

a wide range of applications in network science, molecular biology, feature engineering, computational neuroscience, image processing among other fields. Earlier approaches required hand-engineered representation models to extract and manipulate the graph structural properties. A more recent approach that overcomes the complexity of previous approaches in terms of scalability and computational power is to employ neural network architectures to learn a particular kind of clustering or graph partitioning with minimum information loss. This work detailed the deep architectures, methods, and algorithms proposed by the wider research community and provided with the future research directions in the areas of deep machine learning and community detection.

Deep machine learning approaches simultaneously provide for functionalities of dimensionality reduction and latent variable discovery but have been exploited mostly for simple sequences and information grids such as text and images. However, graphs are far more complex in that they contain complex topological structures, have no specific node ordering, and often contain multimodal and dynamic features. Therefore, very recently, deep neural network architectures that allow for nonlinear transformations of the graph structure have been developed for the task. The challenges are to evaluate these methods on a wide range of datasets including synthetic and real-world datasets, to validate their performance against ground truth and to improve these methods for computational efficiency and scalability. Other major challenges include modeling dynamic graph behavior and temporal graphs, improving interpretability and evaluation against more meaningful benchmarks that emphasize properties useful for an array of tasks while dealing with real-world applications.

References

1. De Meo P, Ferrara E, Fiumara G, Proveti A (2014) Mixing local and global information for community detection in large networks. *J Comput Syst Sci* 80:72–87. <https://doi.org/10.1016/j.jcss.2013.03.012>
2. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486:75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>
3. Fortunato S, Hric D (2016) Community detection in networks: a user guide. *Phys Rep* 659:1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
4. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E—Stat Phys Plasmas Fluids Relat Interdiscip Top* 70(6). <https://doi.org/10.1103/PhysRevE.70.066111>
5. Flake GW, Lawrence S, Lee Giles C, Coetzee FM (2002) Self-organization and identification of web communities. *Computer* (Long Beach, Calif) 35:66–71. <https://doi.org/10.1109/2.989932>
6. Rives AW, Galitski T (2003) Modular organization of cellular networks. *Proc Natl Acad Sci U S A* 100:1128–1133. <https://doi.org/10.1073/pnas.0237338100>
7. Chen J, Yuan B (2006) Detecting functional modules in the yeast protein–protein interaction network. *Bioinformatics* 22:2283–2290. <https://doi.org/10.1093/bioinformatics/btl370>
8. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. *Sci Rep* 2. <https://doi.org/10.1038/srep00336>
9. Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol Lett* 5:558–567. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>

10. Deco G, Corbetta M (2011) The dynamical balance of the brain at rest. *Neuroscientist* 17:107–123. <https://doi.org/10.1177/1073858409354384>
11. Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP (2015) Convolutional networks on graphs for learning molecular fingerprints. In: *Advances in neural information processing systems*. pp 2224–2232
12. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P (2016) Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des* 30:595–608
13. Liu C, Ma Y, Zhao J, Nussinov R, Zhang Y-C, Cheng F, Zhang Z-K (2020) Computational network biology: data, models, and applications. *Phys Rep* 846:1–66. <https://doi.org/10.1016/j.physrep.2019.12.004>
14. Rosvall M, Bergstrom CT (2010) Mapping change in large networks. *PLoS One* 5. <https://doi.org/10.1371/journal.pone.0008694>
15. Bui TN, Jones C (1992) Finding good approximate vertex and edge partitions is NP-hard. *Inf Process Lett* 42:153–159. [https://doi.org/10.1016/0020-0190\(92\)90140-Q](https://doi.org/10.1016/0020-0190(92)90140-Q)
16. Liang Y, Cao X, He D, Chuan W, Xiao W, Weixiong Z (2016) Modularity based community detection with deep learning. *IJCAI Int Jt Conf Artif Intell* 2252–2258
17. Ding Y (2011) Community detection: topological versus topical. *J Informetr* 5:498–514. <https://doi.org/10.1016/j.joi.2011.02.006>
18. Nettleton DF (2013) Data mining of social networks represented as graphs. *Comput Sci Rev* 7:1–34. <https://doi.org/10.1016/j.cosrev.2012.12.001>
19. Abbe E (2018) Community detection and stochastic block models. *Found Trends Commun Inf Theory* 14:1–162. <https://doi.org/10.1561/01000000067>
20. Vishwanathan SVN, Schraudolph NN, Kondor R, Borgwardt KM (2010) Graph kernels. *J Mach Learn Res* 11:1201–1242
21. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
22. Yu W, Zheng C, Cheng W, Aggarwal CC, Song D, Zong B, Chen H, Wang W (2018) Learning deep network representations with adversarially regularized autoencoders. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2663–2671. <https://doi.org/10.1145/3219819.3220000>
23. Kipf TN, Welling M (2016) Variational graph auto-encoders 1–3
24. Berg R, van den Kipf TN, Welling M (2017) Graph convolutional matrix completion
25. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99:7821–7826
26. Karrer B, Levina E, Newman MEJ (2008) Robustness of community structure in networks. *Phys Rev E* 77:46119
27. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D (2004) Defining and identifying communities in networks. *Proc Natl Acad Sci* 101:2658–2663
28. Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. *Phys Rev E—Stat Nonlinear Soft Matter Phys* 80:1–12. <https://doi.org/10.1103/PhysRevE.80.056117>
29. Goyal P, Ferrara E (2018) Graph embedding techniques, applications, and performance: a survey. *Knowledge-Based Syst* 151:78–94. <https://doi.org/10.1016/j.knosys.2018.03.022>
30. Liu X, Wang W, He D, Jiao P, Jin D, Cannistraci CV (2017) Semi-supervised community detection based on non-negative matrix factorization with node popularity. *Inf Sci (Ny)* 381:304–321. <https://doi.org/10.1016/j.ins.2016.11.028>
31. Shinnou H, Sasaki M (2008) Spectral clustering for a large data set by reducing the similarity matrix size. In: *Proceedings 6th international conference language resources evaluation Lr*. pp 201–204
32. Hamilton WL, Ying R, Leskovec J (2017) Representation learning on graphs: methods and applications 1–24
33. Choong JJ, Liu X, Murata T (2018) Learning community structure with variational autoencoder. In: *Proceedings of the IEEE international conference on data mining, ICDM*, pp 69–78. <https://doi.org/10.1109/ICDM.2018.00022>
34. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A* 105:1118–1123. <https://doi.org/10.1073/pnas.0706851105>

35. Kingma DP, Welling M (2019) An introduction to variational autoencoders. *Found. Trends® Mach Learn* 12:307–392. <https://doi.org/10.1561/22000000056>
36. Wang C, Pan S, Long G, Zhu X, Jiang J (2017) MGAE: marginalized graph autoencoder for graph clustering. *Proceedings of the international conference on information and knowledge management. Part F1318*, pp 889–898. <https://doi.org/10.1145/3132847.3132967>
37. Cao S, Lu W, Xu Q (2016) Deep neural networks for learning graph representations. In: 30th AAAI conference on artificial intelligence. pp 1145–1152
38. Azaouzi M, Romdhane L (2018) Ben: an efficient two-phase model for computing influential nodes in social networks using social actions. *J Comput Sci Technol* 33:286–304. <https://doi.org/10.1007/s11390-018-1820-9>
39. Tian F, Gao B, Cui Q, Chen E, Liu TY (2014) Learning deep representations for graph clustering. *Proc Natl Conf Artif Intell* 2:1293–1299
40. Choong JJ, Liu X, Murata T (2020) Optimizing variational graph autoencoder for community detection with dual optimization. *Entropy* 22:1–21. <https://doi.org/10.3390/e22020197>
41. Jin D, Ge M, Li Z, Lu W, He D, Fogelman-Soulie F (2018) Using deep learning for community discovery in social networks. In: *Proceedings of the international conference on tools with artificial intelligence ICTAI*. pp 160–167. <https://doi.org/10.1109/ICTAI.2017.00035>
42. Jin D, Ge M, Yang L, He D, Wang L, Zhang W (2018) Integrative network embedding via deep joint reconstruction. *IJCAI international joint conference on artificial intelligence*. pp 3407–3413. <https://doi.org/10.24963/ijcai.2018/473>

Deep Transfer Learning Model-Based Automated Detection of COVID-19 from X-ray Images and Interpretation of COVID-19 Images Using GLCM Texture Features



Shilpa Ankalaki, Kartikeya Shorya, and Jharna Majumdar

1 Introduction

The novel coronavirus (nCoV) infection proved its existence for the first time in Wuhan, China, and has expansively spread across the world since January 2020 [1]. World Health Organization (WHO) declared the outbreak of nCoV as a “Public Health Emergency of International Concern” on 30 January 2020 [2]. COVID-19 severely affected the countries like China, USA, Italy, Spain, China, Germany, France, Iran and India. It is essential to come up with the automatic detection of COVID-19 using X-ray or CT images which is most useful when there is a shortage of COVID-19 detection kits. Nowadays, deep learning approaches in artificial intelligence have been used in many of the applications like image classification, image segmentation and object detection including medical image segmentation. Deep learning approaches play a major role in the detection of COVID-19 using X-ray or CT image data, and researchers also realized significant discoveries of imaging studies of COVID-19 using deep learning approaches.

This research aims to assess the efficiency of advanced pre-trained convolutional neural networks using the approach of transfer learning. To accomplish this task, 328 samples of COVID-19 and 234 normal chest X-ray images are collected and utilized to train and test the deep transfer models like VGG19, ResNet50 and DenseNet121.

S. Ankalaki (✉)

Department of M.Tech CSE, Nitte Meenakshi Institute of Technology, Bangalore 560064, India

K. Shorya

Department of CSE, Nitte Meenakshi Institute of Technology, Bangalore 560064, India

J. Majumdar

Department of M. Tech Computer Science & Engineering, Head Centre for Robotics Research, Head Centre for Space Research, Nitte Meenakshi Institute of Technology, Bangalore 560064, India

e-mail: jharna.majumdar@nmit.ac.in

The contributions of this work are as follows:

- VGG19, ResNet50 and DenseNet121 deep transfer models are employed to classify the images into COVID-19 and non-COVID-19 image and performance evaluation of these models.
- Employed GLCM features to determine the best feature that distinguish the COVID-19 and non-COVID-19 images.
- Segment the COVID-19 and non-COVID-19 images using best GLCM feature to interpret the COVID-19 and non-COVID-19 images.
- GRAD-CAM algorithm is used along with best performing deep transfer model to interpret the decision of deep transfer model and COVID-19 and non-COVID-19 images.

2 Literature Survey

This section provides a study about state-of-the-art researches which utilised machine learning and deep learning in the field of medical to detect and classify COVID-19 affected X-ray images.

The authors of [3] developed the methodology for the detection of COVID-19 with limited dataset of chest COVID-19 X-ray images using model-based GAN and approaches of deep transfer learning. Researcher's collected 307 images with four different classes namely, normal, pneumonia bacterial, COVID-19 and pneumonia virus. Alex net, Restnet18 and Google Net are the deep transfer models selected to train the dataset. In this research, investigators obtained the best performance using Google Net with accuracy of 99.9%.

The investigators employed real-time object detection system called YOLO and integrate with proposed model Dark Net to accomplish the task of automatic COVID-19 detection [4]. In this study, investigators proposed binary classification to classify the images as no-finding and COVID classes and multi-class classification to classify images into no-findings, COVID and pneumonia classes; and achieved the accuracy of 98.08% and 87.02% for binary classes and multi-class case, respectively.

A comprehensive study has been conducted on the multi-task deep learning methods for COVID-19 detection [5]. In this study, researchers accomplished both classification and segmentation of COVID-19 infected regions using two deep transfer learning approaches namely, inception residual recurrent convolutional neural network for classification and NABLA-N network model for segmentation. These techniques are evaluated on both X-ray and CT scan images.

Detection of COVID-19 using deep learning algorithms along with machine learning algorithm is accomplished by the study in [6]. In this study, researchers employed ResNet50 for feature extraction and SVM for classification.

AI-based structure for COVID-19 detection is demonstrated by the study conducted in [7] by proposing the model called SqueezeNet. SqueezeNet is light

network design is tuned for the COVID-19 diagnosis with Bayesian optimization additive. Using this method, researchers obtained the accuracy of 98.3% for multi-class and 100% for recognition of COVID-19 class individually.

Researchers considered additional clinical symptoms, exposure history and laboratory testing. Artificial intelligence has been used to integrate these information with chest CT findings to early detection of COVID-19 patients [8].

3 Proposed Work

The proposed work consists of four phases those are,

- Detection of COVID-19 Images
- Analysis of GLCM features that distinguish between COVID-19 and normal images
- Detection of COVID-19 infected regions in an X-ray images
- Interpretation of results using heat maps.

3.1 Detection of COVID-19 Images

Transfer learning is an approach in deep learning where we make use of pre-trained models instead of creating models from scratch which will save us lot of time and precious processing power. The reason for using transfer learning is because these pre-trained networks can be used again will also be beneficial if we do not have enough data. In this proposed work, well-known pre-trained CNN like VGG19 network, ResNet50 network and DenseNet121 network are employed to discriminate between normal and COVID-19 infected images. VGG-19 [9] comprises of 16 convolutional layers in five convolutional blocks and three fully connected layers (fc6-8). ResNet is a form deep network based on residual learning. The ResNet-50[10] contains 50 layer. DenseNet is relatively similar to ResNet with some fundamental differences. ResNet uses an additive method that merges the previous layer (identity) with the future layer, whereas DenseNet concatenates the output of the previous layer with the future layer. The DenseNet [11] is divided into dense blocks where a number of filters are different, but dimensions within the block are the same. Transition layer applies batch normalization using down sampling. The number of filters changes between the DenseBlocks, increasing the dimensions of the channel. The growth rate (k) helps in generalizing the first layer. It controls the amount of information to be added to each layer. Figure 2 depicts the pre-trained CNN networks architecture used in this study (Fig. 1).

Proposed work optimized the pre-trained CNNs for COVID-19 dataset using the approach of transfer leaning. In this work, the input layer of the CNN architectures are substituted with new input layer, where the size of images is $224 \times 224 \times 1$. All the models are trained over the dataset for maximum of 15 epochs. During

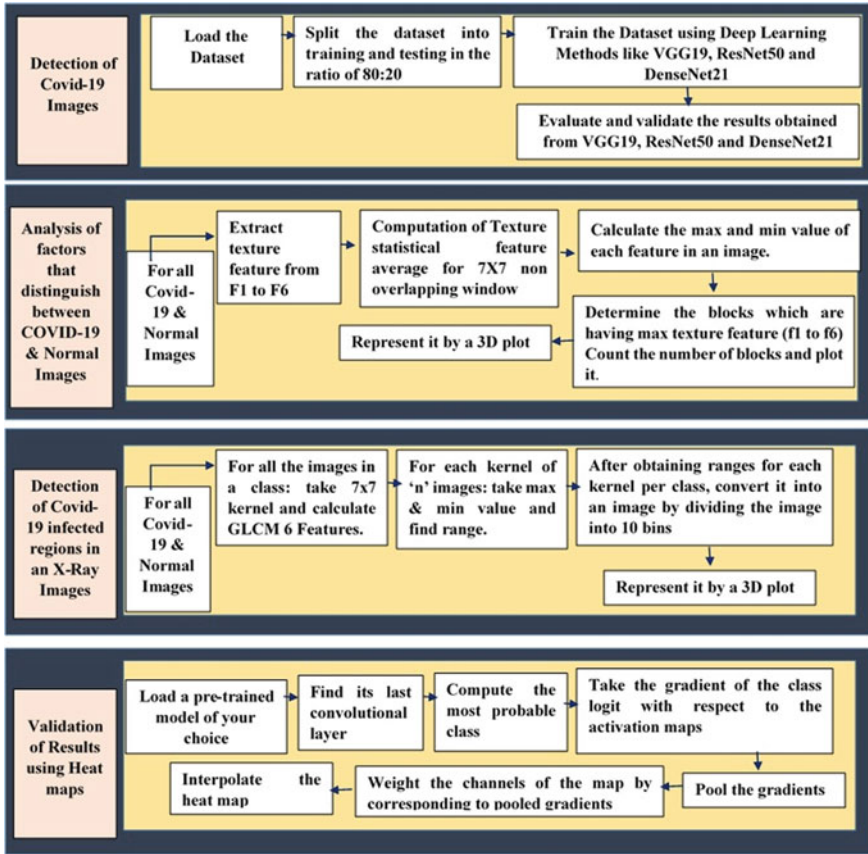


Fig.1 Different phases of proposed methodology

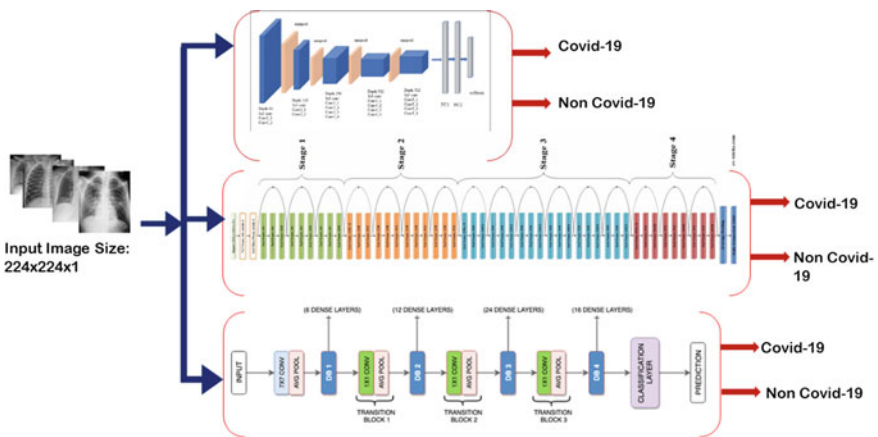


Fig. 2 An overview of pre-trained CNN networks architecture used in this study

training phase, a small portion of the training data is kept separately to be able to use for validation. The ratio of training and validation is 80:20. All networks are trained using the parameters, namely optimizer, SGDM; initial learning rate, 0.01; validation frequency. In this work, instead of using constant learning rate, a learning rate scheduler is used to decay the learning rate. It follows Step Learning Rate Policy with learning rate decays by a constant of 0.1 in every 7 steps and cross entropy loss is used as loss function.

3.2 Analysis of factors that distinguish between COVID-19 and Normal Images

Extraction of Texture Features

The changes of grey values in its vicinity of an image will be reflected by texture features. Such changes are correlated statistically and spatially. One of the well-known feature extraction technique is grey-level-co-occurrence matrix (GLCM). Different combination of grey level intensities is tabulated in GLCM. For every (i, j) in GLCM, the number of occurrences for every pair of levels i and j which are at different $(i1, j1)$ at a distance d is computed. The pixels are described in a way, i.e.

- Adjacent horizontally, P0
- Adjacent vertically, P90
- Adjacent diagonally, P45 and P135.

From the co-occurrence matrix, Haralick [12] proposed 14 features out of which 6 are invariant. The 6 features are shown in Fig. 3.

In this phase, an attempt has been made to determine the factors that distinguishes between COVID-19 and non-COVID-19 images. To accomplish this task, the proposed methodology employed texture feature characteristics. In this study, we have considered GLCM feature, where GLCM represents the image with respect to six features those are: angular second moment (F1), contrast (F2), maximum probability (F3), entropy (F4), homogeneity (F5) and variance (F6). The proceeding section explains method adopted in this work.

Figure 4 summarized following steps.

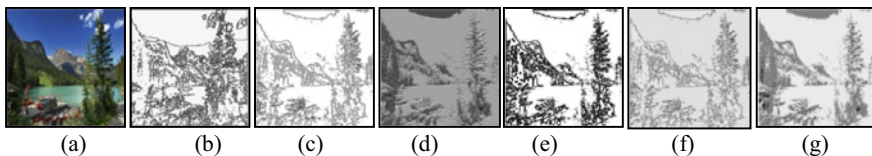


Fig. 3 GLCM texture features **a** input image **b** energy **c** entropy **d** contrast **e** max probability **f** homogeneity **g** variance

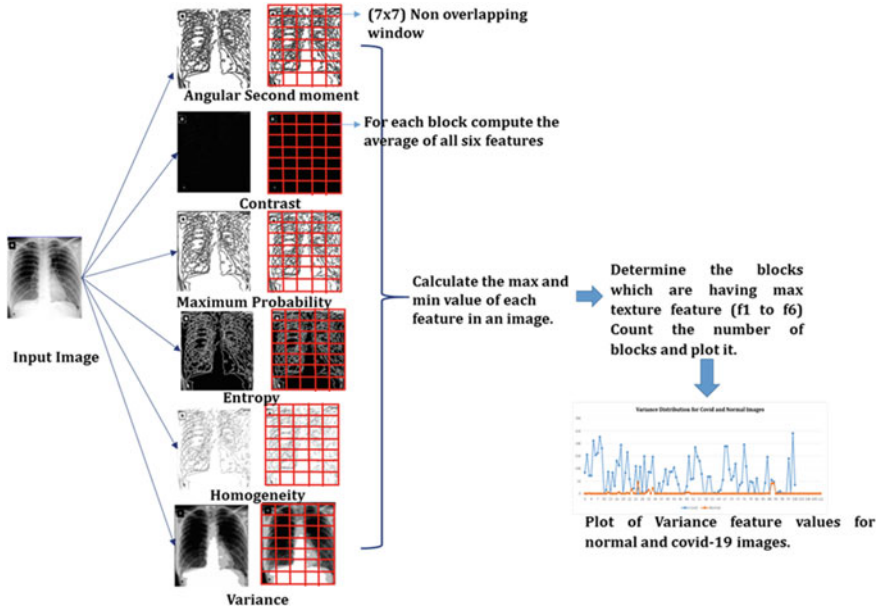


Fig. 4 Analysis of GLCM features that distinguish between COVID-19 and normal images

- Create the dataset which consist of “n” number of COVID and normal images.
- For all images in the class, apply texture feature from F1 to F6.
- Consider the 7×7 non-overlapping window, compute the average of each 7×7 window average in an image for all 6 texture features.
- Plot the features using 3D plot.
- Calculate the max and min value of each feature in an image.
- Determine the blocks which are having max texture feature (f1–f6) except entropy. For entropy, consider the minimum entropy value of an image.
- Count the number of blocks and plot it.

3.3 Detection of COVID-19 infected regions in X-ray Images

In this method, we have considered the 7×7 overlapping window with stride = 1 and performed following steps.

Number of Classes: 2
 Class 0: Normal
 Class 1: COVID-19 affected
 Kernel: 7×7
 Steps:

- Take “n” images from each class

- Take 7×7 kernel and calculate GLCM 6 Features
- Take max and min value and find range
- After obtaining ranges for each kernel per class, convert it into an image by dividing the image into 10 bins (Fig.5)
- Represent the output image.

The entire feature map is divided into 10 bins, and for each bin, a palette colour is chosen from Fig. 5.

Here goes the feature maps obtained after following the previous steps. In Fig. 6a, the input image is an X-ray image of a normal subject. The respective features maps are shown in Fig. 6b–g. Similar feature maps for COVID-19 affected subject is shown in Fig. 7.

Fig. 5 Palette colour of each bin

Bin	Palette
10	
9	
8	
7	
6	
5	
4	
3	
2	
1	

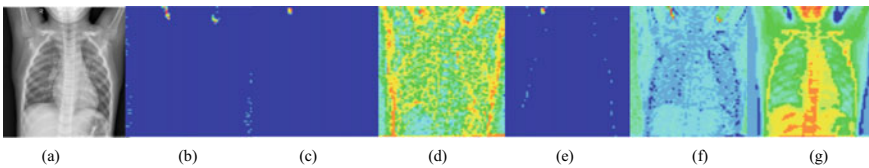


Fig. 6 **a** Input image (normal), **b** energy, **c** entropy, **d** contrast, **e** maximum probability, **f** homogeneity, **g** variance

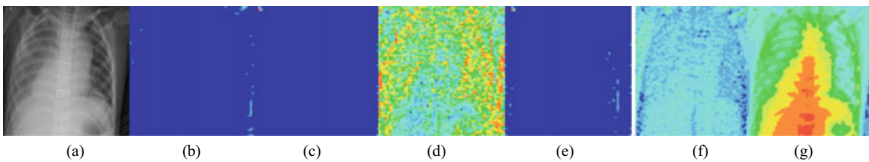


Fig. 7 **a** Input image (COVID-19), **b** energy, **c** entropy, **d** contrast, **e** maximum probability, **f** homogeneity, **g** variance

3.4 Interpretation of Results using Heat maps

Usually, deep learning models are considered as “black box” methods, and many times, we are not able to justify the answers for the questions like which area of an image is examining by the network, which series of neurons activated in the forward-pass during inference/prediction and how the network attained at its final output. All these things causes the interpretability question, i.e. how can we trust the decisions of a model if we cannot properly validate how it arrived there? This interpretability problem is addressed by the method GRAD-CAM [13]. Gradient-weighted class activation mapping (GRAD-CAM), uses the gradients of target concept, passing the gradient information into the final convolutional layer to produce a coarse localization map emphasizing the significant areas in the image for predicting the concept. Steps involved in the GRAD-CAM are summarized below:

- Load the desired convolutional model
- Determine its last convolutional layer
- Compute the utmost probable class
- Logit the gradient of the class with activation maps and pool the gradients; channels of the maps are weighted with respects to the pooled gradients.
- Interpolate the heat map.

4 Results and Analysis

4.1 Detection of COVID-19 Images

The task of detection and classifying the COVID-19 and non-COVID-19 images is accomplished by using the pre-trained VGG-19, ResNet-50 and DenseNet-121 CNN architectures and customised its last dense layer to classify the images into COVID-19 and non-COVID-19 images. Tables 1, 2 and 3 depict the configuration of final dense layer added to VGG-19, ResNet-50 and DenseNet-121 CNN architecture to obtain the classification results using transfer learning.

To train the above models, we have used the following parameters: instead of a constant learning rate, a learning rate scheduler was used to decay the learning rate. It follows Step Learning Rate policy with learning rate decaying by a constant of 0.1 in every 7 steps. Cross entropy loss function and Stochastic gradient optimizer with a momentum of 0.9 are used to train the models. Figures 8, 9, 10 and 11 depict the validation loss and accuracy for VGG19, ResNet50 and DensNet121 custom models, respectively.

Both VGG19 and DenseNet121 CNN architecture achieved an accuracy of 98.80% and accuracy of ResNet50 custom model is 97.65% to classify the images into COVID-19 and non-COVID-19 classes. VGG19 and DenseNet121 custom models are achieved the same accuracy as 98.80%, we have used VGG-19 architecture to generate the heat maps.

Table 1 Custom model of VGG19

VGG19 custom model	
Layers or functions	Parameters
Linear layer	(25,088, 4096)
ReLU	N/A
Dropout	Probability = 0.5
Linear layer	(4096, 1024)
ReLU	N/A
Dropout	Probability = 0.5
Linear layer	(1024, 128)
ReLU	N/A
Linear layer	(128, 2)
SoftMax	Dim = 1

Table 2 Custom model of ResNet50

ResNet50 custom model	
Layers or functions	Parameters
Linear layer	(2048, 1024)
ReLU	N/A
Dropout	Probability = 0.5
Linear layer	(1024, 256)
ReLU	N/A
Dropout	Probability = 0.3
Linear layer	(256, 16)
ReLU	N/A
Linear layer	(16, 2)
SoftMax	Dim = 1

Table 3 Custom model of DenseNet121

DenseNet121 custom model	
Layers or functions	Parameters
Linear layer	(1024, 128)
ReLU	N/A
Dropout	Probability = 0.5
Linear layer	(128, 16)
ReLU	N/A
Linear layer	(16, 2)
SoftMax	Dim = 1

Fig. 8 Validation loss and accuracy for VGG19 custom model

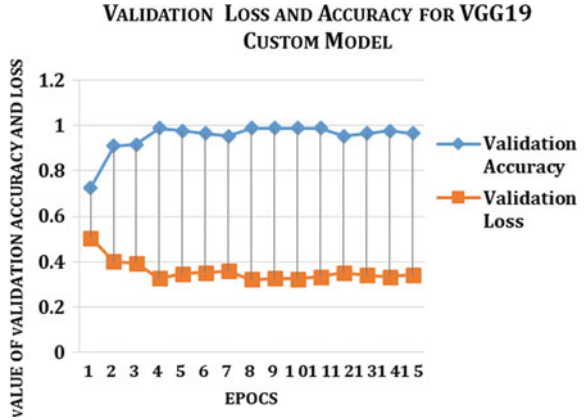


Fig. 9 Validation loss and accuracy for ResNet50 custom model



Fig. 10 Validation loss and accuracy for DenseNet121 custom model

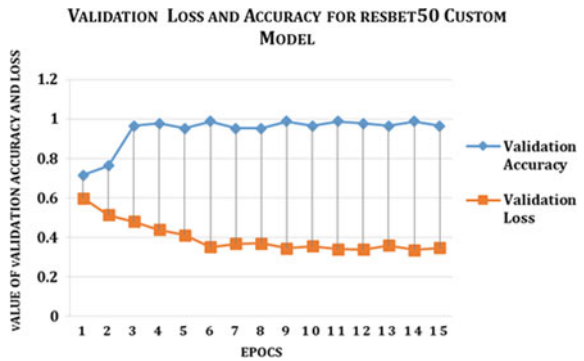
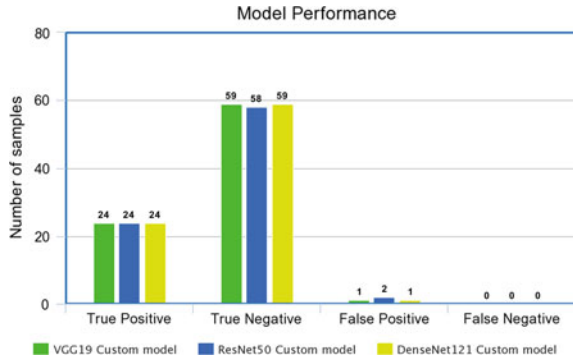


Fig. 11 Model performance of VGG19, ResNet50 and DenseNet121 custom model



4.2 Analysis of GLCM features those distinguish between COVID-19 and Normal Images

This section provides the experimental results of the method described in Sect. 3.2. The detailed analysis has been carried out on 200 images (100 COVID-19 images and 100 non-COVID-19 images) to analyse the GLCM feature distribution in COVID-19 and non-COVID-19 images. In this study, images are divided into 7×7 non-overlapping window for each window, average GLCM features (F1–F6) and determine the blocks which are having maximum and minimum features; and count of these blocks are computed and analysed for all the images. Figures 12, 13, 14, 15, 16 and 17 depict the analysis of images with respect to all GLCM features.

Analysis of the GLCM features depicted in Figures 12, 13, 14, 15, 16 and 17 clearly shows the count of the blocks which are having maximum value of GLCM feature and variation of counts for COVID-19 and non-COVID-19 images. The range of each GLCM features, and its occurrences are depicted in Table 4.

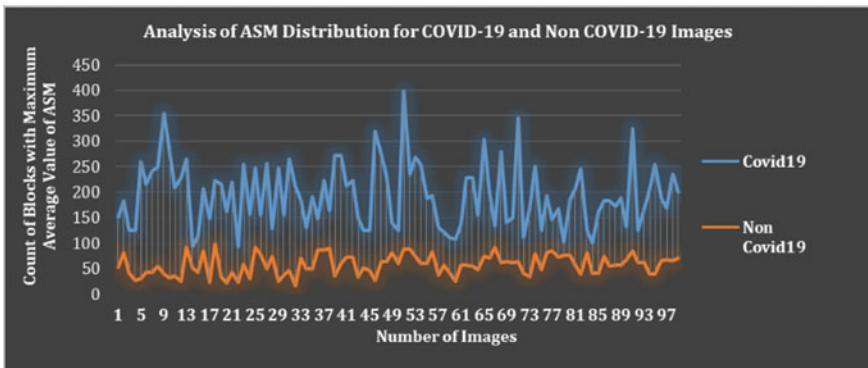


Fig. 12 ASM distribution for COVID-19 and non-COVID-19 images

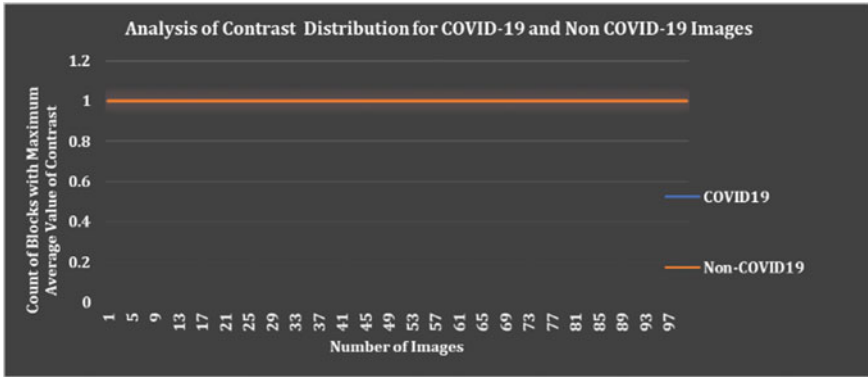


Fig. 13 Contrast distribution for COVID-19 and non-COVID-19 images

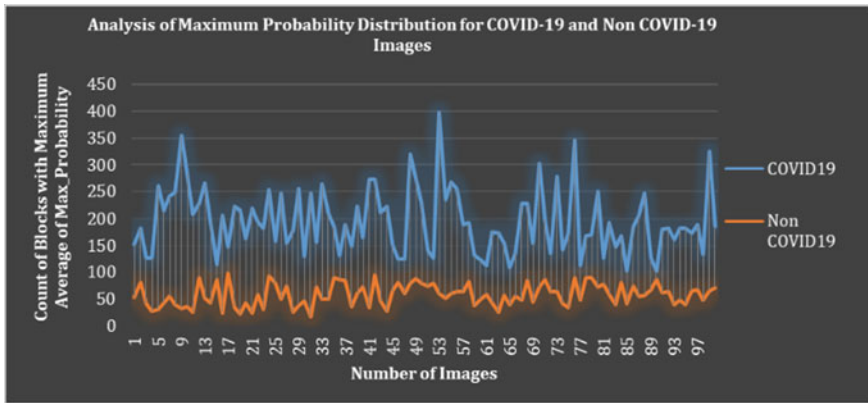


Fig. 14 Max probability distribution for COVID-19 and non-COVID-19 images

From this analysis, we can conclude that variance is the best feature which distinguish between COVID-19 and non-COVID-19 images. This feature is considered for the further analysis to detect COVID-19 infected regions in X-ray Images which are explained in further section.

4.3 Detection of Covid-19 Infected Regions in X-ray Images

On considering feature maps of GLCM, it can be seen that feature “variance” reflects the change in grey values along the vicinity of a pixel (Fig. 18). The distribution of texture feature values is subjectively more visible in variance feature. Below are the results of variance feature maps of GLCM texture for normal and COVID-19 affected X-ray image of lungs. Number of test samples used are 120 normal and 120

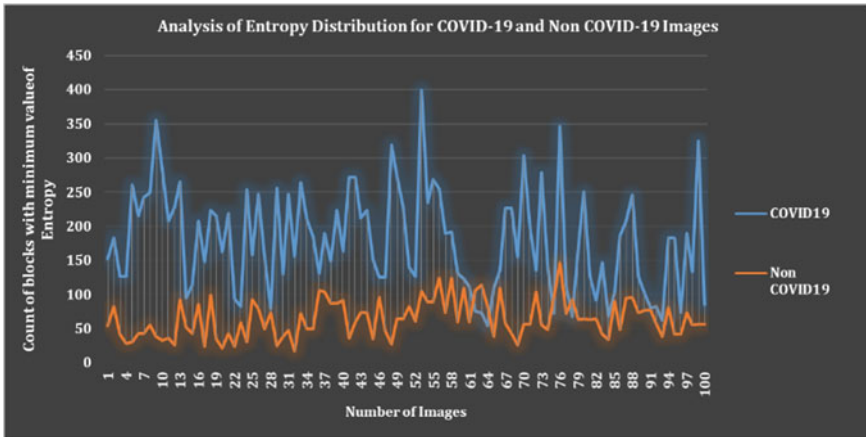


Fig. 15 Entropy distribution for COVID-19 and non-COVID-19 images

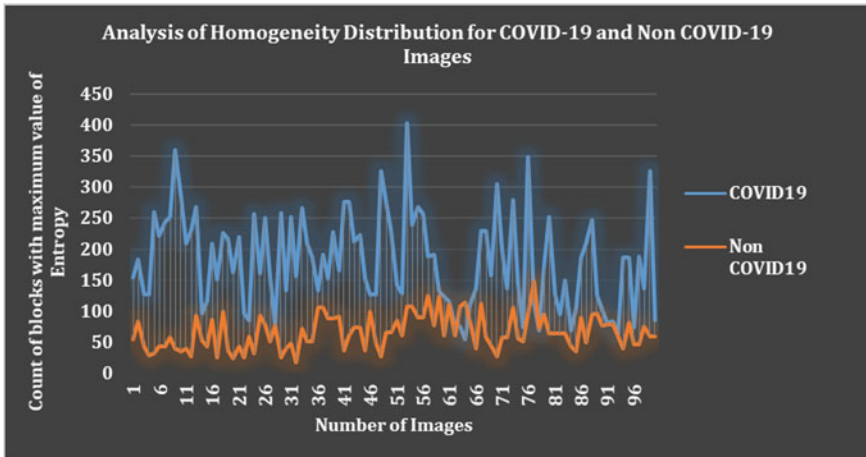


Fig. 16 Homogeneity distribution for COVID-19 and non-COVID-19 images

COVID-19 affected images. Figure 18, displays the output of variance feature map for normal and COVID-19 affected subjects.

It can be subjectively inferred from Fig. 18 that regions inside lungs are too shallow in normal subjects. Fig. 18a, i, m and q, and for COVID-19 affected images Fig. 18c, g, k and l regions have fog/hazy like regions. On comparing the same with variance feature map of GLCM texture, there is a wider range of values in lungs region, i.e. the distribution is more in normal subjects. Fig. 18b, f, j and r, and in case of COVID-19 affected lung region Fig. 18d, l, p and t, there is a narrow range of values which can be interpreted from the palette colour maps.

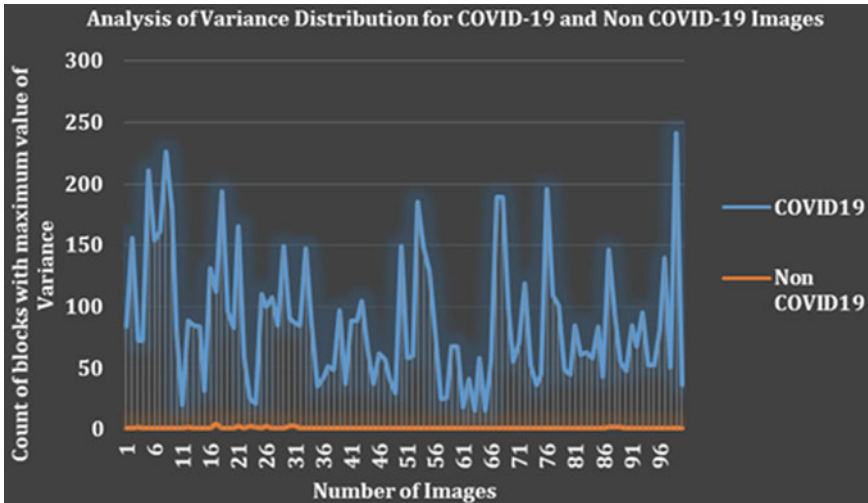


Fig.17 Variance distribution for COVID-19 and non-COVID-19 images

4.4 Interpretation of Results Using Heat Maps

GRAD-CAM algorithm is used to generate the heat maps of COVID-19 and non-COVID-19 images and interpret the images of both the classes and decisions of a model. GRAD-CAM employed VGG19 model to generate the heat maps. Figure 19 depicts the heat maps of non COVID-19 images, and Fig. 20 depicts the heat maps of COVID-19 images.

5 Conclusion

In this work, deep transfer learning models are employed to categorise the chest X-ray images into COVID-19 and non-COVID-19 classes and achieved the good accuracy for all the models. VGG19 and DenseNet121 models achieved 98.80% and ResNet50 achieved the accuracy of 97.65%. The comprehensive study has been conducted and proved that GLCM variance feature gives the better information to distinguish the COVID-19 and non-COVID-19 images, and variance feature is used to interpret these images. GRAD-CAM algorithm applied to VGG19 model to interpret the decisions of the network model. The way forward of this research includes the segmenting the lung region from chest X-rays and removing other artefact such as text and medical device traces on chest X-rays.

Table 4 Statistics of GLCM feature for COVID-19 and non COVID-19 images

GLCM feature	Number of blocks of 7X7 non-overlapping window	Maximum feature value (100 images)	Range of number of blocks with maximum feature value (100 images for each class)
ASM	32 × 36	COVID-19 images: 256	COVID-19 images: 94–399
		Non-COVID images: 256	Non-COVID images: 17–99
Contrast	32 × 36	COVID-19 images: 7.53–245.61	COVID-19 images: 1
		Non-COVID images: 13.8–249.46939	Non-COVID images: 1
Maximum probability	32 × 36	COVID-19 images: 16	COVID-19 images: 102–399
		Non-COVID images: 16	Non-COVID images: 17–99
Minimum entropy	32 × 36	COVID-19 images: –44.3464	COVID-19 images: 54–399
		Non-COVID image: 44.3464	Non-COVID images: 17–147
Homogeneity	32 × 36	COVID-19 images: 16	COVID-19 images: 54–404
		Non-COVID images: 16	Non-COVID images: 17–148
Variance	32 × 36	COVID-19 images: 1126–2756	COVID-19 images: 15–241
		Non-COVID images: 1682–2756	Non-COVID images: 1–3

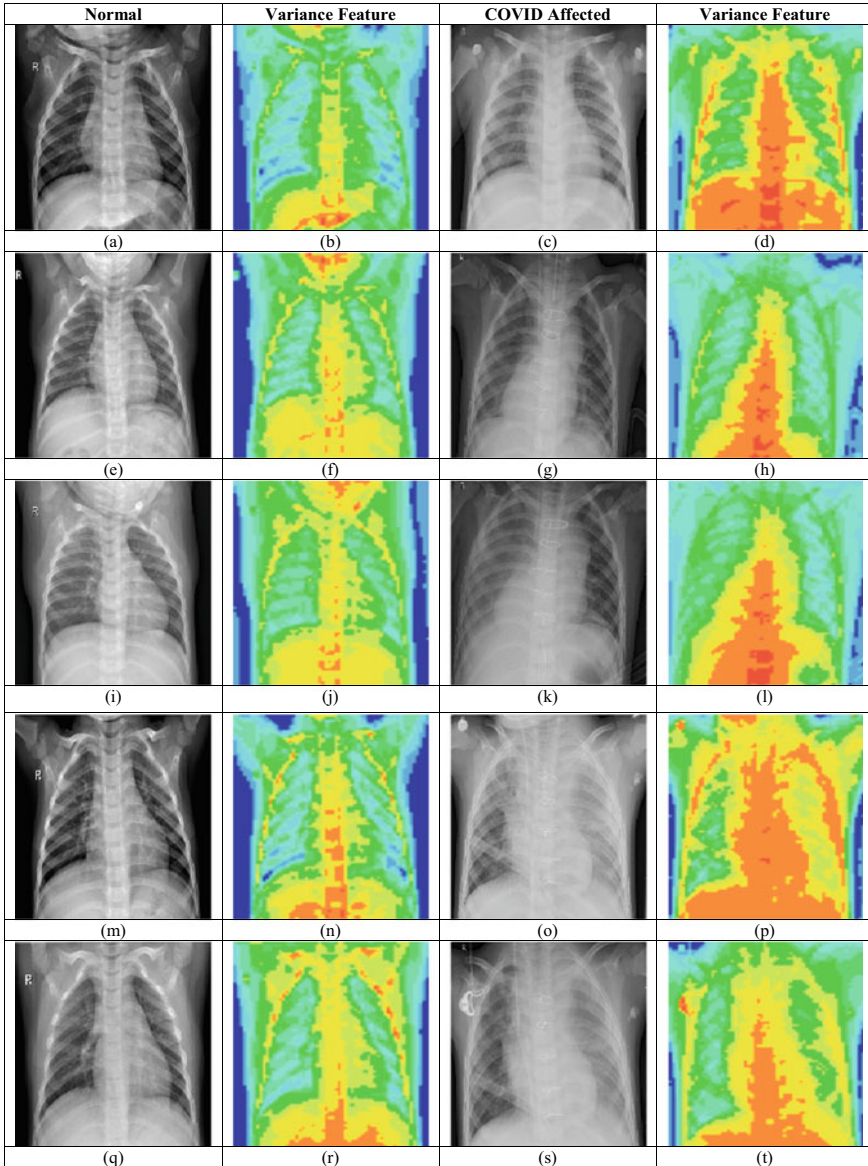


Fig. 18 Feature maps of normal and COVID-19 affected lungs. **a, e, i** and **q** represent normal lung images, and **b, f, j** and **r** represent respective variance feature maps. **c, g, k** and **l** represent COVID-19 affected lung images, and **d, l, p** and **t** represent respective variance feature maps

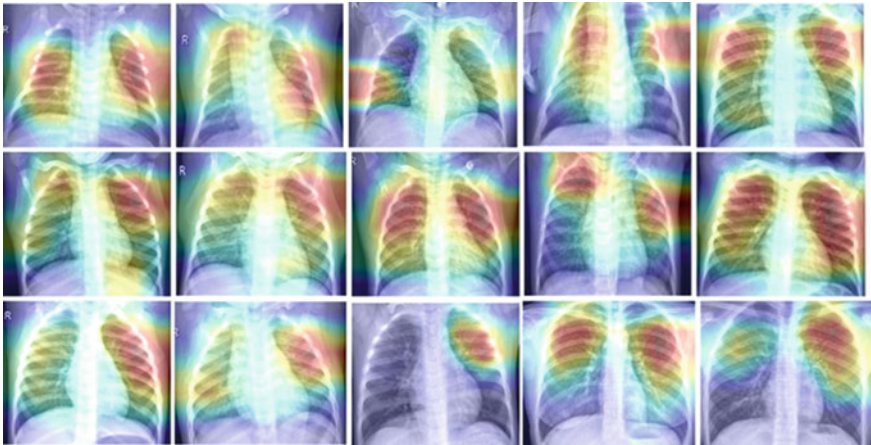


Fig. 19 Heat maps of non-COVID-19 images

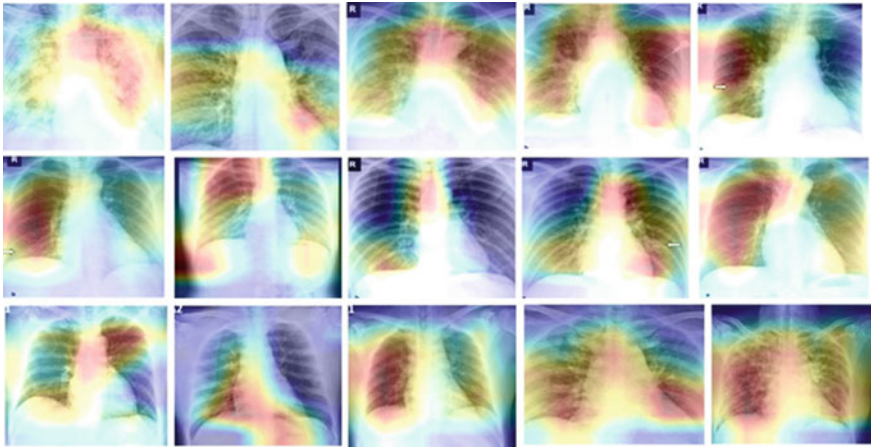


Fig. 20 Heat maps of COVID-19 images

Acknowledgements The authors express their sincere gratitude to the Visvesvaraya Technological University (VTU), Belagavi, Karnataka for their encouragement and giving the opportunity and financial support to carry out the research in the areas of COVID-19, one of the most important research area today.

The authors extend their heartfelt thanks to Prof N.R Shetty, Advisor and Dr H.C Nagaraj, Principal, Nitte Meenakshi Institute of Technology for giving constant encouragement and support to carry out the research at NMIT.

References

1. World Health Organization Novel Coronavirus (2019-nCoV) Situation report-11. 2020. https://www.who.int/docs/default-source/coronaviruse/situationreports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4. Accessed 24 Mar 2020
2. World Health Organization, Novel Coronavirus (2019-nCoV) Situation report-30. https://www.who.int/docs/default-source/coronaviruse/situationreports/20200219-sitrep-30-covid-19.pdf?sfvrsn=6e50645_2. Accessed 24 Mar 2020
3. Loey M, Smarandache F, Khalifa NEM (2020) Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry* 12:651. <https://doi.org/10.3390/sym12040651>
4. Ozturk T et al (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 121 (2020):103792. <https://doi.org/10.1016/j.combiomed.2020.103792>
5. Alom MZ, Shaifur Rahman MM, Nasrin MS, Taha TM, Asari VK COVID_MTNNet: COVID-19 detection with multi-task deep learning approaches. arXiv: 2004.03747 [eess.IV]
6. Sethy PK, Behera SK (2021) Detection of coronavirus disease (COVID-19) based on deep features. Preprints 2020030300. <https://doi.org/10.20944/preprints202003.0300.v1>
7. Ucar F, Korkmaz D (2020) COVIDiagnosis-Net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses* 140:109761. <https://doi.org/10.1016/j.mehy.2020.109761>. PMID: 32344309; PMCID: PMC7179515
8. Mei X, Lee H, Diao K et al (2020) Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med*. <https://doi.org/10.1038/s41591-020-0931-3>
9. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. 2014 arXiv preprint arXiv: 1409.1556
10. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp 770–778
11. Huang G, Liu Z, van der Maaten LWK (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*
12. Haralick RM, Shanmugan K, Dinstein I (1973) Texture features for images classification. *IEEE Trans Syst Man Cybern* 3(6):610–621
13. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2019) Grad-CAM: visual explanations from deep networks via gradient-based localization. arXiv:1610.02391v4 [cs.CV]

Automatic Symptom Extraction from Unstructured Web Data for Designing Healthcare Systems



Priyanka C. Nair, Deepa Gupta, and B. Indira Devi

1 Introduction

In the contemporary times, Internet Web pages are a major storehouse of information for a common user. With an almost ubiquitous accessibility to data on Web pages, the domain offers extensive content on medical subjects, especially based on identification of diseases. The presentation of the content, however, varies vividly across pages and has been arranged based on the page designer's best appreciation of user choices and behavior. Data is available as indexed pages on a search engine as well as non-indexed entities. Data from multiple websites need to be scanned as information related to various kinds of diseases cannot be obtained from a single website. In addition to this, the symptoms for different diseases would be different.

However, all content available as documents embody a typical structure imposed by the Web page and underlying browser technologies. This presents a scope to identify, extract and correlate content on any section of a Web page using an abstract data structure. Once identified, such data can be stored and parsed to generate a DOM tree [1]. This methodology has been widely used in data extraction in disease identification systems that can be used in clinical decision support (widely used in information extraction).

P. C. Nair · D. Gupta (✉)

Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: g_deepa@blr.amrita.edu

P. C. Nair
e-mail: v_priyanka@blr.amrita.edu

B. Indira Devi
Department of Neurosurgery, National Institute of Mental Health and Neurosciences, Bangalore, India

This paper focuses on the design details of an information extraction system primarily focused on symptom extraction from medical websites. The websites and the diseases for which the symptoms have to be extracted is an input to the proposed system. The proposed system finds the symptoms listed in those Web pages and provides it in the csv format for CDSS. In this work, a framework of data-rich region (DRR) detection and text processing has been proposed. The remainder of this paper is organized as follows: Sect. 2 discusses about the literature survey done in the field of information extraction. Section 3 discusses about the architecture and implementation details of the proposed work. Section 4 brings forth the experimental results and the conclusion, limitations and future work of the proposed work are discussed in Sect. 5.

2 Literature Survey

Medical information which is available online is humongous. The information which is available can be deemed useful for many purposes, but it requires the extraction of useful information from the data available online. Information extraction involves the conversion of unstructured data into structured data. Continuous research is ongoing in the area of information extraction in medical domain. The extraction of data from Web pages is fully dependent on the format of the Web documents [2]. The document features can be broadly categorized as text-based, tree-based and visual features. Extraction from text-based features will use approaches using strings or regular expressions. Extraction of Web pages containing html tags usually use document object model (DOM) approach. A tree-like structure is created based on DOM and its elements extracted based on its relative position to other elements.

The working of a Web crawler is mentioned in few works [3, 4]. The list of website URL called seeds is visited by the Web crawler, and it recursively finds all the links mentioned in each of the seeds. The final list of complete URL is called crawl frontier. Extracting information from those links and converting the data to tree structure is the mostly used approach. The content of interest is obtained by navigating through the tree structure. Another research has proposed a framework that can be used irrespective of the domain for information extraction from online Web forums using a tree-based approach [5]. The work focuses on the semantic representation of data and anonymization. The html page of the webpages are retrieved and a DOM (document object model) is created which helps in the extraction of useful data. The data extracted is represented as resource description framework (RDF) graphs and anonymization is done on the data which requires privacy. The framework was successfully tested for extracting medical information related to adverse drugs. A method has been proposed for structuring electronic health records and extracting medical concepts from the structured data [6]. Concept is extracted from EHR using natural language processing techniques and passed on to SNOMED-CT for further validation.

A framework is created to extract posts from various medical forum website which can be used irrespective of the structure of webpages [7]. A semantic-based approach is used for the design. DOM tree is generated from the html page. The identification of data-rich region (DRR) in the DOM tree is achieved by using semantic rules in the form of regular expressions. The tree is parsed repeatedly for identifying the root node having the maximum semantic features. The XPATH to each leaf node of DRR is determined and stored in the template database for extracting information from similar webpages. A work attempts to provide a classification framework for extracting data from online health forums [8]. The sentences extracted from online health forums were manually classified to three categories medication, symptom and background. The sentences were converted to feature vectors and used for training the model. The random forest model was chosen for classification of the sentences. The contribution of various attributes of breast cancer in classifying the instances into different classes is also being explored in [9]. Different classifiers have been applied to the data to classify into different categories. Another work focuses on extraction of multi-disease diagnosis has been achieved by a CDSS developed from severity of the symptoms [10]. Adaptive neuro-fuzzy inference system (ANFIS) has been applied in the work to diagnose the disease. The medication recommendations have been provided using a rule-based system.

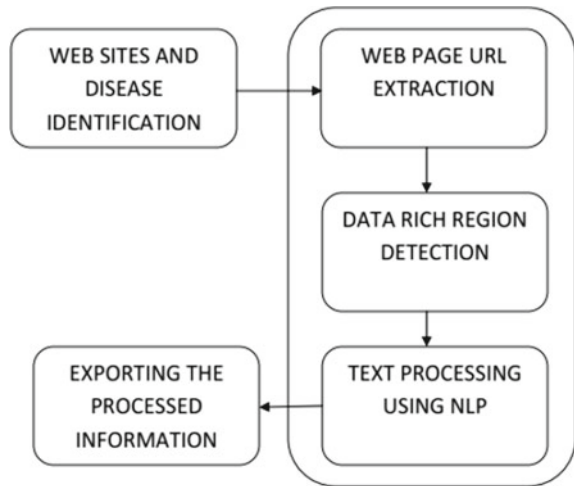
The proposed work gives a framework for symptom extraction from online websites. The work follows the general approach of converting the html pages into a tree structure and then parsing it for symptom data extraction. Preprocessing is done on the extracted to ensure data is in the required format. A novel approach for symptom validation is done with the help of SNOMED-CT data. SNOMED-CT which is a nomenclature for medical terms is used for symptom validation. The next section explains the various components involved in the proposed work.

3 The Proposed System for Symptom Extraction

The proposed system is implemented for extracting symptoms from the Web pages for the set of predefined diseases. The aim of extracting symptoms is to use these information in developing a clinical decision support system. The symptoms extracted will be a template for data procurement from doctors for patient data collection.

The high-level architecture of the proposed system is as depicted in the Fig. 1. The architecture is divided into five major components are Web site and disease identification, Web page URL extraction, data-rich region detection, NLP text processing and data storage. Each of the components has been explained in the following subsections.

Fig. 1 High-level architecture of the proposed system



3.1 Web Site and Disease Identification

Identifying the popular medical websites which forms a basis for the extraction is very important. The websites need to be authentic and are identified based on various existing research works. Websites like Mayo Clinic, WebMD etc., are a few which are selected in this work. The diseases whose symptoms have to be extracted have to be chosen. Most of the commonly occurring diseases, like gastritis, malaria, dengue etc., are identified for the work.

3.2 Web Pages URL Extraction

The URL of the identified websites from Sect. 3.1 is the target for the proposed works data extraction. The diseases which are identified from Sect. 3.1 are considered as the keywords which have to be searched on the defined websites. The flow of execution is depicted in Fig. 2. Any symptom-related information from these websites will be extracted. The input to this phase is the website names. Finding the URL from the websites is achieved through packages specifically designed for URL extraction. Extraction of HTML contents of the website has been performed for all the disease names that have been identified.

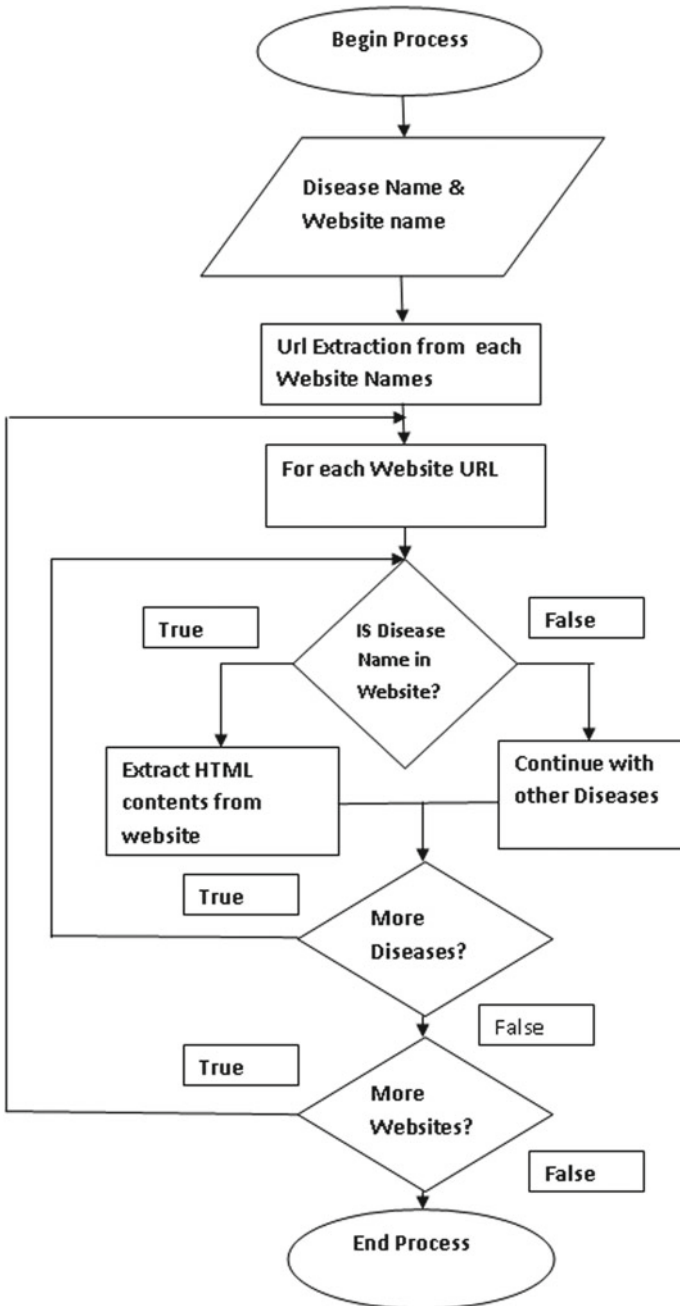


Fig. 2 Flowchart of the html extraction

3.3 Data-Rich Region Detection

The html Web pages has a body section whose data has been extracted. In the proposed work, the data-rich region (DRR) refers to the part of html page where symptom-related information for the diseases identified exists. In order to identify the DRR, initially the html data has been converted to a tree like structure for the easy.

An invalid tag list is created to include those tags which are insignificant and need not be part of the tree structure. A root node is initially created which will be modified to add nodes as they are created. The valid tags are parsed. For each tag, it is appended to its parent node along with the data in the tag. The tag which has text with symptom key word in it is saved separately to mark the starting of any kind of information related to the symptoms (Fig. 3).

The structure of the tree format is as shown in Fig. 4. The next step is to identify the data-rich region. For identifying the data-rich region, the symptom node which was saved in the tree creation phase is used. An in order traversal from the saved node is done to find an unordered list tag. All the information under the unordered list tag is considered as the candidate for DRR. The unordered list tag is chosen as the target. The data extracted from the DRR is stored in an excel.

3.4 Text Processing Using NLP

The excel sheet which was populated from the phase mentioned in Sect. 3.3 has to be preprocessed [11, 12] for getting results in the desired form. The symptom text mentioned in each of the row of excel sheet is the input to the preprocessing stage. The architecture of the text processing is as shown in Fig. 5. The textual data is converted into tokens and the stop words are removed. All the synonyms for the tokens are also considered. SNOMED-CT which is a nomenclature for medical terms is integrated to the system and saved in database [13]. The symptom tokens are validated against the symptoms extracted from SNOMED-CT. This is to ensure that only valid medical terms are selected as symptoms. The selected symptoms corresponding to diseases are populated to excel sheet.

4 Experimental Results and Discussion

The implementation of the framework is done in Python language. The results indicate that this framework can be adopted for any symptom extraction from any websites. (nlp toolkit/csv format). The diseases selected for the proposed framework are as mentioned in Table 1. The diseases were selected based on expert's suggestion. Initially, the websites chosen for symptom extraction were the popular

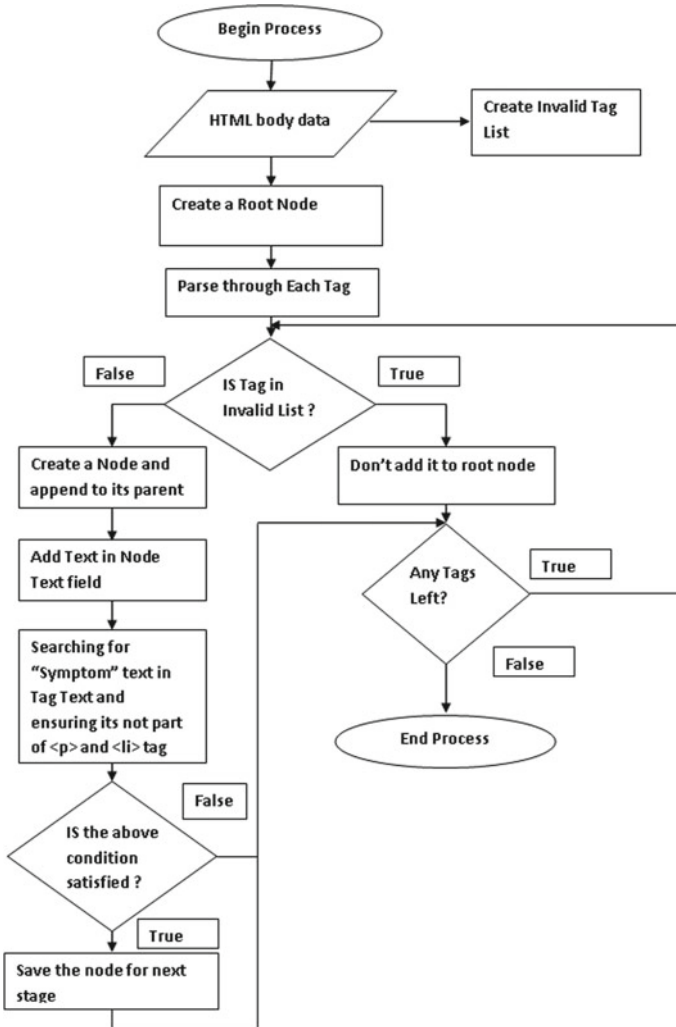


Fig. 3 Tree structure creation from html

medical websites like mayoclinic, medicalnewstoday, webmd, medlineplus, healthline, msdmanuals, patient.info etc. URLs for the websites are retrieved after the Web pages URL extraction phase. The html part of the websites are then extracted and stored in tree structure as mentioned in Sect. 3.3. The snippet of excel output of this phase is as mentioned in Fig. 6. After validating and selecting only valid medical terms identified from SNOMED-CT, the final excel data appears as shown in the Fig. 7. The final excel output data has columns for each disease modeled in the system. Each column has the symptoms extracted from the website with frequency

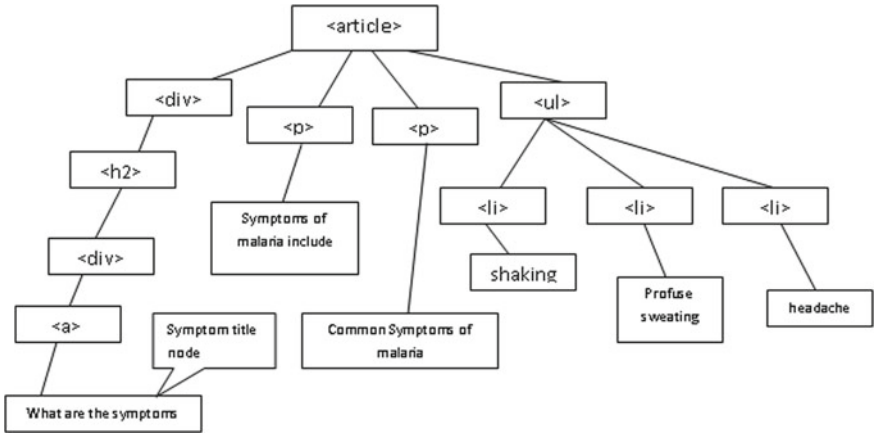


Fig. 4 Tree format of the proposed system

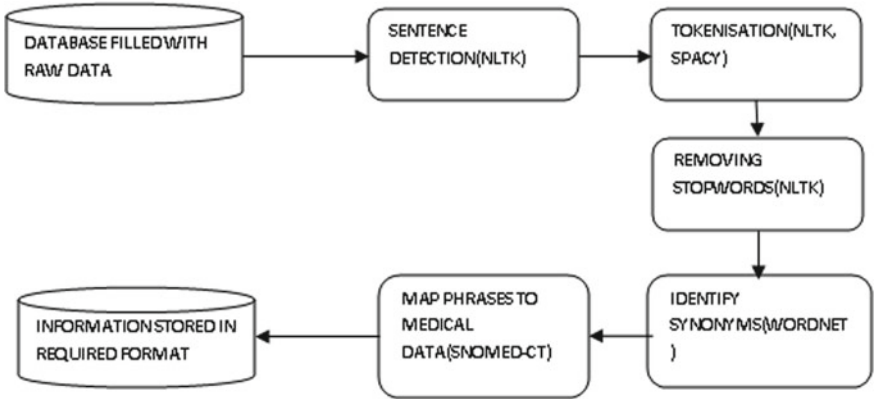


Fig. 5 Symptom text preprocessing architecture of the proposed system

Table 1 Disease table for the proposed method

Disease name	Disease type
Malaria, typhoid, dengue	Fever related
Pneumonia, tuberculosis, bronchitis	Chest infection related
Conjunctivitis, glaucoma	Eye related
Gastritis	Stomach related
Leptospirosis	Bacterial infection
Measles, jaundice	Skin related

malaria	typhoid	dengue	Pneumonia
mayoclinic Ad Choices	mayoclinic Weakness and fatigu	mayoclinic Ac	mayoclinic Ac
mayoclinic Opportunities	mayoclinic Become delirious	mayoclinic M	mayoclinic Lc
mayoclinic High fever	mayoclinic Extremely swollen at	mayoclinic Irr	mayoclinic Co
mayoclinic Sweating	mayoclinic Ad Choices	mayoclinic Bl	mayoclinic Oj
mayoclinic Policy	mayoclinic Abdominal pain	mayoclinic Oj	mayoclinic Pc
mayoclinic Headache	mayoclinic Opportunities	mayoclinic Cc	mayoclinic Co
mayoclinic Vomiting	mayoclinic Rash	mayoclinic Rr	mayoclinic St
mayoclinic Moderate to severe s	mayoclinic Muscle aches	mayoclinic St	mayoclinic Na
mayoclinic Diarrhea	mayoclinic Diarrhea or constipat	mayoclinic Si	mayoclinic Cl
medicalnewstoday fever, headac	mayoclinic Fever that starts low	mayoclinic Pt	mayoclinic Fa
medicalnewstoday a sensation c	mayoclinic Sweating	mayoclinic Bl	mayoclinic Fe
medicalnewstoday sweats, follo	mayoclinic Policy	mayoclinic Pt	medicalnewst
medicalnewstoday seizures son	mayoclinic Lie motionless and e	mayoclinic Bl	medicalnewst
webmd P. falciparum. This is the	mayoclinic Headache	mayoclinic Na	medicalnewst
webmd Kidney failure	mayoclinic Dry cough	mayoclinic Hc	medicalnewst
webmd People who live in areas	mayoclinic Loss of appetite and	mayoclinic Vc	medicalnewst
webmd Being very tired (fatigue)	medicalnewstoday Untreated, it	mayoclinic Di	medicalnewst

Fig. 6 Data-rich region excel output

malaria	typhoid	dengue	Pneumonia	Tuberculosis	Bronchitis
7 - headache	6 - fever	6 - vomiting	5 - fever	6 - fever	6 - fever
7 - fever	5 - fatigue an	4 - rash	5 - chills	5 - weight los	5 - depression
7 - chills	4 - headache	4 - muscle pa	4 - vomiting	5 - fatigue an	5 - cough
5 - vomiting	4 - diarrhea	4 - joint pain	4 - fatigue an	5 - cough	4 - fatigue an
4 - nausea	4 - depression	4 - headache	4 - depression	4 - loss of ap	4 - burn
4 - jaundice	3 - rash	4 - fever	4 - cough	4 - chest pain	3 - wheezing
4 - fatigue and tiredness	2 - weakness	3 - peeling sk	4 - chest pain	4 - breast pai	3 - chills
4 - diarrhea	2 - tic	3 - nausea	4 - breast pai	3 - vomiting	3 - asthma
4 - anemia	2 - nausea	3 - bleeding g	3 - shortness	3 - peeling sk	2 - tic
3 - vomiting blood	2 - delirium	2 - vomiting b	3 - headache	3 - chills	2 - shortness
3 - palpitations	2 - constipati	2 - tic	3 - confusion	2 - sprain	2 - flu
3 - depression	2 - confusion	2 - fatigue an	2 - vomiting b	2 - burn	2 - cold
2 - seizures	2 - chills	2 - burn	2 - nausea	1 - wheezing	1 - vomiting b

Fig. 7 Snippet of disease versus symptom final output

count of the symptoms in all crawled website. The symptoms which are extracted is also validated with the help of experts.

5 Conclusion and Future Work

The work proposes a symptom extraction framework for extracting symptoms related to diseases from medical websites. This framework has incorporated tree construction from html data, identifying DRR and NLP techniques for processing. The novel

idea in this includes SNOMED-CT validation for choosing the symptoms related to disease. The symptoms extracted are also validated with the help of experts (2–3).

The future work includes the design of a clinical decision support system which is used for predicting the probable diseases based on symptoms. The symptoms extracted using this framework will be used for modeling.

References

1. Negm N, Elkafrawy P, Salem BA (2012) A survey of web information extraction tools. *Int J Comput Appl* 43(7):19–27. <https://doi.org/10.5120/6115-8296>
2. Dastidar BG, Banerjee D, Sengupta S (2016) An intelligent survey of personalized information retrieval using web scraper. *IJ Educ Manage Eng* 5:24–31. <https://doi.org/10.5815/ijeme.2016.05.03>
3. Varlamov MI, Turdakov DY (2016) A survey of methods for the extraction of information from web resources. *Program Comput Soft* 42(5):279–291. <https://doi.org/10.1134/S0361768816050078>
4. Naveen P, Nair PC, Gupta D (2020) Predicting the degree of emotional support in an online health forum for HIV using data mining techniques, emerging trends in electrical, communications, and information technologies. Springer, Singapore, pp 81–94
5. Mahto DK, Singh L (2016) A dive into web scraper world. In: 3rd International conference on computing for sustainable global development (INDIACom). pp 689–693
6. Audeh B, Beigbeder M, Zimmermann A, Jaillon P, Bousquet CY (2017) Vigi4Med scraper: a framework for web forum structured data extraction and semantic representation. *PLoS ONE* 12(1):e0169658. <https://doi.org/10.1371/journal.pone.0169658>
7. Barbantan I, Potolea R (2015) Knowledge extraction and prediction from unstructured medical documents. In: *ICT innovations*
8. Kumaresan U, Ramanujam K (2018) Automated scraping of structured data records from health discussion forums using semantic analysis. *Inform Med Unlocked* 10(2018). <https://doi.org/10.1016/j.imu.2018.01.003>
9. Shastri SS, Nair PC, Gupta D, Nayar RC, Rao R, Ram A (2018) Breast cancer diagnosis and prognosis using machine learning techniques. In: *The international symposium on intelligent systems technologies and applications*. Springer, Cham, pp. 327–344
10. Tandra S, Gupta D, Amudha J, Sharma K (2019) A fuzzy-neuro-based clinical decision support system for disease diagnosis using symptom severity. In: *International conference on soft computing and signal processing*. Springer, Singapore, pp 81–98
11. Nair PC, Gupta D, Devi BI, Bhat NR (2019) Automated clinical concept-value pair extraction from discharge summary of pituitary adenoma patients. In: 9th International conference on advances in computing and communication (ICACC). IEEE, pp 258–264
12. Tandra S, Nautiyal A, Gupta D An efficient text labeling framework using active learning model. In: *Intelligent systems, technologies and applications. advances in intelligent systems and computing*, vol 1148. Springer, Singapore
13. Nair PC, Gupta D, Devi BI A survey of text mining approaches, techniques, and tools on discharge summaries. In: *Advances in computational intelligence and communication technology*. Springer, Singapore, pp 331–348

Bug Triaging: Right Developer Recommendation for Bug Resolution Using Data Mining Technique



B. H. Chaitra and K. S. Swarnalatha

1 Introduction

The way toward removing valuable data through data investigation is called data mining. It is otherwise called knowledge discovery. Utilising data mining techniques, we can decrease costs, increase incomes or both. There are two types of data, *clear cut* and *numerical*, utilized for mining reason like whole number, fraction, cost, burn, varchar2 and so on. Data mining is difficult on information that is not numerical or clear cut. The big business information found is mostly non-numerical and non-clear cut. For the achievement of business, extricating information from this unstructured data will be hard; accordingly, it is handled utilizing content mining methods with the goal that it tends to be prepared by data mining calculations furthermore, methods.

Bugs are the programming blunders that reason critical execution debasement. Bugs lead to poor client experience and low-framework throughput. Vast open-source programming improvement undertakings, for example, Mozilla and Eclipse get many bug reports. They as a rule utilize a bug following framework where clients can report their issues which happened in their particular ventures. Every approaching bug report should be triaged. Choosing the most suitable designer to fix another bug report is one of the most critical stages in the bug triaging procedure, and it has a noteworthy impact in diminishing the time taken for the bug fixing process and the expense of the undertakings.

In this paper, naïve Baye's text classifier is used to classify the bugs and assign it to the right developer. This classifier has inbuilt functions to select, extract features

B. H. Chaitra · K. S. Swarnalatha (✉)

Department of Information Science and Engineering, NITTE Meenakshi Institute of Technology, Bengaluru, India

e-mail: swarnalatha.ks@nmit.ac.in

B. H. Chaitra

e-mail: chaitrabh@rvce.edu.in

and classify them. It also pre-processes the data to remove the duplicates. It includes a model for pre-processing called as “Bag of Words”. The model includes removal of stop-words, tokenization and stemming. This classifier includes inbuilt prediction algorithm, and it classifies and randomly assigns the bugs to the developer. Clusters consisting of the developers’ details such as experienced backend developer, experienced UI developer, fresher backend developer and fresher UI developer are included. The classifier classifies the bug based on its severity and priority and checks in the cluster of the suitable developer and assigns it.

2 Bug Triage

Fixing bug reports through the conventional bug triage framework is very tedious and furthermore forces extra expense on the undertaking. For instance, Eclipse has 239 dynamic designers as on January 2011 and 282 adjusted documents on the Eclipse stage venture. Henceforth, numerous explores have been done to make the customary bug task effective and programmed. The triager peruses another bug report, settles on a choice about the bug, and after that chooses the most proper engineer who can resolve the bug. One of the imperative reasons why bug triaging is such a protracted procedure is the trouble in choice of the most able designer for the bug kind. The bug triager, the individual who doles out the bug to a designer, must be mindful of the exercises (or intrigue territories) of the considerable number of designers in the venture. Bug triaging regularly takes two months to resolve a bug. On the off chance that the engineer, to whom the bug report is allotted, could not resolve it, it is appointed to another engineer. This would devour both time and cash. In this way, it is extremely critical on part of bug triager to appoint the bug report to an engineer who could effectively fix the bug without need of any hurling. Consequently, the activity of bug triager is extremely urgent. Bug management and bug fixing is a difficult undertaking in programming advancement. Choosing the perfect individual for the correct activity is the most vital assignment. In conventional framework new bugs are actually relegated to a designer by a specialist supervisor, i.e. human specialists who are costly prompting mistake. A straight forward undertaking ought not to be doled out to a high evaluation engineer as the cost factor will increment. Utilizing information mining strategy, the proposed system will oversee bug characteristic dataset and designer quality dataset utilizing data classification and ontology idea so a bug is allocated to the correct engineer.

3 Triage Process

Triage is a term used in software testing to describe the priority and severity of defects. This is a technique where each bug is prioritised depending up on its severity, risk and frequency. Triage is a French action term trier, which means to sort, isolate,

filter or select. The term triage was utilised amid World War I by French specialists treating the war zone injured at the guide stations behind the front. Those in charge of the expulsion of the injured from a front line or their consideration a while later would separate the unfortunate casualties into three classes: the individuals who are probably going to live, paying little mind to what care they get; the individuals who are probably not going to live, paying little mind to what care they get; those who prompt consideration may have a beneficial outcome in result.

4 Bayesian Classifiers

Bayesian classifier, be contingent on the possibility that the job of a characteristic class, is to anticipate the assessments of highlights for individuals from that class. Models are assembled in classes since they have regular qualities for the highlights. These classes are often called regular sorts. In this segment, the objective element compares to discrete class, which is not really parallel. The thought behind Bayesian classifier is that, if a specialist knows the class, it can foresee the estimations of different highlights. On the off chance that it does not know the class, Bayes' standard can be utilized to anticipate the class given (a portion of) the element esteems. In this Bayesian classifier, the learning specialist manufactures a probabilistic model of the highlights and employs that model to anticipate the characterization of an additional precedent.

5 Naïve Bayes Classifier

A mechanized procedure of discovering some metadata about an archive is termed as text characterization. It is utilized in different territories like record ordering by proposing its classes in a substance the board framework, spam sifting, automatically help work area demands arranging and so on. The naïve Bayes text classifier is utilised in this examination for bug order. Innocent Bayes is a probabilistic classifier, which depends on Bayes' hypothesis with maverick presumption. Gullible Bayes classifiers are direct classifiers that are known for being straightforward yet exceptionally proficient.

6 Working of Bug Life Cycle

Working of the bug life cycle is as shown in the Fig. 1.

- i .Quality analyst finds the deformity.
- ii .Status is set to new.

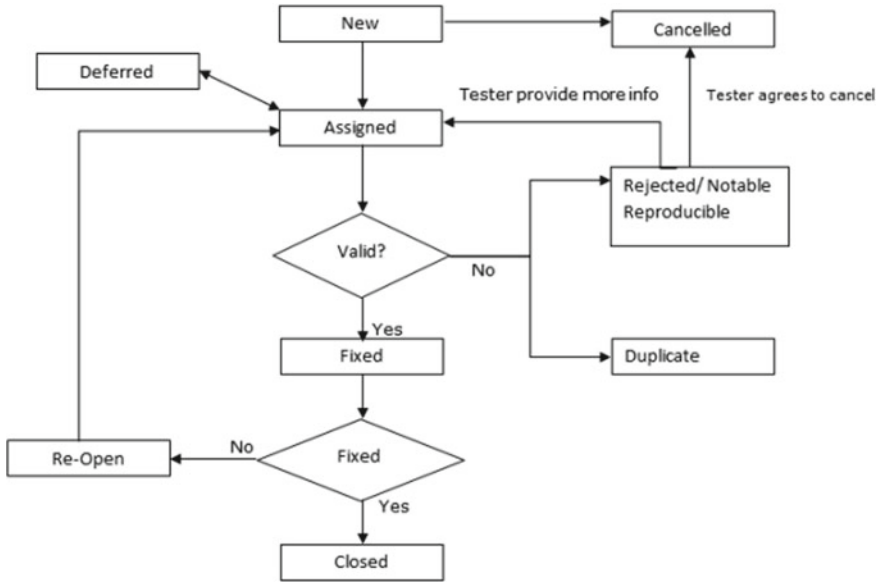


Fig. 1 Working of a bug life cycle

- iii .The project manager is then asked to break down the deformity sent.
- iv .If the deformity is legitimate, only then the project manager selects it.
- v .If the deformity is not substantial, the status is given “Rejected”.
- vi .The project supervisor shares out a status rejected. If the deformity is not rejected, the next step is to check whether it is in degree. Let us consider, there is capacity email usefulness for a similar application, and an issue is discovered with that. However, it is anything but a piece of the present discharge when such deformities are referred as a delayed or conceded status.
- vii .The project lead then checks if a comparative deformity is raised earlier. If yes, the deformity will be shared out as a status copy.
- viii .If the deformity is not raised earlier then an engineer is appointed to begin fixing the code, during which the deformity is allotted a status in-advance.
- ix .A deformity will be appointed as status fixed, once the code is fixed.
- x .The analyser will then re-test the code. The deformity will be closed if the test case passes. In case the experiment fails again, the deformity is re-opened and shared out to the designer.

7 Objective of This Paper

The key objective of this paper is to reduce the man power towards the triage meeting and also to assign a right developer to handle and fix the bug based on the severity and priority of the bug. The right developer refers to the person who is skilled and

experienced based on the bug requirement. As stated, handling a large number of bugs in the repository is a tedious job. The developer who is highly experienced in a particular domain gets a bug with a lowest severity and the developer, who is a fresher gets a highest severity bug, then the expenses on the experienced resource becomes expensive and on the other hand the fresher will neither be able to handle the bug nor he might take a very long time which a company cannot afford. Main objective of bug triage is to assess, organize and appoint the goals of deformities. The concept of bug triage through a manual triage meeting, works in a small company who work on simple projects. But it will be very tedious and takes longer duration in huge firms where the bug repository will be huge, and time is wasted in the meeting data about conference proceedings published by Springer is now being made available as LOD, for an ever increasing number of conferences. A detailed description is available at <http://lod.springer.com/wiki/bin/view/Linked+Open+Data/About>.

8 Scope of This Paper

The extent of proposed fill in as pursues:

- i .Allotting need levels to the new bug report.
- ii .Appointing positioning to the anticipated rundown of engineers.
- iii .For illustration, determination can expel uninformative bug reports.
- iv .Keyword determination can evacuate uninformative words. Watchword determination improves the precision of bug triage.

9 Literature Survey

- Michel V. Godfrey et al. proposed an instance of the careful structure that normally facilitates the bug task. To activate data about the fashioner's ability, the vector space model was utilized from the historical backdrop of the starting late fixed bugs. The vector show is utilized to recover the outline as well as the title from the report to build up a vector which can later be utilized to discover comparative reports by mining the information in the bug store. So as to impact an incredible bug triage to delineate, the creators facilitated study wherein they amassed an examination from the organizers as to past bug fixing learning, their fulfilment with the bug task, paying little personality to whether they are profitable in addition, beyond any doubt about overseeing bugs as of now, and so forth. The general data had given them the major assessments to propose this show [1].
- Hemant Josh et al. exhibited a bug expectation calculation, the reason for which to foresee the quantity of bugs which were to be recognized and revealed in every month. So the bug forecast of any month fundamentally relies on the bug check of forerunner month. This expectation is accomplished through the forecast algorithm actualize in separate paper [2].

- Lei Xu, Lian Yu et al. have advanced the investigate methodology affiliation instructions which determines the connection amid bug fixing arrangements plus bug types. The improper portions of the source code is familiarised by the troubleshoot system. It is exceptionally simple for the designers to rectify if the mistakes are discovered. The decided affiliation rule helps to anticipate documents that normally change together, for example, capacities or factors [3].
- Nicholas Jalbert et al. instantiated a model that naturally displays whether an arriving bug report was unique or copy of an officially in effect report. It spares the engineer's time and endeavours. To anticipate bug duplication, simple framework was utilized which included simple strategies, for example, literary semantics, grouping in charts and surface highlights the point of reference model and test it inside a thing gathering tackling the assistance works out. A programmed bug triage approach was proposed in order to maintain a strategic distance from the costly expense of manual bug triage, which applies content order methods to predict engineers for bug reports. Methodology includes a bug report which is mapped to a report and a related engineer is mapped to the name of the report. During this point, the bug triage was reformed into an issue of content characterization and is consequently comprehended with developed content characterization procedures, for example, naïve Bayes. The after effect of content grouping involves a human triager sharing out new bugs by fusing his/her mastery. For an anticipated rundown of engineers by a classifier, in Engineer Prioritization designers were positioned by the needs. Along these lines, the engineer prioritization is utilized to segregate the engineers with comparative probabilities in the forecast. While in profile-oriented developer recommendation, a methodology where profile was made for every designer in light of their past work. This profile is mapped to a space mapping grid which showed the skill of every designer in their comparing territory [4].

10 Algorithm

Naïve Bayes Text Classifier

A basic probabilistic classifier, the Naïve Bayes classifier [5], depends on Bayes hypothesis with solid and guileless autonomy presumptions. It is one of the most fundamental content arrangement strategies with different applications in email spam discovery, individual email arranging, record order, explicitly unequivocal substance location, and language identification and assumption recognition.

Regardless of the gullible plan and distorted suppositions that this method utilizes, naïve Bayes performs well in numerous mind-boggling genuine issues. Despite the fact that it is frequently beat by different systems, for instance, helped trees, arbitrary woods, max entropy, support vector machines and so forth, naïve Bayes classifier is proficient since it is less computationally serious (in both CPU and memory) and it requires a little measure of preparing information. In addition, the preparation time with naïve Bayes is essentially littler instead of elective strategies. Naïve Bayes

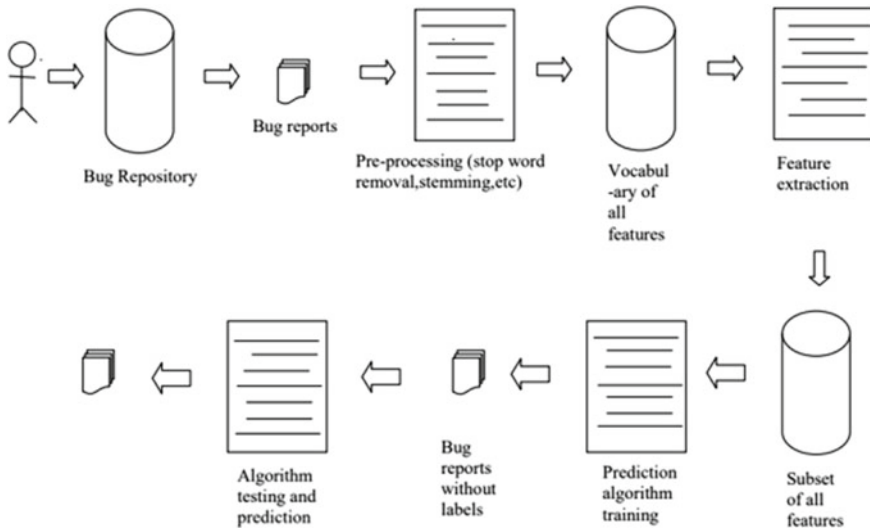


Fig. 2 Bug classification system

classifier is better as far as CPU and memory utilization is concerned. The bug classification system is as shown in Fig. 2.

11 Bag-of-Words Model (BoW)

Machine learning algorithm standouts amongst the most imperative sub-undertakings in example order include extraction and determination; the three fundamental criteria of good highlights are recorded beneath:

- **Salient:** The highlights are essential and important regarding the issue space.
- **Invariant:** Invariance is regularly portrayed in setting of picture arrangement. The highlights are invulnerable to bending, scaling, introduction, and so on. A decent precedent is given by C. Yao and others in Rotation-Invariant Features for Multi-Oriented Text Detection in Natural Images [6].
- **Discriminatory:** The chose highlights bear enough data to recognize well between examples when used to prepare the classifier.

Before fitting the model and utilizing AI calculations for preparing, consider as to how to best speak to a content record as an element vector. A normally utilized model in natural language processing is the so-called sack of words. The thought behind this model truly is as basic as it sounds. Initially, the production of the vocabulary is considered as—the gathering of every unique word that happen in the training set and each word is related with a check of how it happens.

Table 1 Bag of words representation of two sample documents D1D1 and D2D2

	hi	hello	how	U
XD1	1	1	0	0
XD2	1	1	1	1
Σ	2	2	1	1

This vocabulary can be comprehended as a lot of non-excess things where the request does not make a difference. Let D1 and D2 be two records in a training dataset:

D1: hi hello

D2: hi hello how u

Based on the above two records, vocabulary can be written as:

$$V = \{hi : 2, hello : 2, how : 1, u : 1\}$$

The d-dimensional element vectors for the individual reports where the dimensionality is equivalent to the quantity of various words in the vocabulary ($d = |V|$) could be developed utilizing the vocabulary. This is termed as vectorization.

Table 1 includes a model in which one analysis is whether the 0s of the element vectors are twofold tallies (1 if the word happens in a specific report, 0 generally) or supreme checks (how regularly the word happens in each record).

12 Implementation

There are mainly three modules: *Admin*, *Developer* and *Quality Analyst*.

Module 1—Admin

The admin has all the rights and permissions to access the tool. The admin adds the employee details into employee database and maintains. Admin also maintains the Bug Repository/Bug Dataset. When admin adds a new employee details, an e-mail will be sent to the respective employee who consists of a unique password for the employee to login. Only Admin has the right to edit or delete data from the employee database. Admin has an option to pre-process the bug data in the bug dataset.

Module 2—Developer

The developer is the one to whom the bug is assigned. Developer has the option to edit and delete the bug. In this the developer will fix the bug and change the status of the bug. Once the developer receives the bug, he/she opens it and checks. If he/she finds it critical one, then he/she will fix it immediately. At this time, the status of the bug will be “open”. If the developer finds it major or minor, he/she will change

the status either as “pending” or “not fixed”. If the developer changes the status as pending or not fixed then, the reason of the status must be given. Once the bug is assigned, as it will be assigned randomly, they will not know who has developed the part where the bug has been found. So, in order to resolve this problem, there is a chatroom for the developers where they can chat with the other developers and get the proper code from the developer who has actually developed and fix it.

Module 3—Quality Analyst

The quality analyst is the one who adds the bug and sends it to the developer. quality analyst adds the new bug by entering the product in which the bug has occurred, component of the bug i.e., whether it is a web issue or a compatibility issue, bug summary-complete description of the bug, priority of the bug whether it is highest, high, normal, low, severity of the bug whether it is critical, major, minor then, submit. As soon as the bug is added, it will be assigned to the right developer based on his/her skills and experience. Quality analyst can also edit or delete the bugs from the bug list. All the three modules have dashboard in common, which gives the information updated/assigned bug and count of critical, major, normal and minor bugs.

13 Future Enhancements

The major task that has to be addressed in the near future is strategic developers. The knowledge of how the system assigns bug fixing tasks and manipulates task assignment must be understood by the developers. Thus, the assignment of bugs is a manipulation-free and a fair process. Automatic bug triaging for bug fixation reduces the human effort, cost and efficiency, but it can be made even more effective. The developed code can be linked with this tool so that, the developer can retrieve the code, make changes and save it and intimate other developers.

14 Conclusion

Bug triaging is a process where the bugs are assigned to the developers in order to fix them. This paper is all about assigning bugs to the right developer based on his/her experience and skill set. Initially, the bug report or a new bug raise in an application is reported in bug dataset by the tester. By using naïve Bayes text classifier, the bug data is pre-processed in order to remove stop words, tokenization and stemming. These three operations are carried out under the model “Bag of Words”. This classifier has an inbuilt feature to extract the features from the bug dataset. These extracted features are compared with the features present in the clusters. These clusters contain the developer details such as experienced backend developer and the respective skillset, fresher backend developer and the respective skillset, experienced UI developer and the respective skillset and fresher UI developer and respective skillset. The extracted

features of the bug data set is compared with the features of the clusters to check which developer will be suitable for the bug fixation based on bug severity, priority, developer's experience and skillset. If there are developers with same experience and same skillset then, the bug is assigned randomly.

To create clusters, PostgreSQL has a built-in package to create clusters automatically. This tool is almost similar to SQL, on which the data is entered and the respective package to create clusters is imported. Once these packages are imported, the data of the developers is split and it itself forms a cluster.

Naïve Bayes classifier thoroughly carries out the operations stated above automatically. And it also includes prediction and classification algorithms. Naïve Bayes text classifier is extremely easy to implement and understand the concept. It is also flexible and efficient enough to use. This paper assists to reduce the human work and the expenses in the companies that they have to invest more on the human resource for the small work. The study helps to avoid the manual work or the manually conducted triage meetings in the companies which in turn saves the human effort, time and cost.

References

1. Hotho A, Nürnberger A, Paaß G (2005) A brief survey of text mining. *GLDV J Comput Linguist Lang Technol* 20:19–62
2. Hassan AE (2008) The road ahead for mining software repositories. *IEEE Comput Soc* 48–57
3. Diehl S, Gall HC, Hassan AE (2009) Special issue on mining software repositories. *Empirical Softw Eng Int J*
4. Michael OB, Robin GC (2009) A bug you like: a framework for automated assignment of bugs. In: *IEEE 17th international conference*
5. Javed MY, Mohsin H (2012) An automated approach for software bug classification. In: *IEEE sixth international conference on complex, intelligent, and software intensive systems*
6. Yao C, Zhang X, Bai X, Liu W, Ma Y (2013) Rotation-invariant features for multi-oriented text detection in natural images. *PLoS ONE*

Analysing Machine Learning Techniques in Python for the Prediction of Diabetes Using the Risk Factors as Parameters



M. S. Akanksha, Kolachana Vinutna, and M. N. Thippeswamy

1 Introduction

Machine learning is a branch of artificial intelligence, or AI, that helps build intelligent software to enable machines to perform their jobs in a skilful manner of their own. The statistical learning methods and modelling provide much-needed support for the intelligent software that is used to develop machine intelligence [1]. Machine learning constitutes various models that have been defined for formulating the large datasets to obtain a prediction, description, prescription or cognition on the data. With the increase in demand in using these models for various fields like medical, business, politics and education, the importance of ML is increasing exponentially, and its grip is gaining momentum among technologists. There are three types of ML—supervised, unsupervised and reinforcement learning. Supervised learning involves some prior information on the dataset being fed into the model before it does the prediction. Unsupervised learning is a little more complex and predicts the outcome without prior information being fed to it. Reinforcement learning involves many decision-making outcomes in a sequence to obtain a final result, more like a game environment.

Machine learning, being a part of artificial intelligence, often intermingles with other fields like data mining, data analytics, deep learning and computation. There is often a need to rely on these other fields and their analysis on the data, before applying an ML algorithm. In the project, the aim is to use data science and data analytics to understand the dataset before applying the ML algorithms.

M. S. Akanksha (✉) · K. Vinutna · M. N. Thippeswamy
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore, India

M. N. Thippeswamy
e-mail: thippeswamy.mn@nmit.ac.in

Diabetes mellitus or simply diabetes is a term for several conditions involving how the human body converts the consumed food into glucose and thereby runs the bodily functions [2]. Insulin is a crucial hormone produced by the pancreas, which is responsible for the transportation of the converted glucose through the bloodstream to various parts of the body. The condition highlights the improper secretion of insulin by the pancreas, due to which the glucose remains in the blood and eventually increases the blood sugar level. This may cause serious effects on the functioning of the body and may even be life-threatening to some.

There are different types of diabetes, namely prediabetes, type 1, type 2 and gestational diabetes. While prediabetes can be managed and reduced with the help of proper diet or exercise and gestational diabetes goes away after delivery in most cases, type 1 and type 2 diabetes are permanent. Type 1 diabetes develops during childhood due to defects in the pancreas that causes improper production of insulin from an early stage. Type 2 diabetes develops later in life either due to obesity or insulin resistance of cells. For this project, the major focus is on type 2 diabetes due to its widespread nature and ability to further cause conditions like heart diseases and stroke.

This paper uses the well-known Python language to perform the machine learning operations on the dataset. Python is a versatile language and is the best at data classification as it uses simple functions and models to help with the prediction. What makes it better is that it is a language that is easily compatible with multiple platforms and has a robust developer community to share ideas and develop projects. Also, because it can be used for operations to further research on the topic, it would be a good choice to implement the project.

The unique element in this project is the addition of the “Number of hours of physical activity per week” parameter to determine the prediction. This move was made because most people do not understand the influence of physical activity on diabetes and how it helps induce better insulin production in the body. With this parameter, it can also be determined if the ones who have a better physical activity have a lesser chance of diabetes and if increasing the activity would help reduce the adverse effects of the disease. Another unique element of this project is that the dataset information was collected through a survey made specifically by the research team. The data was collected from a sample of random people among all age groups and genders so that the data would be exactly as random as to how the normal worldly population would be.

2 Literature Review

There have been several papers that have concentrated on predicting diabetes using various machine learning algorithms. One such work [3] explains how the classification algorithms such as SVM, naïve Bayes and decision tree classifier work on the Pima Indians diabetes dataset. The accuracy of each algorithm is found out and the naïve Bayes algorithm was chosen as the best algorithm.

Another such proof of work was [4], where the aim was to design a predictive system, to predict the possibility of diabetes based on the age parameter. It predicted at what age a person may suffer from diabetes. The highest accuracy was obtained with the decision tree classifier.

To classify the risk of diabetes prediction on the Pima Indian Diabetes dataset, [5] the decision trees, logistic regression, naïve Bayes and random forest algorithms were tested for their accuracies. Bagging and boosting techniques were used to increase the robustness of the models. In the end, the random forest algorithm gave better results when compared to the others.

There are many such anonymous reviewers whose research helped the team to derive results and come to conclusions.

3 Dataset Description

The data for the dataset was obtained by an extensive survey among Indians. Data was collected from people of different age groups, genders and intensities of the disease. The data was also obtained from diabetes patients who were receiving treatment from hospitals. With widespread marketing of the survey, it was possible to obtain 251 records, representing the general population varying across various age groups. The diabetic records were 40 in total (around 15.94%) and the non-diabetic records were 201 (around 84.06%). As the number of diabetics in the general public range from low to moderate, the dataset was aimed at an almost 20–80 split among diabetics and non-diabetics, respectively.

The parameters for diabetes mellitus are indeed many based on the different tests and medical diagnostics conducted on the person. Various sources list different parameters for diagnostics [6, 7]. The values of the parameters define at what stage the disease is. However, if all the parameters have to be considered for the model, chances are that some patients may have abnormal values for some but normal values for the others. This kind of ambiguity would not result in a comprehensive model. Hence, the analysis has been performed by using only the most important parameters that clearly define the symptoms of diabetes mellitus.

The survey passed onto the sample of the population consisted of the following fields that had to be filled for the training of the algorithm: age, gender, weight (kg), height (ft. and inch), blood pressure, fasting blood sugar, insulin dosage (if diabetic), Hereditary diabetes (yes or no) and number of hours of physical activity per week (ranging from <2.5 h to >5 h). The outcome or the dependent variable established for the algorithms is whether the individual is diabetic or not. Tables 1 and 2 show details of the dataset as obtained from the Google Form.

The parameters were chosen because of their very specific yet direct relation to the onset of diabetes. Age is an important risk factor because type 2 diabetes in India shows maximum effects in the age group of 20–70 years [8]. Gender also affects the onset of the disease as studies have proven that females are more susceptible to type 2 diabetes than males [9]. The height and weight of individuals were obtained

Table 1 A few rows of the dataset as obtained from the Google Form

Timestamp	Outcome	Gender	Age	Weight	Height	BP	Blood sugar	Insulin	Hereditary	Activity time
2020/05/09 5:30:55 pm GMT + 5:30	Yes	Male	50	84.0	5.06	120/80	130.0	35	Yes	More than 5 h
2020/05/09 5:48:09 pm GMT + 5:30	Yes	Male	76	82.0	5.06	90/110	80.0	50	Yes	Less than 2.5 h
2020/05/09 6:28:07 pm GMT + 5:30	Yes	Female	68	61.0	5.04	110/80	140.0	40	No	More than 5 h
2020/05/09 6:39:16 pm GMT + 5:30	Yes	Male	68	85.0	5.07	120/80	300.0	5	No	More than 5 h
2020/05/09 6:40:39 pm GMT + 5:30	Yes	Female	64	92.0	5.03	120/80	220.0	20	No	More than 5 h

Table 2 Data types of the data as obtained from the Google Form

#	Column	Non-null count	Dtype
0	Outcome	251 non-null	object
1	Gender	251 non-null	object
2	Age	251 non-null	int64
3	Weight	251 non-null	float64
4	Height	251 non-null	float64
5	BP	251 non-null	object
6	Blood sugar	251 non-null	float64
7	Insulin	251 non-null	int64
8	Hereditary	251 non-null	object
9	Activity time	251 non-null	object

to calculate the body mass index, another important risk factor. The more the BMI value of the individual, the higher are the chances of abnormal insulin production, and thereby the higher the possibility of type 2 diabetes [10]. Body mass index greater than 25 are said to be overweight and above 30 are obese. Obesity is a definitive indication of diabetes mellitus. Blood pressure has been directly involved with the disease, such that the patients experience high BP and possibly, an irregular fluctuation between high and low BP as well [11]. Fasting blood sugar is a direct indication of the blood sugar level in the blood, a major test that patients have to undergo for diabetes analysis. Insulin dosage value had to be entered by those participants who had been diagnosed with diabetes. Just as an indicator for the possibility of diabetes, the hereditary parameter was chosen. The last parameter for the analysis was the number of hours of physical activity per week. It has been researched and proven in many studies that physical exercise plays an important role to reduce the preconditions and intensity of the disease. Furthermore, physical activity increases the metabolism of the body that reduces obesity and also aids in insulin production, due to which doctors prescribe physical activity in the prediabetes stage [12].

The above parameters were collected from the individuals, but a few modifications were made to the data by introducing new columns to aid with the analysis. The column for corrected height was introduced to convert the given height to cm values. The BMI calculation metrics require the weight to be in kg and height to be in cm. Hence, the modification was made. Two other columns to signify whether the BP was higher or lower than the normal blood pressure (120/80) were added to the dataset.

4 Methodology

The methodology of the paper describes how the analysis was conducted on the dataset to arrive at the result. Figure 1 presents a clear view of the methodology used for the research and analysis presented in the paper.

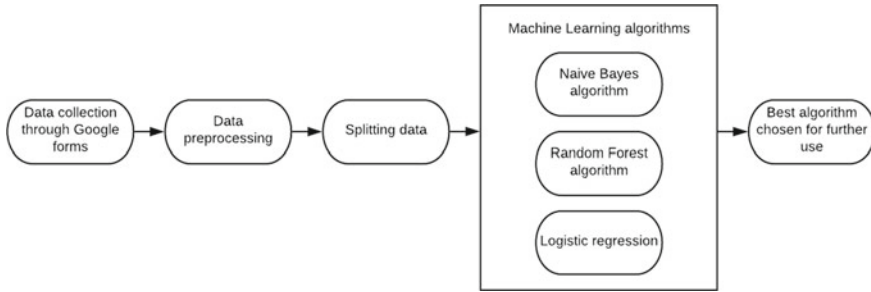


Fig. 1 Design of the machine learning analysis

The information and data from the 251 individuals were collected through a survey using Google Forms. Many more people could be reached out to using the virtual questionnaire form from different parts of India. Data collection was the most essential part of the analysis as it helped to form a training model and then evaluate based on the characteristics.

The data that has been obtained from the survey required to be cleaned and pre-processed to make it conform to the dimensions of the models being used for the analysis. The data obtained from the Google Forms was converted to a CSV file to help with the data analysis and was pre-processed to remove all NULL, NaN, NONE and other inappropriate or unclear information. Also, the data that did not follow the data type of that particular column was cleared off. All the data after pre-processing was converted into numeric data types to ease the process of fitting the data into the algorithms. The final data set that was utilised had only the data that could be used for analysis and no extra or unnecessary information. Table 3 shows a glimpse of the dataset after pre-processing has been performed on it. Table 4 shows the updated data types of the parameters after all the columns are converted to the numeric type. It can be seen that even the object types have been converted either to int or float types after the pre-processing.

After the pre-processing, the entire data was split into two sets—the TRAINING SET and the TEST SET. The training set refers to that partition of the data that would be used to train the model. Model training is performed to let the system know how the information should be processed and what kind of data would produce what kind of outcome. Since the models being used—logistic regression, random forest and naive Bayes are all supervised machine learning algorithms, it becomes essential to use a training Set [13]. Another dataset is obtained after the splitting, called a test set that would be used by the model to perform the prediction, such that it can be verified if the model has learnt how to classify well or not. Test sets and training sets have to be separated, and it must be noted that no record must be present in both the sets [14]. The split performed on the dataset for this research paper puts around 70% of the records into the training set and around 30% of the records into test set. Further verification is also performed to see if the diabetic cases and non-diabetic cases are evenly split between both the sets. An additional round of pre-processing

Table 3 First five rows of the dataset after data pre-processing

Outcome	Gender	Age	Weight	Blood sugar	Insulin	Hereditary	Activity time	CorrHeight	BMI	Higher BP	Lower BP
1	1	50	84.0	130.0	35	1	3.0	152.5524	36.094514	120.0	80
1	1	76	82.0	80.0	50	1	1.0	152.5524	35.235121	90.0	110
1	0	68	61.0	140.0	40	0	3.0	152.5016	26.228958	110.0	80
1	1	68	85.0	300.0	5	0	3.0	152.5778	36.512051	120.0	80
1	0	64	92.0	220.0	20	0	3.0	152.4762	39.571609	120.0	80

Table 4 Data types of each attribute in the dataset after pre-processing

Outcome	int64
Gender	int64
Age	int64
Weight	float64
Height	float64
Blood sugar	float64
Insulin	int64
Hereditary	int64
Activity time	float64
CorrHeight	float64
BMI	float64
Higher BP	float64
Lower BP	int64
dtype:object	

Table 5 Data after splitting

Original true	40(15.94%)
Original false	211(84.06%)
Training true	27(15.43%)
Training false	148(84.57%)
Test true	13(17.11%)
True false	63(82.89%)

is done on the split sets to ensure for one last time that the split did not induce any None, Null or NaN values in the records.

The split data was then fed to each of the three machine learning algorithms being used in the paper, i.e. naïve Bayes, random forest and logistic regression for training. After that, the test set was fed to the model and the outcome is registered. Parameters like accuracy, precision, recall and the confusion matrix were calculated and were compared among the three models to see which one offers the best outcome. The algorithm with the best results can be used for further research on the topic. Table 5 shows a quick analysis of how the data is split among the sets so that a proportionate percentage of positive and negative outcomes are obtained at the end.

5 Data Visualisation

There does exist an inherent relationship between one column of the dataset with the other columns that could help with the analysis being performed. To analyse the usefulness of the interrelation and whether one column makes a significant impact

on another, the tool called data visualisation was used. Data visualisation refers to a graphical representation of the data to offer more understanding and to highlight the behaviour of the data under the influence of certain parameters [15]. At any point in time, it is essential to know the data inside out, and data visualisation aids in establishing just that. Matplotlib and Seaborn are being used as the data visualisation tools. Figure 2 shows the count of the diabetic and the non-diabetic people in the form of a bar chart and a pie chart. Both these plots are plotted using Matplotlib.

Figure 3 shows the feature distribution plot of the different parameters in the dataset. It shows the range of the values that are present in the dataset under the different columns. It gives a good insight as to whether the values are within bounds or not to some extent.

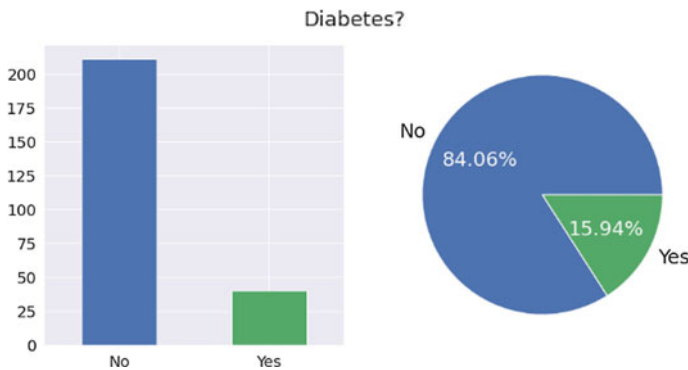


Fig. 2 Bar chart and a pie chart showing the count of the diabetic and non-diabetic people in the dataset

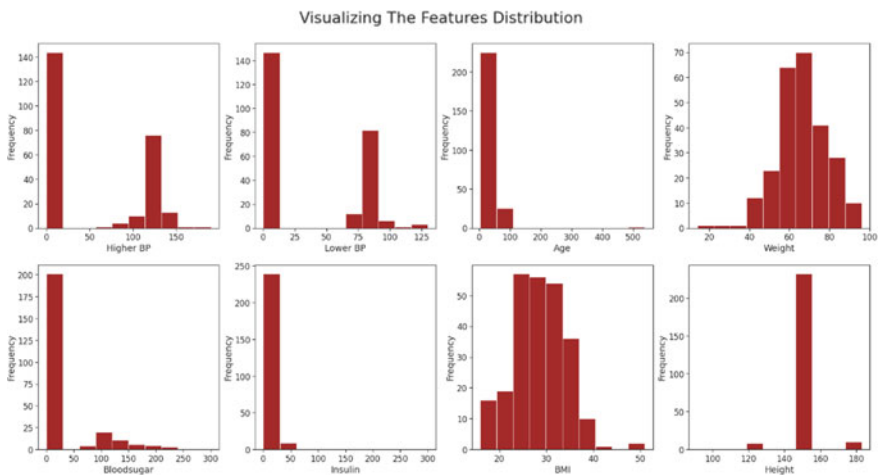


Fig. 3 Visualizing the feature distribution

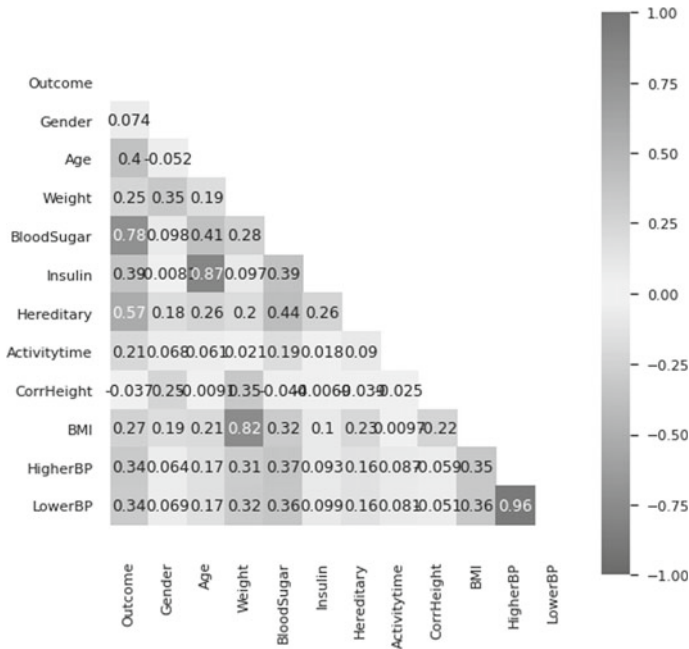


Fig. 4 Heatmap of a correlation matrix which uses a monochromatic scale of colours with Seaborn

Figure 4 shows a heat map of a correlation matrix plotted with Seaborn. A correlation matrix uses coloured cells to show the correlation in the form of a matrix between two attributes in the dataframe [16].

Table 6 shows the correlation matrix in the form of a table with the values depicting to what extent the attributes in the dataframe are co-related to each other.

Figure 5 shows a pair plot which is plotted using Seaborn. A pairplot is used to plot multiple pairwise bivariate distributions in the dataset [17]. To do this, the Seaborn library is loaded and then applied on the dataframe.

The feature analysis using data visualisation proved that the data received was pre-processed in the right manner, and all the insights showed all data conformed to a format. This is essential for the prediction to be performed because irregular data and outliers can affect the performance and efficiency of the predicted results. Also, it showed each feature is related to each other and how much impact the risk factors have among themselves as far as the classification is concerned. Finally, it gives an overview of the attributes and also helps the ML model to train faster by putting all values within a certain limit.

Table 6 Correlation table

	Outcome	Gender	Age	Weight	Blood sugar	Insulin	Hereditary	Activity time	CorrHeight	BMI	Higher BP	Lower BP
Outcome	1.00000	0.073998	0.400782	0.246888	0.783552	0.386682	0.571474	0.211599	0.036943	0.273321	0.343547	0.335927
Gender	0.073998	1.00000	-0.052212	0.345749	0.097791	-0.008093	0.184637	0.068488	0.245475	0.188111	0.064159	0.069420
Age	0.400782	-0.052212	1.00000	0.191274	0.407721	0.868421	0.262227	0.060539	-0.009127	0.206950	0.166482	0.165628
Weight	0.246888	0.345749	0.191274	1.00000	0.280312	0.096726	0.200721	0.021223	0.345493	0.819952	0.307658	0.315392
Blood sugar	0.783552	0.097791	0.407721	0.280312	1.00000	0.392318	0.435589	0.193215	-0.043700	0.321724	0.368205	0.362516
Insulin	0.386682	-0.008093	0.868421	0.096726	0.392318	1.00000	0.255177	0.018183	-0.006944	0.099878	0.092935	0.099381
Hereditary	0.571474	0.184637	0.262227	0.200721	0.435589	0.255177	1.00000	0.090261	-0.038600	0.230271	0.155947	0.155381
Activity time	0.211599	0.068488	0.060539	0.021223	0.193215	0.018183	0.090261	1.00000	-0.024965	0.009686	0.086664	0.081151
CorrHeight	-0.036943	0.245475	-0.009127	0.345493	-0.043700	-0.006944	-0.038600	-0.024965	1.00000	-0.223610	-0.05937	-0.051471
BMI	0.273321	0.188111	0.206950	0.819952	0.321724	0.099878	0.230271	0.009686	-0.223610	1.00000	0.354173	0.357070
Higher BP	0.343547	0.064159	0.166482	0.307658	0.368205	0.092935	0.155947	0.086664	-0.059377	0.354173	1.00000	0.964320
Lower BP	0.335927	0.069420	0.165628	0.315692	0.362516	0.099381	0.155381	0.081151	-0.051471	0.357070	0.964320	1.00000

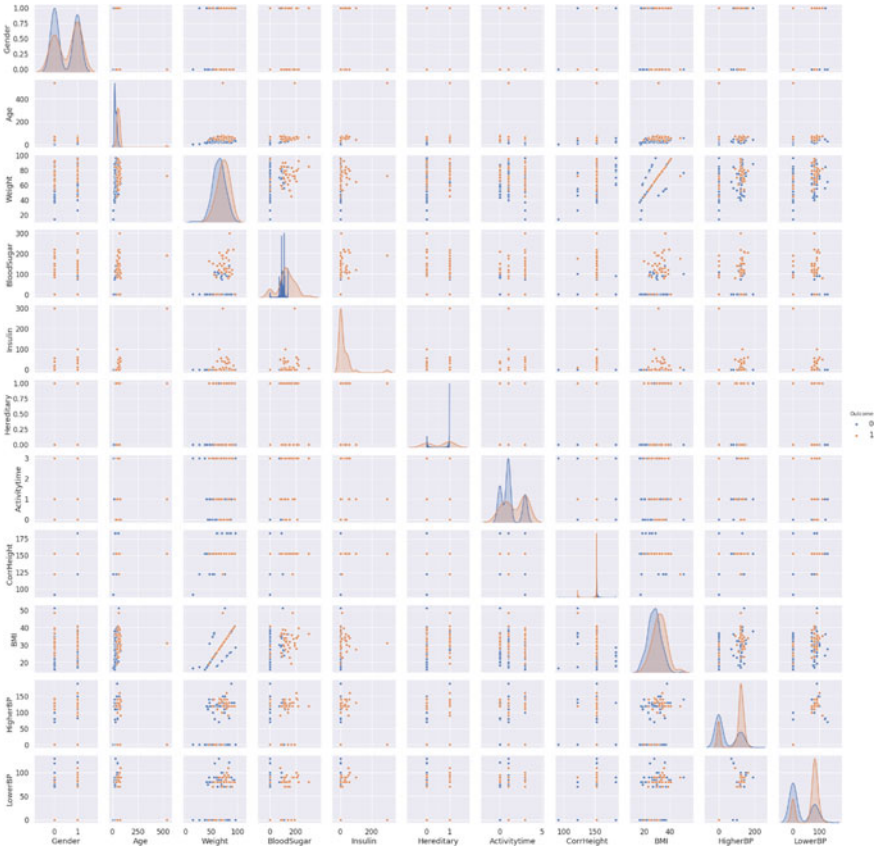


Fig. 5 Pair plot using Seaborn

6 Implementation

Machine learning algorithms were used to analyse the data for a very specific reason—to help predict the onset of diabetes depending on the risk factors that were being used in the dataset. Each algorithm is divided into certain steps that offer continuity and help obtain the results. The steps and procedures would be described under the sections of the algorithms. It must be noted that the method of functionality for all the algorithms are different based on different approaches and equations. This section is divided into three segments with each segment dedicated to one algorithm and the analysis that was used.

6.1 Naïve Bayes Algorithm

Naïve Bayes classifier makes use of the simple Bayes theorem which is based on conditional probability [19]. It helps to analyse how likely an event will occur provided another event has already occurred. Each time a new event occurs, a hypothesis is updated. The Bayes theorem expressed in terms of probability is as follows,

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} \quad (1)$$

where

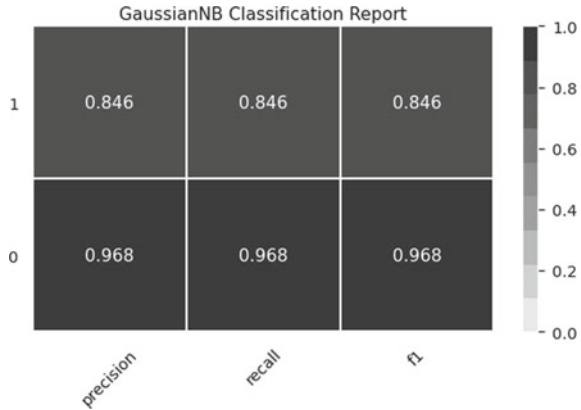
- P denotes the probability
- $P(E|F)$ = Probability of event E (hypothesis) occurring given that F (event) has occurred.
- $P(F|E)$ = Probability of the event F (event) occurring given that E (hypothesis) has occurred.
- $P(E)$ = Probability of event E (hypothesis) occurring.
- $P(F)$ = Probability of event F (event) occurring.

The naïve Bayes is used as a multi-class classification algorithm on the dataframe. New predictions can be made by the naïve Bayes model using the Bayes theorem as shown above. A Gaussian distribution variable is assumed by the model because it is dealing with real-time data. This extension of the naïve Bayes algorithm is called the Gaussian naïve Bayes algorithm. The Gaussian model is imported from Scikit which makes the model training and further prediction easy. The usage of built-in functions for ML in Python assures the right application for all the data types being used and keeps the imperfections of the calculations to the minimum.

The model was first set to a variable using GaussianNB class and then trained using fit() function and the training set as a variable. After the training, the prediction was done on the test set. Finally, a report was constructed on the classification scores that the prediction brought out. For the dataset, the mean accuracy was 94.8%, the mean precision was 86.67% and the mean recall was 82.14%. All these metrics are shown in Table 7 in the Results section. To strengthen the classification that was made, a confusion matrix was constructed for the test set prediction. The results showed that the true positive was 61 entries, the false positive was 2 entries, the false negative was 2 entries and the true negative was 11 entries, where positive stands for non-diabetics and negative is for diabetics. Figure 6 shows the classification report for the two classes (1,0) and their precision, recall and F1 score.

Overall, the Gaussian naïve Bayes algorithm's simplicity makes it the easiest method for classification. Furthermore, it uses very less time for training the dataset and also takes linear time ($O(f)$ where f is the number of features) to perform the predictions as compared to the expensive exponential-type processing [20]. It also does not pose the problem of overfitting its data while predicting the results. This

Fig. 6 Gaussian naïve Bayes classification report



makes it the best algorithm to use to find results in the least time possible, using a non-complicated logic for the analysis.

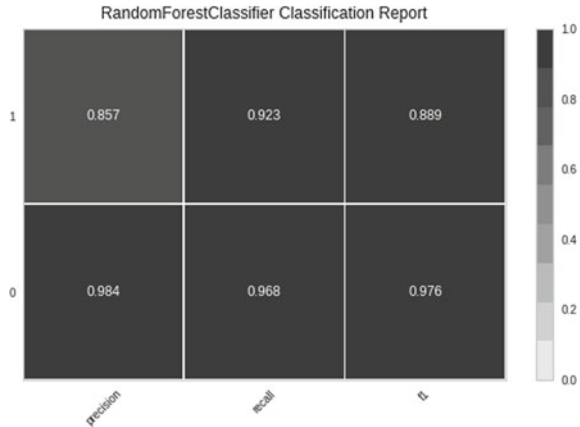
6.2 Random Forest Algorithm

Random forest algorithm creates multiple decision trees with the data available and gets outcomes from each of them. The algorithm is just an extension of the decision tree algorithm which classifies data based on simple yes or no outcomes. It works on the principle that if multiple decision trees are used for prediction, the precision of the result would increase. The best outcome is then chosen based on the majority [21]. Sklearn can be used to work with this model; it measures the importance of a feature by analysing how much the tree nodes that use that feature reduce impurity across all trees. The score is computed automatically for all the features in the dataset after the training of the model. It then scales all the results and finally computes the sum which is to be equal to one [22].

The model was first set to a variable using the random forest classifier class and then trained using the fit() function with the training set as a variable. The prediction was done by the model on the test set after being trained. A report was constructed on the classification scores that the prediction brought out. For the dataset, the mean accuracy was 95.4%, the mean precision was 92.2% and the mean recall was 78.5%. All these metrics are shown in Table 7 in the Results section. The confusion matrix that was constructed to represent the prediction showed that true positive was 61 entries, false positive was 2 entries, false negative was 1 entry and true negative was 12 entries. Figure 7 shows the classification report for the two classes (1,0) and their precision, recall and F1 score.

In general, it can be noted that random forest classification performs a more complex computation than that of naïve Bayes as it constructs multiple decision trees and chooses the best answer among them all. The training takes some more

Fig. 7 Random forest classifier classification report



time depending on the number of trees that are created for the model and so does the prediction ($O(fn)$ where f is the number of features and n is the number of decision trees created) [20]. It can also be observed that the accuracy of the random forest classification is high as it helps strictly classify true positive and true negative values.

6.3 Logistic Regression

Logistic regression is a regression algorithm, i.e. it helps in the prediction of continuous values as compared to the discrete values predicted for classification. It is a supervised machine learning algorithm. A logistic regression function can be described in Eq. 2. It assumes that the data would follow a linear function, usually the sigmoid function like that in Eq. 3, to classify the results. When a threshold value is placed as a condition for the regression, it becomes a classification problem [23].

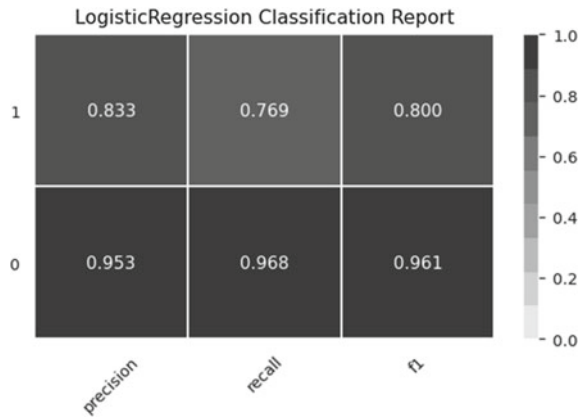
$$y = \alpha_0 + \alpha_1x_1 + \dots + \alpha_nx_n \tag{2}$$

$$F(x) = \frac{1}{(1 + e^{(-x)})} \tag{3}$$

This algorithm is best used for two-class regression. It describes and tries to estimate the deep relationship between the independent variables and the dependent variable. The dependent variable follows Bernoulli distribution, and the estimation is done through the principle of maximum likelihood [24]. The model is imported from the Scikit library to ensure that all the values are conformed to a specific pattern and that the prediction is made as accurately as possible.

The model was fit to a variable using the logistic regression class, with the regularisation parameter c as 0.1. The model was then applied to the training set using

Fig. 8 Logistic regression classification report



the `fit()` function of the class. The prediction was then made on the test set after the training. It could also be observed that by changing the regularisation parameter of the model, different results were obtained. The purpose of the coefficient was to push the independent parameters more towards zero and obtain a regularised outcome. For the dataset, the mean accuracy was 96%, the mean precision was shown to be 88.67% and the mean recall was 85.7%. The confusion matrix created for this algorithm showed that true positive was 61, false positive was 2, false negative was 3 and true negative was 10. All these metrics are shown in Table 7 under the results section. Figure 8 shows the classification report for the two classes (1,0) and their precision, recall and $F1$ score.

Logistic regression generally performs well even when the noise variables and variance is increased in the dataset [25]. The beauty of the algorithm is that though training takes a long and a laborious process, the prediction is very fast and the computation is a linear one ($O(f)$ where f are the features for the prediction) [20]. This is because the logistic regression equation in itself is pretty complex to compute and the added sigmoid function makes it take more time during training. It is also observed that due to this computation while training, the outlier influence becomes very limited and the values are fairly regularised.

6.4 Comparison Metrics for Evaluation

There is a need to identify what evaluation metrics would be correct to establish which algorithm is the best predictor. For that, it was determined that the classification report functions in Python would directly help in the identification. The accuracy, precision, recall and $F1$ score metrics in the classification report were used for the comparison.

Accuracy can be defined as the ratio of the number of correct predictions made by the algorithm to the total number of predictions made. It also defines whether the outcome obtained is close to the accepted value or not [26].

$$\begin{aligned}
 \text{Accuracy} &= \frac{(\text{Number of Correct Predictions})}{(\text{Total number of Predictions made})} \\
 &= \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}
 \end{aligned}
 \tag{4}$$

Precision refers to the ratio of the correct positive outcomes to the total positive outcomes predicted by the classifier, implying that it is a ratio of the true positives to the sum of true positives and false positives. It also implies how likely the classifier would predict the same outcome if used on the same dataset repeatedly [26].

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
 \tag{5}$$

Recall means the ratio of the number of correct positive outcomes to the number of all relevant outcomes, implying that it is the ratio of the true positives to the sum of true positive and false negatives [26].

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
 \tag{6}$$

F1 Score can be defined as the harmonic mean between precision and recall. The greater the *F1* score, the better the model, as it tries to find a common point between being precise and being accurate. The range for this score lies between 0 and 1 to tell how good the prediction of the model is [26].

$$F1 = 2 * \left(\frac{1}{((\text{Precision})^{-1} + (\text{Recall})^{-1})} \right)
 \tag{7}$$

7 Results

The three models were successfully created and the predictions were made based on the dataset values. The output for each model included the confusion matrix and the classification report. The mean results (of Class 0 and Class 1) were then made into a data frame of their own and printed into a matrix format to get better visibility of the comparison. Table 7 reflects the mean scores of each of the models.

Table 7 Comparison of the algorithms in terms of mean accuracy, precision, recall and *F1* score

	Model	Accuracy	Precision	Recall	<i>F1</i> score
0	GaussianNB	0.948864	0.866667	0.821429	0.834936
1	Logistic Regression	0.960227	0.886905	0.857143	0.870879
2	Random Forest	0.954413	0.922619	0.785714	0.835664

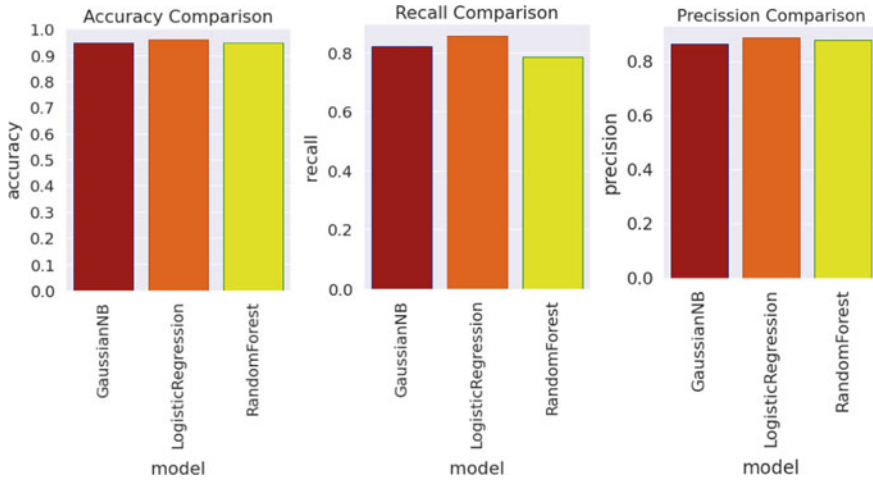


Fig. 9 Comparison of accuracy, precision and recall

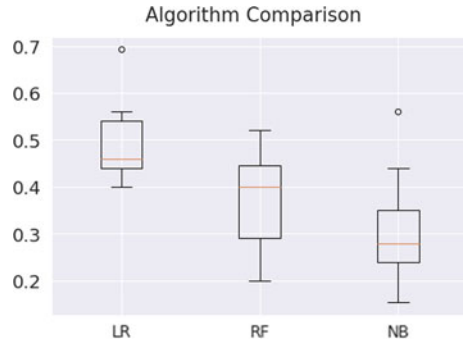
Figure 9 shows a comparison of the three models in terms of mean accuracy, precision and recall. It visualises Table 7 in the form of bar charts using Seaborn.

The tabular and visual comparisons show that logistic regression has the highest accuracy. This is because of its property to classify by regularisation and because it can help nullify the influence of outliers in the data. By comparing even, the recall score, the best algorithm is logistic regression. Recall refers to the ratio of the number of correct predictions returned to the number of correct predictions that should have been made. Also, the *F1* score shows that logistic regression has a higher value of prediction.

By analysing the tabulated scores, the precision of the prediction of the random forest algorithm exceeds that of the others. This means that the predictions made by the random forest algorithm are mostly right, with a very small possibility of misclassification. This may be because multiple decision trees are created and the best answer is chosen, proving that only the most precise tree would be chosen by the algorithm.

To better understand which algorithm to choose between logistic regression and random forest, a box plot was constructed for the algorithms to see the range of the predictions that were made. Figure 10 represents the box plot. The X-axis of the box plot has the algorithms and the Y-axis has the range scores. The graph

Fig. 10 Box plot comparison of the algorithms



was constructed by using the K Fold function, where the dataset and features are split into K consecutive folds and each fold is used once for predicting the results by using the remaining $K - 1$ folds as training data. This function was also used as a cross validator of the predictions made by the algorithms. The resultant box plot brought a clear distinction among the algorithms. The mean line for logistic regression was higher than that of naïve Bayes or random Forest, showing that the average number of correct predictions made by this algorithm was far greater than those made by the others. Also, the range of the predictions—the quartiles and the extremities—had higher positions on the graph compared to the other two algorithms. This further meant that after multiple results were computed by K Fold function, the logistic regression algorithm fared consistently well as compared to the remaining two algorithms.

This proved, very evidently, that the logistic regression algorithm performed the best prediction among the three machine learning models to help predict the onset of diabetes based on the parameters in the dataset.

8 Conclusion

Since logistic regression proved to be the best machine learning algorithm for the classification of our dataset, the future scope of the project would be to test deep learning algorithms on the dataset. Machine learning can go only as far as to classify smaller datasets, but using deep learning would prove beneficial to take a larger number of parameters for classification. Also, because deep learning can create abstract relationships from complex datasets, a whole range of pre morbidities that apply to diabetes mellitus can be added to the dataset. This would further enhance the learning of the model. For example, deep learning could be encouraged to predict the exact type of diabetes in the patient, like type 1, type 2, prediabetes or gestational diabetes, based on the symptoms and other underlying conditions of the person.

This research is focused on choosing the best machine learning model among naïve Bayes, random forest and logistic regression algorithms to predict the onset

of diabetes mellitus in people. A dataset having the parameters—age, gender, blood pressure, fasting blood pressure, number of hours of physical exercise per week, height, weight and others—was used for the research. An analysis of the dataset was made using data visualisation and charts so that clarity was obtained on the relationship between the different parameters in the dataset. The computation was coded in Python language due to its simplicity in computing complex ML algorithms.

The dataset was split into training and test sets that were used to train and test the models, respectively. The same splits were used for each of the three algorithms so that the comparison would be unbiased and reliable. A quick analysis of the predictions was made with the help of confusion matrices and classification scores.

On completing the predictions, it was observed on comparing that logistic regression algorithm provided the best predictions for the dataset and that it had a high precision as well. This proves that logistic regression could help in predicting the onset of diabetes mellitus based on the parameters provided in the dataset.

Acknowledgements The research team would like to thank Dr M.N. Thippeswamy, the research mentor for this paper, for his excellent guidance and insight at every step of the research. The team would especially like to thank all the participants of the survey, for their consent to provide details for the dataset and for complying with the requests that were made. Lastly, special thanks to the college, Nitte Meenakshi Institute of Technology, for providing a wonderful research opportunity to the team and for helping the idea materialise.

References

1. Mohammed M, Khan MB, Bashier EBM (2017) Machine learning algorithms and applications. CRC Press
2. WebMD Diabetes mellitus—types and description. <https://www.webmd.com/diabetes/guide/types-of-diabetes-mellitus>
3. Sisodia D, Sisodia DS Prediction of diabetes using classification algorithms. <https://www.sciencedirect.com/science/article/pii/S1877050918308548>
4. Early predictive system for diabetes mellitus. In: Industrial conference on data mining. Springer, pp 420–427
5. Nai Arun N, Mounghmai R Comparison of classifiers for the risk of diabetes prediction. <https://www.scopus.com/record/display.uri?eid=2-s2.0-84962885345&origin=inward&txGid=92705c75f09b9575155d40abd686ef07>
6. Heart.org Symptoms, diagnosis and monitoring of diabetes. <https://www.heart.org/en/health-topics/diabetes/symptoms-diagnosis--monitoring-of-diabetes>
7. World Health Organisation Screening for type 2 dDiabetes. https://www.who.int/diabetes/publications/en/screening_mnc03.pdf
8. World Health Organisation: India. <https://www.who.int/india/events/world-diabetes-day>
9. Mahajan S, Bharadwaj K, Mahajan R (2019) Gender difference in the risk of developing diabetes mellitus type 2 and oral glucose tolerance test in dental students. *Int J Oral Health Sci* 9(2)
10. Agrawal N, Agrawal MK, Kumari T, Kumar S (2017) Correlation between body mass index and blood glucose levels in Jharkhand population. *Int J Contemp Med Res* 4(8)
11. WebMD Diabetes and high blood pressure. <https://www.webmd.com/diabetes/high-blood-pressure#:~:text=Diabetes%20damages%20arteries%20and%20makes,heart%20attack%2C%20and%20kidney%20failure>

12. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, Horton ES, Castorino K, Tate DF (2016) Physical activity/exercise and diabetes. American Diabetes Association
13. Elite Data Science Model training with machine learning. <https://elitedatascience.com/model-training>
14. Technopedia What is training data? <https://www.techopedia.com/definition/33181/training-data>
15. Tableau Data visualisation beginner's guide. <https://www.tableau.com/learn/articles/data-visualization#:~:text=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data>
16. Display R What is a correlation matrix. <https://www.displayr.com/what-is-a-correlation-matrix/>
17. Seaborn seaborn.pairplot() official documentation. <https://seaborn.pydata.org/generated/seaborn.pairplot.html>
18. Toptal Supervised machine learning algorithms in python. <https://www.toptal.com/machine-learning/supervised-machine-learning-algorithms>
19. Easysol Machine learning algorithms explained—Naive Bayes classifier. <https://blog.easysol.net/machine-learning-algorithms-4/#:~:text=Naive%20Bayes%20Classifiers%20rely%20on,time%20new%20evidence%20is%20introduced>
20. The Kernel Trip Computational complexity of machine learning algorithms. <https://www.thekerneltrip.com/machine/learning/computational-complexity-learning-algorithms/>
21. Datacamp Understanding random forest classifier in python. <https://www.datacamp.com/community/tutorials/random-forests-classifier-python>
22. Built In Complete guide to random forest algorithm. <https://builtin.com/data-science/random-forest-algorithm>
23. Geeks for Geeks Understanding logical regression. <https://www.geeksforgeeks.org/understanding-logistic-regression/>
24. Datacamp Understanding logistic regression in python. <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>
25. Kirasich K, Smith T, Sadler B (2018) Random forest versus logistic regression—binary classification for heterogeneous datasets. SMU Data Sci Rev 1(3)
26. Towards Data Science Metrics to evaluate your machine learning algorithms. <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>

Session-based Personalized Recommender System for Online Shopping



B. R. Sreenivasa, C. R. Nirmala, and M. V. Manoj Kumar

1 Introduction

The development of web and smartphone innovations have brought about changing how individuals shop. Individuals purchase increasingly more items online through the web or mobile application instead of performing traditional-style shopping. Recommendation systems [1] help customers in taking care of data over-burden by suggesting new things (i.e. products) that suit the client's needs (i.e. favourites) and requirements. Recommendation systems gather data on the client's inclinations for the products in a respective area such as an online shopping environment and movie and afterwards endeavour to forecast what different products the customer is probably going to discover significant (i.e. useful). Data related to customers' inclinations might be procured in an explicit manner through likes/dislikes, a rating, etc. and on. Similarly, the data can be procured in an implicit manner by verifiably observing the customer's activities. The well-known RS methods are content-based, collaborative or by combining multiple methods in designing hybrid methodologies. Collaborative filtering (CF) methodologies [2, 3] depend on the collaborative information of historical information of subscribers as well as of product for carrying out forecasting operation. The content-based methodologies endeavour to forecast products using content features likeness/similarities. The hybrid forecasting method [4,

B. R. Sreenivasa (✉)

Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, India

C. R. Nirmala

Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, India

M. V. Manoj Kumar

Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India

5] generally combines multiple methodologies discussed above for building efficient forecasting designs.

The state-of-the-art forecasting method concentrated on modelling customers' preferences and choices of likeness under users' historical behaviour on particular items and in general and consistently disregards the sequence behaviour data. Consider that shopper inclination/choice changes considering different customers' behaviour sequence. In this way, as opposed to considering one sort of behaviour, for example adding to cart, click stream and buying, there exist different behavioural sequences corresponding to products qualities cases with various conduct towards an item. Thus, it is preliminary to incorporate multiple behaviour sequences and forecast collaboratively what a user will purchase, select or prefer in future considering the certain behavioural context. In recent times, a few endeavours have been placed into creating forecast techniques using customer behaviour sequence data [6, 7]. Nonetheless, none of the current techniques is intended for learning sequence information considering a multi-behaviour sequence by our knowledge. Furthermore, if we straightforwardly treat various behaviours towards a product as different components in sequence, or just disregard the variance among behaviour sets, existing forecasting strategies will pose issues in establishing a correlation concerning behaviour sets and product sets. In [8, 9], utilized artificial intelligence (AI) and recurrent neural network (RNN) for session-based product forecasting. RNN has been widely used in the most existing method as a natural choice for addressing the issue of learning behaviour sequences efficiently. The model is applied to various sequence-based forecasting issues in time series forecasting, signal processing (SP), pattern recognition, etc. in a productive manner. In RS, RNN is widely used for modelling a session-based forecasting environment with good results.

This work focuses on building efficient session-based RS for the e-commerce environment [10]. His work endeavours to establish if a set of products looked into by the customer during an ongoing session is probably going to end with a buy. For meeting real-world challenges and circumstances, the above-discussed problems are challenging because the forecasting model needs to learn inter-product dependencies and its relationship. Recent work modelled based on session-based RS [5, 11] and [12] centre around forecasting future top-k products list of the session, as opposed to foreseeing the utilization purpose/goal. This paper accepts that forecasting the purpose of the customers early in the session may aid the number of strategies improving the session result. For example if the model can predict the session purpose (i.e. if the RS model predicts the user leaves the ongoing session without purchase), then the system might provide a certain discount for changing the mood or intent of the customer. Along with this, this paper addresses the cold-start issues for purchasing purpose forecasting in the current ongoing session where customers' purchase history is not available for new entrant products in the e-commerce environment. This circumstance generally occurs in an e-commerce environment when new products are added frequently. However, these issues have not yet been overcome by existing methodologies. For overcoming research challenges, this work presents a time-centric predication model for online shopping environment using a hybrid learning technique. Research contribution as follows:

- This work presented a time-centric prediction model combining both short and long-term behaviour sequences.
- TCP model attains good performance considering hit rate (HR) and mean reciprocal rate (MRR) concerning state-of-the-art RS methods.

The manuscript is articulated as follows. In Sect. 2, some baseline methods are described. In Sect. 3, the proposed time-centric recommendation model for online shopping portal is presented. In Sect. 4, the experiment is conducted using online shopping portal data and the performance of TCP, and the existing recommendation model is evaluated. The conclusion and future work are described in the last section.

2 Literature Baseline Methods

Many research work has been carried out in literature to overcome the shortcomings of traditional recommender system by considering user preferences based on the characteristics of items to provide more accurate recommendations [13], and it is been used in many product-based RS [14], tourisms RS [15], restaurant RS problems [16] and many others. In [17], the author described how the RNN model is used to predict the user next buying pattern based on the previous history log files. To minimize the costs and to increase the prediction accuracy, all older states are combined into single window and keeps only latest states. Quadrana et al. [18] presented a hierarchical recurrent neural network for session-based RS, and it mainly focuses on user identities. Methods based on recurrent neural networks concentrate on session-oriented proposal issues. Out of these techniques, gru4rec is a significant method used for effectively planning for session-based recommendation scenarios [11, 12]. Then, gru4rec constructs gated recurrent units with the help of RNN [19, 20] to predict the likelihood of possible events (e.g. click streams) at the start of every session. The solitary item determines the contribution of the method, which is based on vector in one-hot encoded form and yield. It gives a positioning circulation. Typical GRU layer screens a state that encodes existing items in the same session. While predicting with the help of GRU, the items of a session must be fed of into the system. As far as activation functions, tanh functions to work best for GRU and the ranking layer. Use of RNNs for prediction issues is a characteristic decision. Decision of the loss function and utilization of session parallel mini-batches to accelerate the training phase are key components of the approach.

3 Time-Centric Recommendation Model for Online Shopping Portal Environment

This section presents a time-centric prediction/recommendation (TCP) model for online shopping portal using session-centric behaviour information of customers. For

obtaining a relationship between user current session and past behaviour for recommending items, it is important for modelling both short and long-term behaviour contexts (i.e. time-centric HLCP model [21]). As discussed in [21], the RNN model is efficient in modelling the long-term behaviour of the customer. Thus, this work uses RNN for modelling long-term dynamic behaviour. The architecture of the RNN model is shown in Fig. 1. Similarly, in [21] showed that FFNN is efficient in modelling short-term behaviour context of the customer when compared with RNN. Thus, this work uses FFNN for modelling short-term sequence. The architecture of FFNN is shown in Fig. 2.

For capturing the dynamic behaviour of the customer, this paper uses behaviour-based matrices for obtaining feature sets of various sorts of behaviour sets. Then, the illustration of subscriber v at ℓ is estimated using the following equation

$$i_\ell^v = \mathcal{X}i_\ell^v + \sum_{j=0}^{o-1} \mathcal{D}_j \mathcal{N}_{\ell-j}^v r_{\ell-j}^v, \tag{1}$$

where $\mathcal{N}_{\ell-j}^v \in \mathbb{S}^{e \times e}$ depicts a behaviour-based transition matrix design concerning behaviour on the j th product of subscriber v .

The cold-start problem can be addressed by considering $i_0^v = v_0$. Further, the existing sequence-based model generally neglects continuous session variance (SV)

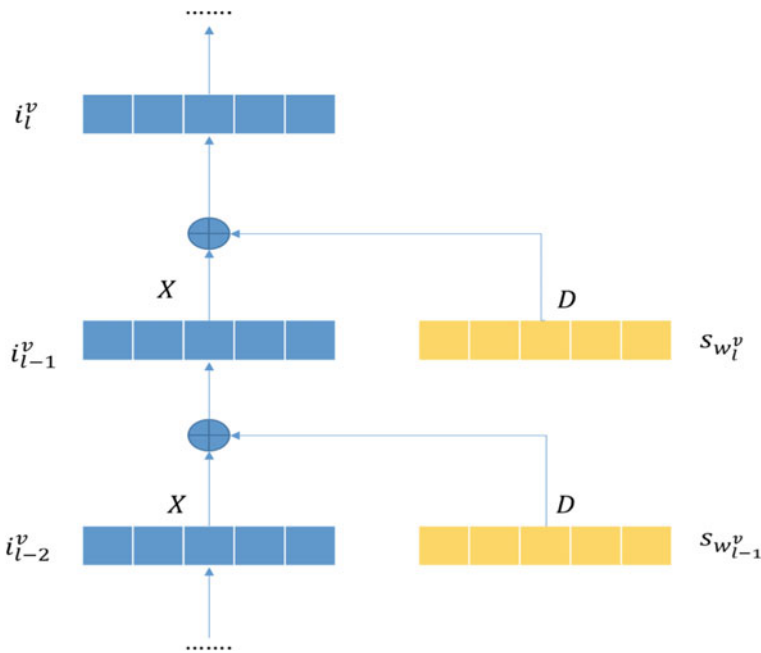
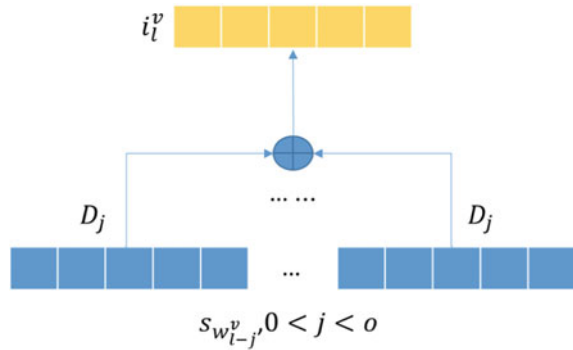


Fig. 1 Architecture of RNN model

Fig. 2 Architecture of the FFNN model



among input feature sets. The session variance feature is very useful in forecasting as short-term session window variances generally have higher effects on future buying than using long-term session variances. Besides, as the behaviour of purchasing certain items are periodic. Thus, the impact of session variances results in more dynamic in such conditions. Considering these conditions, this work improves the HLCP model [21] by incorporating session variances knowledge and model time-centric prediction model.

Using [21], user preference can be learned better considering customer location. However, it is still important to consider incorporating time/session variance data into HLCP. Thus, this work presents an efficient time-centric prediction (TCP) model by substituting location-centric (LC) transition matrices (TM) with session-centric TM's. The TCP model is shown in Fig. 3. From Fig. 3, for a given customer v , the location l is computed as follows

$$i_l^v = Xi_{l-o}^v + \sum_{j=0}^{o-1} U_{u_l^v - u_{l-1}^v} s_{w_{l-j}^v}, \tag{2}$$

where u_l^v depicts the present time, u_{l-1}^v depicts the time of every product of each layer of TCP, and $U_{u_l^v - u_{l-1}^v}$ depicts the time-centric TM of time variance $u_{l-1}^v - u_l^v$ among time u_{l-1}^v and u_l^v . The time-centric TM aids in capturing session-specific effects behaviour on the most recent activity log. Further, Eq. (2) is rewritten similarly with the HLCP model as follows

$$i_l^v = Xi_0^v + \sum_{j=0}^{l-1} U_{u_l^v - u_{l-1}^v} s_{w_{l-j}^v}, \tag{3}$$

where $i_l^v = v_0$ is depicting the preliminary condition of customers. From modelling dynamic behavioural traits, behavioural-centric TM's is used in the TCP model as follows

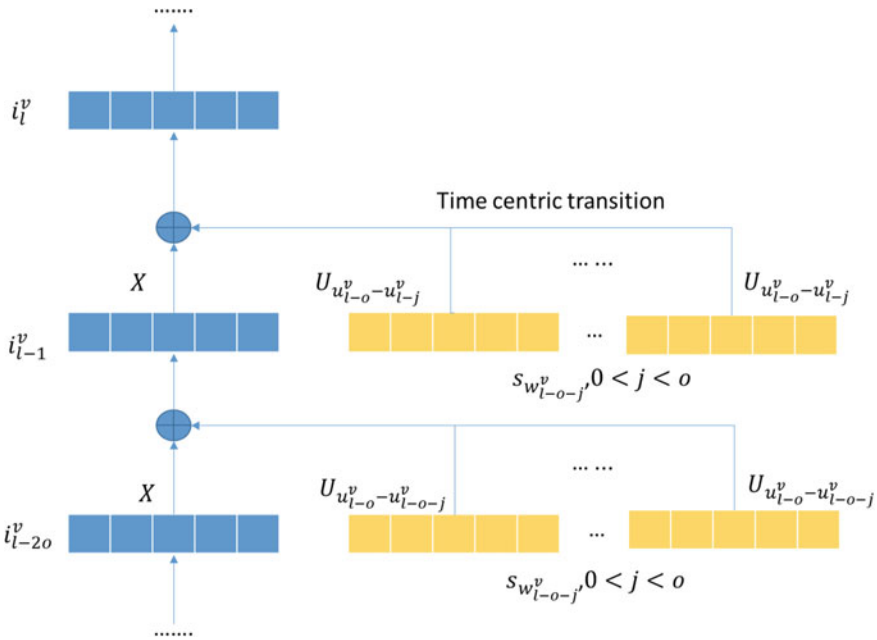


Fig. 3 Architecture of TCP that combines both short and long-term session behaviour sequences

$$i_l^v = X i_{l-o}^v + \sum_{j=0}^{l-1} U_{u_l^v - u_{l-1}^v} N_{c_{l-m}^v} s_{w_{l-j}^v}. \tag{4}$$

Then, carry out forecasting operation whether a customer v will carry out certain behaviour c on particular product w at sequential location $l + 1$ is estimated similarly to HLCP using following equation

$$z_{v,l+1,c,w} = (t_l^v)^U N_{cSw} = (i_l^v + v_v)^U N_{cSw}. \tag{5}$$

The experiment is conducted using an online shopping portal dataset for both proposed TCP and existing recommendation models. The TCP model attains significant recommendation accuracy performance when compared with the existing recommendation model which is experimentally shown below.

4 Experiment Results and Discussions

This section discusses the results of the proposed algorithm and comparison in reference to evaluation benchmark, performance metrics and accuracy based on various standard datasets.

Table 1 Characteristics of the e-commerce datasets

Dataset RSC15	TMALL
Actions	13.42 M
Sessions	1.77
Items	425,348
Timespan in days	91
Action per session	7.56
Unique items per session	5.56
Action per day	149,096
Session per day	19,719

Evaluation Protocol and Performance Measures

The general computational error and in session-based recommendation issues are to produce a list of ranked items that in some structure “matches” a given session starting. What speaks to a decent match relies upon the particular application situation. It could be a set of alternative shopping things in an online business situation or a continuation of a given music listening meeting. The experiment is conducted on the Tmall dataset [11]. This dataset was distributed with regards to the Tmall competition and contains user-interaction logs of the tmall.com site for one year. For Tmall, each split comprises of 30 days of training and 1 day of test information. It is one of china’s greatest Internet business web-based interface. Then, the preliminary job of the RS model is to forecast what the customer will buy in the current ongoing session using time-centric prediction and existing recommendation model. The performance of TCP and existing recommendation (ER) methodologies are evaluated using HR and MRR metrics. The characteristics of the data set used for the implementation of the proposed work is shown in Table 1. The Hit Rate and Mean Reciprocal Rank performance for different cases using Tmall dataset is shown in Table 2. Table 3 represents the comparison of of proposed and the existing work. Figures 4 and 5 represents the Mean Reciprocal rank and Hit Rate of the proposed work for Top 20 and 10 recommendations respectively with different iterations.

(a) *Hit rate/Accuracy outcome achieved by TCP and ER methodologies:*

Table 2 Hit rate (HR) and mean reciprocal rank (MRR) for a list length of 10 and 20 obtained for the e-commerce TMALL datasets considering different cases

Case	Metrics	Metrics MRR@20	Metrics HR@20	Metrics MRR@10	Metrics HR@10
Case1	Proposed HTCP	0.2976	0.5307	0.1969	0.4589
Case2		0.2983	0.5319	0.1977	0.4568
Case3		0.2989	0.5341	0.1984	0.4562
Case4		0.2986	0.5353	0.1982	0.4569
Average		0.2984	0.5347	0.1978	0.4571

Table 3 Result comparison with proposed work and existing work in terms of HR and MRR

Metrics	Metrics MRR@20	Metrics HR@20	Metrics MRR@10	Metrics HR@10
Proposed HTCP	0.2984	0.5347	0.1978	0.4571
Existing gru4rec (ER)	0.1852	0.4038	0.0986	0.2119

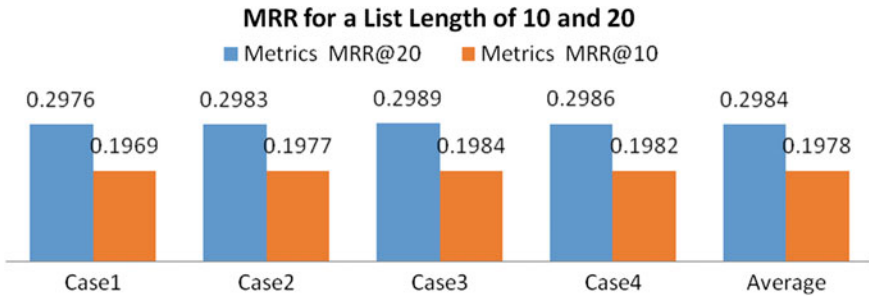


Fig. 4 Mean reciprocal rank for the list length of 20 and 10 with different cases for Tmall dataset

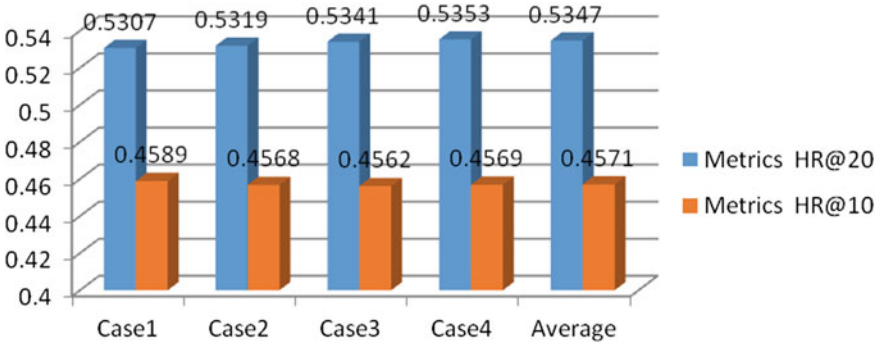


Fig. 5 Hit rate for the list length of 20 and 10 with different cases for Tmall dataset

The accuracy outcome (i.e. HR) of TCP and existing recommendation methodologies is evaluated in this section. The accuracy outcome is evaluated considering top-10 k recommendation as shown in Fig. 6. The TCP achieves an HR outcome of 0.4571, and ER (i.e. existing system (ES)) method achieves an HR outcome of 0.2119. From the result attained, it shows the TCP method achieves much superior HR outcome when compared with ER methodologies considering the top-10 k recommendation. Further, the accuracy outcome is evaluated considering top-20 k recommendation as shown in Fig. 7. The TCP achieves an HR outcome of 0.5347, and the ER method achieves an HR outcome of 0.4038. From the result attained, it shows the TCP method achieves much superior HR outcome when compared with ER methodologies considering the

Fig. 6 Hit rate performance considering the top-10 k recommendation

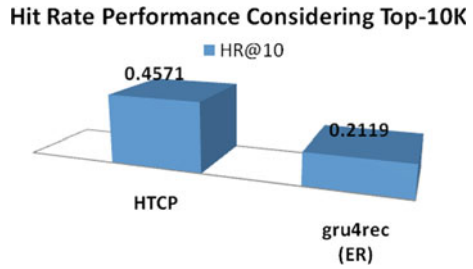
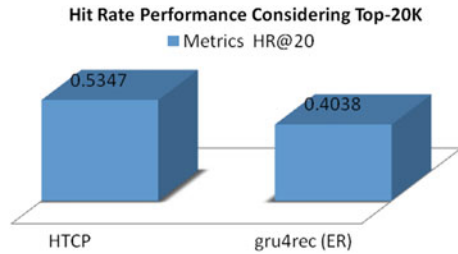


Fig. 7 Hit rate performance considering the top-20 k recommendation



top-20 k recommendation. The overall result attained considering the top-10 k recommendation and top-20 k recommendation shows TCP achieves much superior HR outcomes.

- (b) *Mean reciprocal rate/Accuracy outcome achieved by TCP and ER methodologies:*

The accuracy outcome (i.e. MRR) of TCP and existing recommendation methodologies is evaluated in this section. The accuracy outcome is evaluated considering top-10 k recommendation as shown in Fig. 8. The TCP achieves an MRR outcome of 0.1978, and the ER method achieves an MRR outcome of 0.0986. From the result attained, it shows the TCP method achieves much superior MRR outcome when compared with ER methodologies considering the top-10 k recommendation.

Fig. 8 Mean reciprocal rate performance considering the top-10 k recommendation

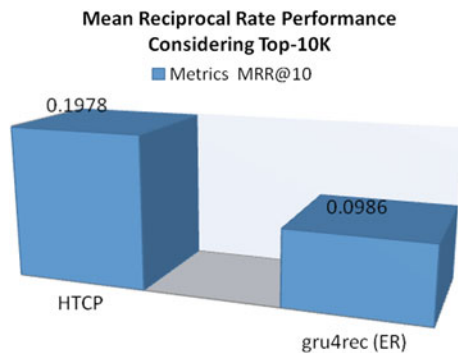
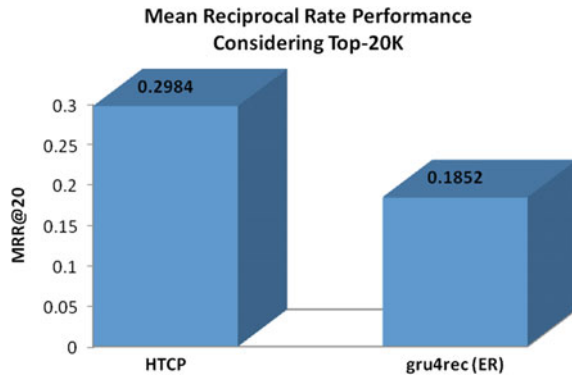


Fig. 9 Mean reciprocal rate performance considering the top-20 k recommendation



Further, the accuracy outcome is evaluated considering top-20 k recommendation as shown in Fig. 9. The TCP achieves an MRR outcome of 0.2984, and the ER method achieves an MRR outcome of 0.1852. From the result attained, it shows the TCP method achieves much superior MRR outcome when compared with ER methodologies considering the top-20 k recommendation. The overall result attained considering the top-10 k recommendation and top-20 k recommendation shows TCP achieves many superior MRR outcomes. TCP achieving better HR and MRR will aid in improving better profitability for the e-commerce environment.

5 Conclusion

Having the option to predict the users short-term interest for an online session is an exceptionally applicable issue practically speaking, which has brought expanded intrigue additionally up in the academic field as of late. Despite the fact that various diverse algorithmic methodologies were proposed throughout the years, no standard benchmark datasets and baseline algorithm exist today. In this work, we have looked at some of the recent and computationally complex calculations for session-based product recommendation. The experiment investigations on various diverse datasets show in many cases one of the simpler methods is able to outperform even the most recent methods based on recurrent neural networks in terms of the prediction accuracy. The work showed the effectiveness of using hybrid forecasting combining RNN and FFNN for forecasting user's choices in the current session. It is seen by combing RNN and FFNN, the hybrid model is efficient enough to minimize the cold-start and data sparseness problems in time-centric recommendation system. The experiment is carried out for evaluating the outcome achieved by the proposed TCP concerning existing recommendation methodologies. The overall result attained describes that HR outcome of 0.4571 and 0.2119 achieved by TCP and ER methodologies, respectively. Then, the HR outcome of 0.5347 and 0.4038 is achieved by TCP and ER methodologies for a list length of 10 and 20, respectively. Along with, MRR outcome

of 0.1978 and 0.0986 and 0.2984 and 0.1852 is achieved by TCP and ER methodologies for a list length of 10 and 20, respectively. Result attained in this work shows that the proposed TCP model is very efficient when compared with ER methodologies considering HR and MMR evaluation metrics. Thus, will aid in improving better productivity (i.e. better sales resulting in good profit) of the e-commerce environment. Further work will study performance evaluation considering varied datasets and also consider refining mathematically the TCP model.

References

1. Jesús Bobadilla D, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowledge-based systems* 46:109–132
2. Herlocker JL, Konstan JA, Borchers A, Riedl J (2017) An algorithmic framework for performing collaborative filtering. In *ACM SIGIR Forum*. vol 51. ACM, pp 227–234
3. Ricci F, Rokach L, Shapira B (2015) Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, pp 1–34
4. Jannach D, Ludewig M, Lerche L (2017) Session-based item recommendation in e-Commerce: on short-term intents, reminders, trends, and discounts. *User Model User-Adapt Interact* 27(3–5):351–392. <https://doi.org/10.1007/s11257-017-9194-1>
5. Yu F, Liu Q, Wu S, Wang L, Tan T (2016) A dynamic recurrent model for next basket recommendation. In: *Proceedings of the 39th international ACM SIGIR*
6. REF: Hernández S, Álvarez P, Fabra J, Ezpeleta J (2017) Analysis of users' behavior in structured e-commerce websites. In *IEEE Access* 5:11941–11958
7. Yu X, Jiang F, Du J, Gong D (2017) A user-based cross domain collaborative filtering algorithm based on a linear decomposition model. *IEEE Access* 5:27582–27589
8. Fu M, Qu H, Yi Z, Lu L, Liu Y (2018) A novel deep learning-based collaborative filtering model for recommendation system. *IEEE Trans Cybernet*. <https://doi.org/10.1109/TCYB.2018.2795041>
9. Mu D, Guo L, Cai X, Hao F (2018) Query-focused personalized citation recommendation with mutually reinforced ranking. *IEEE Access* 6:3107–3119
10. Ludewig M, Jannach D (2018) In: *Evaluation of session-based recommendation algorithms*. arXiv preprint [arXiv:1803.09587](https://arxiv.org/abs/1803.09587)
11. Hidasi B, Karatzoglou A (2017) Recurrent neural networks with top-k gains for session-based recommendations. arXiv preprint [arXiv:1706.03847](https://arxiv.org/abs/1706.03847)
12. Hidasi B, Karatzoglou A, Baltrunas L, Tikk D (2015) Session-based recommendations with recurrent neural networks. arXiv preprint [arXiv:1511.06939](https://arxiv.org/abs/1511.06939)
13. Hassan M, Hamada M (2017) A neural networks approach for improving the accuracy of multi-criteria recommender systems. *Appl Sci* 7:868
14. Palanivel K, Sivakumar R (2011) A study on collaborative recommender system using fuzzy-multicriteria approaches. *Int J Bus Inf Syst* 7:419–439
15. Jannach D, Gedikli F, Karakaya Z, Juwig O (2012) Recommending hotels based on multi-dimensional customer ratings. In: *Information and communication technologies in tourism*. Vienna, Austria, Springer, pp 320–331
16. Sanchez-Vilas F, Ismoilov J, Lousame FP, Sanchez E, Lama M (2011) Applying multicriteria algorithms to restaurant recommendation. In *Proceedings of the 2011 IEEE/WIC/ACM international conferences on webintelligence and intelligent agent technology*, Lyon, France, 22–27 August 2011, pp 87–91
17. Adrana M, Karatzoglou A, Hidasi B, Cremonesi P (2017) Personalizing session-based recommendations with hierarchical recurrent neural networks. In: *Recsys* pp 130–137

18. Wu S, Ren W, Yu C, Chen G, Zhang D, Zhu J (2016) Personal recommendation using deep recurrent neural networks in NetEase. In ICDE. pp 1218–1229
19. Malte L, Jannach D (2018) Evaluation of session-based recommendation algorithms. *User Model User-Adapted Interact* 28.4–5:331–390. Crossref. Web.
20. Greenstein-Messica A, Rokach L, Friedman M (2017) Session based recommendations using item embedding. In: *Proceedings of the 22nd international conference on intelligent user interfaces*. ACM, pp 629–633
21. Sreenivasa BR, Nirmala CR (2019) Hybrid location-centric e-Commerce recommendation model using dynamic behavioral traits of customer. *Iran J Comput Sci*. <https://doi.org/10.1007/s42044-019-00040-3>

A Time Series Cryptocurrency Price Prediction Using LSTM



B. Aditya Pai, Lavanya Devareddy, Supriya Hegde, and B. S. Ramya 

1 Introduction

The fast improvement of cryptocurrency standards during the most recent decade is one of the most disputable and questionable developments in the cutting-edge worldwide economy [1, 2]. Huge variances in the conversion scale of digital forms of money and their high instability, just as the absence of legitimate guidelines of their exchanges in many nations brought about noteworthy risks related to a venture into crypto resources [3]. This has prompted speculative research about their place and role in the foreseeable future [4]. Hence, the issue of creating fitting strategies and models for the very purpose of foreseeing costs for the cryptocurrency is important both for established researchers and for analysts, financial specialists and brokers.

A comparative analysis of the ARIMA forecasting properties with recurrent neural networks (RNNs) was conducted by Rebane and Karlsson [5] for cryptocurrencies such as Ethereum (ETH), Ripple (XRP), Litecoin (LTC), DASH, NEM (XEM), Siacoin (SC), Monero (XMR) and Stellar (STR). The results showed that ARIMA models have poorer forecasting properties than neural networks.

A study [6] reveals that macroeconomic and financial indicators have a relatively minute correlation with the composition of bitcoin prices. It is observed that variables like transaction volume affected the supply and demand of bitcoin. Thus, in our approach, we define this as a time series forecast problem which can be solved using LSTM and root means square error (RMSE) as an evaluation metric for closing prices where sequences of prices are “remembered” by the model during the process.

B. Aditya Pai · L. Devareddy · S. Hegde · B. S. Ramya (✉)
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore, India
e-mail: ramya.bs@nmit.ac.in

2 Related Work

Countless research has been conducted of the change of Bitcoin value using Time Series Analysis [7]. Research on Bitcoin value utilizing twitter sentiment information has been conducted by Lönnö and Stenqvist [8]. VADER—valence aware dictionary and sentiment reasoner which is a “consolidated vocabulary-based way to deal with measuring the person tweet assumption force and afterward assembled the scores into time-arrangement” has been considered here. Their model resulted in a precision of about 79%.

Linear SVM and Neural Network related research has been conducted by Kim [9]. The after effect of his trial deciphers that “the estimation of the upper bound and the kernel parameter has a delicate job in the presentation of Support Vector Machine’s expectation.” Their research resulted in an accuracy of 64.75%.

A suitable model that can predict the market price of Bitcoin is proposed in [10] using a few statistical models. A combination of MA and AR models has resulted in the ARIMA model. It uses the result of the autocorrelation charts where it finds out the order and then fits the data integrating MA and AR observations, resulting in better accuracy than the others. With AR and MA, the Bitcoin price is predicted with accuracy of 89.24 and 87.58%, respectively, while the ARIMA model resulted in an accuracy of 90.31%.

To analyze the price dynamics of Ethereum, Bitcoin and Ripple, authors in [11] propose a long short-term memory (LSTM) recurrent neural network and an artificial neural network (ANN). It was found that LSTM tends to rely more on short-term memory, while ANN tends to rely more on long-term memory, which tells us that—“efficiency of LSTM to utilize useful information hidden in historical memory is stronger than ANN. However, given enough historical information ANN can achieve a similar accuracy, compared with LSTM.”

In Lipo Wang and Edwin Sin’s paper [12], genetic algorithm, an artificial neural network ensemble approach, explores the relationship between the prices of Bitcoin on consecutive days. We infer that—“the ensemble method, GASEN, was able to perform well for the classification task with consistent accuracy of around 58% to 63%. With a simple trading strategy, the ensemble was able to obtain promising results in making profit of close to 85% in return.”

The paper by Sean McNally, Simon Caton and Jason Roche [13] aims to predict Bitcoin prices in USD using different models. They found that—“LSTM achieved the best accuracy, whereas the RNN achieved all-time low RMSE. The ARIMA prediction performed poorly in terms of accuracy and RMSE. Deep learning models like the RNN and LSTM are manifestly effective for Bitcoin prediction with the LSTM more capable for recognizing longer-term dependencies.”

3 Proposed System

We outline the problem on a one-to-one sequence problem and use a neural network with stacked long short-term memory (LSTM) cell design to predict the costs of cryptocurrencies. LSTMs have a foothold over standard feed-forward neural networks and recurrent neural networks [14, 15]. This is often attributable to their property of selectively remembering patterns for long durations of time [16]. We tend to train a model consisting of a network of LSTM cells stacked over one another and outline a pipeline for the deployment of the model. The “close” values of cryptocurrency values are taken as the sequential input for the model. Alternatively, the volume/open/high/low/market cap will be chosen subjects to the appliance of the model.

The proposed system consists of the subsequent phases:

Phase 1: Identify the currencies most talked about and identify their historical data containing following attributes—opening price, highest value, lowest value, closing value, volume and market cap.

Phase 2: Data pre-processing and EDA.

Phase 3: Create train-test-validation data partitions.

Phase 4: Build a predictive model using LSTM architectures.

Phase 5: Test the model and tune hyper parameters.

Phase 6: User integration and visualization with an interface.

Figure 1 describes the proposed system for predicting cryptocurrency prices using LSTM.

4 Data Pre-Processing

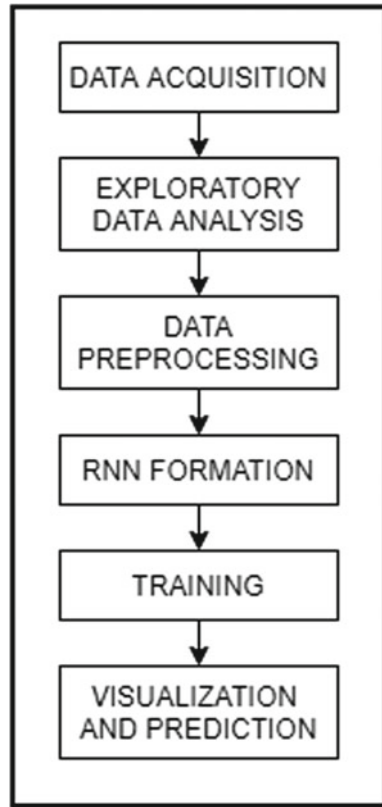
Many machine learning algorithms endeavor to discover trends in the information by looking at the highlights of data points [17, 18]. In any case, there is an issue when the highlights are on radically various scales. To solve this, we use scalar. The main objective of scalar is to make each data point have a similar scale so each component is equally significant.

We will be using the min–max scalar is one of the most well-known approaches to standardize information. For each element, the base estimation of that element gets changed into a 0, the most extreme value gets changed into a 1, and all other weighted points get changed into a decimal somewhere in the range of 0 and 1.

The formula for min – max scalar is : $m = (x - x_{\min}) / (x_{\max} - x_{\min})$

Min–max scalar is sensitive to the presence of outliers. We use this trait in order to not lose information about our data, i.e., the data’s volatility.

Fig. 1 Flow of control for proposed system



5 Model Architecture

The model basically recreates a network to work like the neurons in our brain. The network is trained to give predictions based on being trained on existing data. Deep neural networks have been used in accounting of economic markets. In our study, we use an overarching profound learning model to break down and anticipate cryptocurrency value elements through long short-term memory. The LSTM incorporates three layers, each having ten nodes. Each LSTM cell state contains three gates—forget gate, input gate and an output gate. LSTM controls the removal or addition of data through the gates to accomplish the function of either ignoring or committing memory (Fig. 2).

Figure 3 Mathematical Shown above is the complete architecture of RNN developed during this project. We have used LSTM cells with Adam optimizer and MSE to calculate error. It contains 4 layers. The first layer is the input layer. It acts as the entry point of the cryptocurrency data into the RNN. The second and third layers are LSTM layers. Its input space dimensionality is 1 and 256, respectively, while the output space dimensionality is 256 for both. They use hyperbolic tangent (tanh) as

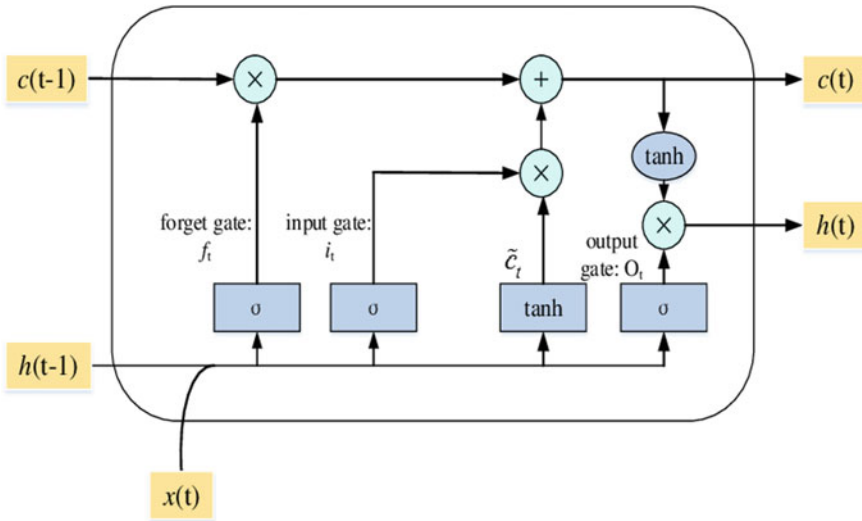


Fig. 2 Single LSTM cell architecture

Fig. 3 Mathematical model of LSTM

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\vec{C}_t = \tanh(W_C \cdot [h_t, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \vec{C}_t$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = O_t * \tanh(C_t)$$

the activation function and sigmoid as the activation function for the recurrent step. The fourth and final layer is a dense layer. It computes the dot product of its input matrix and weight matrix, and then adds to it the bias, producing the prediction value (Fig. 4).

The following graphs show mean squared error vs epoch number plots obtained after training the model for Bitcoin, Ethereum, Litecoin and Ripple time series data, dating back to 2017 (Figs. 5, 6, 7, 8).

Fig. 4 Architecture of the RNN

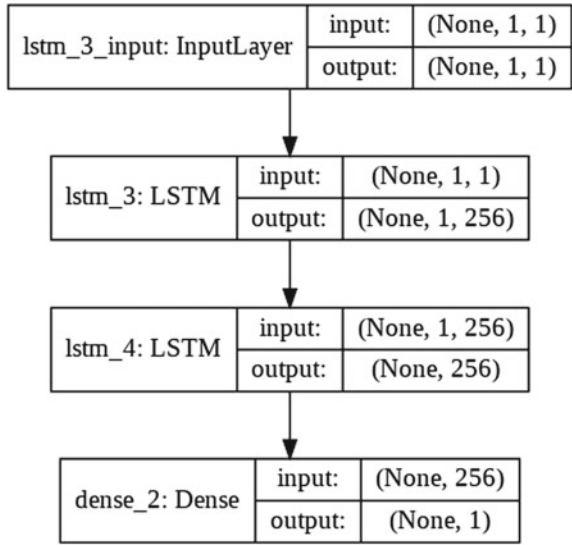


Fig. 5 MSE graph for bitcoin (BTC)

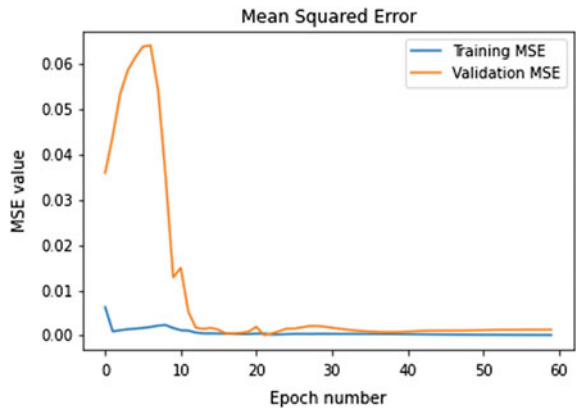


Fig. 6 MSE graph for ethereum (ETH)

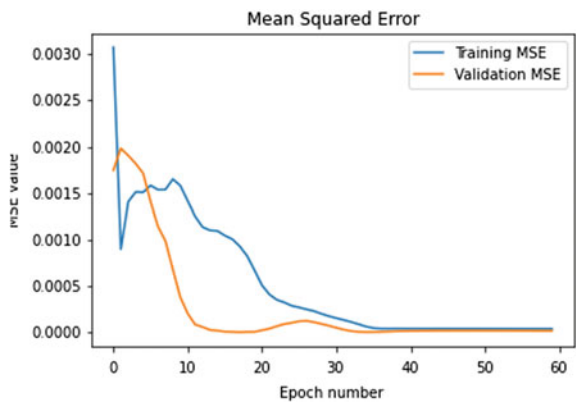


Fig. 7 MSE graph for litecoin (LTC)

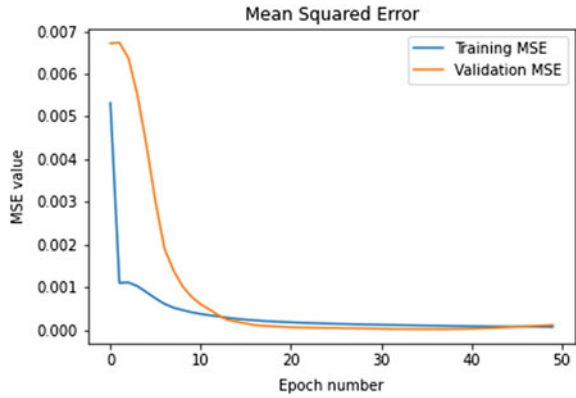
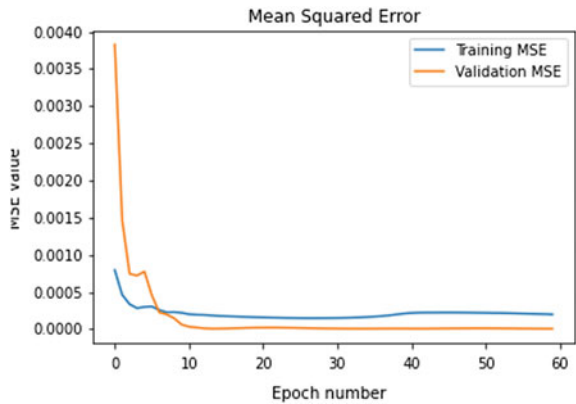


Fig. 8 MSE graph for ripple (XRP)



6 Results

After training the 4 datasets using the proposed model, the following results were observed. In terms of mean squared error (MSE), the LSTM model shows superior performance when compared with an ANN model in general while predicting cryptocurrency values. The MSE values obtained after training datasets for Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) and Ripple (XRP) are $1.4925 e^{-04}$, $6.0898 e^{-05}$, $5.0990 e^{-05}$, $2.4237 e^{-04}$, respectively.

The comparison of results existing between support vector regression (SVR), neural network (NN), decision tree regression (DTR) [19], and our proposed LSTM model has been tabulated below (Table 1):

The following graphs showcase the predicting capabilities of the proposed model (Figs. 9, 10, 11, 12).

Table 1 Comparison of results of existing and proposed LSTM model

Cryptocurrency	RMSE			
	SVR	NN	DTR	LSTM (Proposed model)
Bitcoin (BTC)	0.0209	0.0161	0.017	$1.49e^{-04}$
Ethereum (ETH)	0.0163	0.0149	0.465	$6.09e^{-05}$
Ripple (XRP)	0.037	0.0115	0.467	$2.42e^{-04}$

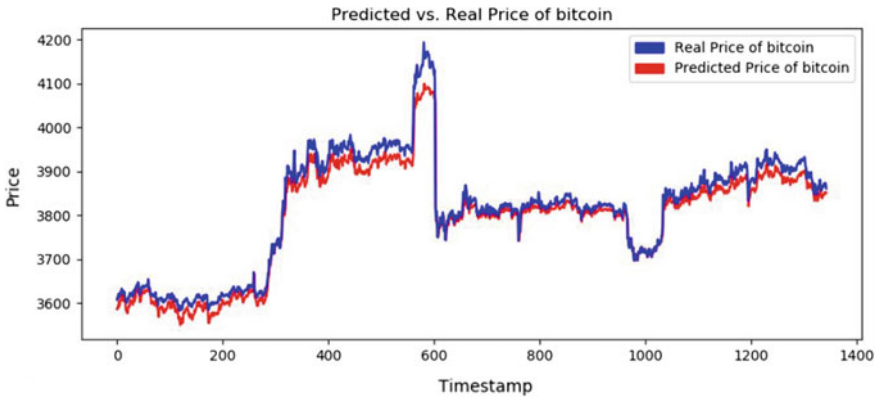


Fig. 9 Prediction graph for bitcoin (BTC)

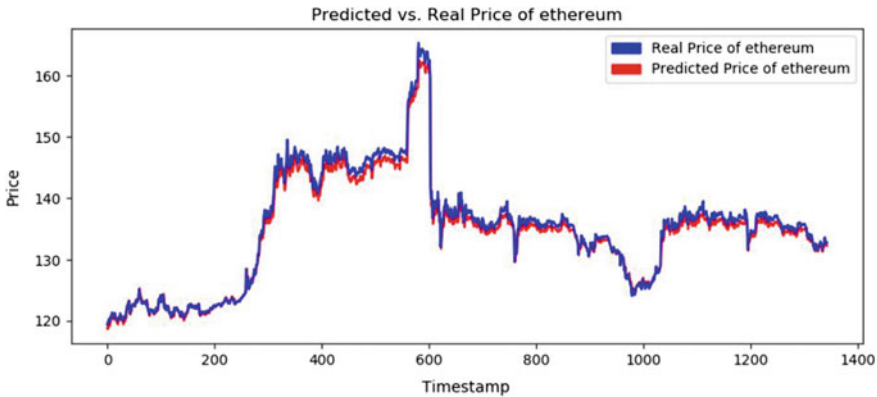


Fig. 10 Prediction graph for ethereum (ETH)

7 Conclusion and Future Work

Cryptocurrency price forecasting is an important component of automation in forex and algorithmic trading [20]. A model that predicts prices with minimal error can be integrated with software’s and used in accordance with trading strategies [21].

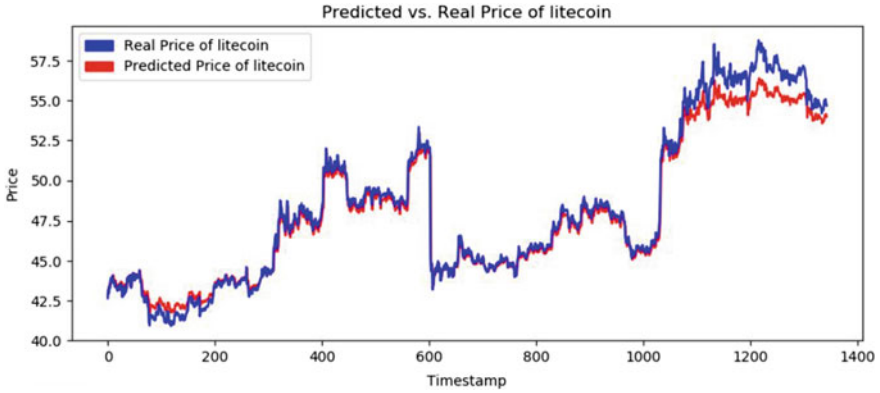


Fig. 11 Prediction graph for litecoin (LTC)

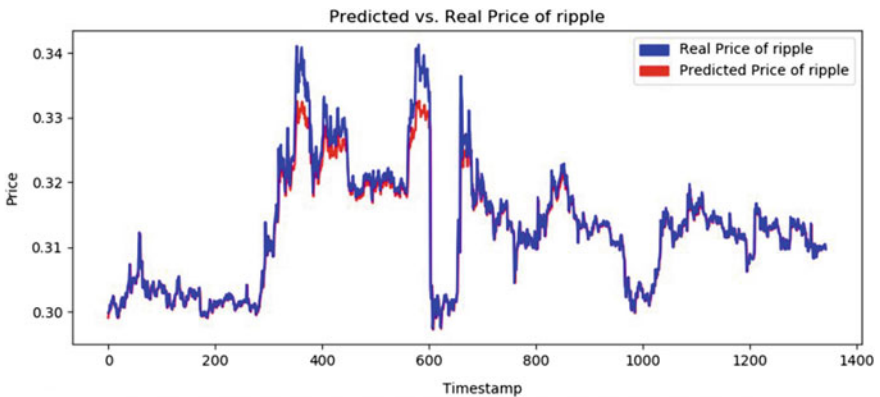


Fig. 12 Prediction graph for ripple (XRP)

Moreover, this forecasting model that uses LSTM can also be applied to other features like close prices, market capitalization, etc.

This model can be deployed as a service on the cloud or even as a rest API that is scalable according to usage and market fluctuation. The existence of this model deems useful for multiple economic frameworks by increasing explain ability and reducing risk in cryptocurrency domain.

References

1. Aggarwal A, Gupta I, Garg N, Goel A (2019) Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction. In: 2019 Twelfth international conference on contemporary computing (IC3), Noida, India

2. Misnik A, Krutalevich S, Prapakpenka S, Borovykh P, Vasiliev M (2019) Impact analysis of additional input parameters on neural network cryptocurrency price prediction. In: 2019 XXI international conference complex systems: control and modeling problems (CSCMP), Samara, Russia, 2019
3. Wimalagunaratne M, Poravi G (2018) A predictive model for the global cryptocurrency market: a holistic approach to predicting cryptocurrency prices. In: 2018 8th international conference on intelligent systems, modelling and simulation (ISMS), Kuala Lumpur, Malaysia
4. Bai C, White T, Xiao L, Subrahmanian VS, Zhou Z (2019) C2P2: a collective cryptocurrency up/down price prediction engine. In: 2019 IEEE international conference on blockchain (Blockchain), Atlanta, GA, USA, 2019.
5. Rebane J, Karlsson I (2018) Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: a comparative study. In: SIGKDD Fintech'18, August 2018, London, UK. https://fintech.kdd2018.a.intuit.com/papersDSF2018_paper_papapetrou.pdf
6. Ciaian P (2015) The economics of Bitcoin price formation. *Appl Econ* 48:1799–1815
7. Karasu S, Altan A, Saraç Z, Hacıoğlu R (2018) Prediction of bitcoin prices with machine learning methods using time series data. In: 2018 26th signal processing and communications applications conference (SIU), Izmir
8. Stenqvist E, Lönnö J Predicting Bitcoin price fluctuation with Twitter sentiment analysis. unpublished
9. Kim K (2003) Financial time series forecasting using support vector machines. *Neurocomputing* 55(12):307319
10. Roy S, Nanjiba S, Chakrabarty A (2018) Bitcoin price forecasting using time series analysis. In: 21st International conference of computer and information technology, IEEE
11. Yiyiing W, Yeze Z (2019) Cryptocurrency price analysis with artificial intelligence. In: 5th International conference on information management, IEEE
12. Sin E, Wang L (2017) Bitcoin price prediction using ensembles of neural networks. In: 13th International conference on natural computation, fuzzy systems and knowledge discovery. IEEE
13. McNally S, Roche J, Caton S (2018) Predicting the price of bitcoin using machine learning. In: 26th Euromicro International conference on parallel, distributed, and network-based processing, IEEE
14. Rathan K, Sai SV, Manikanta TS (2019) Crypto-currency price prediction using decision tree and regression techniques. In: 2019 3rd international conference on trends in electronics and informatics (ICOEI). Tirunelveli, India
15. Almasri E, Arslan E (2018) Predicting cryptocurrencies prices with neural networks. In: 2018 6th International conference on control engineering and information technology (CEIT), Istanbul, Turkey
16. Jay P, Kalariya V, Parmar P, Tanwar S, Kumar N, Alazab M (2020) Stochastic neural networks for cryptocurrency price prediction. *IEEE Access* 8:82804–82818
17. Rane PV, Dhage SN (2019) Systematic erudition of bitcoin price prediction using machine learning techniques. In: 2019 5th international conference on advanced computing and communication systems (ICACCS), Coimbatore, India
18. Alessandretti L, ElBahrawy A, Maria Aiello L, Baronchelli A (2018) Anticipating cryptocurrency prices using machine learning. *Complexity* 8983590:16
19. Pintelas E, Livieris IE, Stavroyiannis S, Kotsilieris T, Pintelas P (2020) Investigating the problem of cryptocurrency price prediction: a deep learning approach. In: Maglogiannis I, Iliadis L, Pimenidis E (eds) *Artificial intelligence applications and innovations. AIAI 2020. IFIP advances in information and communication technology*, vol 584. Springer, Cham
20. Liang J, Li L, Chen W, Zeng D (2019) Towards an understanding of cryptocurrency: a comparative analysis of cryptocurrency, foreign exchange, and stock. In: 2019 IEEE international conference on intelligence and security informatics (ISI), Shenzhen, China
21. Mohanty P, Patel D, Patel P, Roy S (2018) Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices. In: 2018 7th international conference on reliability, infocom technologies and optimization (Trends and Future Directions) (ICRITO), Noida, India

Smart Bus Stop Reminder with Connected Bus Stops



C. Shashi Kumar, Venkat Charan, K. R. Suhas Gowda,
N. Sushmitha Gowda, and M. R. Sowmya

1 Introduction

As a means of public transportation, buses are widely used in our country. Passengers are facing many problems in bus management where they are still using the manual ticket generation method. They have to wait for the conductor to issue the ticket for everyone on the bus. Manual ticket generation is the biggest problem for conductors during peak hours. To overcome this situation, a web and android application has been developed where the passengers can use it easily and efficiently to travel during the peak and normal hours.

The passenger provided an NFC tag where it can be tapped on the NFC reader. As soon, the NFC reader is tapped the details of the passenger are available to the conductor in this application. The conductor does not have to carry the NFC scanner always with him. If he holds an NFC-based phone, the passenger can just tap his card to the conductor's phone. This reduces ticket issuing and when they are alighting down from the bus the NFC tag has to be tapped again. When the passenger tap on the first time, source (Board) location will be taken and the second time tapping results in the destination (alight) location. The amount will be calculated based on the distance traveled between pick up and drop off of the passenger. The amount will be automatically debited from the passenger which is recharged by the admin. Admin can add the customer and set the new landmarks and he can add extra buses during peak hours or he can remove the buses whenever required. Admin recharges the passenger's account online and can keep track of the buses and its location.

C. Shashi Kumar · V. Charan · K. R. Suhas Gowda · N. Sushmitha Gowda · M. R. Sowmya (✉)
Department of CS & E, Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: sowmya.mr@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_51

2 Literature Survey

FoisalMahediHasan et al. [1]

“RFID-based ticketing for public transport system: Perspective Mega city Dhaka, 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 6, pp. 459-462, 2016. The paper-based public transport ticketing system, prevailing in the Mega city Dhaka (Bangladesh), introduces severe malfunction in the system, a malicious argument among the public, corruption, and most of all traffic jams. This paper suggests a much more public-friendly, automated system of ticketing as well as the credit transaction with the use of RFID-based tickets.”

Hua-Ling [2]

“Origin-Destination Demands Estimation in Congested Dynamic Transit Networks”, International Conference on Management Science & Engineering (14th), 2017. This paper investigates the problem of estimation of time-dependent passenger origin-destination (OD) matrices in congested transit networks where real-time updated passenger counts and prior OD matrices are available. A bi-level programming model is proposed for the dynamic estimation of the passenger OD matrix. The upper level minimizes the sum of error measurements in dynamic passenger counts and time-dependent OD matrices, and the lower level is a new schedule-based dynamic transit assignment model that can determine simultaneously the dynamic average travel costs and route choices of passengers in congested transit networks. The lower-level problem can be formulated as a variational inequality problem. A heuristic solution algorithm is adapted for solving the proposed bi-level programming model. Finally, a numerical example is used to illustrate the applications of the proposed model and solution algorithm.”

3 Architectural Design

(A) Admin application

Admin will have the authority to add new buses or to delete buses. He can create the passengers' application and monitor buses. The admin can add a passenger by filling basic information like name, age, phone number, and address. After adding the details, NFC tag is generated, and it is given to the passenger. He can create a login id and password for each bus. The conductor can use the application to know the details of the passengers when they board the bus by tapping the NFC tag to the NFC reader twice that is, while boarding and alighting the bus. Admin can keep track of the buses, the number of passengers traveled every day, the total amount collected, and bus location. A new bus and bus stops can be added to the route, can be deleted from the route as and when required. This information will be available in the conductor application as well. Extra buses to certain places can be added during peak hours and can be removed during a normal hour (Fig. 1).

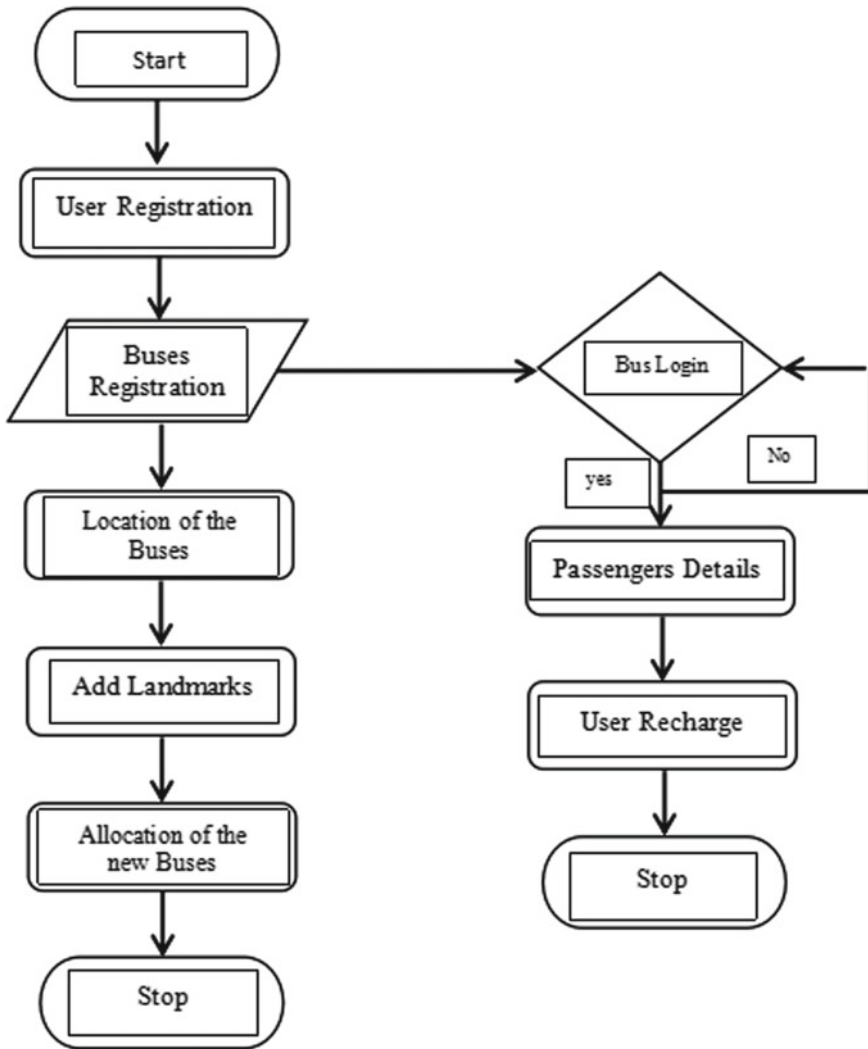
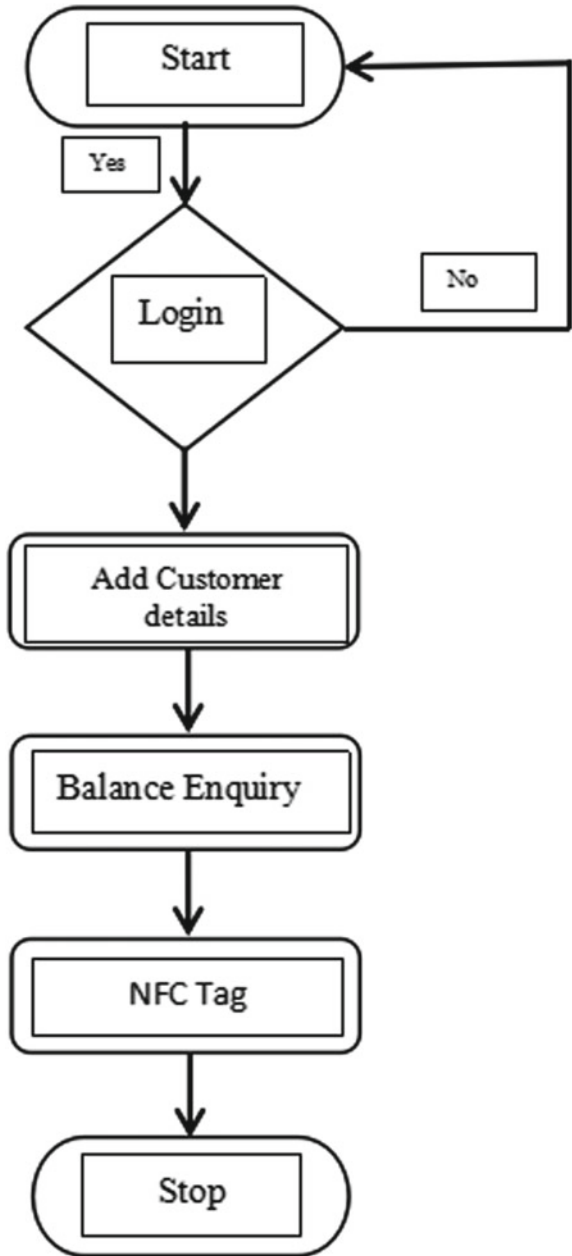


Fig. 1 Admin and read application

(B) Passengers application

Admin creates the passengers application. After its creation, the passenger can get their own NFC tag by filling basic details, and then, this information has to be added to the NFC tag. After creating it successfully, passengers should log in to check balance enquiry on this application. Once this is been set up by the passenger the NFC tag can be used on the NFC reader, while boarding and alighting the bus. If this is done successfully, the conductor will get the information about the passengers, and also paper waste can be reduced (Fig. 2).

Fig. 2 Admin write application



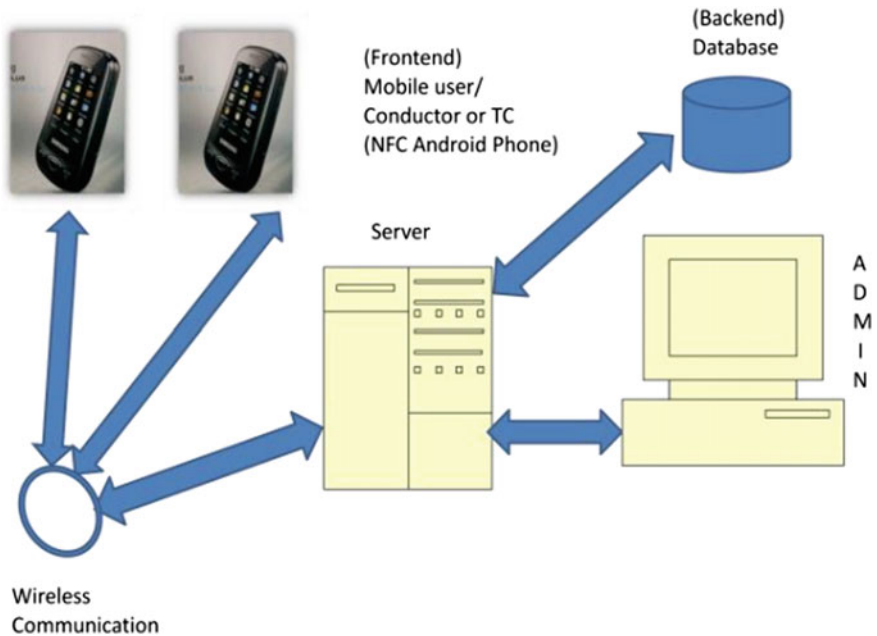


Fig. 3 NFC application block diagram

4 Methodology

This system consists of one web admin application, two android applications, and wireless communication (NFC tag). All of these are connected to a centralized server. In the database data is stored, where all the passengers' details and bus locations, and the total amount which is added to through the application will get stored. One application is to write data to the NFC tag, and the other is used to read the NFC tag information. Admin can recharge the passengers. All the data are stored in the database (Fig. 3).

5 Problem in the Existing System

The existing system has a drawback in manual ticket issuing. The conductor should calculate the amount to issue tickets to the passengers. From the year 2014, ETM machine (electronic ticketing machine) is used for generating the ticket by setting the pickup location and drop location.

- A. **Waiting for the conductor to take the ticket.**

If the bus is full, each passenger has to wait for their turn to receive a ticket. If the TC officer boards the bus in between then the passenger who is waiting may face a penalty. When the buses are full, some people may travel without a ticket. The passenger may also face issues when the conductor stops the bus on the way to issue tickets.

B. No refund of the balance.

If the passenger is not tendering the exact amount and give more amount to the conductor then there are chances of passengers end up giving money more than required or wait until the conductor gets money to give them the change or they may forget sometimes and alight the bus or may also get into a situation where money has to be divided between two or more passengers.

C. Excessive waste of paper.

When the conductor issues the ticket, there are chances of losing it which may cause trouble during checking. By not issuing tickets, paper can be saved and also littering.

6 Proposed System

By considering the drawbacks of the existing system, an application is developed where the user can easily use it after the admin issues an NFC tag. Admin collects all basic information from passengers before issuing the NFC card. The passenger needs to submit a government-issued ID card while submitting basic documents. With the use of this NFC tag, the ticket can be easily generated by tapping it to the NFC reader when the passenger boards the bus. As the ticket can be generated within seconds to the passenger, the ticket issuing problem can be solved. While alighting from the bus, the NFC tag is tapped again which helps in calculating the distance traveled and the amount will be debited automatically. In this pandemic situation, passengers should maintain social distancing so, while entering or by exiting from the bus it will be easy work for a conductor to tap and get the information from passengers even when there is a rush at the bus stop. NFC card is generated by admin. He has complete information about the passenger. When the passenger taps his card on the conductor's device, the information is accessed and sent to an admin immediately through a conductor's device. The passenger is notified by a normal text message to their registered mobile number. This e-ticket generation saves paper, and other issues faced by the passengers when they tender more money. This application helps the passenger travel easily and also helps the bus management to keep track of the amount collected and the list of passengers.

Implementation

Modules implementation

A. User registration

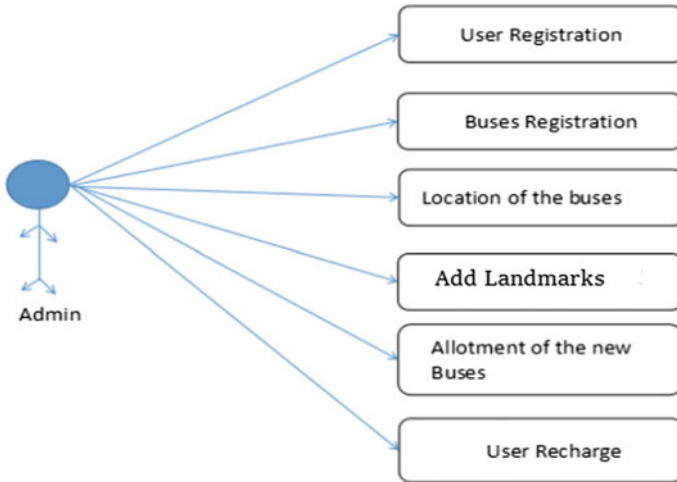


Fig. 4 Admin web application

- B. Buses registration
- C. Location of the buses
- D. Add landmarks
- E. Allotment of new buses
- F. NFC tag (Fig. 5)

Admin can register the user, bus, keep track of the bus location, landmark, allotment of new buses, bus stops, adding, deleting buses and bus routes, add and delete passengers, and recharge customer for the required amount.

Admin adds the user by filling the user registration form which requires the details of the user. This creates an NFC tag. Admin creation will permanently register to the NFC tag so that the passengers can use it after adding any other details if required. The data addition or deletion will be notified to the passenger to their valid contact number. This confirms the successful registration and usage of tags, while traveling. The admin will create an individual login for the buses based on the bus number which helps in tracking of the bus and also gets the list of the passengers traveled daily (Fig. 4).

Conductor android application is used to get the passengers list when the passengers get on the bus and taps on the NFC reader, for every bus, there is a unique login id based on the bus number. With this, the list of the passengers and the amount can be easily verified /collected (Fig. 5).

Here user gets the application by the admin. The user can get the NFC tag by admin or on his own by using the NFC writer. The user requires the NFC writer to create the NFC tag. NFC writer helps in storing the passenger details on the NFC device. After the user obtains the NFC tag commuting through buses will be easy. This application also allows the passenger to check the remaining balance in their account. The amount has to be recharged by admin (Fig. 6).

Fig. 5 NFC read application

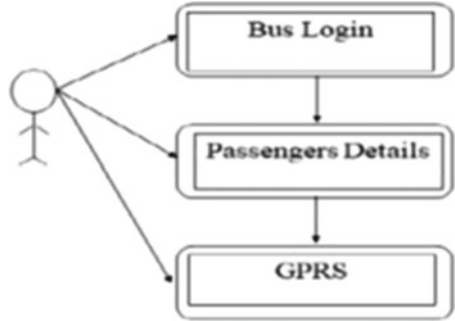
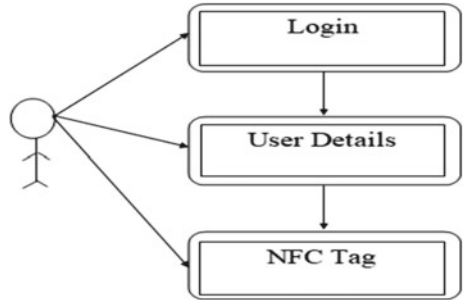


Fig. 6 NFC write application



7 Result

A. Web Application Admin

See (Figs. 7, 8)



Fig. 7 Admin login page. Admin has to log in to the application to perform various operations



Fig. 8 This is used to enter passenger details. This is used to add new passengers, edit existing passengers, and to delete a passenger

Figure 9a–d Admin is going to create the bus login id. This page is used to add a route, delete route. This gives the details of the bus like the bus number, the total number of passengers traveled, the amount collected, and existing routes (Figs. 10, 11 and 12).

B. Android Application

See Figs. 13, 14, 15 and 16.

8 Conclusion

The smart bus stop remainder with connected bus stops helps in the easy transaction and it also provides the passengers hassle-free traveling on the bus. With the help of NFC tags and the reader, the passenger details, distance traveled by the passenger can be easily obtained. For each bus, there will different (unique) login IDs so that tracing and tracking of the passengers is possible with the help of the passengers’ list and the amount collected by the passengers. Admin can track the bus and set the landmark where ever it is required and admin can even add the bus or delete the bus by looking at the previously collected passenger list at a certain time. This application makes passengers travel comfortably. During this pandemic situation, e-ticket helps the passengers to travel safely by avoiding a person to person contact, and also helps in tracking of passengers.



Fig. 9 a View route b Add routes c Existing routes d Landmarks

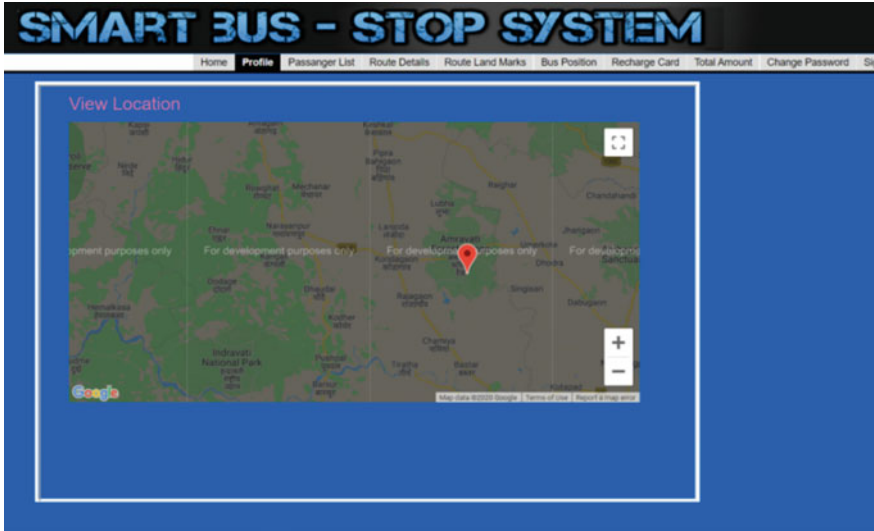


Fig. 10 This is used to keep track of the bus location by the admin. This information is updated every 10 s

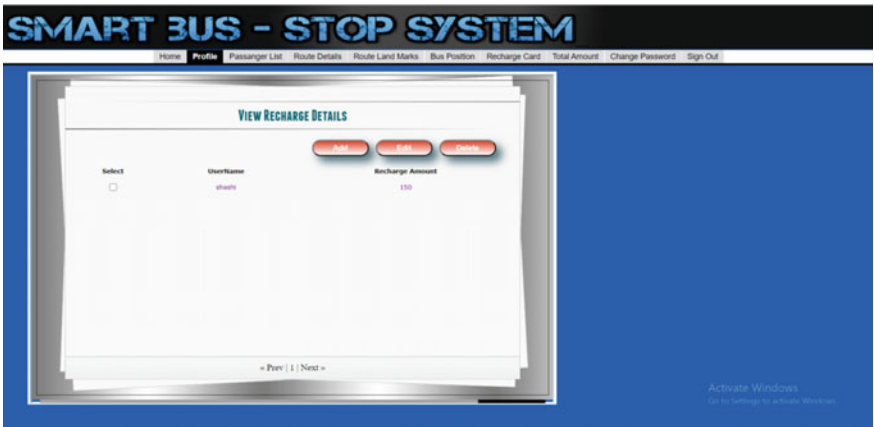


Fig. 11 This is used by the admin to recharge the passenger account and to view the remaining balance

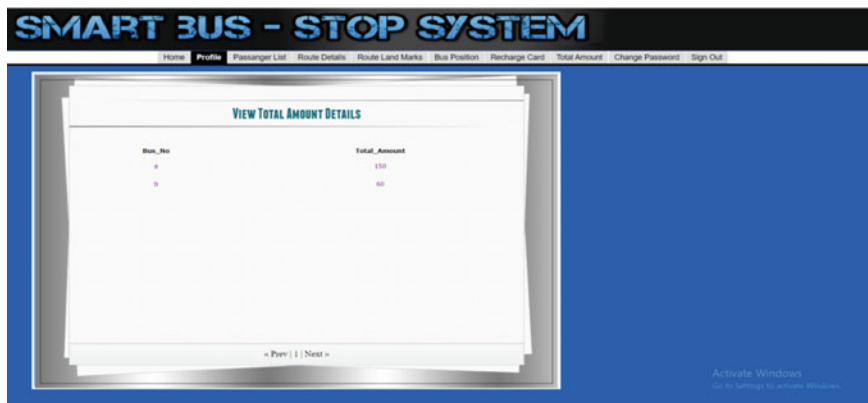


Fig. 12 This page gives the bus number and the amount collected by the passengers traveled on the bus

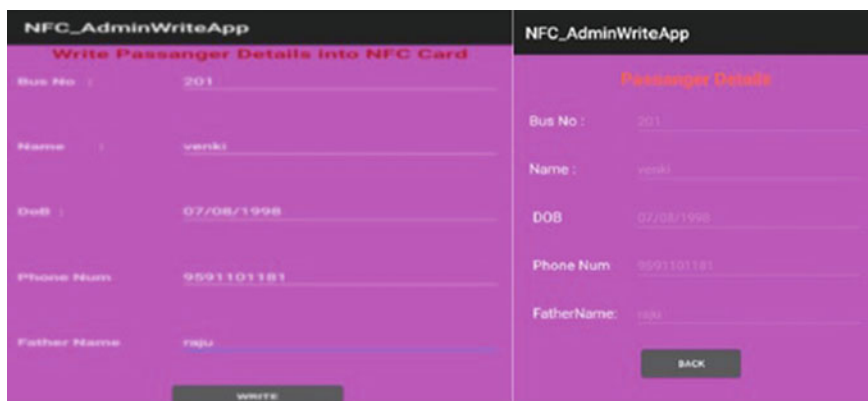
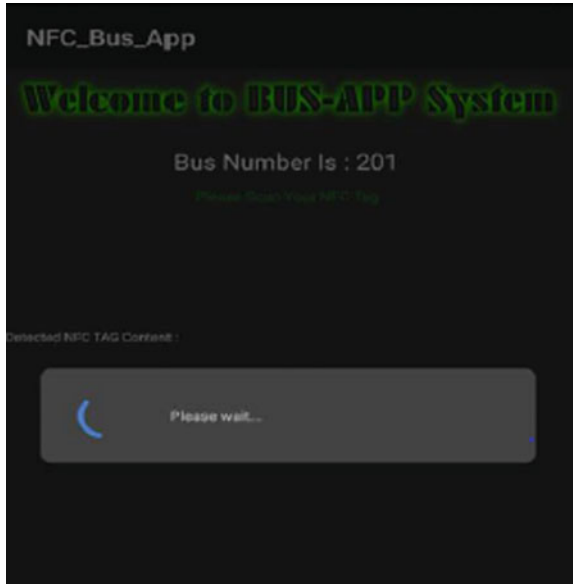


Fig. 13 Android write application: admin will the user details to the NFC tag. admin can view the NFC tag details

Fig. 14 NFC read application where the user will tap when boarding and alighting the bus



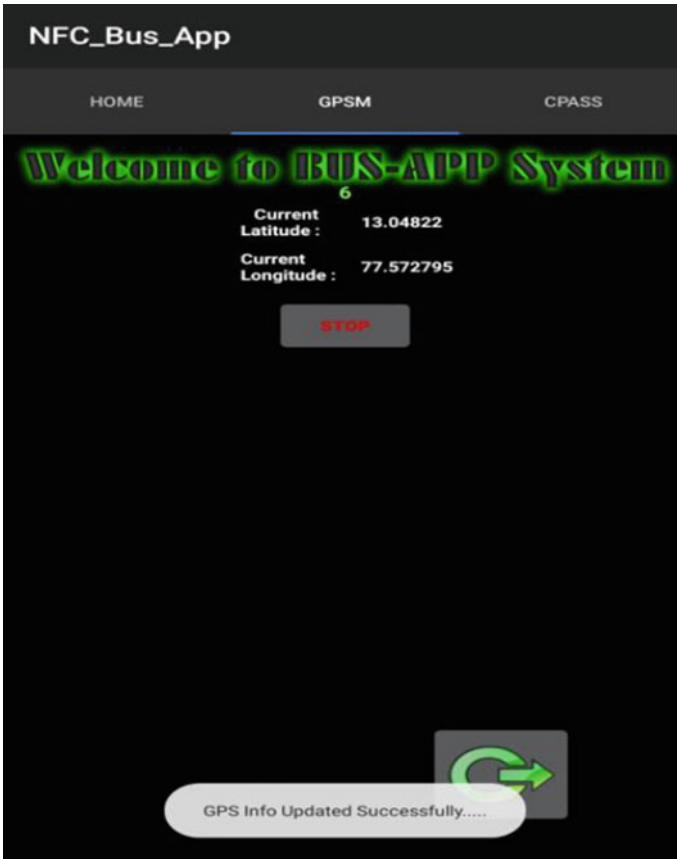


Fig. 15 To obtain the location and the location of the bus will be sent to the admin every 10 s

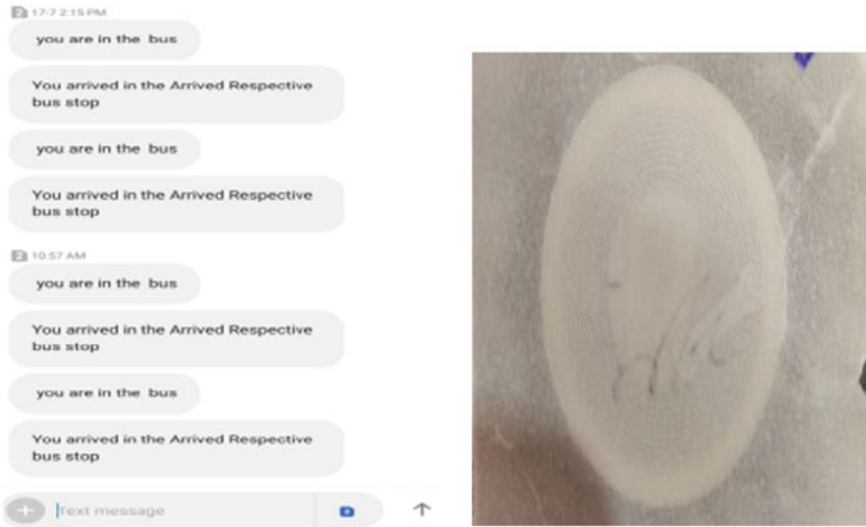


Fig. 16 Passengers will get the message when boarding and alighting the bus and NFC tag

References

1. Hasan FM et al (2016) RFID-based ticketing for public transport system: perspective megacity Dhaka. In: 3rd IEEE international conference on computer science and information technology (ICCSIT), vol 6, pp 459–462
2. Hua-Ling R (2017) Origin-destination demands estimation in congested dynamic transit networks. In: International conference on management science and engineering (14th)
3. Kumar (2015) RFID based embedded system for vehicle tracking and prevention of road accident. *Int J Eng Res* **1**(6). Near field communication: ECMA international standardizing information and communication systems
4. Das R (2008) NFC-enabled phones and contactless smart cards 2008 2018. *Card Technol Today* **20**
5. NFC in Public Transport-January (2011) <http://www.nfc-forum.org> Transport Infrastructure. In: Fourth international workshop with focus on near field communication 2012 4th international workshop on NFC. IEEE. 978-0-7695-4678-0/12 \$26.00 © 2012. <https://doi.org/10.1109/NFC.2012.1413>

Routing of Flits in Parallel Input Interface Scenario in a Generalized Network-On-Chip Framework Using Wormhole Flow Control Algorithm



R. Uma, H. Sarojadevi, and V. Sanju

1 Introduction

In network-on-chip (NoC) [1–6], the packets can be transmitted using three routing mechanisms which describes how the packets are processed and forwarded to the destination node by the intermediate nodes. In store-and-forward, before forwarding the message to the selected adjacent nodes it will be stored in the intermediate nodes. Here the packets are not allowed to get transmitted out of a node until received fully. Due to this, extra delay is incurred in the process. When a packet reaches at an intermediary node and its route towards destination is open, it can get transferred to the outgoing path without waiting for the complete data to get deposited in the intermediate node. A storage mechanism is needed to encounter a busy path. In virtual cut-through routing, a physical path is formed between the sender and receiver nodes during the circuit establishment phase. In the packet transmission phase, the packet is transferred down the route to destination. In wormhole routing strategy [7], a packet is partitioned into several small units for communication. The path is managed by the header flit (or flits). When a module receives the header flit(s) of a message, a look up table gets created in the node and the look up table is used for managing the header information. Header flits decide on the next path based on the routing algorithm they used. The header is followed by the remaining data units in a systematic order. For reaching the destination, the flits first move along the vertical direction and then travel

R. Uma (✉) · H. Sarojadevi
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: uma.r@nmit.ac.in

H. Sarojadevi
e-mail: sarojadevi.n@nmit.ac.in

V. Sanju
Reva Institute of Technology and Management, Bangalore, India
e-mail: sanju.v@reva.edu.in

towards horizontal direction. Interleaving of all the flits of a message with the flits of other messages is not allowed. All the data units need to stay in adjacent channels of the network since most of the flits does not contain any routing information. When the header unit of a message is stopped from traversing, then all the other data units of a message stop proceeding and prevent the progress of any other message which requires the same routes for movement. Wormhole switching strategy reduces the memory bandwidth in the nodes required to route the messages. Data units of the messages can be stored in small buffers. Wormhole routing of flits through parallel input interface scenario in a mesh network topology on NoC framework using Verilog simulation has been explored in this paper.

2 Motivation

In the NoC scenario, the routing strategies [8] ‘store-and-forward’ and ‘virtual cut-through’ do not address the issues of transmission delay and buffer utilization at the same time. These issues can be addressed using wormhole routing strategy. The motivation of this work is to route the flits in parallel input interface in a generic framework using wormhole routing strategy, such that a system with better performance, less transmission delay and limited buffer utilization can be attained.

3 Analysis

3.1 Routing Strategy

Wormhole Routing: Fig. 1 shows flowchart for the transmission of flits in a network using wormhole routing. A particular node, termed current node, at a time may receive flits from other nodes. Alternatively, if the current node itself is sending the packets to other nodes, it first converts the packets to flits. A current node then checks whether its address is equal to the destination node address. If so, then convert the flits received to form the packet. Otherwise after identifying the output side the node, check whether the output side is busy or not. If not busy it sends the flits forward, or else it stores the flits in input buffer and transmits at a later time when the next node is free.

When the size of the data in the packet is changed, the number of flits formed would be directly proportional to the size. The flit header captures the information about the size of the data which is carried by the flit. This makes sure that at the time the flits are received, it would be possible to fully receive and recreate the packet from all the flits received without any loss of data. In wormhole routing, smaller buffer size is required and latency is least.

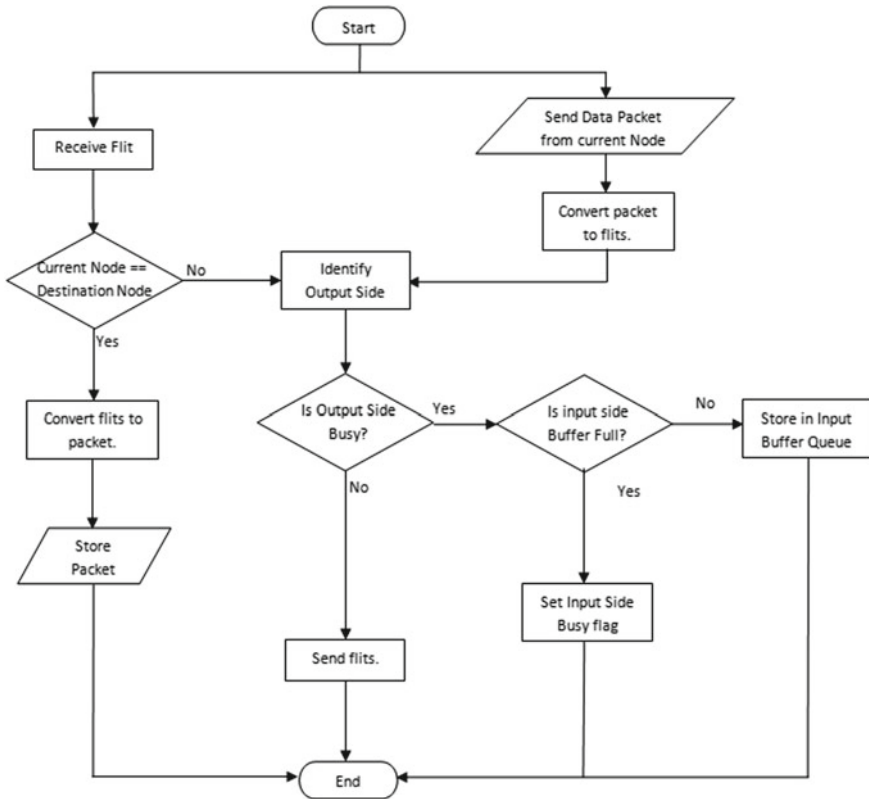


Fig. 1 Wormhole routing

Virtual cut-through Routing: Figure 2 shows the flowchart for the transmission of packets in a network using virtual cut-through routing [9]. Situation may be that either the packets get received by the current node from other nodes or the current node itself is sending the data packet. The current node checks whether its address is equal to the destination node address. If current node itself is the destination node, then packet can be stored. Otherwise after identifying the output side the node checks whether the output side is busy; if not, it will send the packets, else store in the input buffer and then later transmit the packet.

Store-and-Forward Switching: Fig. 3 shows the flowchart for the transmission of packets in a network using store-and-forward mechanism. In store-and-forward [10] routing the packets get stored completely and verifies that the last packet has also received before transmitting it to another node. Further, the node checks whether the current module itself is the target module. If so, the data will be stored in the current module itself. Otherwise the output side is identified and checks whether the output side is busy, if not transmit the packets otherwise store it in input buffer and later gets transmitted.

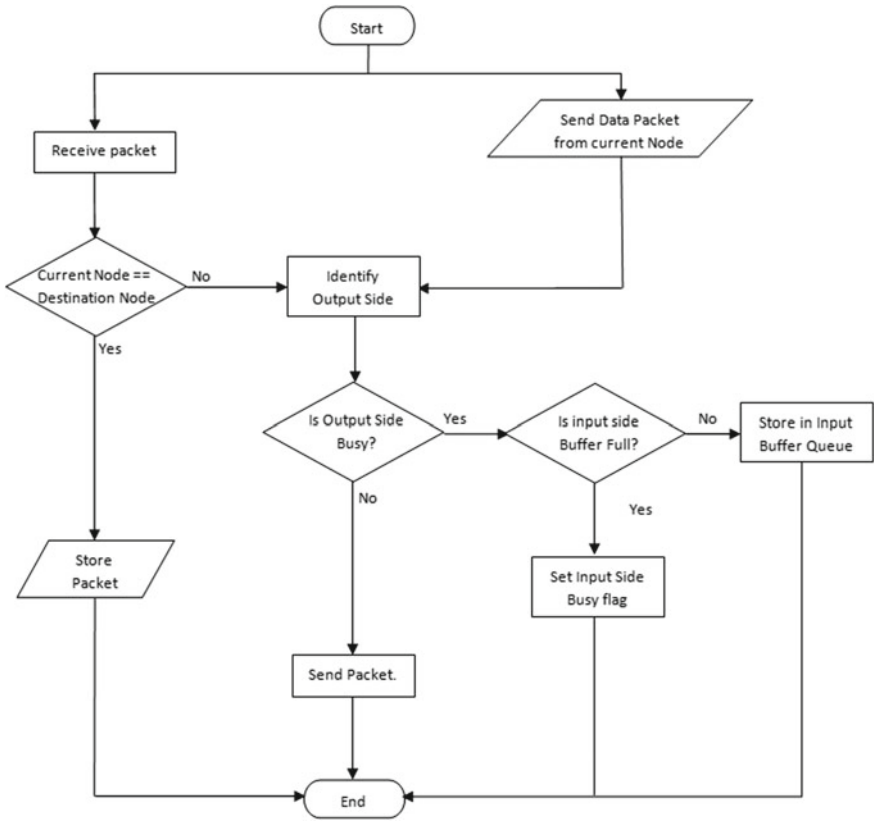


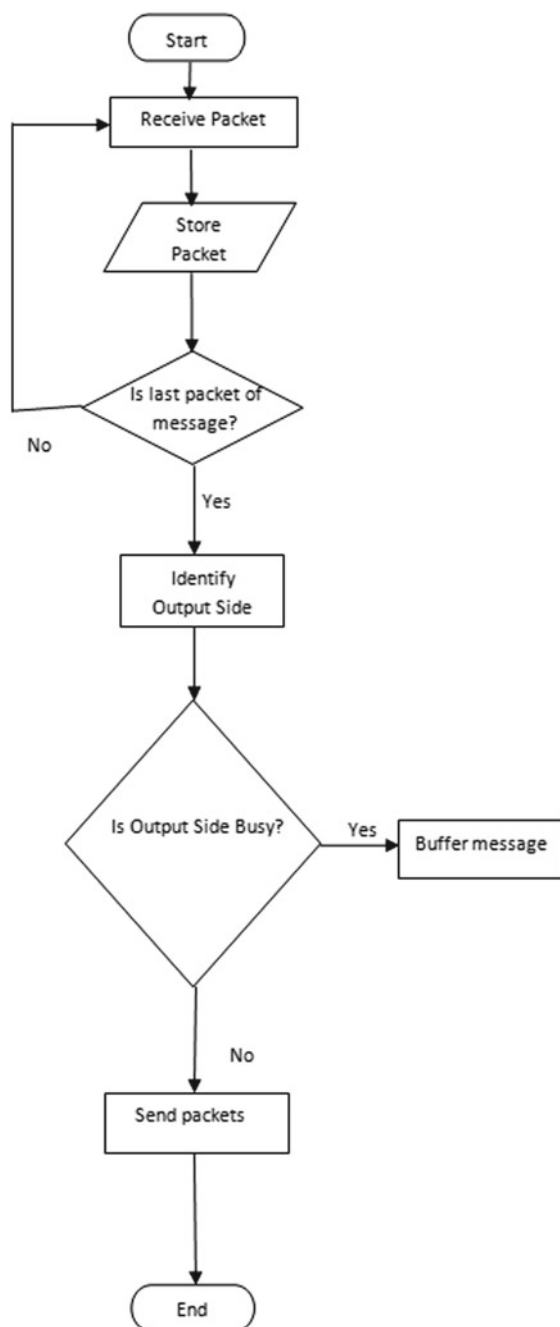
Fig. 2 Virtual cut-through routing

The routing strategies such as store-and-forward as well as the Virtual cut-through had either a latency or a reliability problem. In both these cases, there is a need to have larger buffer size to store the full packet information as required.

In wormhole routing as the messages are divided into flits, there is only a need of small buffer size which would be for a flit, as compared to the other routing strategies like store-and-forward and virtual cut-through. Figure 4 represents the time taken by a packet to reach the destination using different strategies.

4 Design

The framework and the internal components are discussed in this section.

Fig. 3 Store-and-forward routing

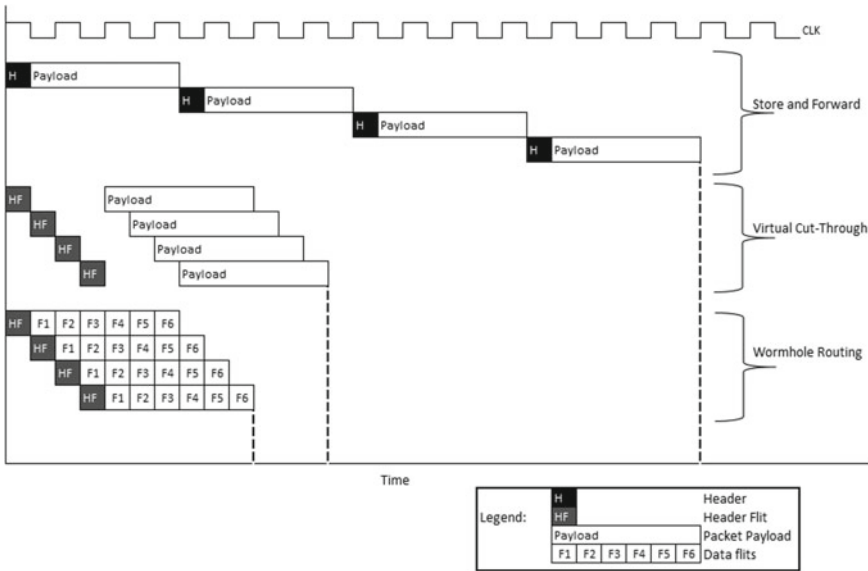
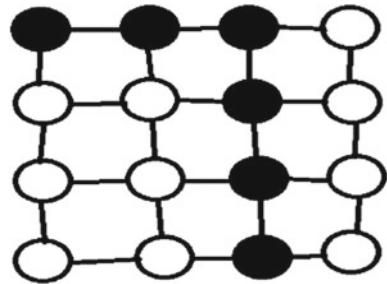


Fig. 4 Packet transfer using different routing strategies

Fig. 5 Framework for 2D mesh topology



4.1 Framework

The framework is indicated in Fig. 5. A 2D mesh is used for realizing X–Y routing, and the modules are placed in a matrix form.

4.2 Addressing Scheme

To identify different modules, an addressing scheme relative to the matrix form is used. For a $P \times P$ module generic framework, the addressing can be represented as $(0,0)$ starting from the top left corner and $(p-1, p-1)$ at the bottom right corner.

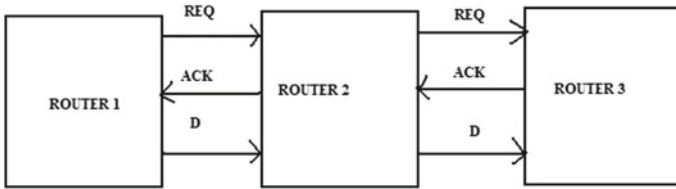


Fig. 6 Interconnection between modules

Fig. 7 Packet format

DEST <i>M Bits</i>	SRC <i>M Bits</i>	DATA <i>K Bits</i>
------------------------------	-----------------------------	------------------------------

Fig. 8 Flit format

Tail flit <i>M bits</i>	Data flits <i>N bits</i>	Header flit <i>M bits</i>
-----------------------------------	------------------------------------	-------------------------------------

4.3 Interconnect

The interconnection between the modules is represented in Fig. 6. Each module has request, an acknowledgement and a data (D) channel for connecting every other module.

4.4 Packet Format

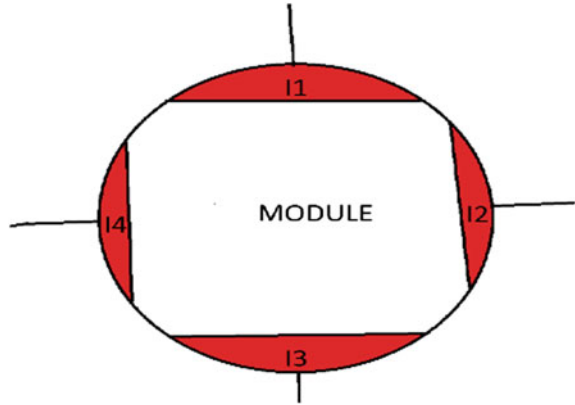
Different modules will start exchanging the packets across the networks during the computation process. The format of packet is specified with the address of the target node followed by the address of the sending node and data. The number of modules in the system determines the size of sender and receiver address. Figure 7 describes the packet format used in this framework.

Header, data and tail flits are the three flit headers of the flit format shown in Fig. 8.

4.5 Module Design

On receiving a request, the four interfaces of the router can transmit the packets simultaneously. Figure 9 shows the skeleton of a module.

Fig. 9 Skeleton of a module



4.6 Design Interface

For receiving and sending flits, two circular buffers are used in the design of interface. When the flits reach the interface, a temporary register is used for storing and transmitting the flits. The send buffer also uses the same structure. Based on the routing logic, the flits which is stored in the send buffer is moved to the next interface unit and then towards the target address. The control logic is used for verifying whether the bits are busy or not and checks the status of bits in the interface. Figure 10 shows the interface design [11] of a module.

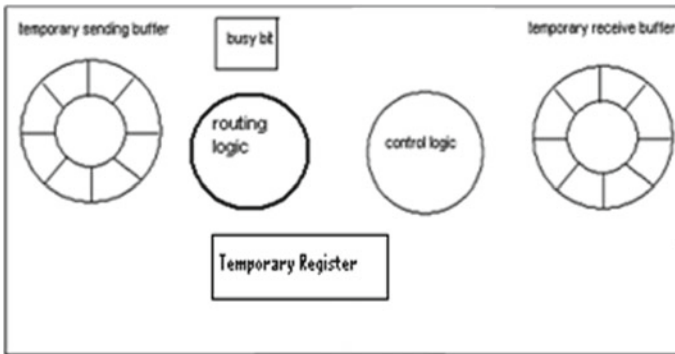


Fig. 10 Interface design

5 Verilog Simulation of Routing the Flits

One of the most frequently used languages in digital hardware design is Verilog and is an especially useful event-driven logic simulator. It is generally used because it solves a problem with good simulation speed that stays to improve. Figure 11 shows the simulation results of creating a packet. The coordinates (X, Y) of the destination address get concatenated with the coordinates (X, Y) of the source address. A packet is created by concatenating the data with the source and destination address.

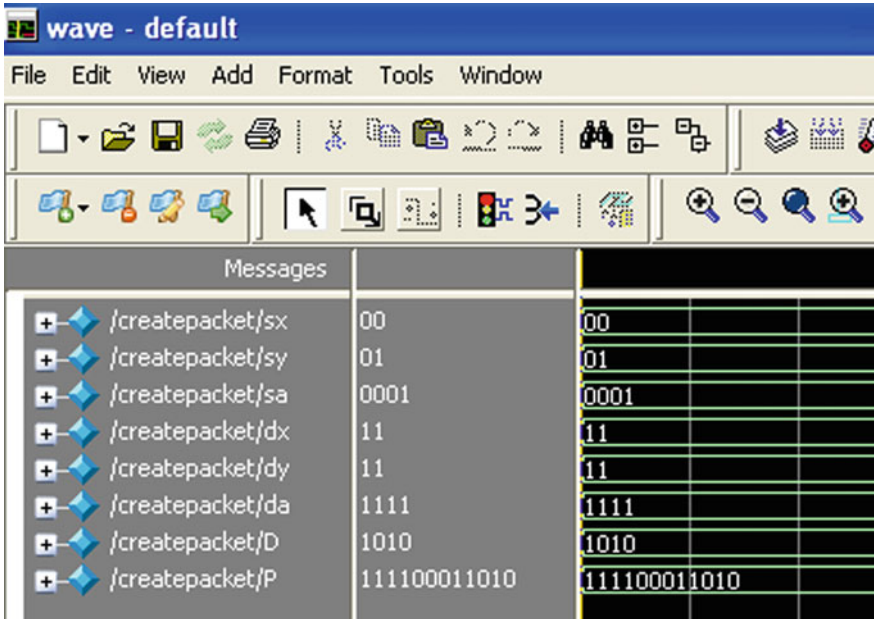


Fig. 11 Creating a packet

5.1 Pseudocode for Creating a Packet

```

module createpacket(destAddr,sourceAddr,Data,Packet);
    input destAddr, sourceAddr;
    input Data.
    output Packet.
    always @ (destAddr and sourceAddr)
    Packet= {destAddr, sourceAddr, Data};
    // concatenating the data with the source and
    // destination address
    Endmodule
    
```

5.2 Routing Flits Parallel in Interfaces

Figure 12 shows the transmission of flits parallel to all interfaces. In the first clock cycle, flits get transmitted through top interface to right-send buffer of right interface, and from right interface to down send buffer of the down interface, and from down interface to left-send buffer of the left interface and through left interface to top send buffer of top interface.

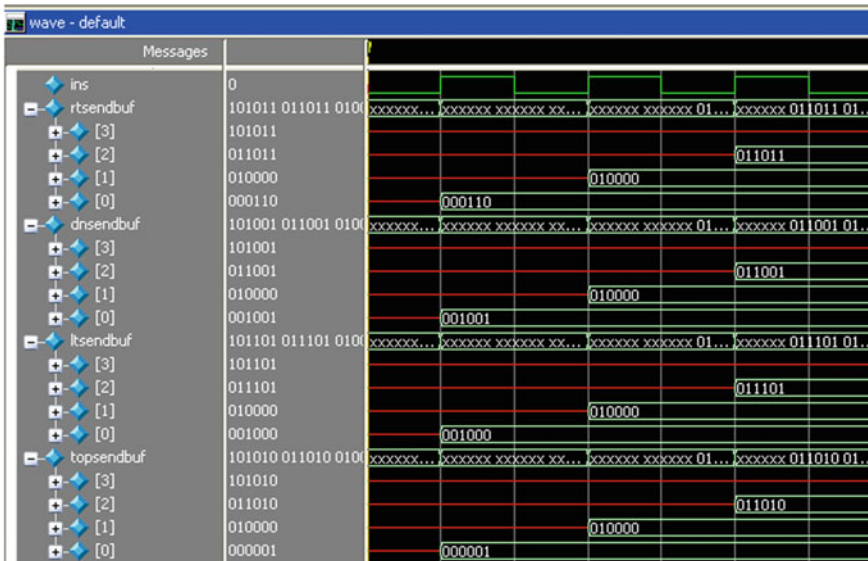


Fig. 12 Transmission of flits parallel to all interfaces

Because of parallel transmission [12], buffers of all the interfaces become full in four clock cycles.

Figure 13 shows the transmission of flits to the receive buffer of an interface unit when the target buffer is busy. Here top interface sends flits to the right-send buffer of the right interface unit, right interface unit sends flits to down send buffer of the down interface unit, down interface sends flits to the left-send buffer of the left interface unit. When left interface unit tries to send flits to the right-send buffer of right interface, since right-send buffer is busy it gets stored in the receive buffer of the left interface.

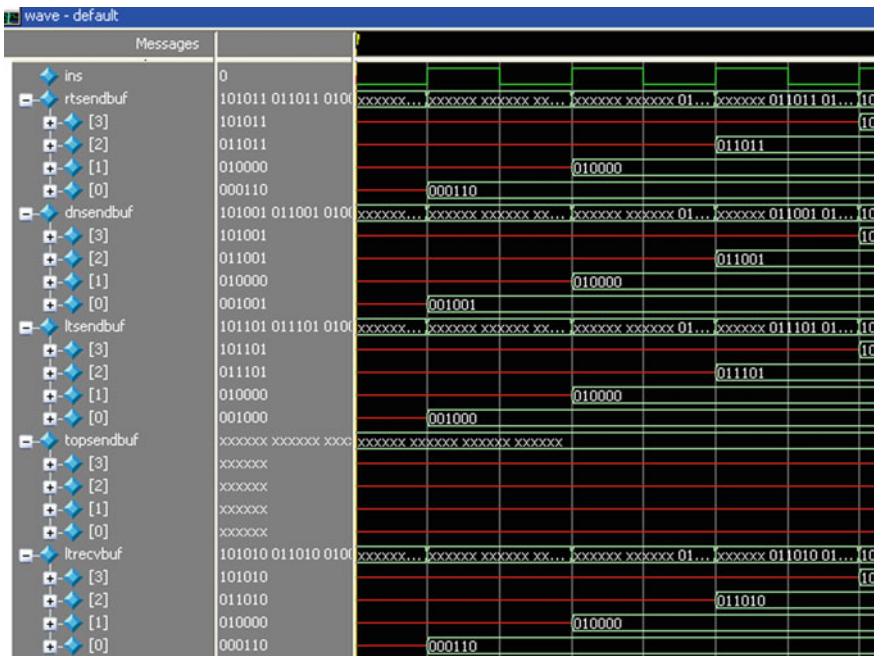


Fig. 13 Transmission of flits to the receive buffer of an interface

5.3 Pseudocode for Routing the Flits in Parallel

```

repeat(4)
{
// Prepare the flit for the current node.
stored = PrepareFlit(1).
// Route the selected side input.
stored = NodeRouting (Flit);
// Switch between input sides if Parallel
// operation required.
if (isParallel != 0)
    inputSide = (inputSide + 1) % 4;
$display($realtime);
}

function NodeRouting;
{
//Store the received packet in Temp register of input
// side.
stored = StoreToTempReg(inputSide, flit);
//Extract the header of the flit to flitHdr
case(flitHdr)
{
HeaderFlit :
{
//When header flit comes, find the destination
// and update the LUT.
}
DataFlit :
{
//Store in the destination as obtained from
// the LUT
Stored = StoreInDestBuffer(inputSide, tempDestSide);
}
TailFlit:
{
//Reset the LUT info for the side.
//Reset the outputBusy Flag
// Store in the destination as
//obtained from the LUT
stored = StoreInDestBuffer(inputSide, tempDestSide);
}
}
}
endfunction

```

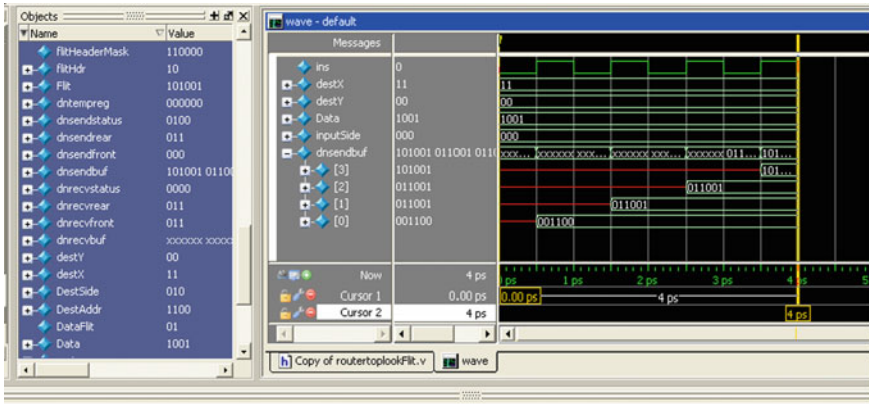


Fig. 14 Time measurements of single packet transfer scenario

6 Time Measurement

6.1 Single Packet Transfer Scenario

Figure 14 shows the simulation of processing time required for a full packet, which in this case are divided into 4 flits, through one module. It takes 4 picoseconds for the full packets with 4 flits to get processed within a module. The transmission of a packet from one module to another module primarily is dependent on the processing time within a module and so the total time required to traverse from one module to another module is equal to the number of traversals required to reach the target node from the sender node multiplied by the processing time period taken by each module. To calculate the time period taken for a module to process a packet, the total time for the simulation is 4 ps which are executed in steps of 1 ps each. The system tasks \$real-time and \$display are used to measure and display time. The IDE shows the transfer time required for single packet transfer scenario.

In single packet transfer scenario, the time taken to fill one interface buffer = 4 ps and the total time taken to fill 4 interface buffers = $4 \times 4 = 16$ ps.

6.2 Parallel Flit Transfer Scenario

Simulation of parallel transfer of flits is shown in Fig. 15. In a parallel scenario, the worst-case situation is that there are incoming packet units from all the 4 inputs and all these packet units are going to 4 different outputs. The simulation result shows that the total time consumed for filling all the interface buffers is 4 ps, which is less time compared to its serial counterpart because it is parallel.

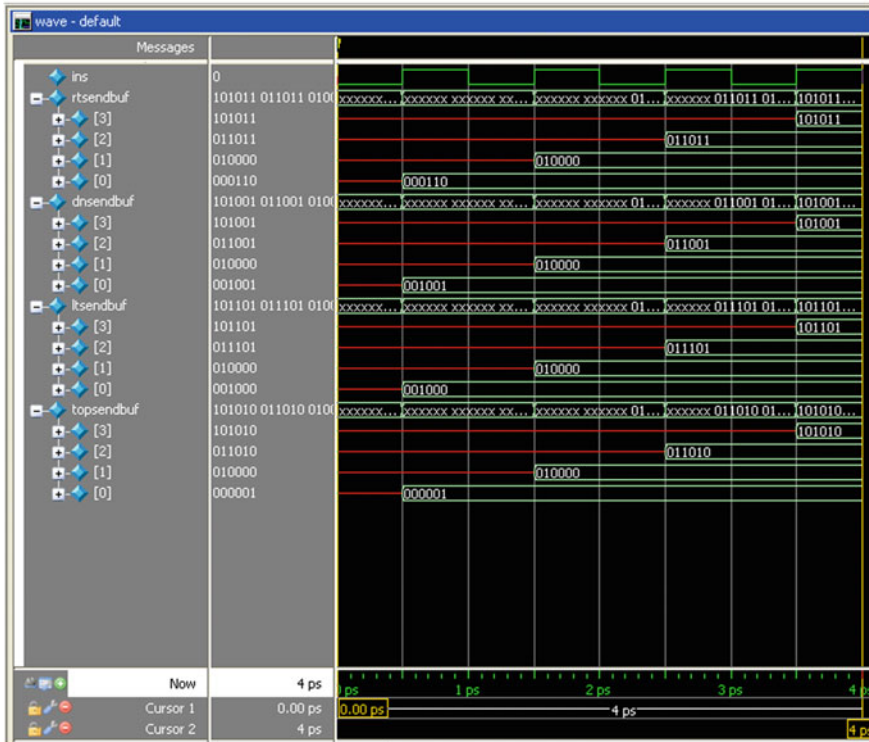


Fig. 15 Time measurements of parallel transfer scenario

Figure 15 shows the parallel transfer scenario. The system tasks \$real time and \$display are used for measuring and displaying time. The IDE has the display feature to show the transfer time require for parallel transfer scenario.

7 Conclusion

The Verilog simulation used in this work is to simulate a wormhole routing-based network-on-chip systems. In wormhole routing strategy, the advantage is that when a packet needs to be sent, it gets divided into flits, which are smaller chunks of data in comparison with the whole packet. This reduces the memory requirement in each of the nodes involved in the communication compared to store-and-forward and virtual cut-through routing. Simulation for routing the flits in a parallel interface transfer scenario is done using Verilog. In the above case study, when comparing single transfer scenario and parallel transfer scenario, it is clear that in single transfer scenario 4 cycles are needed for the transfer of a full packet towards single interface, whereas in parallel transfer scenario only 4 cycles are needed for the transfer of full

packets towards four interfaces. Verilog is used for this simulation, as it provides the features to solve a problem with good simulation speed.

Acknowledgements We thank our Director, Principal, HoD, Guide and all staffs of Nitte Meenakshi Institute of Technology, Bangalore, for their support in completing this work.

References

1. William J. Dally, Brian Towles, "Route Packets Not Wires: on chip interconnection network", DAC 2001 June 2001.
2. Hemani A, Jantsch A, Kumar S, Postula A, Öberg J, Milberg M, Lindqvist D (2001) Network on a chip: an architecture for billion transistor era. DAC 2001
3. Bjerregaard T, Mahadevan S (2006) A survey of research and practices of network on-chip. ACM Computing Survey March 2006
4. Rantala V, Lehtonen T, Plosila J (2006) Network on chip routing algorithms. TUCS Technical Report No 779, August 2006
5. Ali M, Welzl M, Zwicknagl M (2008) Networks on chips: scalable interconnects for future systems on chips. In: 4th European conference on circuits and systems for communications. 03 September 2008
6. Sanju V, Chiplunkar NN, Baby BY (2009) Design of a generic network on chip frame work for store and forward routing for 2D mesh topology. In: International conference on emerging trends in electronic and photonic devices and systems
7. Sanju V, Koushika C, Sharmili R, Chiplunkar N, Khalid M (2014) Design and implementation of a network onchip-based simulator: a performance study. Int J Computat Sci Eng 9(1/2)
8. Wang P, Ma S, Lu H, Wang Z (2014) A comprehensive comparison between virtual cut-through and wormhole routers for cache coherent network on-chips. IEICE 1–12
9. Wehbe T, Wang X (2014) Efficient buffer design and implementation for wormhole routers on FPGAs. Springer International Publishing Switzerland, pp 233–239
10. Hesham S, Rettkowski J, Göhringer D, Mohamed A, Ghany AE (2015) Survey on real-time network-on-chip architectures. Springer International Publishing Switzerland
11. ZhuanSun Z, Li K, Chen G (2017) Multipath routing algorithm for application-specific wormhole NoCs. Concurrency Computat: Pract Exper 1–12
12. Levitina LB, Rykalovab Y (2019) Computer interconnection networks with virtual cut-through routing. In: 14th International conference on future networks and communications (FNC), August 19–21

Application of Fine Decision Tree Machine Learning Algorithm to Predict the Subclinical Mastitis in Cow Milk Using Prototype E-nose



M. J. Anand, V. Sridhar , and Ramasamy Ravi

1 Introduction

Mastitis is a common disease that afflicts dairy animals; Treatment is expensive and causes severe damage to the dairy industry. Mastitis causes a huge decrease in productivity with a huge impact on the quality and quantity of milk. Mastitis is the common and widely reported disease among dairy cattle which causes economic loss due to severe effect on the quantity and quality of milk produced. The prevention of mastitis is done by conducting training from veterinarians to the dairy workers which can reduce the spread of sub and clinical mastitis. Bovine mastitis has a high incidence worldwide and has become a very expensive challenge to dairy industries [1].

Nowadays, milking machines are available, can provide assistance during milking which is not adequate, the system must comprise automatic on line measurement of important parameters of the milk drawn from each cow as it is being milked—for example, milk parameters, such as, amount of milk, fat content, and somatic cell count (SCC) from each cow [2, 3]. The clinical mastitis is complex, it can be identified or detected by using SCC count and using VOC components in the milk [4]. There are many prototype electronic nose has been developed for detection of mastitis diseases in cow milk using different classifier such as Principle Component Analysis (PCA), Artificial Neural Network (ANN), Support Vector Machine (SVM) [5, 6, 15–18].

M. J. Anand (✉)
PES College of Engineering, Mandya, India

V. Sridhar
NMIT, Bangalore, India

R. Ravi
Food Science, Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, USA
e-mail: rravi@TNstate.edu

In mastitis, there are 2 types—clinical mastitis which is visible to the eyes, and the subclinical mastitis which is confirmed only after a test. California Mastitis Test (CMT) is one such diagnostic technique, which is much faster than the SCC count and is reliable. [7]. Other traditional testing methods are time-consuming and need expensive instrumentation. Embedded system-based monitoring unit monitors unwanted concentration of VOCs present in the headspace of raw milk. TGS gas sensor is used to detect VOC components. [8, 9]. Mastitis is mainly due to pathological changes in the organ tissue of the mammary gland that has an effect on the standard and amount of milk. If mastitis contaminated milk is consumed, humans are affected by several diseases such as tuberculosis, sore-throat, Q-fever [10].

Lack of awareness, lack of markers for later detection, delays in subclinical inflammation detection, unhealthy milking practice in milk production system are considered to be one of the major economic assets and disorders in this sector that will cause large economic losses to the country.

2 Materials

For conducting the experiment to determine subclinical mastitis disease in cow milk, hundred samples of HF cow milk were obtained from nearby villages of Mandya. The milk was then separately poured into sterilized 500 ml beaker and was placed in a rotation incubator at 15 °C and 200 revolutions per minute (rpm). Once the samples were obtained, pH of the milk was recorded using pH meter and prototype E-nose was used for sensing the VOCs in the milk are also recorded at room temperature (26 °C).

3 Experimentation

The proposed embedded system shown in Fig. 1 consists of a sensor array chamber, MSP430 microcontroller, and data acquisition system together with a display for visual indication of the mastitis disease in the cow. The sensor array was composed of commercially available Figaro gas sensors: TGS 813, TGS 2620, TGS 832 and

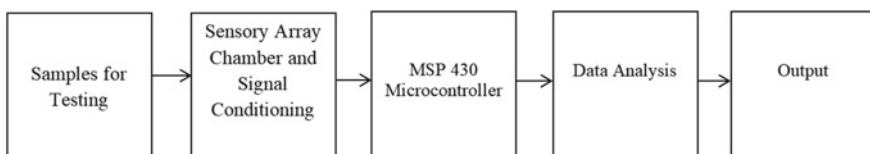


Fig. 1 Block diagram of proposed prototype electronics nose for detection subclinical mastitis disease

temperature, humidity sensor. These sensors are used to detect acetic acid, acetaldehyde, and ethanol for detection of Mastitis disease affected milk. The sensor array responses are acquired and analyzed using the data acquisition tool. While taking measurements, milk samples were placed below the sensor array in the sensor chamber.

The infected cows were identified with the help of veterinary doctors, where all the infected cows were under the regular treatment for subclinical mastitis. It was identified from the somatic cell count of the milk of the infected cow's udder are above 300,000 cells/ml by Bacteriological diagnostic lab.

Milk sample 50 ml is placed in a sensor array chamber and waited for few minutes until the sensor response becomes stable. If the output response of sensors TGS 813 (Acetic acid), TGS 832 (Ethanol), and TGS 2620 (Acetaldehyde) reads in the range of 0.135–0.137 V, 0.045–0.059 V, and 0.00325–0.00549 V, respectively, then the milk is considered as without subclinical mastitis disease with a pH value of the milk sample under test in the range of 6.32 – 6.50. The responses of the prototype E-nose (sensors) for ten raw milk samples without mastitis diseases are plotted as shown in Fig. 2. The above range of voltage of each sensor is stored as lookup table in the system memory. We could obtain similar set of response for all the raw milk samples. Similarly, for the mastitis infected cow milk samples, the output response of sensors TGS 813, TGS 832, and TGS 2620 reads in the range of 0.26–0.479 V, 0.20–0.501 V, and 0.00169–0.00329 V, respectively, and the pH value of the mastitis affected milk sample under test lies in between 6.78 and 7.10. The responses of the prototype E-nose for milk samples affected by mastitis diseases are plotted as shown in Fig. 3.

During the test, for identification of raw milk or infected cow milk, the each sensor output correspondingly should match with the range of look-up table values

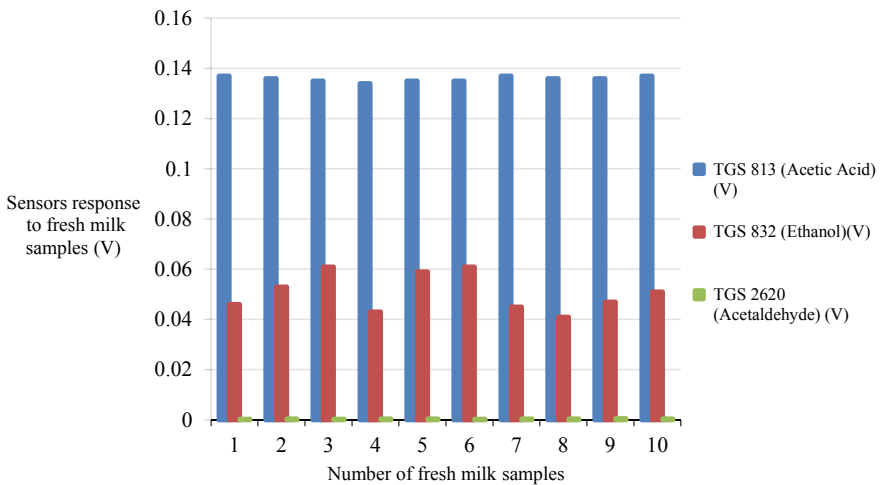


Fig. 2 Sensor array responses for raw milk samples without mastitis diseases

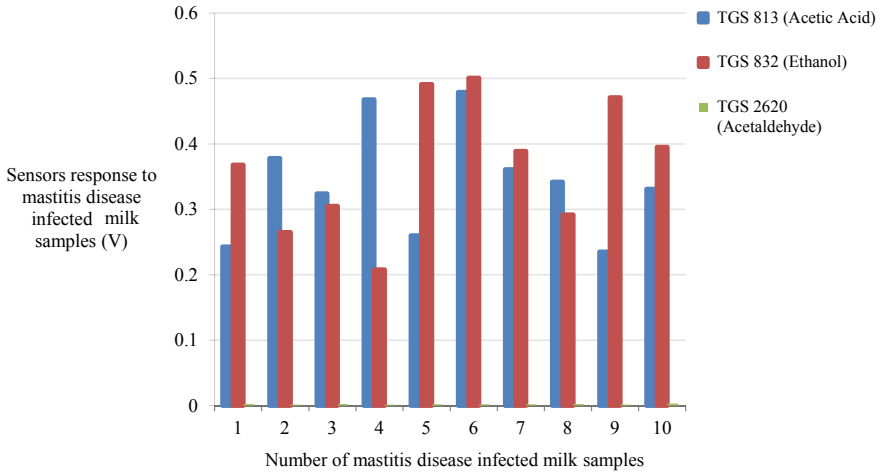


Fig. 3 Sensor array responses for raw milk samples with mastitis diseases

as specified and pH value should be in the range of 6.32 to 6.5 or 6.78 and 7.10. If the somatic cell counts (SCC) < 200,000 cells/ml in raw milk or SCC > 300,000, then the milk is considered as raw milk or mastitis infected milk.

4 Fine Decision Tree Machine Learning Algorithm

The trained network model 1 is developed in the fine decision tree Machine learning algorithm. The predicted output is matched with the targeted output and gives results as ‘1’ is assigned as for mastitis disease infection free cow milk and ‘2’ is assigned as for mastitis diseases infected in cow milk.

The training data set was used to train the fine decision tree algorithm to produce a trained model as model1. Fig. 4 shows scatter plot providing a visual representation of the correlation between the two variables and it gives the clear separation of ‘1-correct’ assigned for cow’s milk samples without infected shown in blue color (dot), ‘2-correct’ is assigned for cow’s milk affected by mastitis disease samples shown in purple color (dot). Figure 5 shows classification accuracy of 97 and 100% for cow milk without and with infected mastitis diseases and also overall accuracy of 98.5%.

After the decision tree algorithm generates the trained model, test inputs are applied to the trained model to predict the output. These outputs are compared with the expected output, and then the overall accuracy of predictive model of decision tree algorithm is 98.5%.

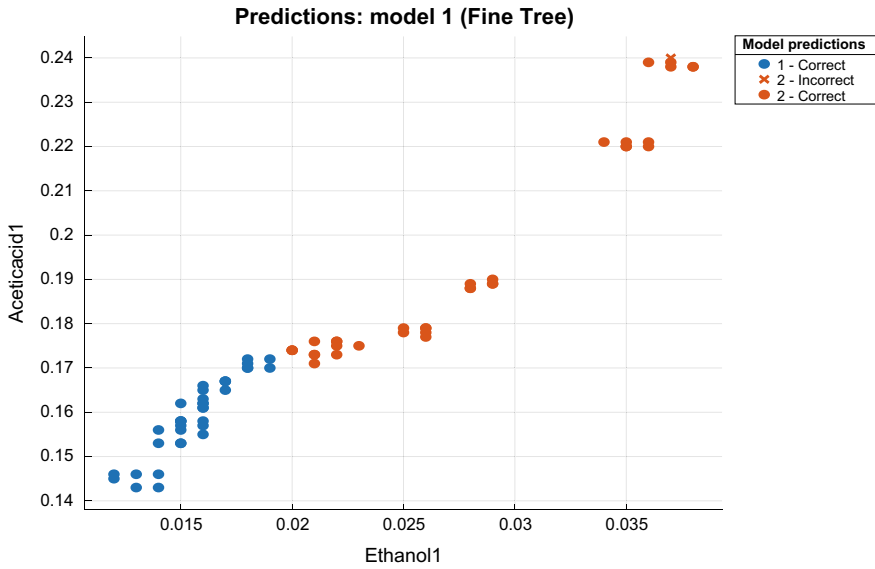


Fig. 4 Scatter plot of fine decision tree algorithm for detection of mastitis diseases in cow milk

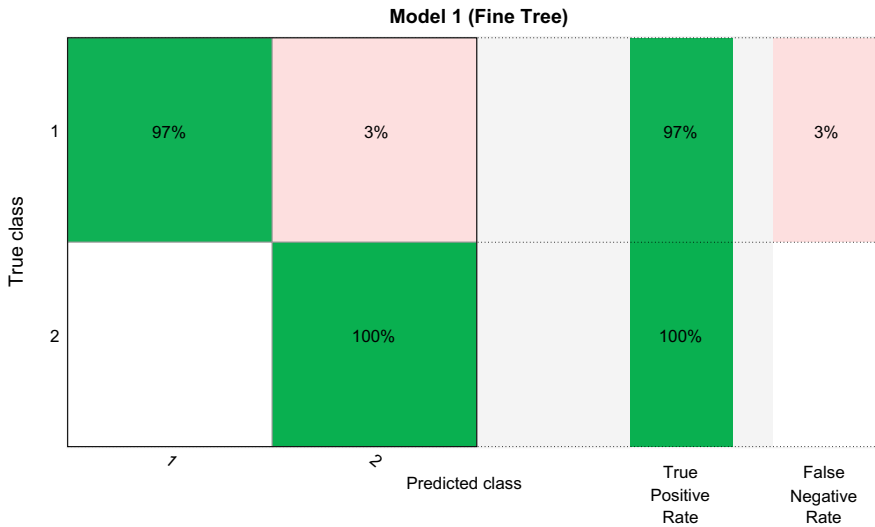


Fig. 5 Confusion matrix of predicted output of fine decision tree algorithm for detection of mastitis diseases in cow milk

5 Result and Discussion

In this experiment, fine decision tree machine learning algorithms were implemented for cow milk infected with mastitis diseases detection. The fine decision tree confusion matrix of Fig. 5 shows that first row of the matrix provides class 1 predicting 97% correctly as cow milk without mastitis diseases, second row of the matrix provide class 2 predicting 100% correctly as cow milk with mastitis diseases. The overall accuracy of predictive model is 98.5% for detecting the cow milk without and with mastitis diseases.

Figure 2 provides the pattern of the cow milk samples without mastitis diseases and also composition equivalent voltage in this range for acetic acid (0.038–0.041 V), acetaldehyde (0.044–0.045 V), and ethanol (0.05–0.058 V). For example, it may indicate acetaldehyde—0.044–0.045 V (3.2 mg/l); ethanol 0.05–0.058 V (1.7 mg/l) and acetic acid—0.038–0.041 V (1.8 mg/l). Similarly, the sensors record following responses, in samples of milk which have come from cows with mastitis disease—acetaldehyde (12 mg/l)—(0.89–0.9) V, ethanol (10 mg/l)—(0.79–0.82) V, and, acetic acid (9 mg/l)—(0.6–0.65) V [1]. These increases in acetic acid, acetaldehyde, and ethanol in cow milk infected with mastitis diseases are due to the multiplication of bacteria *Streptococcus aureus* (acetaldehyde), *Escherichia Coli* (acetic acid), *Streptococcus dysgalactiae* (ethanol) in milk which in turn increases these gases in the milk.

Table 1 shows the cow milk with mastitis diseases samples of 50 and cow milk without mastitis diseases samples of 50 using fine decision tree algorithm. Accuracy is found to be 99%, which indicates the overall performance of the classifier for detecting cow milk without and with mastitis diseases. Sensitivity is found to be 100% cow milk with mastitis diseases and Specificity is found to be 98.03% correctly identifies cow milk without mastitis diseases.

In order to validate the proposed system, the pH meter is used to find the pH values of the cow milk samples in order to detect the milk is infected with mastitis diseases or not. From this Fig. 6, it has been concluded that the pH value of mastitis infected raw milk is (7.5) higher than the pH value of raw milk. Therefore, the proposed system is able predict the mastitis diseases detection in cow milk with the overall accuracy of fine decision tree algorithm is 98.5% compared to the other classifiers such as SVM, PCA, and ANN [11, 12].

Table 1 Sensitivity and specificity of mastitis diseases detection in cow milk

Test samples		Test outcome		Total
		Test positive	Test negative	
Mastitis disease cow milk	Positive	49 (TP)	01 (FP)	50 (PP)
Cow milk without mastitis disease	Negative	00 (FN)	50(TN)	50 (NP)

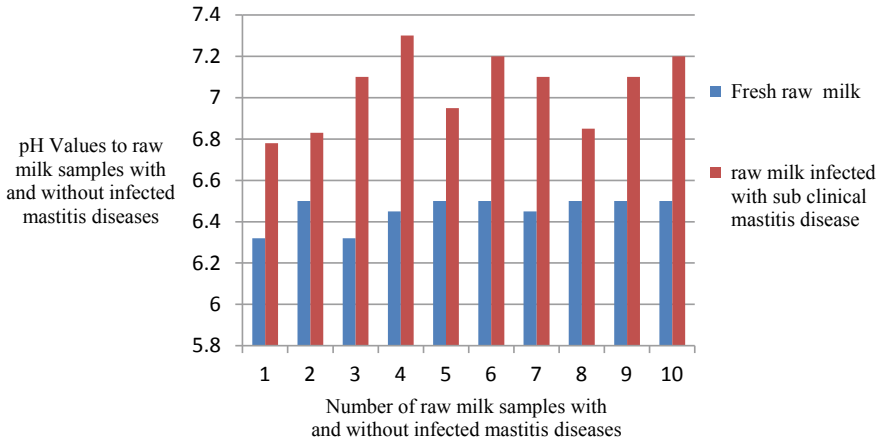


Fig. 6 pH Sensor responses to the cow milk and mastitis diseases infected milk samples using pH meter

6 Conclusion

The present research work was carried out to design, develop a prototype electronic nose system to differentiate the subclinical mastitis affected milk using machine learning algorithm. The main intention of the work is to develop a simple device which can be used by the unskilled people in a non-laboratory setup to differentiate the subclinical mastitis affected milk due to various causes. The major beneficiary of the proposed system is dairy industry, and the system will help to make quantitative analysis of the milk at the site of milking, milk collection centers, procurement, and distribution centers and help the dairy industry to prevent loss by the subclinical mastitis in cow and maintain the quality of the milk. In order to make proper prediction about the quality of milk, machine learning algorithms were used in order to validate the outcome of the quantitative analysis carried out the experiment which resulted in 98.5% validation.

References

1. Liera I, Castro C, Mancera (2017) Production of milk and bovine mastitis. *J Adv Dairy Res* 5(2):1000174
2. Lima RS, Danielski GC, Pires ACS (2018) Mastitis detection and prediction of milk composition using gas sensor and electrical conductivity. *Food Bioproc Technol* 11:551–560
3. Kumar N, Manimaran A, Kumaresan A, Sreela L (2016) Episodes of clinical mastitis and its relationship with duration of treatment and seasonality in crossbred cows maintained in organized dairy farm. *Veterinary World* (2016)
4. Inalpulat M, Kizil U, Bilgucu E (2016) E-Nose identification of milk somatic cell count. *J Graduate School of Natural Appl Sci* 2(1):22–35, Çanakkale Onsekiz Mart University

5. Sarnobat SK, Mali AS (2016) Detection of mastitis and monitoring milk parameters from a remote location. *Int J Electr Electron Comput Sci Eng* 3(6):2454–1222
6. Gomes F, Saavedra MJ, Henriques M (2016) Bovine mastitis disease/pathogenicity: evidence of the potential role of microbial biofilms. *Federation of European microbiological societies (FEMS). J Invest Sci Pathogens and Disease* 74
7. Nakov D, Hristov S, Andonov S, Trajchev M (2014) Udder-related risk factors for clinical mastitis in dairy cows. *Vetinary Arhiv* 84(2):111–127
8. National Academy of Agricultural Sciences (NAAS) “Mastitis Management in Dairy Animals” (2013) Policy Paper No. 61, New Delhi
9. Sharma N, Singhand NK, Bhadwal MS (2011) Relationship of somatic cell count and mastitis: an overview. *Asian-Aust J Animal Sci* 24(3):429–438
10. Grillo GJ, Perez MA, Anton JC, Ferrero FJ (2002) Direct Evaluation of the fresh-milk somatic cell concentration (SCC) through electrical permittivity measurements. In: *IEEE Instrumentation and measurement technology conference anchorage, AK, USA*
11. Mammadova N, Keskin E (2013) Application of the support vector machine to predict subclinical mastitis in dairy cattle. *The Scient World J*
12. Cavero D, Tolle K-H, Henze C, Buxad C, Krieter J (2008) Mastitis detection in dairy cows by application of neural networks. *Livestock Sci* 114(2–3):280–286

Detection of Cyberbullying on Twitter Data Using Machine Learning



A. Sandesh, H. V. Asha, and P. Supriya

1 Introduction

Considering the era of Information Technology, social networking plays a vital role in everyone lives. Web has become a part and parcel of our lives for effective communication and establishing substantial relationships in the society. Larger sector of the population communicates through online in more diverse ways. It has made our life easy as well as tough. These enhancements provide an open platform for communication where users can create their profile to interact with others irrespective of their location and boundaries. Despite the fact that web utilization is innocuous and the advantages of advanced correspondence are apparent, the opportunity and namelessness of the experienced online users make youngsters open to cyberbullying causing significant dangers [1, 2]. Cyberbullying is provocation done using innovation and could vary from disturbing text messages to frightful posts online on long-range informal communication destinations. Another study found that one out of three youngsters has encountered digital dangers whilst utilizing the Internet [3]. Digital crooks are using Internet media as a stage for doing various kinds of cybercrimes such as phishing, spamming, spreading rumours, harassing, and cyberbullying. In particular, precisely, because of the ongoing improvement of online correspondence and web-based life, cyberbullying and gossip spreading have become a significant issue.

A. Sandesh (✉) · H. V. Asha · P. Supriya

Faculty, Department of Computer Science and Engineering, NITTE Meenakshi Institute of Technology, Bangalore, India

H. V. Asha

e-mail: asha.hv@nmit.ac.in

P. Supriya

e-mail: supriya.p@nmit.ac.in

According to the survey by Annapolis [4], about 42% of children have been tormented on the web. National Crime Prevention Council (NCPC) appraises that about 43% of young people have been harassed on the web [5]. Amongst this rate, being disregarded and affronted were the most widely recognized types of cyberbullying. Estimates show that about nine out of every ten school students have been victimized online and have been hurt emotionally. Some of the websites allow one student to bash on another student. Every four out of ten school student's passwords have been stolen, and the bully uses this to block the students from using their own accounts and send messages or a post posing as the original account holder. About one fourth of the kids who are active online have received mean or threatening emails. Recently, it is been observed that cyberbullying has caused major damage both psychologically and emotionally, and it is been said that cyberbullying has had similar effects as a real-life bullying outcome. However, real-life bullying can be curbed, but often, there is no escape for cyberbullying. Chat rooms are considered as the primary location where about 56% of the victimization happen due to cyberbullying with primary targets being girls [6]. The most talked and controversial game, named Blue Whale, which caused the suicide of many adolescent's, is the perfect example for cyberbullying.

Harassing is anything but another marvel, and cyberbullying has showed itself as computerized innovations. Today, the most popular social networking sites like Facebook, Twitter, Snap Chat, Instagram, and instant messaging platforms like WhatsApp have revolutionized the mode of communication making the world smaller for everyone to communicate irrespective of long distances. They have offered a great platform for social interaction and are helping in setting up new connections and keep up existing companionships over a long run [7, 8]. However, these online platforms increased the danger of kids being defied with compromising circumstances including prepping or explicit transgressive conduct, signs of gloom and self-destructive musings, and cyberbullying.

As shown by [9–11], cyberbullying may contrarily affect the casualty's confidence, scholastic accomplishment, and enthusiastic prosperity. [12] Study has made a finding which explains that self-detailed impacts of cyber bullying lead to negative impacts on school evaluations and sentiment of bitterness, outrage, dread, and discouragement. Cyberbullying in outrageous cases can prompt self-hurt considerations and even self-destructive endeavours. The survey done by many researchers shows that cyberbullying has become a major cause of concern. Cyberbullying detected in an early stage helps in protecting the youngsters from harassments and supporting their mental well-being. Successful detection is completely dependent on effectively monitoring the online content. However, the quantity of material over the web becomes a limitation making it unfeasible to screen all the produced content physically. And hence, intelligent systems which are capable of processing the large amount of data and able to automatically signal the potential threats are required to address this problem. These systems if used effectively can react rapidly and keep undermining circumstances from escalating. Recent research [13] says present youths have well-accepted the programmed observing, as long as effective strategies for follow-up are worked out considering the privacy and autonomy of the individual

are protected. Parental control tools [14] tend to be utilized to square unsuited or unfortunate substance, and furthermore, some social locales use watchword-based control apparatus. Unfortunately, some of these efforts have failed to detect the indistinct and elusive appearance of cyberbullying where no explicit vocabulary is made use of. All these have increased the need for a system which is capable of self-learning and go beyond the technique of spotting a threat keyword. At this time of crisis of COVID-19 where everything is happening virtual, people are intended to use the social media for all communication ranging from online classes to work meetings to conferences. Individuals in all age group amidst this pandemic are tended to use online platforms increasing the threat of bullying. Formulating canny and self-learning frameworks that go past catchphrase spotting is of significance and needs of great importance for shielding individual interest. At this juncture, it is unavoidable to build a model which may give fruitful cyberbullying predication results. However, building this model relies upon numerous components and the features extracted from the social networking sites.

Previously, acquisition of information related to individuals and society was an impossible task, but today's era of big data gives us access to required information via Social Media (SM). SM is the primary source of data of any individual irrespective of it being professional data or personal data. The application of Machine Learning (ML) algorithms to social media information helps in exploiting the authentic information to foresee the eventual fate of a wide scope of uses. ML algorithms provide a chance to viably foresee and identify negative types of human conduct, for example cyberbullying. Many theorems and various multidisciplinary and interdisciplinary techniques are merged to provide an adroit examination of information on human conduct. These theorems are also designed to distinguish and confine individual's destructive behaviour involving complicated approaches. The availability of abundance of data has made ways for many novel computational methods, inquiring new research questions and opportunities to inter-disciplinary approaches. Traditional methods such as factual strategies have had confinements regarding scale and exactness. These techniques are mainly based on organized type of information on human conduct. They can operate well in small-scale human networks. Application of these approaches to hefty online social network has caused major limitations with respect to scale and extent of data. Hence, optimized methods to detect and restrain aggressive behaviours considering the content and size of network on complex systems need to be constructed. Therefore, strategies that address both perspectives (content and network) ought to be advanced to recognize and control forceful conduct in complex frameworks.

Further sections in this paper are organized as follows: Sect. 2 presents the required background about cyberbullying and its associated work. The projected work is presented in Sect. 3 which briefs about its architectural design, data collection, data annotation, and feature extraction. Section 4 discusses about the result obtained on implementation of the system. Finally, conclusion and future scope are presented in Sect. 5.

2 Background and Literature Survey

Cyberbullying is characterized as utilization of web-based social networking, email, phones, instant messages, and internet locales to undermine, bother, humiliate, or socially avoid somebody of lesser force [15–17]. Anonymity on the virtual world promotes cyberbullying from unknown individuals. The act of cyberbullying amongst former peers and associates aware of individual private information is of major concern in the present society as this can be used to attack or revenge against associates. The abundant audience size provided by social media bestows to the lopsidedness amongst cyberbullies and their victims. Present day youths are digital natives spending maximum time on Internet-enabled devices as they believe in being virtually connected to all locations and individuals across the globe [18]. And, ‘turning off the phone’ to abstain from tormenting is not considered as a suitable arrangement which can prompt further isolation [19]. Study in ML has proven to be promising area of work in terms of detection of cyberbullying. However, an extensive review on ML research for cyberbullying detection features a few constraints which are presented in [20]. It also gives bits of knowledge on the general procedure for cyberbullying detection and above all reviews the approach.

Preceding the advancement of correspondence innovations, it is been observed that social association has evolved within small cultural/social boundaries/limits [21]. Ongoing exploration in correspondence has helped development of technologies that rises above the fleeting and spatial constraints of customary correspondence. With advent of technology, online correspondence has made a move towards user-driven advancements, for example SM sites, websites, online virtual networks, and web-based sharing platforms. With this transformation, activities involving antagonism and violence have also emerged completely [22]. Individual’s negative attitude on SM and their revenge seeking behaviour pose a new challenge [22, 23]. Authors of [24] have presented about cyberbullying by stressing on the point of data being hampered due to lack of authentic data for studying. They conducted a broad survey by selecting a group of people for their project and collecting the data directly across different platforms. Latent Semantic Indexing (LSI) is used in [25] for uncovering cyberbullying in labelled assortment of comment posts from Formspring.me. Their framework significantly outperforms the gauge with an exceptional query and is not reliant on dictionary of harassing terms. A solitary demonstration of hostility or mortification may cause trouble and embarrassment for the person in question in the event that it is shared or preferred by a huge crowd [26]. According to the concept of ‘snowball effect’ considered in [27], one post might be rehashed or disseminated by others, so it gets out of control for the underlying harasser and has bigger impacts than was initially expected. Nobara et al. [28] perform loathe speech identification on Yahoo Finance and news data with the help of supervised learning. Kayes et al. [29] find that users will in general ban the harsh substance posted on Yahoo Answers in overwhelmingly right manner. Paper [30] distinguished the aggressive and cyberbullying words from YouTube recordings and cybercrimes as one of the models in cyberbullying and online provocation. In [34], authors have presented 22

automated cyberbullying detection analyses. Also, their work is supported by two datasets experimented to test existing practises.

The results of [35] have shown that inclusion of personality and sentiment has increased the cyberbullying detection.

3 Proposed Work

Although, the recent developments and innovations in information communication technologies have given way for social media platforms to suggest communication opportunities, and it also increased the susceptibility of younger generation individuals getting into a threatening situation online. Successful prevention of cyberbullying on social media platform depends upon the satisfactory recognition of conceivably hurtful messages. The objective of the presented research is to automate a cyberbully detection model that distinguishes destructive behaviour related to the security of individuals, hate comments, harmful, incorrect, or any form of mean content about other users in social media platforms by modelling the textual data generated by bullies, sufferers, and onlookers of online bullying.

Figure 1 represents the general architecture of the cyberbullying detection model. It consists of two major segments: first being the big data analytic and second, the machine learning. A huge amount of user-content-related data generated from social media platform are first pre-processed using big data analytic and feature engineered to clean up and reshape the data to bring it to a form that is more suitable for fitting the machine learning model. The complete pre-processed data are further split into training set for construction of the model and test set for evaluation of the model. Once the evaluation of the model is complete, the model is deployed for real-time simulation. After deployment, the model is monitored constantly to evaluate its performance against the newly generated mutated data. Depending upon various performance parameters against the newly generated mutated data, the model may have to be retrained in order to improve its performance against the newly generated data.

Data Collection: Data is one amongst the vital parts of ML-based algorithm. However, the data on its own do not carry any implicit meaning unless certain information or implication is mined. Text data extracted on SM platform in the form of posts, comments, or opinions are used in modelling the cyberbullying detection model. Data extracted are employed for training and testing for improving the accuracy of the model. Most of the machines learning models consider generalized data. But, for better accuracy, it is better to consider the unlabelled real data. Till recent years, data on social media were considered insignificant. Big data analytics has changed the scenario. What is considered important is that does the data extracted from SM represent any activities or not [20]. A corpora used for construction of the cyberbullying model can be constructed by collecting data from standard Twitter API which returns tweets specific to the specified query [31].

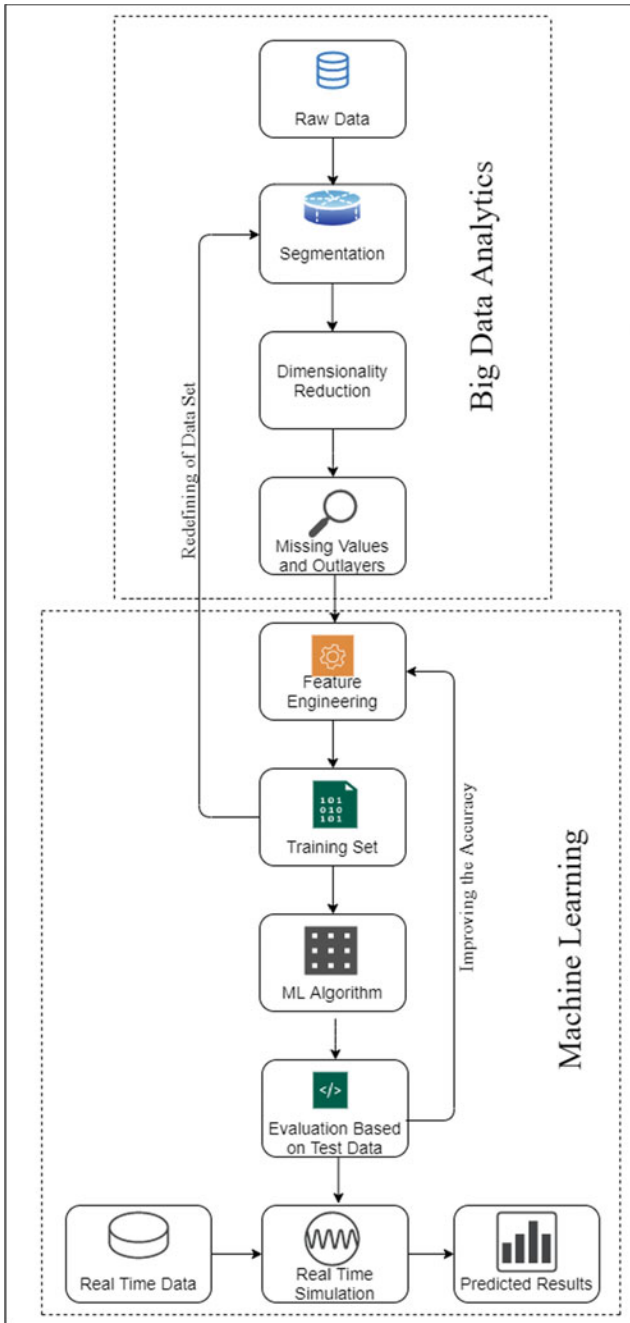


Fig. 1 Architecture of the proposed system

Data Annotation: Building an automatic cyberbullying system is not an easy task since cyberbullying itself is a wide research area. As already mentioned, cyberbullying can be in any form: verbal, implicit, explicit, or even it might be non-verbal. It is vital to consider the sociolinguistic aspects and what encompasses cyberbullying in developing the system for detection [32]. The present annotation scheme annotates a given post as cyberbullying or non-cyberbullying by considering the occurrence of certain keywords that implicate hate, aggression, distress, negativity, and so on. The dataset used for the construction of cyberbullying prediction model in the current implementation is obtained from GitHub repository [33] which comprises of a dataset consisting of data generated from the Twitter API and annotated in accordance to the above-mentioned schema. The dataset considered consists of 184,354 records, which is further classified into two nominal categories, cyberbullying being labelled as 1 and non-cyberbullying being labelled as 0.

Feature Engineering: In this phase, all the data in the form of posts are converted to mathematical forms suitable for ML algorithm. Current proposed model has used tokenization, count vectorizer, and Inverse of Document Frequency (IDF). Tokenizing alludes to separating an enormous collection of text into smaller lines, words, or in any event, making words for a non-English language. Count vectorizer is utilized to change a corpus of text to a vector of term/token checks. It additionally gives the ability to pre-process the content information before producing the vector representation making it an exceptionally adaptable element representation module for text. IDF is the inverse of the report recurrence which quantifies the information on the term t . Qualities determined for IDF will be extremely low for the most happening words, for example stop words. This provides a comparative weightage for individual words.

4 Result Analysis

The ‘Waikato Environment for Knowledge Analysis (Weka)’ is utilized to apply an assortment of classifiers to the issue of identifying cyberbullying on twitter information. The results obtained from the model constructed using the twitter API dataset of on ML algorithms are tabulated in Table 1. Whilst implementing, TP and FP rate, precision, recall, and F-score values were considered. The results are tabulated for varying data size (N). The values in bold reflect the best score in the respective column.

Since the data are obtained from social media which do not fall under any defined pattern of data, several ML algorithms, namely JRip, Naïve Bayes, Logistic Regression, SMO, and J48 Decision Tree were examined rather than using a single algorithm. To evaluate the results, True Positive (TP), False Positive (FP), Precision, Recall, and F-score are considered. True Positive is calculated by dividing the number of correctly classified tweets between the total samples taken. False Positive is calculated by dividing the number of tweets whose classification is wrong by the total number of samples. Precision is the ratio of factor of the relevant instances to the

Table 1 Comparison of results

	Weka tool accuracy	TP rate	FP rate	Precision	Recall	F-score
<i>N</i> = 10,000						
Naive Bayes	79.65	0.797	0.797	0.797	0.797	0.797
Logistic regression	77.6	0.776	0.754	0.694	0.776	0.716
SMO	81	0.81	0.74	0.834	0.81	0.738
JRip	80.1	0.801	0.779	0.841	0.801	0.717
J48 decision tree	80.1	0.801	0.779	0.841	0.801	0.717
<i>N</i> = 20,000						
Naive Bayes	80.225	0.802	0.802	0.802	0.802	0.802
Logistic regression	80.05	0.801	0.784	0.73	0.801	0.722
SMO	80.625	0.806	0.785	0.833	0.806	0.724
JRip	80.475	0.805	0.792	0.843	0.805	0.72
J48 decision tree	80.475	0.805	0.792	0.843	0.805	0.72
<i>N</i> = 30,000						
Naive Bayes	79.2	0.792	0.792	0.792	0.792	0.792
Logistic regression	78.8167	0.793	0.787	0.836	0.793	0.703
SMO	79.5833	0.796	0.775	0.811	0.796	0.71
JRip	79.3333	0.793	0.787	0.836	0.793	0.703
J48 decision tree	79.3333	0.793	0.787	0.836	0.793	0.703

retrieved instances. Recall is the ratio of the relevant instances retrieved to the total relevant instances. F-score combines the values of recall and precision. Table 2 gives a comparison of the values got from both Weka tool and implementation model. It can be observed that the implementation model has got better accuracy. Figure 2 gives comparative analysis of all ML algorithms implemented in the current model.

Table 2 Comparison of Weka and implemented results

<i>N</i> = 184,354				
Class label	Precision	Recall	F-measure	Accuracy
Weka tool results				
Non-cyberbullying	0.801	0.998	0.889	80.025
Cyberbullying	0.359	0.004	0.008	
Implementation results				
Non-cyberbullying	0.8053	0.9995	0.8919	80.4736
Cyberbullying	0.8139	0.0089	0.0176	

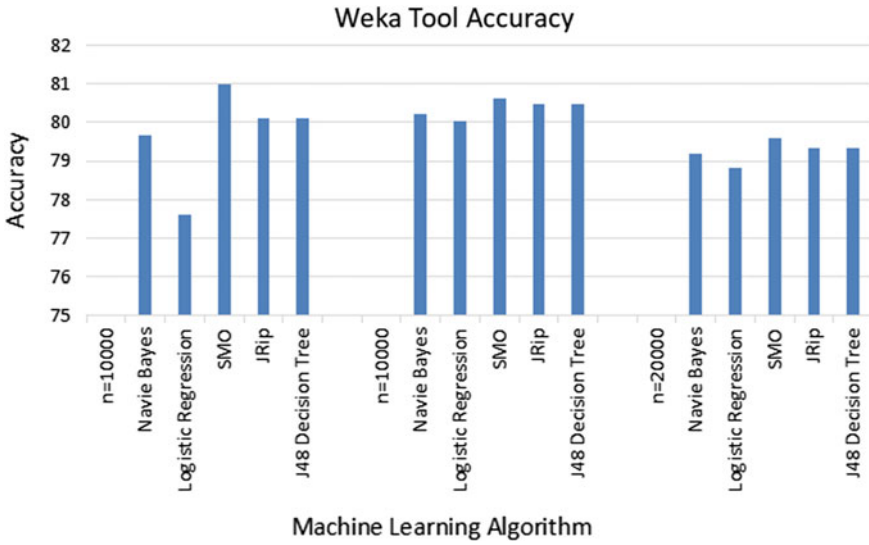


Fig. 2 Graphical representation of the result

5 Conclusion

The presented work has focused on detecting cyberbullying in SM. In particular, Twitter data are considered for implementing the model. The study highlighted the different ways of cyberbullying, its effects, and the various consequences of cyberbullying. Considering the current situation, the study has emphasized on the need of an automatic cyberbullying detection model. Various ML algorithms were implemented to recognize the best method for developing the model by extracting the features of data. It has briefly explained the data collection, data annotation, and feature extraction which were used in implementing the model. For this purpose, accuracy, F-measure, and precision, recall are considered, which help to identify the area under the curve function for modelling the behaviour in cyberbullying. The proposed model has few limitations which can be considered for future enhancement. Authentication of data is lagging since the data are extracted from third party. The user’s profile is not considered for assessing cyberbullying, and system is not checked for cross platforms. The analysis is based on dictionary of bullying terms rather than semantic index. Further, personality and sentiment/emotion of a person can be considered for better accuracy.

References

1. Sonia L et al. (2011) Risks and safety on the internet: the perspective of European children: full findings and policy implications from the EU kids online survey of 9–16-year olds and their parents in 25 countries.
2. Tokunaga RS (2010) Following you home from school: a critical review and synthesis of research on cyberbullying victimization. *Comput Human Behav* 26.3: 277–287
3. Bullying statistics. <http://www.bullyingstatistics.org/content/cyber-bullying-statistics.html>
4. Annapolis. <https://www.annapolis.gov/908/Facts-About-cyberbullying>
5. Chris M (2014) Cyberbullying, trends and tudes.“ NCP. org
6. <https://sites.google.com/a/cypanthers.org/cease-cyber-bullying/statistics>
7. Mckenna KY, Bargh JA (1999) plan 9 from cyberspace: the implications of the internet for personality and social psychology. *Personal Soc Psychol Rev* 4(1):57–75
8. Gross EF, Juvonen J, Gable SL (2002) Internet use and well-being in adolescence. *J Soc Issues* 58(1):75–90
9. O’Moore M, Kirkham C (2001) Self-esteem and its relationship to bullying behaviour. *Aggressive Behav* 27(4):269–283
10. Fekkes M, Pijpers FIM, Fredriks AM, Vogels T, Verloove-Vanhorick SP (2006) Do bullied children get ill, or do ill children get bullied? a prospective cohort study on the relationship between bullying and health-related symptoms. *Pediatrics* 117(5):1568–1574
11. Cowie H (2013) Cyberbullying and its impact on young people’s emotional health and well-being. *The Psychiatrist* 37(5):167–170
12. Price M, Dalgleish J (2010) Cyberbullying: experiences, impacts and coping strategies as described by australian young people. *Youth Stud Australia* 29(2):51–59
13. Royen V, Kathleen et al. (2015) Automatic monitoring of cyberbullying on social networking sites: from technological feasibility to desirability. *Telematics Info* 32.1:89–97
14. Parental Sites (e.g. NetNanny, <https://www.netnanny.com/>)
15. What is cyberbullying? <https://www.stopbullying.gov/cyberbullying/what-is-it/index.html>
16. Willard NE (2007) In: Cyberbullying and cyberthreats: responding to the challenge of online social aggression, threats, and distress. Research Press, Champaign
17. Olweus D (1993) *Bullying at school: what we know and what we can do* Blackwell Publishing, Malden
18. Amanda L et al. (2015) In: *Teens, technology and friendship*. vol 10. Washington, DC, Pew Research Centre, Book
19. Lynne E, Kontostathis A (2012) Reclaiming privacy: reconnecting victims of cyberbullying and cyber predation. In: *Proceedings of the reconciling privacy with social media workshop, held in conjunction with the 2012 ACM conference on computer supported cooperative work*. ACM
20. Al-Garadi MA, Hussain MR, Khan N, Murtaza G, Nweke HF, Ali I, Mujtaba G, Chiroma H, Khattak HA, Gani A (2019) Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. *IEEE Access* 7:70701–70718
21. Quan H, Wu J, Shi Y (2011) Online social networks and social network services: a technical survey. In: *Pervasive communication handbook*. Boca Raton, FL, USA, CRC Press, pp 4
22. Peterson JK, Densley J (2016) Is social media a gang? toward a selection, facilitation, or enhancement explanation of cyber violence. *Aggression Violent Behav*
23. BBC (2012) Huge Rise in Social Media. <http://www.bbc.com/news/uk20851797>
24. April E, Demoll D, Edwards L (2020) Detecting cyberbullying activity across platforms. In: *17th international conference on information technology–new generations (ITNG 2020)*. Springer, Cham, pp 45–50
25. Bigelow JL, Edwards A, Edwards L (2016) Detecting cyberbullying using latent semantic indexing. In: *Proceedings of the first international workshop on computational methods for cybersafety*. pp 11–14

26. Dooley JJ, Cross D (2009) Cyberbullying versus face-to-face bullying: a review of the similarities and differences. *J Psychol* 217:182–188
27. Slonje R, Smith PK, Frise,” A. The Nature of Cyberbullying, and Strategies for Prevention”, *Computers in Human Behaviour*,29(1), PP:26–32,2013.
28. Chikashi N et al. (2016) Abusive language detection in online user content. In: *Proceedings of the 25th international conference on world wide web*. pp 145–153
29. Imrul K et al. (2015) The social world of content abusers in community question answering. In: *Proceedings of the 24th international conference on world wide web*. pp 570–580
30. Karthik D, Reichart R, Lieberman H (2011) Modeling the detection of textual cyberbullying. In: *Fifth international AAAI conference on weblogs and social media*
31. <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>
32. Cynthia H et al. (2018) Automatic detection of cyberbullying in social media text. *PLoS One* 13.10
33. <https://github.com/apeksha104/Cyberbullying-Detection-in-Tweets>
34. Hugo R et al. (2019) Automatic cyberbullying detection: a systematic review. *Comput Human Behav* 93:333–345
35. Vimala B, Khan S, Arabnia HR (2020) Improving cyberbullying detection using Twitter users’ psychological features and machine learning. *Comput Secur* 90:101710

Quora Question Pairs Using XG Boost



Abhishek Chunamari, M. Yashas, Aparajitha Basu, D. K. Anirudh,
and C. S. Soumya

1 Introduction

Quora is a question–answer-based platform, where millions of people ask, contribute and edit every day. No co- incidence, similar questions with the same intent are asked quite regularly on the platform. In order to build a fine-quality knowledge base, it’s important that we ensure each unique question exists on Quora only once. In order to maintain the quality of the platform, it is important that one does not write the same answer to multiple versions of the same question, and readers should be able to find a single canonical page with the question they’re looking for. The Quora Question pair based on machine learning using natural language processing (NLP) and XGBoost algorithm will eradicate the duplicates and minimize inessential contents on the site using a real-world dataset of question pairs.

Key reasons to select Quora Question pair:

- Several questions with the similar objective cause seekers to spend their valuable time searching for the best answered question.
- Also, the contributor has to respond to numerous questions with the same intent.

In order to have a strong foundation, writers shouldn’t have to write the same answer to multiple versions of the same question, and readers should be able to find a single canonical page with the question they’ve been looking for. For example, we’d consider questions like “how to call using whatsapp?,” “how can whatsapp be used for calling?,” and “what are the different ways can one communicate using whatsapp?” to be duplicate questions because they all have the same intention. To prevent duplicity and redundancy from existing on the site, the paper talks about a

A. Chunamari · M. Yashas · A. Basu · D. K. Anirudh · C. S. Soumya (✉)
Department of ISE, Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: soumya.cs@nmit.ac.in

machine learning-based natural language processing system to automatically identify when questions with similar meaning have been asked numerous times.

Some of the major improvements are:

- A single canonical page
- Reduced redundancy
- Duplicity eradicated
- Finer experience for the writer/contributors and seekers

2 Motivation

There are billions of people existing on Earth and over 100 million people visit Quora every month, so coincidentally similar questions are asked on the platform. Numerous questions with the same intent can cause seekers to spend more time and energy to find the suitable answers to their questions; also writers need to answer multiple times for the similar question.

Canonical page systems are sought as they provide a better experience to active seekers and writers. To prevent duplicates from existing on Quora, we've developed machine learning-based natural language processing systems to easily identify when questions with the similar meaning have been asked multiple times [1].

3 Methodology

The methodology that has been followed for the implementation is shown in the following flow diagram (Fig. 1).

The dataset used for the implementation is provided by Quora which was released as their first public dataset. The dataset contains training dataset of 4 lakh question pairs and a testing dataset of 23lac question pairs which are separated in a tabular format.

Fields of the Dataset:

- *id*—unique identifier for the question pair
- *qid1*: Is the unique identifier for the first question
- *qid2*: Is the unique identifier for the second question
- *question1*: Is the full unicode text of the first question
- *question2*: Is the full unicode text of the second question
- *is duplicate*: labels 1 if questions are duplicates, 0 otherwise

Training dataset is the actual dataset which is used in machine learning to train the program or models to perform the desired actions. This data is fed to different algorithms in order to develop the machine or train the machine to perform automatic actions. The dataset is divided into training data and testing data in the ratio of 70:30.

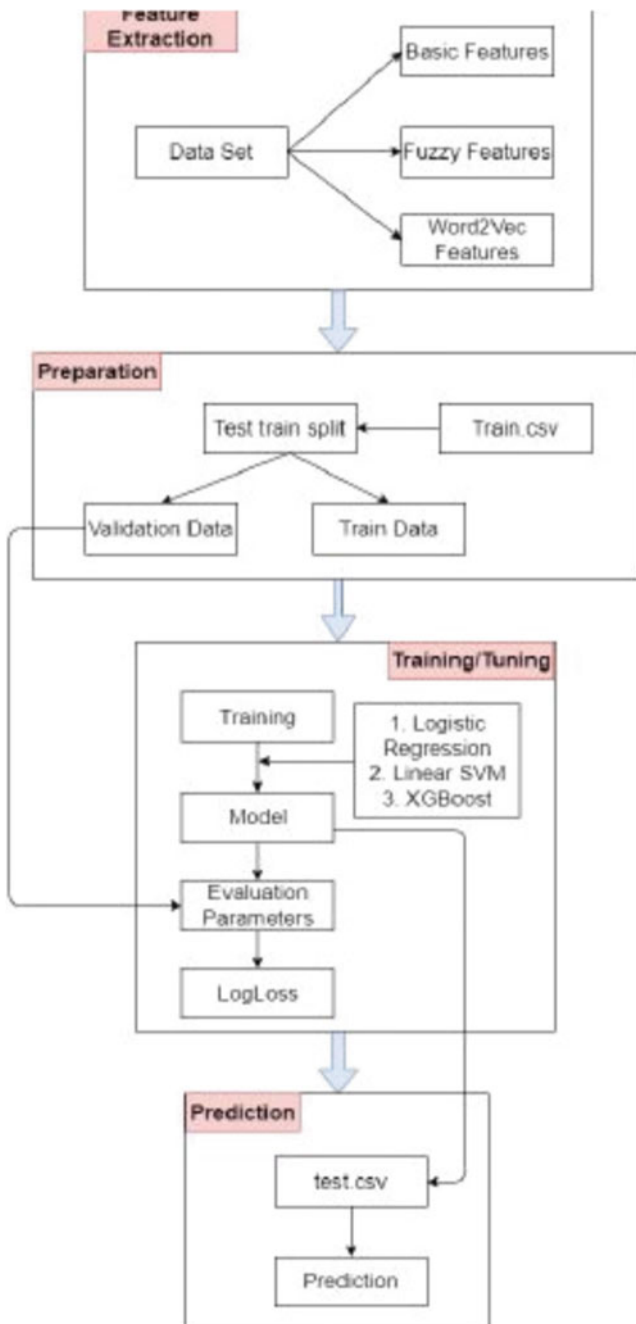


Fig. 1 Flow diagram of implementation

Pre-processing is the process that is used to convert the available dataset which is in a raw format into a clean dataset. The pre-processing of data has various steps. The whole process involves the removal of html tags, correction of typing errors, removal of punctuations, replacing unreasonable words by a common word for example hasn't is replaced with has not, removal of stop words such as like also, also lemmatization which involves for example do, did, done is replaced with do [2].

Feature extraction is the process of extraction of different features from the dataset such as basic feature extraction and fuzzy feature extractions.

Basic feature extraction includes extraction of simple features of sentences such as

- length of the questions
- common words in the question pair
- total number of words in the questions
- sum and difference of the frequency of the qids

Fuzzy feature extraction involves the extraction of the fuzzy features from the sentences. The edit distance is used also called Levenshtein distance which is the minimum edits operations that are needed to change a particular string to another string

- *Simple fuzzy ratio*: it is the simplest fuzzy feature which involves calculating the edit distance between the two strings. It is used to find the minimal difference between the two strings such as punctuations, misspelled words [3].
- *Partial fuzzy ratio*: This is a fuzzy feature which is used where every small detail in the sentence isn't considered. For example, when we don't need to consider the stop words or punctuations partial fuzzy ratio is used. This is mainly used to find substrings.
- *Token set ratio*: Token set ratio is mainly used when we don't consider the number of times a word is repeated in the string.
- *Token sort ratio*: Token sort ratio is mainly used when the order of the words in the sentence doesn't matter. This is useful in plagiarism checks.
- *Features based on—Word2vec* is a process used to vectorize the large corpus of text. The algorithms need to be fed with numbers or vectors and do not take input in a string format. Word2vec is fed with an input of large corpus of text, and the output is a set of vectors of hundreds of dimensions. The vectors in the vector space are positioned in such a way that the vectors or words that share a common context are located in a much more proximity. We have used the google news vectors 300d to create a vector space where a vector of every word is created and are later merged using the sent2vec function. The distances are then calculated based on the vector of sentences -Euclidean, cosine, city block, Canberra, Murkowski, Bray Curtis, Jaccard.
- *Word mover Distance*—Word mover distance is used to measure the distance between 2 sentences or documents even where there are no common words among them. Even though the sentences have no common words WMD measures the dissimilarity among the sentences. Word mover distance uses the word vector embedding method. It treats the sentences as weighted point cloud of embedded

words. WMD uses both semantic and syntactic approach to measure the similarity between the sentences. It calculates the dissimilarity among the sentences by measuring the distance that every word vector has to move in order to change from one sentence into another.

- **Training classifier**, the training classifier used is the XG Boost algorithm. XG boost stands for extreme gradient boosting. XG Boost is a decision tree-based machine learning algorithm that uses a gradient boosting framework. It has high execution speed and high model performance. It implements gradient boosting decision tree algorithm. XG boost algorithm has proved to be much more efficient and useful for data which is in a structured or a tabular form. It is a decision tree-based algorithm. It is an algorithm with the perfect software and hardware optimization techniques to produce results using less computing resources and less amount of time. It uses a technique where it analyses the previous mistakes and adds new models in order to correct those mistakes. It is an ensemble where it predicts the errors of existing models, creates new models and adds them as a new prediction until a final prediction is made.

Training classifiers used to compare with XG boost are Logistic regression and Linear SVM models.

- *Logistic regression*—Logistic regression uses the statistical method for making prediction as an output. It computes whether an output occurs or not. It predicts the probability of the occurrence of a binary event such as a certain class or event existing such as pass or fail, alive or dead, healthy or sick. This is further extended to model several classes of events such as determining whether a question pair is duplicate or not.
- *Linear SVM*—Linear SVM stands for Linear Support Vector Machine. Support vector machines are supervised learning algorithms that can be used for calculation and regression problems or analysis. The algorithm separates the data along the line in the hyperplane in to two classes. Support vectors of the data are those vector points that are closest to the line/hyperplane, which are the critical elements of the dataset. The hyperplane is basically a line that separates and classifies data. SVM can be used in text classification. Given a set of data, training classifiers each marked as belonging to either of the categories, a SSVM training algorithm can build a model that can assign new data or examples to either of the categories [4].
- *XG Boost*—XG Boost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XG Boost is an implementation of gradient boosted decision trees designed for speed and performance. It is a perfect combination of software and hardware optimization techniques to yield superior results using less computing resources in the shortest amount of time [5].
- *TF-IDF*—We are now going to try to improve, by using something called TF-IDF (term-frequency- inverse-document-frequency). This means that we weigh the terms by how uncommon they are, meaning that we care more about rare words existing in both questions than common one. This makes sense, as for example

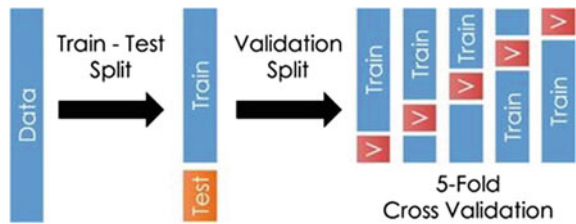
we care more about whether the word “exercise” appears in both than the word “and” - as uncommon words will be more indicative of the content.

4 Test Data

This data is used to determine the accuracy of the output (result). This data is matched with the output of the model, which had used training data. If the match is higher, then that means that the accuracy is higher and, if the match is lower, then the accuracy of the model is lesser compared to the others.

The dataset that was given to us is split into Train data and Test data. The split percentage is usually done randomly but we have divided it in 70 (Train): 30 (Test) ratio. The Train data is also split into different parts by applying the cross-validation process. This helps the model by making it more accurate. Below are some screenshots of some test data, which is different from the training data but has the same dimensions, as the parent dataset is the same [2] (Figs. 2 and 3).

Fig. 2 Working of 5-fold cross validation



id	A	B	C
1	test_id	question1	question2
2	0	How does the Surface Pro himself 4 compare with iPad Pro?	Why did Microsoft choose core m3 and not core i3 home Surface Pro 4?
3	1	Should I have a hair transplant at age 24? How much would it cost?	How much cost does hair transplant require?
4	2	What but is the best way to send money from China to the US?	What you send money to China?
5	3	Which food not emulsifiers?	What foods fibre?
6	4	How "aberystwyth" start reading?	How their can I start reading?
7	5	How are the two wheeler insurance from Bharti Axa insurance?	I admire I am considering of buying insurance from them
8	6	How can I reduce my belly fat through a diet?	How can I reduce my lower belly fat to one month?
9	7	By scrapping the 500 and 1000 rupee notes, how is RBI planning to fight against issue black money?	How will the recent move to declare 500 and 1000 denomination lewin illegal will curb black money?
10	8	What are the how best books of all time?	What are some of the military history books of all time?
11	9	After 12th years old boy and I had sex with a 12 years old girl, with her consent. Is there anything	Can a 14 old guy date a 12 year old girl?
12	10	What is the best slideshow app for Android?	What are the best app for android?
13	11	What services are from Google, Facebook, YouTube, Viber, Twitter, YouTube, Instagram, Skype, Wiki, etc	What social network like Google, Facebook, WhatsApp, Viber, Twitter, YouTube, Instagram, Skype, Wiki, etc
14	12	What if a cricket hits a batsman's helmet and then goes to the boundary?	Should carbonated red balls and 8 yellow balls. If 5 balls are drawn what is the probability of getting 2 red ba
15	13	Just how do you learn fruity loops?	How do Fruity Wrappers work?
16	14	Why does Batman get kill in Batman v Superman?	In Batman v Superman, why reduce Lex Luthor pit Superman against Batman?
17	15	When can I buy a SpaceX stock?	Should I sell or buy INRD stock?
18	16	Is it gouging and price fixing?	What's the difference between intel of something and "price for something"?
19	17	Can a vacuum cleaner concentrate suck your eye out if it is pressed against your face?	Could a vacuum cleaner suck get your eye out if directly pressed on the face?
20	18	I am 20 years old and I still a problem with pimples. What is the remedy?	I am 20 years old and still have acne. It seems more like inner mcent What is some known cure?
21	19	What is it ai living in the middle class?	Why middle class?
22	20	How matter at MIT? Will performing poorly in 11 grade affect my chance?	I have passed 5 AP tests with scores trump 5. Can I apply to UC Berkeley at 11 grade? If I got rejected, does it
23	21	What possible with SAT percentile between 85 and 90?	Is it possible that a person getting below 90 percentile in weres CAT cracked IIT/KAT/INMAT/INAP?
24	22	What are the differences between clients and servers?	What is the difference between a server and a database?
25	23	I want to eat hacking where should I start?	If I want to learn processor what should I do?
26	24	Why do people like Hrithik Roshan tax much?	Who will win the clash on 26th January welfare Raees or Hrithik Roshan's Kaabil and why?
27	25	What should be the first computer table language I learn?	Which language should I x4 as a first programming language?
28	26	What is a coffee and how to eat high on Monocin on an empty stomach with a low heart index. Is the absence of Rases a high dose?	
	test		

Fig. 3 Sample dataset

```
In [71]: pretty_table = PrettyTable()
pretty_table.field_names = ["Model", "Test Accuracy"]
pretty_table.add_row(["Logistic Regression Hyperparameter", Accuracy_Logistic_Regression_Hyperparameter])
pretty_table.add_row(["Linear SVM Hyperparameter", Accuracy_Linear_SVM_Hyperparameter])
pretty_table.add_row(["XGBoost", Accuracy_XG_Boost])
print(pretty_table)

+-----+-----+
|          Model          | Test Accuracy |
+-----+-----+
| Logistic Regression Hyperparameter | 0.7744666666666666 |
| Linear SVM Hyperparameter          | 0.7456333333333334 |
| XGBoost                            | 0.8199333333333333 |
+-----+-----+

In [99]: d_test = xgb.DMatrix(X_test)
p_test = bst.predict(d_test)
sub = pd.DataFrame(p_test)
sub.index = X_test.index
sub.columns = ["p_test"]
sub.to_csv("xgbresults.csv")
```

Fig. 4 Comparative test results

5 Conclusion

See Fig. 4.

As shown above, the XG Boost algorithm out performs the Linear SVM and Logistic regression. Logistic regression has an accuracy of 77%, Linear SVM has the least accuracy of 74% and XG boost algorithm has the highest accuracy of 81%.

References

1. Broder A (1997) On the resemblance and containment of documents. In: Proceedings of the compression and complexity of sequences 1997, SEQUENCES'97, Washington, DC, USA. IEEE Computer Society
2. Kim Y (2014) Convolution neural networks for sentence classification. In: Proceedings of the 2015 Conference on empirical methods for natural language processing, Doha, Qatar, pp 1746–1751
3. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: Proceedings of international conference on learning representations, ICLR 2013, Scottsdale, AZ, USA
4. Bogdanova D, dos Santos C, Barbosa L, Zadrozny B (2015). Detecting shingling MLP CNN LSTM LSTM + CNN Accuracy 0.6657 0.7263 0.8027 0.8107 0.8105 Precision 0.5151 0.5878 0.7102 0.6862 0.7004 Recall 0.7297 0.7245 0.7349 0.8441 0.7994 F1 0.6039 0.6490 0.7223 0.7570 0.7466 semantically equivalent questions in online user forums. In: Proceedings of the 19th conference on computational language learning, Beijing, China, July 30–31, pp 123–131
5. Wang Z, Hamza W, Florian R (2017) Bilateral multi-perspective matching for natural language sentences. <https://arxiv.org/abs/1702.03814>

A New Kaiser-Bessel Constant Modulus Technique for Smart Antenna Beamforming



K. S. Shashidhara, Veerendra Dakulagi, Jasmineeth Kaur, Kim Ho Yeap, Mandeep Singh, and Ratneshwar Kumar Ratnesh

1 Introduction

Smart antennas have long been an attractive solution to a plethora of problems related to signal detection, estimation and beamforming [1]. The smart antenna system which consists of an array of antenna elements with signal processing capabilities can overcome the directivity and beamwidth limitations of a single antenna element, and when it combined with methods from statistical detection and estimation and control theory, a self-adjusting or adaptive system emerges [2, 3]. Mainly, two types of configurations are used in practice, namely the switched beam system and adaptive array system [4]. The features of the switched beam system are (1) forms

The original version of this chapter was revised. In the affiliation of co-author “Ratneshwar Kumar Ratnesh”, the department has been changed. The correction to this chapter is available at https://doi.org/10.1007/978-981-16-1342-5_83

K. S. Shashidhara
Department of ECE, Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: shashidhar.ks@nmit.ac.in

V. Dakulagi (✉) · J. Kaur
Department of ECE, Guru Nanak Dev Engineering College, Bidar, Karnataka 585403, India

K. H. Yeap
Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Kampar, Malaysia
e-mail: yeapkh@utar.edu.my

M. Singh
Department of E&CE, National Institute of Technology Karnataka, Surathkal, India
e-mail: mandeep.singh@nitk.edu.in

R. K. Ratnesh
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology (MIET), Meerut, India
e-mail: ratneshwar.ratnesh@miet.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022, corrected publication 2022

N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790, https://doi.org/10.1007/978-981-16-1342-5_56

fixed multiple beams which have high sensitivity in certain directions. (2) It produces narrow beams with fixed directions. (3) They use simple algorithms for beamforming. (3) The interaction between the base station and the mobile unit is moderate. (4) The complexity and cost are less in this method. (5) It is very easy and simple to integrate with available existing cellular systems. (6) The capacity and coverage of this system are high as compared to the classical antenna systems. The features of the adaptive array system are: (1) this system steers the main beam towards the signal of interest (SOI) and the nulls in the direction of interfering signals. (2) This system requires a digital signal processor (DSP) for implementation. (3) This system uses complicated algorithms for direction finding and adaptive beamforming. (4) It is difficult to realize in existing systems and also expensive. (5) Interaction between a base station and a mobile unit is high due to continuous beam steering. (6) Interference suppression capability is more in this system.

2 Related Work

Adaptive beamforming algorithms have been used for decades in smart antenna technology for beamforming [5]. There are numerous adaptive beamforming algorithms and their variants are available in the literature. Among most of the methods, the least mean square (LMS) and its variants are the most simple and well-known methods. The LMS algorithm is computationally less complex and can be easily used in any smart antenna-based wireless communication systems. Nevertheless, the chief disadvantages of this algorithm are (1) it requires a reference signal for beamforming and converging. (2) It requires a minimum of 70 iterations before it can produce the beam in the indented direction. In certain applications, especially in radar and sonar blind beamforming algorithms that do not depend on the training sequence are required [6]. The most commonly preferred beamformer for such an application is the constant modulus algorithm (CMA). The CMA has been studied and exploited in the smart antenna beamforming for many years. This algorithm exhibits remarkably good performance with a constant envelope and does not use any pilot signal for the beamforming [7].

Adaptive beamforming algorithms were studied at the mobile base station using transmit diversity configuration [8]. An experiment was conducted using a 7-element circular array system at the mobile base station. A potential 4G communication system using a smart antenna is reported first time in [9]. Smart antenna-based direct sequence (DS) CDMA (CD-CDMA) for various bandwidths 1.25 and 5 MHz in difference channels environments is studied in [10]. It is reported that the wireless system using DS-CDMA for wider bandwidth presented better performance as compared to a narrow bandwidth. Increasing the throughput using smart antenna technology in a wireless communication system is presented in [11]. It is also discussed that the packets transmission and reception management is a key compensate for maximizing the system capacity. This concept used in the smart antenna system at a remote station. A smart antenna system using joint time processing is presented in

[12]. This increased the performance of the smart antenna system and can be used for 3G mobile communications. New beamforming algorithms have been developed in [13]. These algorithms were able to control the radiation pattern of multiple base stations. They have modified the broadband beamforming algorithm to improve the system capacity and beam pattern at multiple base stations [14, 15]. This solved the problem of normal smart antenna base stations when their distance becomes shorts.

In this work, we improve the convergence time of the classical CMA by modifying the weight update equation and then we apply the first time Kaiser-Bessel window to suppress the SLL. The experimental results are compared with the previous methods.

3 Problem Formulation

3.1 Array Signal Model

We consider the L antenna element uniform linear array (ULA) separated by $d = \lambda/2$ receive M user signals such that ($M < L$) from unknown directions $\theta = (\theta_1, \theta_2, \dots, \theta_M)$. Here the λ is the wavelength of the received signal. The array output of the ULA is expressed as

$$y(n) = \mathbf{w}^H \cdot \mathbf{x}(n) \tag{1}$$

Here, $\mathbf{x}(n)$ is the received signal vector and $\mathbf{w} = [w_1, w_2 \dots w_M]^T$ is the weight of each antenna element. $(.)^H$ represents the conjugate transpose.

The received signal in the presence of interference and the noise is expressed as

$$\mathbf{x}(n) = s_o(n) \mathbf{a}(\theta_o) + \mathbf{i}(n) + \mathbf{n}(n) = \mathbf{A}(\theta)\mathbf{s}(n) + \mathbf{n}(n) \tag{2}$$

where $\mathbf{a}(\theta_o)$ is the required steering vector of the received signal. $\mathbf{s}(n)$ represents the transmitted signals vector. $\mathbf{A}(\theta) = \mathbf{A} = [\mathbf{a}(\theta_o), \mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_{M-1})]$ is the array manifold. $\mathbf{i}(n)$ is the vector of interference received at the ULA and $\mathbf{n}(n)$ is the influence of noise vector with zero mean.

3.2 The CMA

The Godard CMA exploits a gradient-based technique for adaptive beamforming. Most of the algorithms require a reference or training signal for smart beamforming. The CMA computes array weights and forms adaptive beams without the use of training sequences. Thus, it is also called as blind beamformer. The weight update expression of this approach is

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \mu e^*(n) \times (n) \tag{3}$$

where the step-size is denoted by μ and ε represents the mean square error (MSE) which is written as

$$MSE = e = \left(y(n) - \frac{y(n)}{|y(n)|} \right) \quad (4)$$

The array output using the CMA is

$$y(n)_{CMA} = \mathbf{w}(n)^H \mathbf{x}(n) \quad (5)$$

Since CMA is a blind beamformer, it does not require any reference signal to track the desired signal and is the most suitable beamforming method for the radar application. However, the main downside of classical CMA is it exhibits more side lobes. Because of this lot of energy will be wasted in the system. Hence, reducing SLL is indeed required.

4 Proposed Methods

4.1 Improved CMA

The slow convergence time of the classical CMA algorithm can be improved significantly by properly modifying the weight update expression of this approach. The modified weight update expression is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{2\mu_n}{\|\mathbf{x}_n\|^2} \mathbf{x}_n \left(y_n - \frac{y_n}{|y_n|} \right) \quad (6)$$

4.2 Hanning Window-Based CMA (H-CMA)

In this method, we used the Hanning window concept to suppress the sidelobe level, and it is represented as

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{2\mu_n}{\|\mathbf{x}_n\|^2} \mathbf{x}_n \left(y_n - \frac{y_n}{|y_n|} \right) \quad (7)$$

Here, w_n is the weight of the Hanning window, and its value is given as

$$\mathbf{w}_n = \begin{cases} 0.5 \left(1 - \cos\left(\frac{2\pi n}{L}\right) \right), & 0 \leq n \leq L \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

4.3 Hamming window based CMA (HM-CMA)

Similar to H-CMA algorithm, we applied the Hamming window to the improved CMA algorithm, and it is expressed as

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{2\mu_n}{\|\mathbf{x}_n\|^2} \mathbf{x}_n \left(y_n - \frac{y_n}{|y_n|} \right) \tag{9}$$

Here, \mathbf{w}_n is the weights of Hamming window, and its value is given as

$$\mathbf{w}_n = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{L}\right), & 0 \leq n \leq L \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

4.4 The Proposed Kaiser-Bessel CMA (KB-CMA)

In this approach, unlike, H-CMA, HW-CMA, and KW-CMA, we combine the effect of Kaiser and Bessel window technique to reduce the SLL as

$$\mathbf{w}(n + 1) = \mathbf{w}(n) - \left[\frac{2\mu}{\|s(n)\|^2} (L, \beta_o) s(n) \xi \right] \tag{11}$$

$$\mathbf{w}(n) = \frac{I_0 \left[\pi \zeta \sqrt{1 - \left(\frac{n}{N/2}\right)^2} \right]}{I_0[\pi \zeta]} \quad n = 0, 1, \dots, N/2 \quad \zeta > 1 \tag{12}$$

where ζ is the parameter used to control the shape of the radiation pattern and it is given by empirical relations as: Finally, the array output of this method is

$$y(n)_{(KB-CMA)} = \mathbf{s}(n)^T \mathbf{w} = \mathbf{w}^H \mathbf{s}(n) \tag{13}$$

In the simulation, the ULA antenna array configuration is considered, the number of antennas used is 10, 20, and 30, the number of snapshots are 100, the spacing between the antennas is $\lambda/2$, the direction of user is 0_o . Figure 1 shows the convergence plot for H-CMA and K-CMA and also the radiation pattern of CMA. It is observed that, the convergence of both improved CMA and HW-CMA and K-CMA are almost the same, they converge within ten iterations. Figures 2, 3, and 4 show the radiation pattern of proposed H-CMA, HW-CMA, and KW-CMA for 10, 20, and 30 antenna elements, respectively.

The detailed performance of all three methods are analyzed and tabulated in the following (Tables 1, 2, 3 and 4).

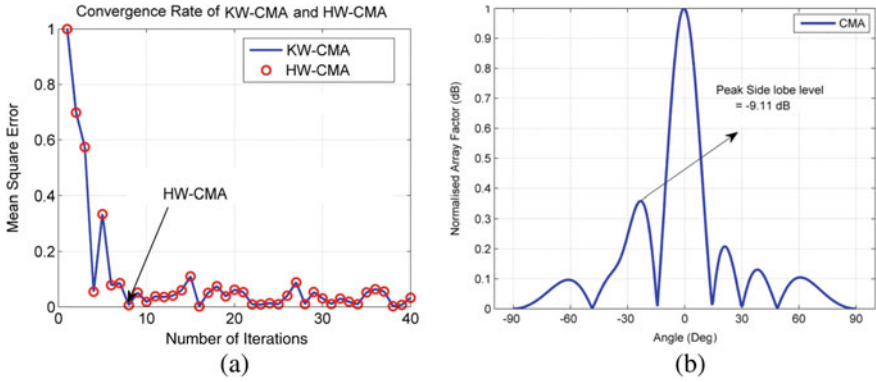


Fig. 1 a Convergence of HW-CMA & KW-CMA b Radiation pattern of classical CMA

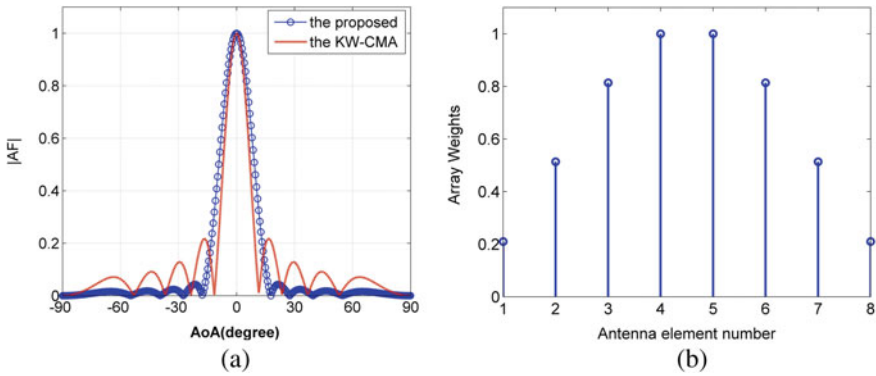


Fig. 2 KB-CMA for $L = 20$ a Radiation pattern b Array weights

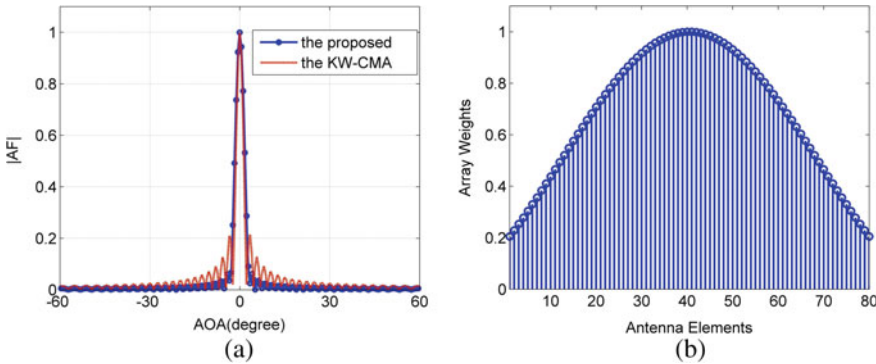


Fig. 3 KB-CMA for $L = 80$ a Radiation pattern b Array weights

Table 1 Performance analysis of H-CMA [14]

Angle of arrival	L	d	HPBW	SLL dB
10°	8	0.45λ	39°	-32 dB
10°	15	0.45λ	25°	-32.5 dB
10°	50	0.45λ	16 °	-35 dB

Table 2 Performance analysis of HW-CMA [14]

Angle of arrival	L	d	HPBW	SLL
10°	8	0.45λ	40°	-40 dB
10°	15	0.45λ	24°	-42 dB
10°	50	0.45λ	15°	-43 dB

Table 3 Performance analysis of KW-CMA [15]

Angle of arrival	L	d	HPBW	SLL
10°	8	0.45λ	36°	-80 dB
10°	15	0.45λ	18°	-82.1 dB
10°	50	0.45λ	10°	-88 dB

Table 4 Performance analysis of proposed method

Angle of arrival	Antennas	d	HPBW	SLL (dB)
10°	8	0.45λ	26°	-90
10°	15	0.45λ	22°	-92
10°	50	0.45λ	8°	-96

5 Conclusion

In this work, we proposed a new adaptive beamforming method based on the CMA algorithms. Simulation results clearly show that the improved CMA converges within ten iterations, whereas about 100 iterations are required for classical CMA to converge. Hence, the proposed method has ten times improvement as compared to the classical method. This factor makes the algorithm a suitable candidate for advanced wireless communication systems. To suppress the SLL, we have applied the Kaiser-Bessel window to the proposed algorithm. From the detailed analysis, we note that H-CMA, HW-CMA, KW-CMA, and the proposed KB-CMA at $L = 50$, respectively, have the SLL -35 , -42 , -88 , and -96 dB. This illustrates that the proposed method has an improvement of about 63.54, 55.20, and 8.4% improvement over the H-CMA, HW-CMA, and KW-CMA, respectively.

References

1. Godara LC (1997) Application of antenna arrays to mobile communications.II. Beamforming and direction-of-arrival considerations. *Proc. IEEE* 85:1195–1245
2. Gershman AB, Nemeth E, Böhme JF (2000) Experimental performance of adaptive beamforming in a sonar environment with a towed array and moving interfering sources. *IEEE Trans Signal Process* 48:246–250
3. Frost OL III (1972) An algorithm for linearly constrained adaptive processing. *Proc IEEE* 60:926–935
4. Buckley KM, Griffiths LJ (1986) An adaptive generalized sidelobe canceller with derivative constraints. *IEEE Trans Antennas Propagat.* 34:311–319
5. Zhang S, Thng IL (2002) Robust presteering derivative constraints for broadband antenna arrays. *IEEE Trans Signal Process* 50:1–10
6. Jiang B, Sun CY, Zhu Y (2004) A new robust quadratic constraint beamforming against array steering vector errors. In: *ICCCS June, Chengdu, China*, pp 765–768
7. Elnashar A, Elnoubi SM, El-Mikati HA (2006) Further study on robust adaptive beamforming with optimum diagonal loading. *IEEE Trans Antennas Propagat* 54:3647–3658
8. Li J, Stoica P, Wang Z (2003) On robust Capon beamforming and diagonal loading. *IEEE Trans Signal Process* 51:1702–1715
9. Li J, Stoica P, Wang Z (2004) Doubly constrained robust Capon beamformer. *IEEE Trans Signal Process* 52:2407–2423
10. Vorobyov SA, Gershman AB, Luo ZQ (2003) Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. *IEEE Trans Signal Process* 51:313–324
11. Shahbazpanahi S, Gershman AB, Luo ZQ, Wong KM (2003) Robust adaptive beamforming for general-rank signal models. *IEEE Trans Signal Process* 51:2257–2269
12. Xu C, Feng G, Kwak KS (2001) A modified constrained constant modulus approach to blind adaptive multiuser detection. *IEEE Trans Commun* 49:1642–1648
13. Li L, Fan HH (2000) Blind CDMA detection and equalization using linearly constrained CMA. In: *ICASSP June, Istanbul, Turkey*, pp 2905–2908
14. Md Bakhar (2017) Efficient blind beamforming algorithms for phased array and MIMO radar. *IETE J Res*
15. Veerendra, Alagirisamy M (2019) Kaiser window based blind beamformers for radar application. *IETE J Res (Taylor & Francis)* <https://doi.org/10.1080/03772063.2019.1689188>.

Generic Security Risk Profile of e-Governance Applications—a Case Study



B. S. Kumar, V. Sridhar, and K. R. Sudhindra

1 Introduction

The technology landscape and architecture of e-Governance applications have changed significantly with the adoption of new technologies like Cloud, containers, APIs, micro services, single-page applications, RESTful Web service, mobile applications, etc. Day-to-day governance is made more vibrant by adopting and integrating the application with SMAC (social media, mobile, analytics and Cloud). Further, the delivery of e-services is made cashless, paperless and presenceless in a seamless manner through a set of APIs provided by Aadhaar, e-Sign, digital locker, UPI, etc. The emergence of said new technologies, mode of access and use of third-party libraries and frameworks have enhanced the criticality and scope of evaluating the security risks in e-Governance application. There are number of critical security threats at application layer which could potentially pose significant risks to an organization's information if not handled properly. As good practice, it is required to determine the security risk profile of e-Governance application in the order of criticality and prevalence. For the purpose of this study, e-Governance application security risk profile is determined in three steps. To begin with, the most critical top application security issues specific to e-Governance is identified by studying and analyzing the security standards and vulnerability listing by industry [1] and community-driven projects such as OWASP [2], WASC [3], SANS/CWE [4], etc. Secondly, risk value associated with each of the identified top application security

B. S. Kumar
P.E.S College of Engineering, Mandya, India

V. Sridhar (✉)
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: venusridhar@yahoo.co

K. R. Sudhindra
B.M.S. College of Engineering, Bangalore, India

issues is estimated based on the methodology specifically devised for e-Governance applications. In the last step, security risk profile is derived by summarizing the top application security issues with risk ratings.

2 Step 1—Identification of Top Application Security Issues Specific to E-Governance Applications

There are numerous application issues/vulnerabilities which could be exploited causing security risks. Instead of listing out and evaluating applications based on numerous vulnerabilities, it is appropriate to study the pattern of security mistakes that generally occur in application design and categorize them according to the type of security risk it can result. An attempt is made to shortlist the type of application vulnerabilities based on criticality and prevalence. Accordingly, list of critical and high security risks as listed by Microsoft and community-driven projects such as OWASP, SANS/MITRE and WASC are referred and analyzed. This would enable testers to evaluate based on these limited number of top security issues categories found across many Web applications. This will also ensure uniformity in creating application security profiles which can be used to determine the security strength of an application.

Based on analysis and comparison of security risks of the above three community-based and industry sources, the following twelve security issues are shortlisted. This will ensure that most of security issues of all four sources are covered and are considered for further analysis for e-Governance applications. However, the sequences in

Table 1 Identified Top 12 application security issues in e-Governance

SI No.	Security issues/vulnerabilities
1	Injection
2	Broken authentication
3	Sensitive data exposure
4	XML external entity
5	Broken access control
6	Security misconfiguration
7	Cross-site scripting
8	Insecure deserialization
9	Using components with known vulnerabilities
	Insufficient logging and monitoring
10	Insufficient logging and monitoring
11	Cross-site request forgery
12	Unvalidated redirects and forwards

which the issues are listed in Table 1 are not based on prioritization and severity ratings of risks.

3 Step 2—Risk Assessment and Estimation of Risk Associated with Top 12 Application Security Risks in e-Governance

Risk rating methodology [5] devised specifically for e-Governance application is adopted to analyze and estimate the risk associated with each of the identified top application security issues. The severity of the risk is estimated based on their impact and likelihood of its occurrence. The approach for estimating risk severity is briefly discussed below:

- Likelihood is estimated by identifying factors under two main aspects, viz. threat agent and vulnerability. The factors considered in terms of threat agents are the skill level of the attackers, their motive to perform the attack and the population size of the attackers. Similarly, the factors considered from vulnerability perspective are how easy to detect the vulnerability, how easy to exploit and its prevalence in e-Governance application. The risk rating for each of these factors with values is detailed in the methodology.
- Impact is estimated by considering two types of impacts, viz. technical and business impacts. Factors considered for technical impact are centered on traditional security aspects such as loss of confidentiality, integrity, availability and accountability. Factors considered for business impact estimation includes financial damage, reputation damage and privacy violation. In both type of impacts, the risk ratings for each of these factors with values are detailed in the methodology.
- After assigning the rating values for each of the likelihood and impact factors, the average of the assigned ratings is calculated to determine the overall likelihood and impact. The overall risk is computed by taking the product of scores obtained from likelihood and impact.

4 Risk Estimation for Injection Vulnerability: A Case Study

As case study, risk estimation from e-Governance perspective is illustrated for injection vulnerability as this is listed as the number-one Web application security risk in the OWASP Top 10 and also its attack surface is enormous.

4.1 Injection

Injection attacks are to a wide class of attacking vectors which allow untrusted data input in a program and is processed as a command or query hampering the normal execution of the program. In e-Governance context, where information and privacy are of utmost importance and quality of service is paramount, this vulnerability poses a serious threat to modern e-Governance applications.

Attack Vectors. This reflects the context by which vulnerability exploitation is possible. There are different attack vectors and the exploitation of these vectors demands an understanding of the method being used for the attack. Following are the list of most commonly used vectors for injection attacks:

SQL Injection (SQLi). This methodology involves tampering of user provided data with SQL commands that can read or modify data from a database. This will open the door for the sensitive data to be exposed to the attacker and grants database access to some external entity. The complete ownership of data of the application is at stake and a comprehensive data breach could happen. Potential impact could lead to information disclosure, data loss, data theft, loss of data integrity, denial of service and full system compromise.

Code Injection. Certain elements of an application depend on the user actions to execute operating system commands. If a person is well versed in the OS architecture, some application code can be fed as system command on behalf of the Web user. This indeed poses threats and if privilege escalation is done, everything about the server, application, database and servers in the internal network could be compromised.

CRLF Injection. An attacker can tamper the response of a server to a Web client and add intentional carriage return and line feed (CRLF) character stream to split an HTTP response header with arbitrary script or command. The sophistication and impact surface of this attack purely depends on the implementation of the application.

IMAP/SMTP Injection. Various big enterprises use a dedicated mail server for information sharing. If the server is not hardened or if there is some room for the external IMAP/SMTP statements to be injected in the server because of improper sanitization, this attack gains its ground and if left untreated, sensitive information may get leaked to the outside world.

ORM Injection. Object relation mapping is a methodology to make database transaction simpler and act as a bridge between objects in code and database. ORM tools generate queries so that user does not struggle to frame complex queries by himself as it leaves a room for some open link for external entity to attack. But sometimes, these ORM tools do not sanitize user data which can lead to injection attacks. Unfortunately, developer cannot be aware as these ORM tools comes embedded inside a framework and internal details are not exposed to the outside world.

LDAP Injection. This is an attack vector to exploit Web applications that form LDAP (lightweight directory access protocol) statements from user-supplied input. When

a Web application fails to properly sanitize user-supplied input, an attacker can inject LDAP statements to execute arbitrary LDAP commands including granting permissions and modifying the contents of an LDAP tree. Potential impact could lead to information disclosure, authentication bypass, privilege escalation, etc.

Command Injection. This injection attack happens when an attacker gains access to execute system level commands through a vulnerable application. An attacker can manipulate certain input data in the Web interface to execute certain unauthorized system commands. This attack can leverage privilege escalation vulnerabilities which may lead to full system compromise.

XPath Injection. XPath Injection attacks occur when a website uses user-supplied information to construct an XPath query for XML data. Potential impact could lead to bypass authentication or access information without needing proper authorization.

4.2 Likelihood and Impact Estimation

Rationalization for estimating the likelihood for threat agents and vulnerability factors and impact for technical and business factors is given in Table 2. Based on risk rating for likelihood and impact, the overall risk for Injection security risk is estimated.

5 Security Risk Profile of E-Governance Application

Risk associated with all other remaining eleven top application security issues are estimated based on risk rating methodology [5] and case study illustrated for injection vulnerability at Sect. 4.

Based on the detailed estimation of likelihood risk factors (threat agent and vulnerability factors) and impact risk factors (technical and business factors) associated with each identified top application security issues, the summary of top e-Governance application security issues with risk ratings in the order of criticality and prevalence is given the Table 3.

The following inferences can be drawn from security risk profile based on the estimation of overall risk score:

- Security misconfiguration issue is most widespread (9), likelihood of occurrence is highest (7.3) and also has highest impact (7.2) with overall risk score of 52.8. The impact to the application varies, and it depends on the nature of the misconfiguration. The impact due to this high security risk may lead to unauthorized access to system data and functionality resulting in data theft or modification and system compromise.

Table 2 Estimation of likelihood and impact

SI No	Factors	Rationalization	Category for rating	Rating
<i>Threat agent</i>				
1	Skill level	Knowledge in SQL, LDAP. XPath, OS commands. XML parsers. ORM. SMTP Headers, etc. is desired skill level for exploiting	Programming skills	6
2	Motive	Motive is to cause data loss, corrupt data, denial of access, system compromise, etc	Shutting down/disruption of services 'Invasion of privacy	7
3	Size of Threat	Public at large (mainly hackers) including external users, internal	Hackers	8
		users could attack using any of the threat vectors listed above	Public (G2C)	
<i>Vulnerability</i>				
1	Detectability	It is easy to discover. Scanners and fuzzers can be used by attackers to find injection flaws in any of the threat vectors listed above	Easy to detect	8
2	Exploitability	Exploitability is easy as any source of data can be an injection vector	Exploitation techniques known	8
3	Prevalence	This vulnerability is widespread as this is invariably exploited by attacker using different attack surface and automated tools	Moderately Widespread	5
<i>Technical impact</i>				
1	Loss of confidentiality	Injection attack can result in data loss or corruption or disclosure to unauthorized parties	Disclosure of significant amount of sensitive/ strategic information	8

(continued)

Table 2 (continued)

SI No	Factors	Rationalization	Category for rating	Rating
2	Loss of integrity	All data could be modified or deleted	Modification of any file or information resulting in total loss of integrity	8
3	Loss of availability	This may result in denial of access and sometimes lead to complete host take over resulting in disruption of service	Loss of some services/reduced performance of critical services	5
4	Loss of accountability	Through system and application logs it is difficult to trace	Difficult to Trace	5
<i>Business impact</i>				
1	Business impact	Successful attack can result in full system compromise, account impersonation, information disclosure, authentication bypass, data loss, data theft, etc	Financial loss. Reputation Damage and Privacy Violation	8
	Overall likelihood		7	
	Overall impact		6.8	
	Overall risk		47.6	

- The injection security issue has second highest risk score of 47.6 with high impact risk rating of 7.0. Probability of occurrence stands second highest due to its easy vulnerability detection and exploitability and also large population of attackers and wide class of attack vectors. This is most dangerous Web application attacks and can result in data theft, data loss, loss of data integrity, denial of service and system compromise.
- Though insufficient logging and monitoring issue has third highest security risk score of 40.6, this issue is both application and operational related issue. From application aspect, it is provisioning of logging and detection security mechanism of auditable events such as failed logins, access control failures, validation failures and high value transactions events. However, insufficient monitoring, escalation and alerting active attacks in real time or near real time is basically operational security issue.
- Sensitive data exposure security issue is next highest risk issue with score of 38. As the name indicates, damage due to this issue could be destructive and dreadful keeping in view of regulatory and legal compliance requirement of protecting

Table 3 e-Governance application security risk profile

Application security risk	Likelihood			Vulnerability				Average		Impact					Overall risk (likelihood impact)		
	Threat agent			Threat agent				Average		Technical impact						Business Impact	Average
	Skill level	Motive	Size of threat agent	Detectability	Exploitability	Prevalence	Loss of confidentiality	Loss of integrity	Loss of availability	Loss of accountability	Business Impact						
Security misconfiguration	6	7	8	7	7	9	7.3	8	8	5	7	7.2	52.8				
Injection	6	7	8	8	8	5	7	8	8	5	8	6.8	47.6				
Insufficient logging and monitoring	6	6	8	7	7	8	7	5	5	7	7	5.8	40.6				
Sensitive data exposure	7	7	8	4	4	8	6.3	8	5	2	7	6	38				
Broken access control	6	7	7	4	4	5	5.5	8	8	5	7	6.6	36.3				
Cross site scripting (XSS)	6	7	8	7	8	4	6.7	5	5	2	7	5.2	34.7				
Using components with known vulnerabilities	4	7	8	7	7	5	6.3	5	5	5	7	5.4	34.2				
XML External Entities (XXE)	6	6	6	4	7	5	5.7	5	5	8	7	6	34				
Cross site request forgery	6	4	8	8	7	7	6.7	5	5	2	7	4.8	32				
Insecure deserialization	6	7	6	4	4	5	5.3	5	5	5	7	5.4	28.8				
Unvalidated redirects for forwards	4	4	8	7	7	4	5.7	5	5	2	7	4.8	27.2				

(continued)

Table 3 (continued)

Application security risk	Likelihood			Vulnerability				Average		Impact				Overall risk (likelihood impact)	
	Threat agent			Technical impact				Business		Average					
	Skill level	Motive	Size of threat agent	Detectability	Exploitability	Prevalence	Loss of confidentiality	Loss of integrity	Loss of availability	Loss of accountability	Business Impact	Average			
Broken authentication	6	4	8	8	7	5	5	5	2	2	7	4.2			26.6

privacy in today's context. Most government departments possess sensitive data like passwords, credit card numbers, health information, and details regarding court cases, strategic and defense information and other personal information which requires extra protection.

The above four high-risk security issues are analyzed from impact perspective. However, it is required to be ensured by stakeholders that all such identified high-risk security issues are addressed and mitigated throughout the application life cycle.

6 Conclusion

There are numerous application vulnerabilities which could be exploited causing security risks. Standard list of critical security vulnerabilities are already identified by industry and community-driven projects such as OWASP, SANS/MITRE and WASC. However, though knowing and evaluating vulnerabilities based on standard list is important, but knowing the actual associated risk from business perspective is just as important [3]. Hence, instead of evaluating applications based on standard listed vulnerabilities, it is appropriate to estimate and understand the associated risk from e-Governance security perspective.

The security risk profile in Table 3 provides summary of overall risk score estimated for each top twelve e-Governance security risk issues. This profile can be used as baseline by stakeholders to mitigate the high-risk issues to acceptable level by identifying and initiating remediation process with appropriate security controls spanning the application life cycle. This includes architects implementing security controls through threat modelling during design, developers performing secure code review during implementation and testers conducting vulnerability/penetration testing during application security assessment. This approach will not only produce hack resistant, secure e-Governance Web applications, but it enables the stakeholders to dynamically develop preventive mitigation strategies for potential vulnerabilities.

References

1. Microsoft Corporation. Improving Web Application Security: Threats and Countermeasures, Ver 1.0, 2003. <https://www.microsoft.com/en-in/download/details.aspx?id=1330>. Retrieved on 10 May 2020
2. The OWASP Foundation. OWASP Top 10 – 2017 The Ten most critical Application Security Risks. https://owasp.org/www-pdf-archive/OWASP_Top_10-2017_%28en%29.pdf.pdf. Retrieved on 10 May 2020
3. Web Application Security Consortium. WASC Threat Classification, version 2.00, Jan 2010. http://projects.webappsec.org/f/WASC-TC-v2_0.pdf. Retrieved on 16 May 2020
4. CWE Community. CWE Top 25 Most Dangerous Software Errors 2019. https://cwe.mitre.org/top25/archive/2019/2019_cwe_top25.html. Retrieved on 16 May 2020

5. The OWASP Foundation. OWASP Risk Rating Methodology. https://owasp.org/www-community/OWASP_Risk_Rating_Methodology. Retrieved on 16 May 2020
6. Common Vulnerability Scoring System (CVSS). Common Vulnerability Scoring System: Specification Document, Ver 3.1. https://www.first.org/cvss/v3-1/cvss-v31-specification_r1.pdf. Retrieved on 10 May 2020
7. Web Application Security Consortium. Web application security statistics 2008. <http://projects.webappsec.org/f/WASS-SS-2008.pdf>. Retrieved on 10 May 2020
8. Pandya DC, Patel NJ (2017) Study and analysis of E-Governance Information Security (InfoSec) in Indian Context. IOSR J Comput Eng (IOSR-JCE) 19(1):04–07. e-ISSN: 2278–0661, p-ISSN: 2278–8727
9. Shailendra Singh SK (2011) E-Governance: information security issues. In: International conference on computer science and information technology (ICCSIT'2011), Pattaya, Dec 2011, pp120–124
10. Teodoro N, Serrão C (2011) Web application security: improving critical web-based applications quality through in-depth security analysis. In: International conference on information society (i-Society 2011). London, pp. 457–462
11. Hassan RG, Khalifa OO (2016) E-Government—an information security perspective. Int J Comput Trends Technol (IJCTT) 36(1)
12. Joshi A, Tiwari H (2012) Security for e-governance. J Inf Oper Manage 3(1):pp-254–257. ISSN: 0976–7754 & E-ISSN: 0976–7762,
13. Ranga Rao V (2016) Government web application security: issues and challenges—a case of India. Int J Comput Appl Technol Res 5(10):619–626. ISSN:-2319–8656
14. Zhou Z, Hu C (2008) Study on the E-government security risk management. IJCSNS Int J Comput Sci Netw Secur 8(5)

Recent Intelligent Optimisation Algorithms for Isolated Microgrids: A Review



H. R. Sridevi , Shefali Jagwani , and H. M. Ravikumar

1 Introduction

Renewable energy sources are gaining a lot of importance in recent years as they provide an alternative to fossil fuels. The different renewable sources of energy are wind, photovoltaic, hydro, biomass, geothermal, tidal energy, etc. Renewable energy generation is increasing every year as it is cost competitive and also pollution free. The global contribution of non-conventional sources from the year 2013–2019 is shown in Fig. 1 [1]. In the year 2019, the installed capacity of new renewable power was more than 200GW. Out of this, 115 GW was from solar PV followed by wind power of around 60 GW and hydropower around 16 GW. The remaining were from biomass, geothermal and solar thermal power. Access to electricity to remote areas can easily be provided by the use of these sources. Microgrid is a new concept that enables RE integration. These are low and medium voltage distribution networks comprising of distributed energy sources including storage devices and loads. The basic structure of microgrid is as shown in Fig. 2.

2 Concept of a Microgrid

The classification of DG sources can be made as follows (i) variable speed (example—wind) (ii) high speed (example—microturbines) (iii) direct energy conversion (example—photovoltaic and fuel cells). Also, classification of microgrids

H. R. Sridevi (✉) · S. Jagwani · H. M. Ravikumar
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: sridevi.hr@nmit.ac.in

S. Jagwani
e-mail: shefali.jagwani@nmit.ac.in

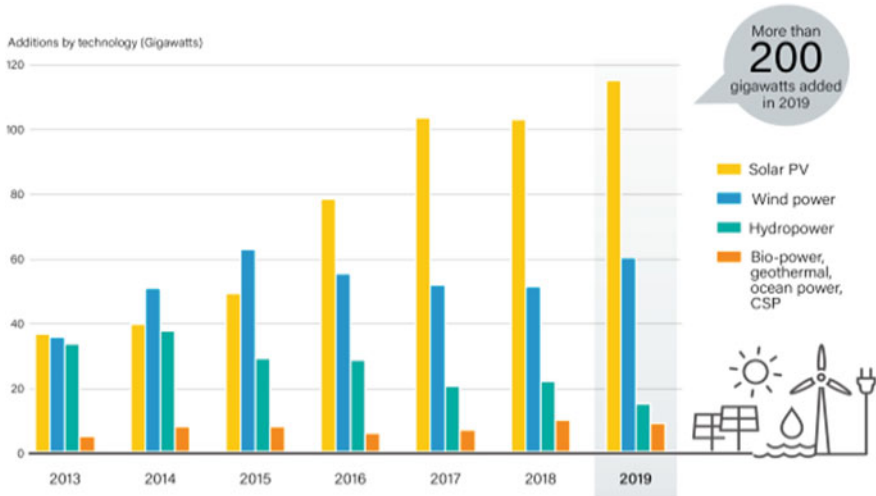
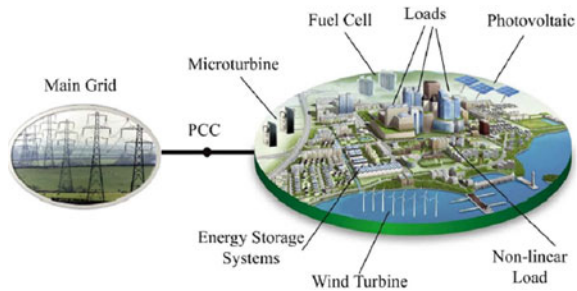


Fig. 1 Annual addition of renewable energy from 2013–2019

Fig. 2 Basic structure of a microgrid



can be considered as AC microgrid and DC microgrid. A typical AC and DC microgrid is as shown in Fig. 3 [2]. For an AC microgrid, AC bus connects the distributed generation and loads. DC type DER and energy storage devices are connected to AC bus via inverters. DC loads are connected to the AC bus via AC–DC converter. Connection of DERs and loads is to a DC bus in case of a DC microgrid. AC type DERs are connected to the DC bus via AC–DC converter.

The use of RES will make the need of power electronic interfaces in the distributed generation units significant. Control systems of the power electronic interfaces in the microgrids ensure the perfect functioning of microgrids. Also, the intermittent behaviour of RES leads to increased challenges for the power system. The modes of operation of a microgrid includes: (i) grid connected mode (ii) islanded mode. In the first mode, the microgrid is connected to the point of common coupling which maintains the microgrid at a constant frequency and allows import and export of power between the two. Whereas in case of islanded mode, microgrid operates independently by generating power locally. Here, voltage and frequency are controlled

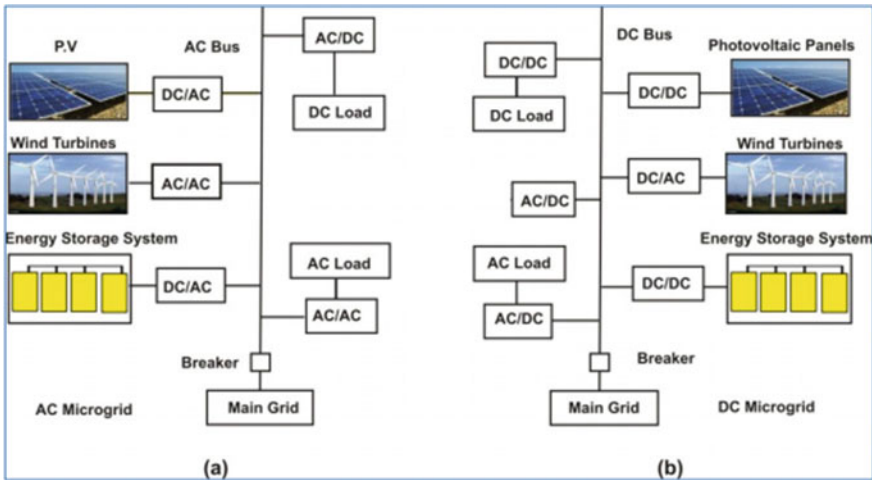


Fig. 3 A typical AC and DC microgrid

by the DGs through VSIs. In an islanded microgrid, the power sharing among various DG units is a difficult situation to handle Hannan [3]. Therefore, an optimal microgrid management is required which falls under the category of mixed integer non-linear programming. To handle and solve these problems in a microgrid, optimisation techniques should be employed.

3 Optimisation Techniques

Optimisation refers to finding the most optimal solution from a set of possible solutions meeting all the necessary constraints. Optimisation problems may be constrained or unconstrained. Solutions to problems encountered in the power system area is quite difficult to solve. In specific when we talk about a microgrid, the main issues are with respect to operation, sizing, power control and scheduling. To address such real-time problems, heuristic techniques are employed. In some situations, the most efficient solution obtained using heuristic techniques are unrealisable; hence, meta-heuristic techniques are developed. These are based on behaviour of nature which are either population based or trajectory based. Some of the soft computing techniques are genetic algorithms, particle swarm optimisation, artificial bee colony, grey wolf, grasshopper optimisation, firefly algorithm, squirrel search, whale optimisation, etc. [4].

In these soft computing techniques, the objective function used to find the best solution of all possible solutions is known as fitness function. In other words, fitness function tells us how close a solution is to its optimal value or how fit a solution is. The error criteria techniques which are performance indices are (i) integral absolute

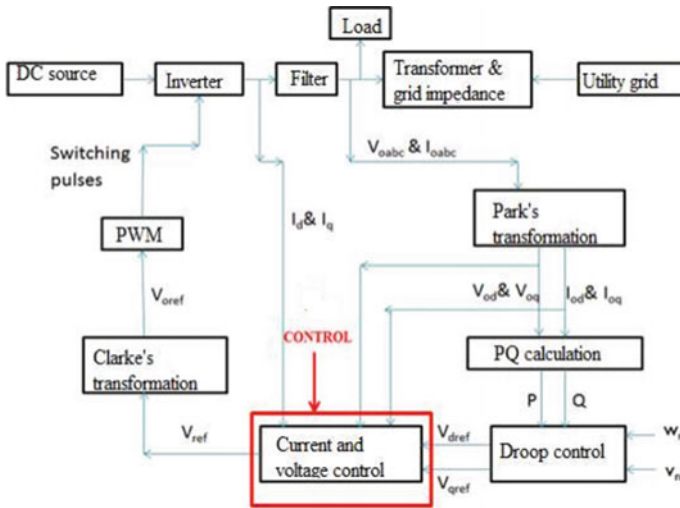


Fig. 4 Block diagram of the islanded mode control [6]

error (IAE) (ii) integral square error (ISE) (iii) integral time square error (ITSE) (iv) integral time absolute error (ITAE) [5].

Optimisation can be implemented in the microgrid to obtain the best performance taking into account all the constraints. In islanded mode of microgrid, an optimisation problem is formulated for design of controllers, filters and power sharing coefficients. The control objectives in the microgrid operating in autonomous mode are generally obtained using PI controller for power and current control loops. The power controller is meant for voltage–frequency (power control mode) required to obtain a maximum balance of power. The current control loop ensures good quality of the inverter output current. The power control loop provides a reference signal to the current controller and also regulates output power along with providing a refined reference signal to the current controller. The Fig. 4 shows the block diagram of the islanded mode control.

The DC output of the DER units are converted to AC of required voltage and frequency using an inverter. The inverter output is connected to the filter to reduce the unwanted harmonics. The three-phase voltage and current of inverter output is transformed to two phase using Park’s transformation. The real and the reactive power is then calculated. The droop control is used for sharing power between DGs in islanding operations. The reference values are fed to the current and voltage controller. The current and voltage controllers are conventionally PI controllers [7]. Clarke’s transformation [8] is then performed and the pulses are generated by applying suitable PWM technique to fire the switches of the inverter. Proper design of the controllers is needed to avoid instability in the microgrid, especially in islanded mode, or in other words, for reliable operation of the microgrid, a robust voltage and frequency control mechanism is essential. Intelligent searching algorithms can be applied for real-time self-tuning of the controllers [9, 10]. These algorithms are evolutionary, heuristic

and non-classical which show better results when compared to other conventional methods. Various algorithms used by researchers in their work include PSO, GA, ANN, ant colony algorithm, etc. The algorithms are used to solve the objective function to maximise the contribution of renewable sources by minimising the power loss, cost, etc. in the microgrid. The study of multi-objective optimisation problems shows superior performance by combining intelligent optimisation algorithms with adaptive techniques [11, 12]. This paper reviews the recent algorithms used for the islanded mode of microgrids.

4 Latest Algorithms Used for Microgrids in Islanded Mode

Some of the recent optimisation algorithms implemented by authors in the last few years are squirrel search algorithm, whale optimisation algorithm, grasshopper optimisation algorithm, firefly optimisation algorithm and grey wolf optimisation algorithm [4]. This section presents the review of these algorithms used in the last few years of this decade for the control of microgrids in islanded mode. Table 1 presents the detailed analysis of these algorithms.

4.1 *Squirrel Search Algorithm*

It is inspired from dynamic foraging pattern of flying squirrels by gliding. In hot weather conditions, these squirrels glide tree to tree in the woods in search of food. (Fig. 5). These creatures change their locations while searching for acorn nuts to maintain their energy requirements in the warm weather. Then, they start searching for hickory nuts which can be stored for cold weathers. As these squirrels are more active in hot weather, they usually store for cold weather during this time. On the basis of this behaviour, squirrel search algorithms can be used to model mathematical problems [14].

4.2 *Whale Optimisation Algorithm*

This technique follows the hunting pattern of humpback whales. It uses the attacking mechanism of whales known as a bubble-net feeding method to catch their prey. It uses a spiral bubble-net in order to hunt the food by generating distinctive bubbles in circular or a '9'-shaped form as shown in Fig. 6. The mathematical model can be created in three steps: encircling the prey, bubble-net attacking mechanism and then searching for prey. Once the model is obtained, WOA is applied to obtain the optimum solution [18].

Table 1 Recent intelligent optimisation techniques for microgrids in islanded mode

Authors and reference numbers	Optimisation techniques	Objective	Outcomes
Durairasan [13], Zheng et al. [14], Jain et al. [15]	Squirrel search algorithm	<ul style="list-style-type: none"> • Voltage and frequency control • Optimal power dispatch 	<ul style="list-style-type: none"> • As seasonal monitoring condition is incorporated, there is better and efficient search space investigation • Premature convergence • It gets confined to a local optimal result while handling difficult issues
Durairasan [13], Qazi et al. [16], Qazi et al. [17], Mirjalili and Lewisa [18]	Whale optimisation algorithm	<ul style="list-style-type: none"> • Voltage and frequency control • Optimal energy management 	<ul style="list-style-type: none"> • Avoiding local optima, it results in achieving better convergence speed • Duration of computation is more particularly for sophisticated problem
Touqeer et al. [19, 20]	Grasshopper optimisation algorithm	Power flow regulation	<ul style="list-style-type: none"> • Easy implementation and high accuracy • Unstable convergence speed • Easily gets trapped into the local optimum
Yang [21], Chaurasia et al. [22]	Firefly optimisation algorithm	<ul style="list-style-type: none"> • Voltage and frequency control • Microgrid operation optimisation • Optimal sizing 	<ul style="list-style-type: none"> • Speed of convergence is high in obtaining the global optima • Adaptability to form hybrid techniques by integrating with other optimisation techniques • To begin the iteration a fair initial solution is not needed • Falls into several local optima
Mirjalili et al. [23]	Grey wolf optimisation	Energy management	<ul style="list-style-type: none"> • Higher performance; • Lower computing Time • Quick searching • Poor local searching capability

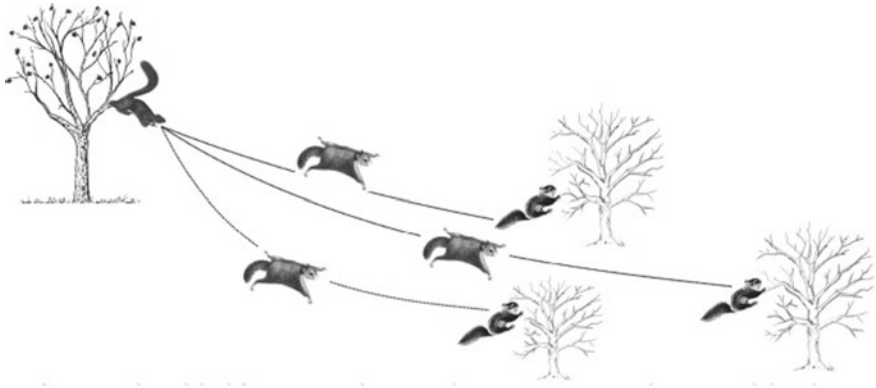


Fig. 5 Concept of squirrel search algorithm [15]

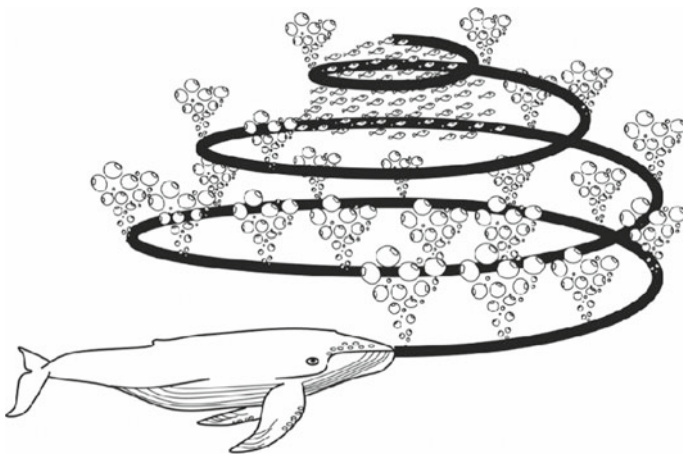


Fig. 6 Attacking mechanism of whale

4.3 Grasshopper Optimisation Algorithm

This recent algorithm is a nature-inspired optimisation technique which resembles the movement of an adult grasshopper in a swarm. This swarm is considered to be among the largest swarms in the world which consists of millions of grasshoppers. They migrate over large distances and may be a nightmare for farmers as they eat almost all vegetation. They follow both the search methods: exploration (move abruptly) and exploitation (move locally). The likely solution of the optimisation problem is determined by the location of the grasshoppers in the swarm. Depending on the grasshopper's position, it may experience three forces with respect to neighbouring grasshoppers: attraction, repulsion and neutral. Figure shows that the space

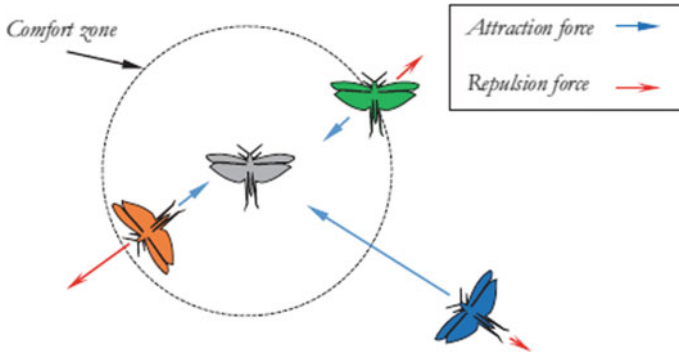


Fig. 7 Concept of grasshopper algorithm

has three sub-area: comfort zone, inside and outside of it as shown in Fig. 7. The main advantages of GOA include lesser control parameters, adaptive search arrangement with gradient-free technique [24].

4.4 Firefly Algorithm:

The light produced by fireflies is a unique phenomenon due to bioluminescence known as flashes. The pattern of flashes serve two fundamental functions: to communicate with their partners and to appeal the prey (Fig. 8). Additionally, it acts as a caution for them. Flashes are produced so that it is in conjunction with the objective

Fig. 8 Firefly algorithm [15]

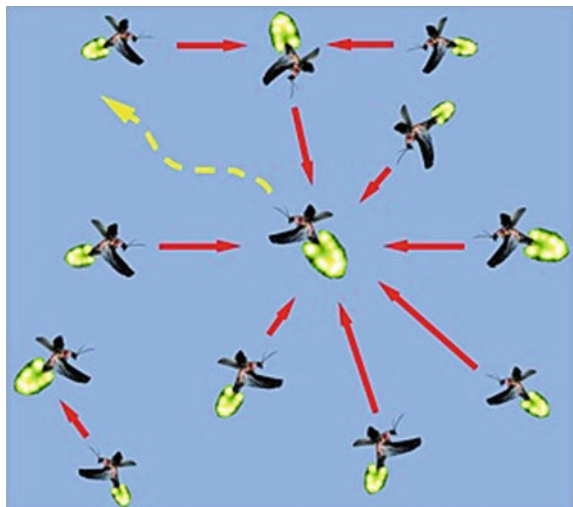


Fig. 9 Group hunting by grey wolves



function. This algorithm is mainly inspired by PSO and has some concepts similar to bacterial foraging algorithms.

However, FA has adjustable visibility, more versatility in attraction, better agility resulting in more efficient exploring of the search space [21].

4.5 Grey Wolf Optimisation:

Grey wolf optimisation was introduced by S. Mirjalili in 2014. It is based on the hunting techniques of grey wolves. These animals mostly wish to stay together as a team of 5–12. There are two interesting aspects of these wolves: first, they follow hierarchical levels (alpha, beta, omega and delta wolves) and second, the group hunting. The main steps of this optimisation technique involve social hierarchy, encircling prey, hunting and finally attacking the prey (Fig. 9) [23].

Table 1 shown gives a snapshot of the recent intelligent optimisation techniques for microgrids in islanded mode.

5 Conclusion

This work reviews the latest optimisation techniques for the islanded microgrids. Five newly introduced algorithms were reviewed with their advantages, limitations and challenges in microgrids. As these algorithms are problem dependent and not unique for a particular problem, it is very difficult to favour one particular algorithm. Thus, depending upon the nature of the problem, one can choose an algorithm. This work will be helpful for the researchers working in the microgrid domain.

References

1. Renewables 2020 Global Status Report. <https://www.ren21.net/gsr-2020>
2. Microgrid Chapter 3. <https://shodhganga.inflibnet.ac.in/bitstream/10603/142737/3/12.%20chapter%203.pdf>
3. Hannan MA, Tan SY, Al-Shetwi AQ, Jern KP, Begum RA (2020) Optimised controller for renewable energy sources integration into microgrid: Functions, constraints and suggestions. *J Clean Prod*. <https://doi.org/10.1016/j.jclepro.2020.120419>
4. Jumani TA, Mustafa MW, Alghamdi AS, Md Rasid M, Alamgir A, Awan AB (2020) Swarm intelligence-based optimization techniques for dynamic response and power quality enhancement of AC microgrids: a comprehensive review. *IEEE Access* 8:75987–76001. <https://doi.org/10.1109/ACCESS.2020.2989133>
5. Fang H, Chen L, Shen Z (2007) Comparison of different integral performance criteria for optimal hydro generator governor tuning with a particle swarm optimisation. In: *Lecture notes in computer science book series*, vol 4490. Springer, Berlin, Heidelberg, pp 1186–1189. https://doi.org/10.1007/978-3-540-72590-9_180
6. Control of converters in islanded and grid connected modes with proposed droop controller. https://shodhganga.inflibnet.ac.in/bitstream/10603/196012/1/11_11_chapter%203.pdf
7. Al-Saedi W, Lachowicz SW, Habibi D, Bass O (2012) Power quality enhancement in autonomous microgrid operation using Particle Swarm Optimization. *Electr Power Energy Syst* 42(2012):139–149. <https://doi.org/10.1016/j.ijepes.2012.04.007>
8. Chattopadhyay S, Mitra M, Sengupta S (2011) Clarke and park transform electric power quality, pp 89–96. doi.org/https://doi.org/10.1007/978-94-007-0635-4_12
9. Al-Saedi W, Lachowicz SW, Habibi D, Bass O (2013) Voltage and frequency regulation based DG unit in an autonomous microgrid operation using Particle Swarm Optimization. *Electr Power Energy Syst* 53:742–751. doi.org/<https://doi.org/10.1016/j.ijepes.2013.06.002>
10. Mohammadi FD (2017) Modeling, simulation and decentralized control of Islanded Microgrids. Graduate Thesis, Dissertations & Problem Reports. <https://researchrepository.wvu.edu/etd/7113>
11. Suchetha C, Ramprabhakar J (2018) optimization techniques for operation and control of microgrids—review. *J Green Eng* 8(4):621–644. <https://doi.org/10.13052/jge1904-4720.847>
12. Munira B (2018) Performance optimisation of standalone and grid connected microgrid clusters. PhD Thesis, Curtin University
13. Durairasan M, Balasubramanian D (2020) An efficient control strategy for optimal power flow management from a renewable energy source to a generalized three-phase microgrid system: a hybrid squirrel search algorithm with whale optimization algorithm approach. *Trans Inst Meas Control*. <https://doi.org/10.1177/0142331220901628>
14. Zheng T, Luo W (2019) An improved squirrel search algorithm for optimization. In: *Complexity*, vol 2019, Hindawi. <https://doi.org/10.1155/2019/6291968>
15. Jain M, Singh V, Rani A (2019) A novel nature-inspired algorithm for optimization: squirrel search algorithm. *Swarm* 44:148–175. <https://doi.org/10.1016/j.swevo.2018.02.013>
16. Qazi SH, Mustafa MW, Sultana U, Mirjat NH, Soomro SA, Rasheed N (2018) Regulation of voltage and frequency in solid oxide fuel cell-based autonomous microgrids using the whales optimisation algorithm. *Energies* 11:13–18. <https://doi.org/10.3390/en11051318>
17. Qazi SH, Bin Mustafa MW, Soomro S, Larik RM (2017) An optimal current controller for photovoltaic system based three phase grid using whales optimization algorithm. In: *IEEE conference on energy conversion (CENCON)*, Kuala Lumpur 15–20. <https://doi.org/10.1109/CENCON.2017.8262450>
18. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
19. Touqeer AJ, Mustafa MW, Md Rasid M, Mirjat NH, Leghari ZH, Salman Saeed M (2018) Optimal voltage and frequency control of an islanded microgrid using grasshopper optimization algorithm. *Energies* 11:3191. <https://doi.org/10.3390/en11113191>

20. Touqeer AJ, Mustafa MW, Md Rasid M, Mirjat NH, Baloch MH, Salisu S (2019) Optimal power flow controller for grid-connected microgrids using grasshopper optimization algorithm. *Electronics*, 8:111. <https://doi.org/10.3390/electronics8010111> www.mdpi.com/journal/electronics
21. Yang XS (2009) Firefly algorithms for multimodal optimization. In: Watanabe O, Zeugmann T (eds) *Stochastic algorithms: foundations and applications*. SAGA 2009 lecture notes in computer science, vol 5792. Springer, Berlin, Heidelberg, pp 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
22. Chaurasia GS, Singh AK, Agrawal S, Sharma NK (2017) A meta-heuristic firefly algorithm based smart control strategy and analysis of a grid connected hybrid photovoltaic/wind distributed generation system. *Sol Energy* 150:265–274. <https://doi.org/10.1016/j.solener.2017.03.079>
23. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer: theories, variants and applications. *Adv Eng Softw* 69:46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
24. Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. *Adv Eng Softw* 105:30–47. <https://doi.org/10.1016/j.advengsoft.2017.01.004>

Design and Analysis of Dual-Band Monopole Antenna Using Star EBG Structures



Prasanna G. Paga, H. C. Nagaraj, R. Tejas, and V. R. Sanath

1 Introduction

In recent years, monopole antennas integrated with planar EBG structures have gained tremendous importance for wireless communication applications due to their ability to improve gain and return loss of the Antenna. There has been a significant increase in the number of users accessing wireless communication. To cope up with increase in the users, the antennas must be small and compact and should give good bandwidth coverage [1]. Moreover, multi-band antennas have become very much useful in the field of communication as they support more frequency bands. In Shelke and Deshmukh [1], microstrip slotted type patch antenna is designed which supports IEEE 802.11a WLAN band, i.e. 5.10–5.50 and 5.85–6.25 GHz. Slotted type microstrip patch produced good results in terms of return loss, gain, VSWR and bandwidth. Return loss was recorded as -20.45 dB for 5.47 GHz and -29.59 dB for 6.1 GHz. In Abdulmohsen Al Shaikhli and Sadeq Abdulhadi Jalal [2], double patch antenna was designed using CSRR and DGS scheme for WLAN and ISM band applications. Also, complementary split ring resonator (CSRR) was employed in the ground plane to increase the performance parameters of the antenna. The measured return loss was -39 dB at 5.8 GHz operating frequency. Dual-band monopole antenna brings out many advantages as monopole exhibits omnidirectional pattern, hence minimizing the losses. One such antenna is proposed in Elwi et al. [3], for RF energy harvesting to serve itself for the applications of IoT. The antenna operates at 2.45 GHz and provided the conversion efficiency of 35%. Other

P. G. Paga (✉) · H. C. Nagaraj · R. Tejas · V. R. Sanath
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of
Technology, Bangalore, India
e-mail: prasanna.paga@nmit.ac.in

H. C. Nagaraj
e-mail: principal@nmit.ac.in

than wireless applications, antenna also serves itself in mobile applications. In Noori et al. [4], triple-band monopole antenna is designed for mobile applications. The proposed antenna covered frequency range from 2.3 to 4.1 GHz. This wide band covered Bluetooth, mobile WiMAX and WLAN bands. Another triple-band slot type microstrip patch antenna is designed in Shi et al. [5], which covered 0.9, 1.8 and 2.4 GHz. The simulated gain and return loss for 0.9 GHz were found to be 8.783 dBi and -11.618 dB, respectively. The corresponding values of gain and S11 were 7.903 dBi and -21.533 dB respectively, for Wi-Fi band.

Electromagnetic band gap (EBG) structures are periodic cells realized by the arrangement of dielectric substrate and metallic conductors. In recent years, EBG structures have been very useful in RF applications due to their special properties [6]. In Raghavaraju et al. [7], double EBG was incorporated around circular monopole antenna, which resulted in enhanced bandwidth and gain at 1.5–3.6 GHz and 4.8–15 GHz operating frequencies which is suitable for GSM/Bluetooth/WLAN/Satellite applications.

In recent years, many dual-band antennas are designed for wearable devices. These devices operate in ISM bands. One such dual-band antenna of size 35×20 mm² is proposed in Kadry et al. [8], wherein acceptable gain, high efficiency and omnidirectional pattern was obtained at ISM band. EBG structure also serves its uses in wearable devices, and the same is proposed in Ashyap et al. [9] with EBG-FSS (frequency selective surface) for medical body area applications to reduce the SAR. In Alqadami et al. [10], compact EBG is employed to reduce the mutual coupling between the bands.

2 Proposed Antenna

The proposed antenna design is as shown in Figs. 1 and 2. Figure 1 shows the top view of the dual-band monopole antenna without EBG structure whereas Fig. 2 shows the

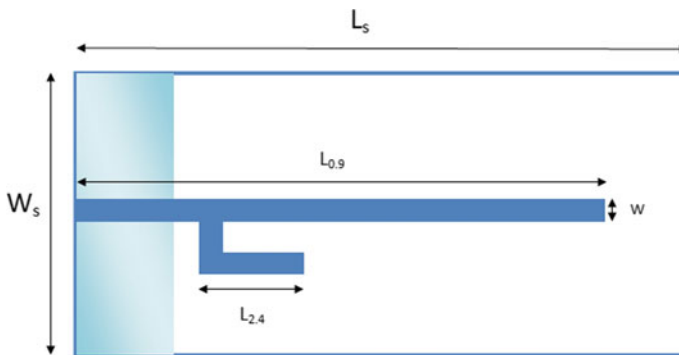


Fig. 1 Top view of proposed dual-band monopole antenna

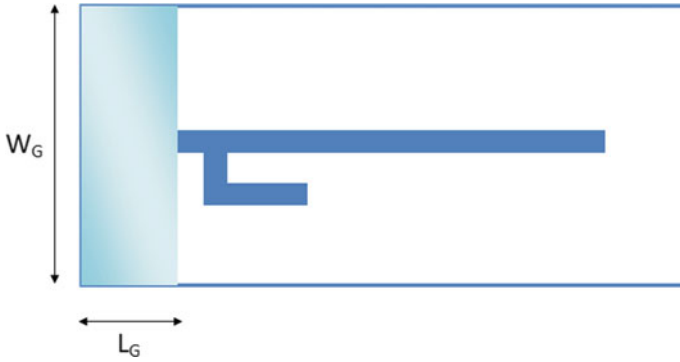


Fig. 2 Ground plane view of proposed dual-band monopole antenna

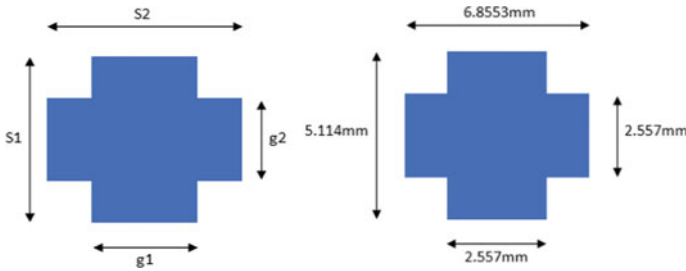


Fig. 3 Star-shaped EBG cell of size $6.8533 \times 5.114 \text{mm}^2$

ground plane view of the same. The long patch, i.e. $L_{0.9}$, is calculated to operate at 0.9 GHz frequency, and short patch, i.e. $L_{2.4}$, is calculated to operate at 2.4 GHz frequency.

In Fig. 3, length = 5.114 mm and width = 6.8553 mm. Star EBG is a modified form of square-shaped EBG cell. That is, all the four edges are etched off.

Star EBG cells, as shown in Fig. 3, are incorporated around the patch of 0.9 and 2.4 GHz in order to enhance the performance parameters of the antenna. Dual-band monopole antenna with star EBG structures is shown in Figs. 4 and 5. As indicated, EBG structures surround the patch of the antenna.

The complete specifications of the proposed dual-band monopole antenna with EBG structure is shown in Table 1.

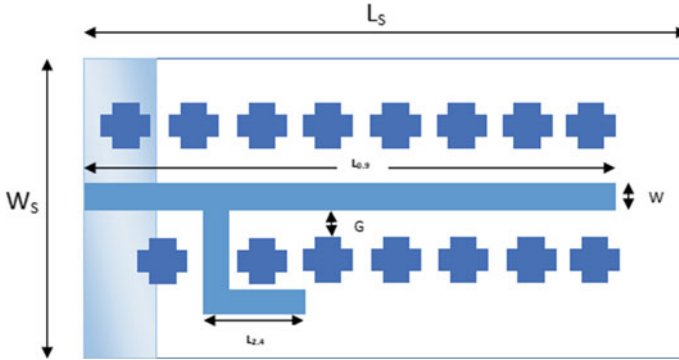


Fig. 4 Top view of dual-band antenna with star EBG structure

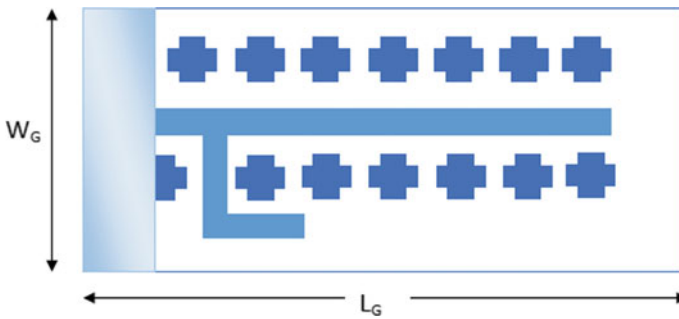


Fig. 5 Ground plane view of dual-band antenna with star EBG structure

Table 1 Antenna design specifications

Dimensions	Value (mm)
Length of the substrate, L_S	100
Width of the substrate, W_S	40
Length of the ground plane of antenna without EBG, L_G	23
Length of the ground plane of antenna with star EBG, L_G	12.5
Width of the ground plane, W_G	40
Width of the feed	3
Length of the patch for 0.9 GHz, $L_{0.9}$	88.9
Length of the patch for 2.4 GHz, $L_{2.4}$	12
Length of star EBG, S_1	5.114
Width of star EBG, S_2	6.8553
Gap between the patch and EBG cell, G	0.8645
Spacing between EBG cells	0.8

3 Antenna Design Consideration

3.1 Design Methodology

In the proposed work, a dual-band antenna has been designed on a FR4 substrate of permittivity 4.4 and thickness 1.6 mm to tune between GSM 900 and Wi-Fi (2.4 GHz) frequency band. The longer conductor is designed to resonate at GSM 900 frequency band, i.e. at 0.9 GHz, while the shorter strip is coupled to resonate at 2.4 GHz frequency band that happens to be the Wi-Fi applications' working frequency. To improve the performance parameters star-shaped EBG structure has been printed along the length of the conductors. Here, monopole class of antenna structures are investigated that helps in overcoming the space constraint characterized by a partial ground plane.

3.2 Design Equations

For 0.9 GHz operating frequency,

$$L_{0.9} = \lambda_g/4 \quad (1)$$

For 2.4 GHz operating frequency,

$$L_{2.4} = \lambda_g/4 \quad (2)$$

$$\lambda_g = \lambda_0/\sqrt{\epsilon_{\text{reff}}} \quad (3)$$

where λ_g = guide wavelength corresponding to GSM band, i.e. 0.9 GHz.

ϵ_{reff} = effective permittivity of the substrate.

λ_0 = free space wavelength.

The effective permittivity is calculated using the expression,

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12 \frac{h}{W} \right]^{-1/2} \quad (4)$$

where ϵ_r = relative permittivity, h = thickness of the substrate, W = width of the monopole strip.

The length and width of ground plane is calculated using,

$$L_G = 6h + L_{2.4} \quad (5)$$

$$W_G = 6h + W_{2.4} \quad (6)$$

Specific absorption rate (SAR) is given by,

$$\text{SAR} = \frac{\sigma E^2}{m_d} \quad (7)$$

where σ = conductivity of the material.

E = electric field.

m_d = mass density.

4 Simulated Results

The detailed simulated results of the dual-band monopole antenna with and without star-shaped EBG structure is discussed below.

Figure 6 shows the return loss (S_{11}) plot of dual-band monopole antenna without EBG structures. The return loss of -17.9963 dB is obtained for GSM 900 band and -12.05881 dB is obtained for Wi-Fi band.

With the introduction of EBG structures, the resonating frequency is shifted closely towards intended operating frequencies, and also the return loss is improved. The return loss comparison plot of both without and with EBG is shown in Fig. 7. The simulated values of return loss of antenna with EBG structures are -28.07146 dB for GSM band and -14.67345 dB for Wi-Fi band.

The 2-D radiation pattern plot for GSM and Wi-Fi band is shown in Fig. 8a and b. The angular axis represents angle in degrees, and the vertical axis represents gain in dB. The gain of 6.29939 dBi and 6.76136 dBi was obtained for GSM900 band and Wi-Fi 2.4 GHz band, respectively.

Star EBG structures improved the gain of the antenna at both the bands as shown in Fig. 9a and b. The radiating structure resulted in a gain of 6.3616 dBi and 7.1963 dBi for GSM 900 and Wi-Fi 2.4 GHz frequency bands respectively.

Fig. 6 Return loss plot of dual-band monopole antenna without EBG structures. Horizontal axis represents frequency (GHz) and vertical axis represents S_{11} (dB)

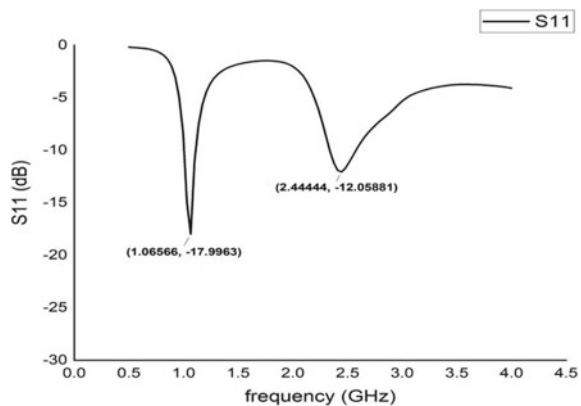


Fig. 7 Return loss plot of dual-band monopole antenna with and without EBG structures. Horizontal axis represents frequency (GHz) and vertical axis represents S11 (dB)

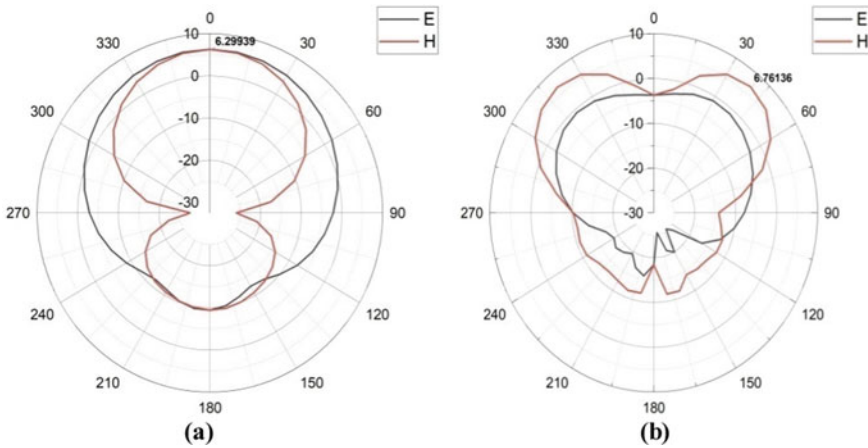
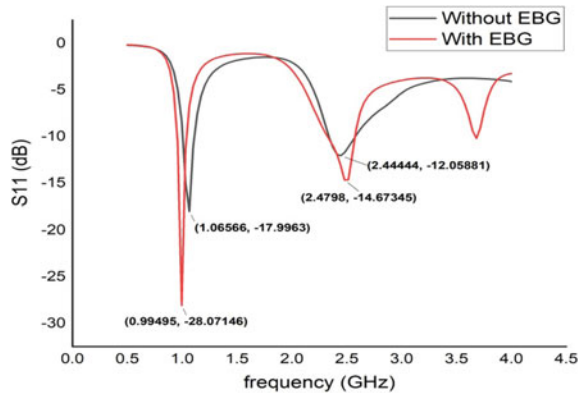


Fig. 8 2-D radiation pattern of dual-band antenna without EBG structure resonating at GSM 0.9 GHz band and Wi-Fi 2.4 GHz band

From Figs. 9 and 10, we can infer that the EBG structures increased the gain by nearly 1% (i.e. 0.98%) and 6.43% for 0.9 and 2.4 GHz bands, respectively.

Bandwidth coverage is also one of the important parameters of the antenna. Figures 10 and 11 shows the bandwidth coverage of the designed without and with EBG structured antenna for both the bands. It is evident from Fig. 10 that the bandwidth coverage for GSM 900 band is found to be 107.7 MHz and the same for Wi-Fi 2.4 GHz band is 240.1 MHz.

From the obtained simulated results, we can say that the EBG structures have increased the performance parameters of the antenna.

Moreover, altering the ground plane, spacing between EBG cells and also the size of the EBG cells affect the performance of the antenna. In this paper, we have also simulated the antenna by varying the specifications mentioned.

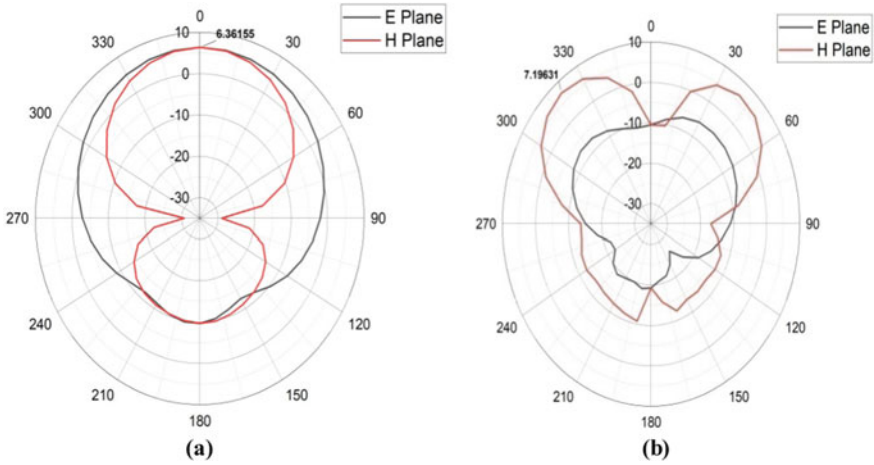
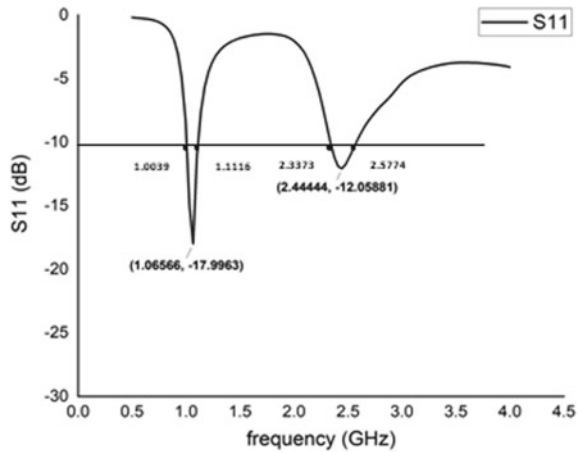


Fig. 9 2-D radiation pattern of dual-band antenna with star EBG structures resonating at GSM 0.9 GHz band and Wi-Fi 2.4 GHz band

Fig. 10 Return loss plot of dual-band antenna (without EBG structure) -10 dB impedance bandwidth extending from 1.0039 to 1.1116 GHz for 0.9 GHz band and 2.3373–2.5774 GHz for 2.4 GHz band



From Fig. 12, it is evident that the best return loss was obtained for spacing = 0.8 mm between consecutive star EBG cells. All the other spacing conditions satisfied $S_{11} < -10$ dB condition, but the best result was obtained for 0.8 mm spacing. Also, the length of the ground plane also accounts for the performance of the antenna. Ground plane (L_G) was decreased by 1 mm, and the best result was obtained for $L_G = 12.5$ mm and it is evident from Fig. 13.

The better S_{11} was obtained for $L_G = 9.5$ mm at GSM band, but it showed poor S_{11} in Wi-Fi 2.4 GHz band. Hence, we chose $L_G = 12.5$ mm as it showed better results in terms of S_{11} at both the bands.

Fig. 11 Return loss plot of dual-band antenna (with EBG structure) – 10 dB impedance bandwidth extending from 0.9555 to 1.0407 GHz for 0.9 GHz band and 2.3247–2.5750 GHz for 2.4 GHz band

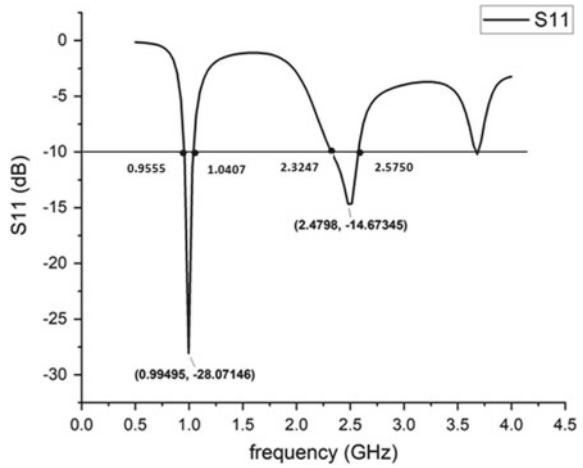


Fig. 12 S11 plot of dual-band antenna (with EBG structures) for different spacing between consecutive EBG cells

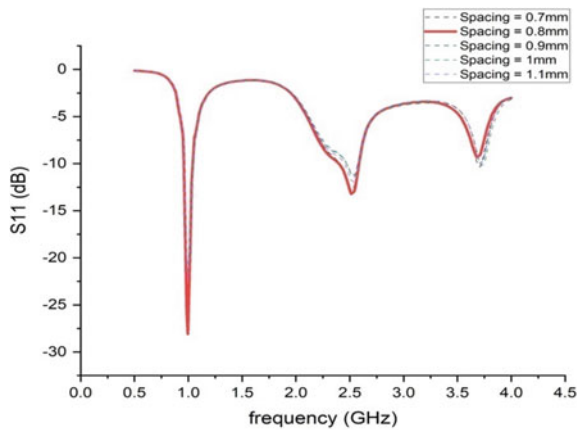
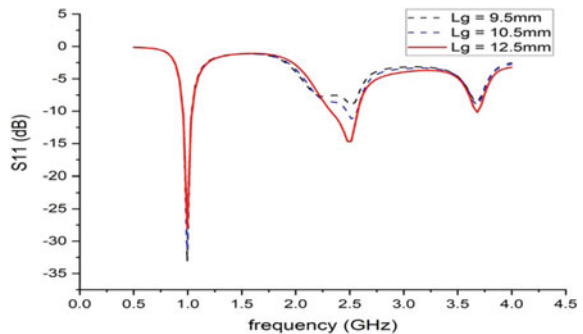


Fig. 13 S11 plot of dual-band antenna (with EBG structures) for different ground plane length



Also, the EBG cell size made a difference in the performance of the antenna as shown in Fig. 14.

It is also important to view an antenna in terms of gain and bandwidth. Table 2 shows performance parameters of the antenna with EBG when the spacing between the antenna and the cell, ground plane length and the EBG cell size were varied (Figs. 15, 16, 17 and 18).

Fig. 14 S11 plot of dual-band antenna (with EBG structures) for different star EBG cell size

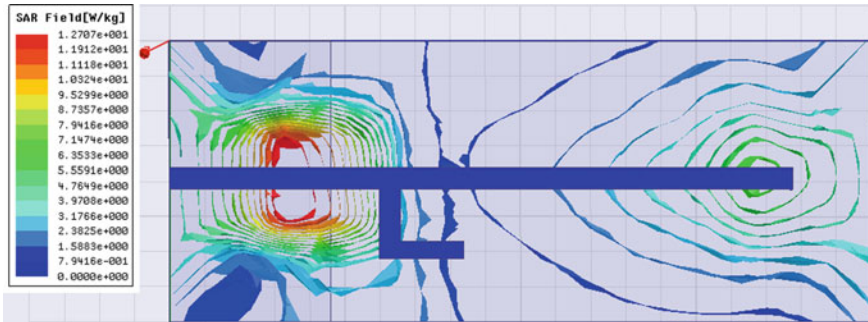
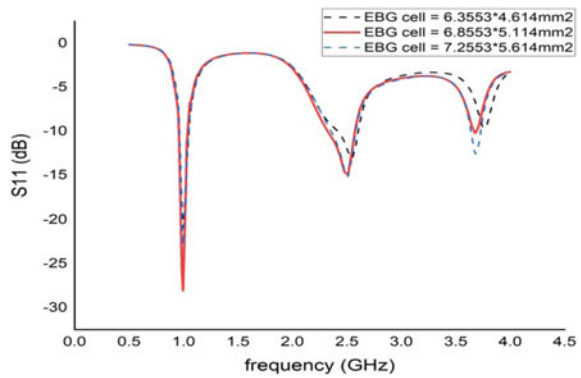


Fig. 15 SAR value of the antenna without EBG at 0.9 GHz indicating 12.7 W/Kg

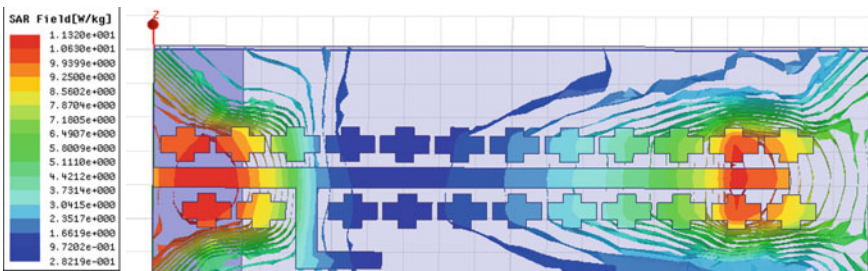


Fig. 16 SAR value of the antenna at 0.9 GHz with EBG indicating 11.32 W/Kg

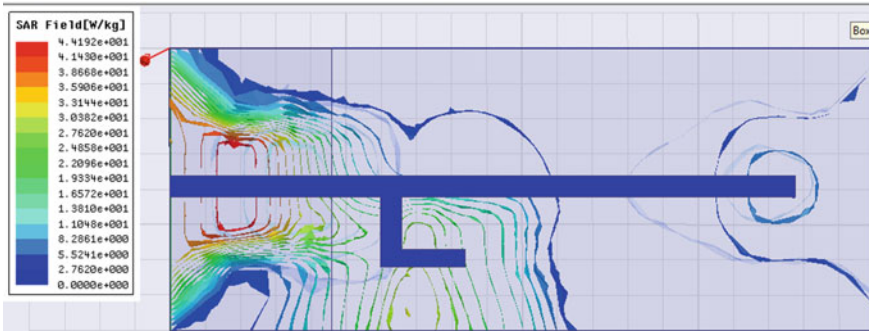


Fig. 17 SAR value of antenna at 2.4 GHz without EBG indicating a value of 44.19 W/Kg

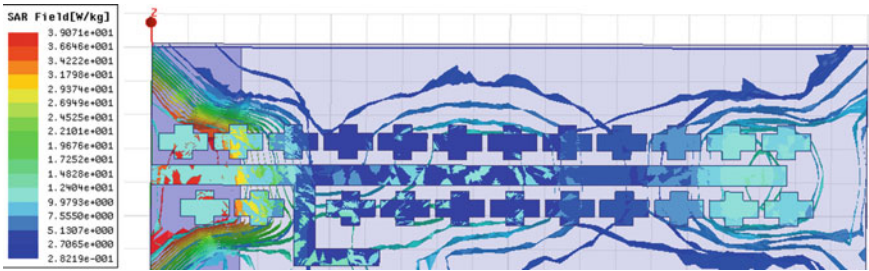


Fig. 18 SAR value of antenna at 2.4 GHz with EBG indicating a value of 39.07 W/Kg

Table 2 Performance parameters of the antenna for different antenna design specifications

Varying parameters	Value (mm)/(mm ²)	Return loss (s11)		Gain		Bandwidth	
		GSM band (dB)	Wi-Fi band (dB)	GSM band	Wi-Fi band	GSM band	Wi-Fi band
Spacing	0.7	-27.5432	-12.8545	6.3630dB	7.2867dB	83.7MHz	170.3MHz
	0.8	-28.0715	-14.6735	6.3616dB	7.1963dB	85.2MHz	250.3MHz
	0.9	-21.7024	-11.0835	6.2941dB	7.1548dB	88.5MHz	103.5MHz
	1	-22.0731	-11.2911	6.3022dB	7.2319dB	88MHz	109.8MHz
	1.1	-22.3827	-11.7894	6.3364dB	7.1575dB	88.1MHz	132.6MHz
Ground plane (L_G)	9.5	-33.0968	-9.0812	-	-	-	-
	10.5	-31.3091	-11.1736	6.3374dB	7.2857dB	82.6MHz	95.5MHz
EBG cell size (S1*S2)	6.3553*4.614	-20.9568	-13.0328	6.3278dB	7.6469dB	88.6MHz	190MHz
	7.2553*5.614	-22.8409	-15.0371	6.2781dB	7.3002dB	88.4MHz	227.6MHz

Table 3 Simulated Results of dual-band monopole antenna with and without EBG structure

Parameter	GSM—900 MHz without EBG	GSM—900 MHz with EBG	Wi-Fi—2.4 GHz without EBG	Wi-Fi—2.4 GHz with EBG
Return loss	−17.9963 dB	−28.0715 dB	−12.05881 dB	−14.6735 dB
Gain	6.29939 dB	6.3616 dB	6.76136 dB	7.1963 dB
Bandwidth	107.7 MHz	85.2 MHz	240.1 MHz	250.3 MHz
SAR (W/Kg)	12.7	11.3	44.19	39.07

5 Results and Discussions

The proposed dual-band monopole antenna with star-shaped EBG structure as shown in Fig. 4 can be used for GSM and Wi-Fi band applications. The geometrical specifications of the EBG cell and the antenna were reported in Fig. 3 and Table 1, respectively. The effect on performance parameters of the antenna when there are changes in the specifications are represented in Table 2. Simulated results of with and without EBG are represented in Table 3 along with specific absorption rate.

6 Conclusion

The above results infer that star-shaped EBG structure enhanced the performance parameters of the antenna by improving the return loss, gain and bands and bandwidth coverage. Without EBG, the GSM 900 band antenna attained a return loss of −17.9963 dB, gain is 6.29939 dB and bandwidth of 107.7 MHz from Table 3. And as for Wi-Fi band, the corresponding return loss values were −12.05881 dB, gain of 6.76136 dB and bandwidth of 240.1 MHz were obtained. With the integration of star-shaped EBG cells surrounding the antenna, there has been a significant improvement in the return loss of −28.0715 dB, gain of 6.3616 dB and bandwidth is 85.2 MHz for GSM 900 band. The corresponding parameters in the Wi-Fi band were: return loss equal to −28.0715 dB, gain of 7.1963 dB, bandwidth of 250.3 MHz. The corresponding SAR values reported were 11.3 W/kg and 39.07 W/kg at GSM 900 and Wi-Fi with EBG which are comparatively less compared to 12.7 W/kg and 44.19 W/kg for the structure without EBG, respectively, inferring the EBG structures can be used for reducing the SAR of the antenna, making it suitable for GSM and Wi-Fi applications.

References

1. Shelke SA, Deshmukh VU (2016) Dual-band microstrip patch antenna for wireless applications. *Int J Sci Res Sci Technol (IJSRST)* 2(2):334–337

2. Abdulmohsen Al Shaikhli NQ, Sadeq Abdulhadi Jalal A (2018) Miniaturized double-patch antenna design for WLAN communication with CSRRDGS. In: 2018 third scientific conference of electrical engineering (SCEE). Baghdad, Iraq, pp 226–229. <https://doi.org/10.1109/SCEE.2018.8684165>
3. Elwi TA, Almukhtar Tawfeeq O, Alnaiemy Y, Ahmed HS, Lajos N (2018) A UWB monopole antenna design based RF energy harvesting technology. In: 2018 third scientific conference of electrical engineering (SCEE). Baghdad, Iraq, pp 111–115. <https://doi.org/10.1109/SCEE.2018.8684112>
4. Noori O, Chebil J, Khan S, Habaebi MH, Islam MR, Saeed RA (2012) Design and analysis of triple-band microstrip patch antenna with h-shaped slots. In: 2012 international conference on computer and communication engineering (ICCCE). Kuala Lumpur, pp. 441–445. <https://doi.org/10.1109/ICCCE.2012.6271226>
5. Shi X, Wang C, Zhao C, Zhang Y, Zhao Y, Xi X (2018) Triple-band metamaterial-loaded small monopole antenna for mobile applications. In: 2018 12th international symposium on antennas, propagation and EM theory (ISAPE). Hangzhou, China, pp. 1–3. <https://doi.org/10.1109/ISAPE.2018.8634049>
6. Bhavarthe PP, Rathod SS, Reddy KTV (2019) A compact dual band gap electromagnetic band gap structure. *IEEE Trans Antennas Propag* 67(1):596–600. <https://doi.org/10.1109/TAP.2018.2874702>
7. Raghavaraju A, Ramakrishna TV, Madhav BTP, Bhavani (2019) Investigation on EBG structured CPW fed CM antenna for WiMAX, WLAN applications. In: 2019 international conference on vision towards emerging trends in communication and networking (ViTECoN). Vellore, India, pp. 1–4
8. Kadry M, Atrash ME, Abdalla MA (2018) Design of an ultra-thin compact flexible dual-band antenna for wearable applications. In: 2018 IEEE international symposium on antennas and propagation & USNC/URSI national radio science meeting. Boston, MA, pp 1949–1950
9. Ashyap AYI et al (2018) Highly efficient wearable CPW antenna enabled by EBG-FSS structure for medical body area network applications. *IEEE Access* 6:77529–77541. <https://doi.org/10.1109/ACCESS.2018.2883379>
10. Alqadami ASM, Nguyen-Trong N, Bialkowski K, Abbosh A (2019) Mutual coupling reduction in wideband electromagnetic medical imaging antenna array using compact electromagnetic band gap. In: 2019 IEEE international symposium on antennas and propagation and USNC-URSI radio science meeting. Atlanta, GA, USA, pp 999–1000

Database-Cloud Technology Framework Approach for Academia and Industry



A. Balachandra , B. R. Kiran , B. S. Ramya , and N. Madhu 

1 Introduction

Data, information, and record management (involving acquisition, creation, retrieval, update, delete, and storage) is an integral component of a patient medical record system, as it has been in the past, in the present time and will continue in the future. The approach in the past to manage patient data and information records was using written manual and human mental logs. These have proved in many instances incorrect due to human and system errors caused during acquisition, entry, recall, communication, record management, and operational use. The design approach, methodology, and the ways in which technology is used will continuously change, as driven by the state-of-the-art technology, doctor, and patient requirements. This will demand designers, manufacturers, service providers, users, and doctors to adopt and use state-of-the-art automatic/semi-automated database management system. Generally, doctor's medical treatment approach involves: data and information acquisition (past data/investigation results), analysis, consultation with peers, and medical experts while investigating symptoms and medical condition. This need imposes doctors or medical institutions to manage patient's history data with the purpose for reviewing

A. Balachandra (✉)

Information Science and Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: balachandra.a@nmit.ac.in

Dbali IT Solutions, Mysore, India

B. R. Kiran · B. S. Ramya · N. Madhu

Information Science and Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: kiran.br@nmit.ac.in

B. S. Ramya

e-mail: ramya.bs@nmit.ac.in

N. Madhu

e-mail: madhu.n@nmit.ac.in

at a later time using new technology and tools by creating an ecosystem. In the recent times, approach and methodology for managing patient history on database and providing data service through cloud interface have evolved due to the advancement in the state-of-the-art database and cloud technology. The availability of such technology and tools has enabled doctors, hospitals, and patients to adopt new state-of-the-art systems and technology for effectively managing patient medical condition data regularly with history.

A few of the major concerns during management of medical condition data for providing treatment are: simplicity in design of user interface for data entry automatically or manually, speed of data retrieval, complexity in database management, operational use, life of record, and accuracy of data. Present medical system for managing medical condition and treatment has evolved with the advent of computer and database technology which has made patient record (data) management extremely simple, medical history easily maintainable and instantly retrievable with no limits to geography demonstrating high level of data integrity and accuracy. This is also supported by high speed of Internet connectivity in recent times. In many situations, it is necessary and important for a patient to manage data and information in the absence of medical supervision. Self-management of medical condition involves self-monitoring and periodically measuring few physiological parameters and recording them. Obtaining such measurement can involve invasive or non-invasive method with manual or automatic ways of data collection at regular intervals. It is a matter of known fact that patient data obtained through self-managing medical condition manually are prone to error (s) caused by human, process, and/or instrument (s)/system (s) used during measurement. The data and information obtained must be recorded correctly and managed in suitable system (s) with the purpose for use by doctor for medical review and treatment at a later time.

This paper has evolved through industry project work in a student product development environment. A solution to an industry problem statement was developed by student groups over two years in two engineering institutes. During the first attempt in providing a solution to the problem statement, the student team was able to develop a sound understanding and provide a blueprint of solution framework. During the second iteration when the second student team was given ready blue print of solution framework, they were able to provide a limited operational system solution with the recommended technology and as directed. During the third cycle, the students were given a limited operational system solution as a reference to develop a related more complex application, and students developed a complete solution with front-end integrated with cloud-database accurately with guidance. Further, they were able to use the database system to develop dataset/data frame and run machine learning algorithms. The third student product development team was able to discover and apply the learning to develop more complex technology-related solution in current and growing future trend areas. This was a novel discovery we came across which motivated us to propose an approach to develop framework (s) for student higher order learning and design thinking. In contrast, industries have technology and tools evaluation division or teams where past solutions are reviewed and prepared for

the future product and prototype needs. Also in industry project, post-mortem tasks result in solution frameworks for the future project and solution needs.

2 Literature Survey

The concept of the cloud-database interface is a recent advancement which has been extended for application and use in medical area. Research has been in progress since few years on cloud-based systems and suitable factors such as availability, integrity, security, flexibility, backup, recovery for medical applications. The literature survey was conducted with objective of identifying research, development, and advancement in database-cloud integration with medical product, and the results are as follows: [1] mentions an overview of components of cloud interface and how they can be integrated with medical device for developing medical solutions emphasizing the design challenges, complexity, steep learning curve and how to leverage existing technology, clouds computational power, and performance attributes; [2] provides how the cloud can be integrated with many different medical systems (devices) in hospital environment for various data intense management on a database with automatic streaming, enabled database operations, security for shared data online with complete interoperability. The paper mentions how cloud can be successfully implemented with encryption for data security in neonatal care; [3] mentions some facts on cloud-based system (s) availability, and that acceptance for subscription-based model is increasing and the benefits of integrating medical device over the cloud for online access by physicians to make informed decision. This source also mentions how of design, providing examples of meta-level implementation, features, and areas where it is suitable.

The work presented in this paper is a framework approach for development and integration of database-cloud with medical product for data management/services, where data integrity and security are of primary concern. It is imperative that design of such systems and data communication over the Internet requires critical design thinking with high level of data security built into the product while following the regulatory guidelines and standards. Our proposed approach is unique and serves as a well-suited candidate for critical design development and evaluation of medical product/solution with pre-defined framework in student and industry environments.

3 Motivation

Developing a medical solution or a product for a problem involves: (1) understanding the need, requirements, and wants, (2) selecting suitable technologies, tools, and understanding their integration complexity, (3) applying regulatory and standard's guidelines, (4) evaluation of feasibility and risks, and (5) developing prototype for customer or design for manufacturing. Firstly, difficulty in developing the solution

lies in the time lag required to select and understand the technology, its capability, and suitability while addressing the complexity. Additional time and effort are expended in knowing its flexibility in integrating it with other related environments to successfully provide the solution to the problem. Following this, a generalized solution model is first developed; later, customizations are added based on the requirements. A generalized technology framework with suitable integrating environments if available will be very useful to develop an initial version of the solution by speeding up work; saving time from wasted effort on re-learning and trying possible way of integrating, using technology and tools. The problem is that in most situations, a technology framework is not made available even at times when a complete solution/system is deployed. There is a need to prepare and make such framework (s) available for student projects for developing higher order and design thinking, to foster innovation and for pedagogical teaching.

4 Proposed Model

The objective of this paper is to suggest an approach for preparing “reference framework (s)” for patient data management (with cloud-database interface) which can be used for faster and simpler solution development. This approach can be used in industry environment at specific time during the product development life cycle time line to speed up the development process by obtaining an understating of concept/prototype feasibility with knowledge of risks. Solution architects and developers/designers can focus on innovation and inventing product/services solutions for customer and service excellence with more assured quality product (s). In academic environment, this allows creating an early stage solution architecture for a problem statement using available “solution reference framework (s) and data model (s),” which can be customized further for specific requirements. This approach can foster higher order thinking, design thinking, and innovation in students enabling faster, effective learning.

To accommodate requirements from above discussion, a reference framework is provided for integrating database-cloud services. The database-cloud framework proposed provides the following design capabilities for developing an architecture solution and services for medical data management system:

- (1) Receive data through simple user interface by automatic and manual ways.
- (2) Organizing data received from automatic or manual ways into database.
- (3) Perform (CRUD) operations on the database.
- (4) Preparing and organizing data such as computation, statistical description, trend, analysis, machine learning, and prediction.
- (5) Facilitate provision of data services and other related services by integrating database with cloud.
- (6) Enabled communication between patient, doctor, hospital/clinic/laboratory through device/others.

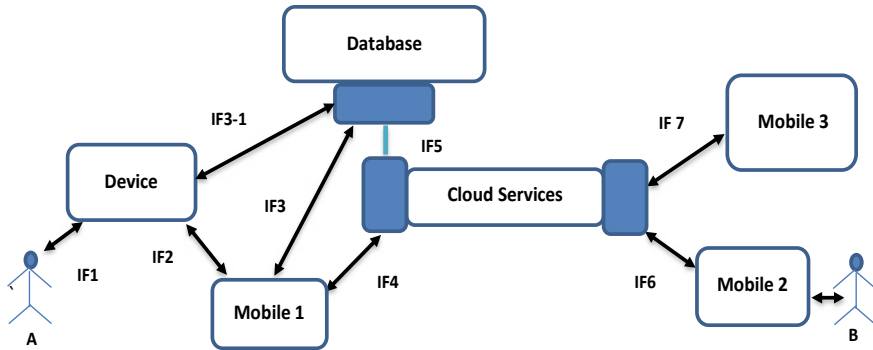


Fig. 1 Database-cloud reference framework

A reference database-cloud framework is as shown in Fig. 1. The components of database-cloud technology framework components are described below:

1. Medical Device/Patient/Doctor/Laboratory data acquisition and communication: Data can be obtained at regular intervals automatically or through manual ways from either or all of the sources. This data obtained from the device/patient/doctor/laboratory personnel manually or automatically can be transferred to database and perform related operations in any one of the following communication modes:
 - (i) From the device directly to the database through a Wi-Fi connectivity.
 - (ii) From the device to the patient registered mobile through a Bluetooth connectivity and from mobile to the database through a Wi-Fi connectivity.
 - (iii) From the device to another device registered with the patient through a Bluetooth connectivity and to the database through a Wi-Fi connectivity.
 - (iv) From the registered mobile to the database through a Wi-Fi connectivity.
 - (v) Avail cloud services from registered mobile (s).
 - (vi) Data communication from the device to the registered patient mobile and back to the device through Bluetooth connectivity, and such data transfer to the database through a Wi-Fi connectivity.
2. Database: Organizing data acquired from automatic or manual ways into database. Enabling write (save), read, search, update, delete.
3. Cloud Services (Database-cloud): Database integrated with cloud for providing following services:
 - (i) For retrieving data, analysis, and organizing data back into the database.
 - (ii) Providing data services though cloud to doctors/patients and registered others.
 - (iii) Capability to update database with treatment/prescription and details provided by the doctor through cloud services.

4. Patient Device(s): Data from patient device(s) stored onto the database for managing history data, communication, and operations through patient mobile
5. Doctors Device: Doctors device (mobile) configured for the doctor to communicate with
 - a. The patient mobile through Wi-Fi.
 - b. Cloud messaging service.
 - c. Cloud services (e.g., data analysis).
 - d. The database through the cloud.
6. Hospital/Clinic/Laboratory: Registered device(s) and mobile(s) are configured for communication with doctor, patient, and within their group.

4.1 Data Model for Proposed System

Medical data management system must demonstrate high level of data integrity, data security, and secured data communication over the network. The database and cloud platforms must be flexible to allow data model definitions on:

1. How data (with data type definitions) can be received, organized (in structured or unstructured modes) managed, and retrieved.
2. Retrieval of data from database for analysis and to see trends.
3. Use cloud services and communication over interface/links to users and devices.
4. Display data from database on device, user mobiles, and systems through known user interface.

The data model solutions must also be able to configure synchronization of data for receiving real-time data, perform interim storage, handle structured or unstructured mode of storage, and organize for easy retrieval from the database. Data are to be organized in unstructured/structured modes for the specific application and need based on the technology architecture. Though in fact the implementation of the solution is very simple for unstructured data management, it may require additional processing to reorganize data for creating a structure for use in relational database systems. This application approach is suitable for early stage product development and during technology phase. Figure 2 shows proposed data model.

This proposed approach using pre-developed framework (s) enables higher level thinking, innovation, and superior learning on product technology and integration development, further for customization through focused effort. This approach is suitable for solution preparation in student and industry environment fostering innovation.

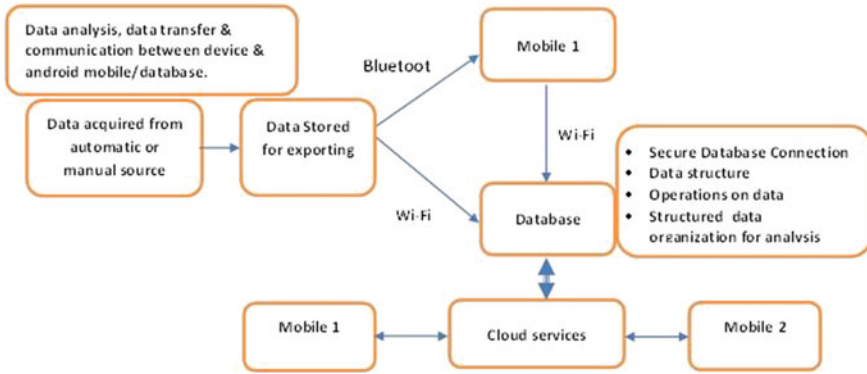


Fig. 2 Data flow between components in the proposed model

5 Database-Cloud Technology Deployment

The deployment of solution uses the above database-cloud proposed framework. The database-cloud technology framework can be used for implementation in alternative ways using available cloud-database technology platforms for database/cloud architecture description, performing operations description, data communication on the Internet, and at all times, regulatory compliance must be followed. In view of the available state-of-the-art of database and cloud technology, there are no database-cloud models for patient data management that can be used as reference models for developing or implementing solution (s). Such frameworks are not available in the academic environment for focused and knowledge learning and skill development. ***There is a need to provide standard frameworks to develop quality and regulatory compliant systems considering factors such as reliability, confidentiality, security, integrity, data accuracy, and scalability.***

A firebase technology-based solution implementation using above proposed database-cloud reference framework is presented in this section. Firebase is a fully managed platform for building iOS, Android, and web apps that provides automatic data synchronization, authentication services, messaging, file storage, analytics suitable for building a prototype mobile backend service also. Firebase API allows sync and store data in real time easily. Firebase hosted on secure sockets layer (SSL) is a typical security technology used to establish encrypted link between a server (host) and client (browser). Firebase is a backend as a service or BAAS, meaning it is a NoSQL data store that is in the cloud that clients can access directly in real time with real-time database, Cloud storage, authentication, hosting, and many other services. [4–7]. Figure 3 demonstrates an implemented database-cloud solution using proposed database-cloud framework.

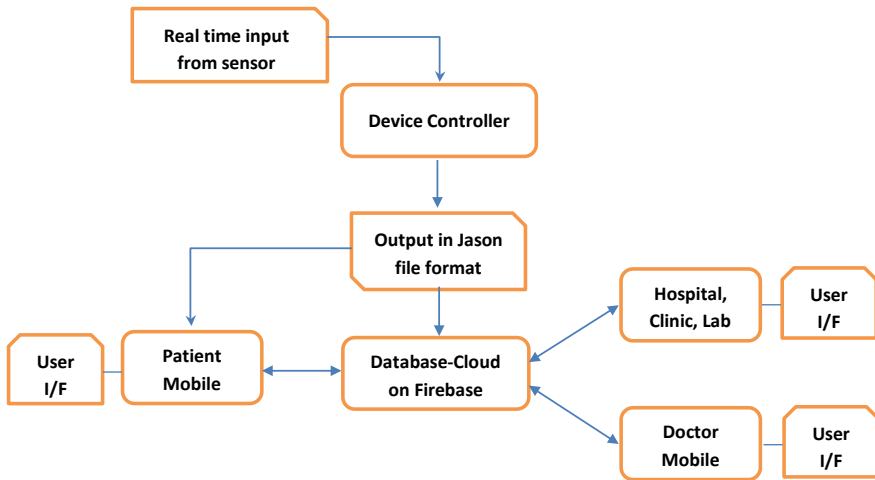


Fig. 3 Database-cloud solution using database-cloud reference framework

6 Conclusions

The proposed technology/data framework approach provides a simple, reliable, and secure way to develop and deploy database-cloud solutions in less time duration while expending less effort. It is well-suited for the first-time innovators, start-up companies, and new companies. This enables analysts, designers, and manufacturers to develop prototype product and services solutions with knowledge of risk and feasibility very early during the product development life cycle. Also, knowledge of required compliance to regulatory and standards can be gained during the development process. In academic environment, the proposed reference framework approach enables faster understanding of the problem statement in the solution domain/space along with high-level implementation details. This will enable students and academicians to invest time in the core areas of interest and research. Solution development can focus on innovation and inventing unique customizable solutions rather than expending effort and time to understand the technology, tools, domain and how to put them all altogether and technical resources. Results from our findings show that the use of such framework approach practices is beneficial in both academics and industry; this provides good credence for extending in other technology areas and applications.

Acknowledgements I, A. Balachandra, along with other authors of this paper wish to thank and express our deep sense of gratitude to Dr. N R Shetty, Director, Dr Nagaraj, Principal, Dr Sanjay, HOD, Department of ISE, from Nitte Meenakshi Institute of Technology, Bangalore, and Dr Vijay Kumar, B P HOD, Department of ISE, and Dr Lingaraju, Professor, from Ramaiah Institute of Technology for their support through industry-academia collaboration which has made this paper publication possible. I wish to thank participating students and all the co-authors and reviewers who

have been very helpful, patient, and providing their valuable contributions all through the time of the preparation of this paper.

References

1. <https://www.mddionline.com/software/cloud-considerations-medical-device-makers>, MDDI Qmed
2. A cloud-based healthcare infrastructure for medical device integration: the bilirubinometer case study. In: 2016 12th IEEE/ASME international conference on mechatronic and embedded systems and applications (MESA)
3. <https://www.medicaldevice-network.com/comment/impact-cloud-computing-medical-devices/>, The impact of cloud computing on medical devices. Verdict
4. <https://en.m.wikipedia.org/wiki/Firebase>
5. <https://en.m.wikipedia.org/wiki/IOS>
6. <https://medium.com/@reactsharing.com/5-reasons-to-not-use-firebase-for-a-big-project-81b543c77e8c>
7. <http://firebase.com>

Comparative Study of Directive-based Programming Models on CPUs and GPUs for Scientific Applications



C. Navya , H. A. Sanjay , and Sanket Salvi 

1 Introduction

The world of high performance computing is an interesting field of study from two decades. As the need for large and complex applications have been increasing across many industries, the need for high performance computing has also increased. With the need to give more assets to the users, there is the chance of utilizing all equipment assets proficiently. However, the CPU technology is not capable of scaling in performance sufficiently to address the demand. GPUs can give amazing execution by utilizing the available GPU cores. Graphics processing unit (GPU) can be used for parallel programming. Parallel programming is a process of dividing complex computational tasks into smaller tasks that can run concurrently. The main objective of HPC is to achieve better performance for large and complex scientific applications by adopting parallel programming paradigm. Multi-cores CPUs will improve the performance of the HPC application up to some extent and cannot be proficiently extrapolated forward in time. In both cases, it is difficult to overcome the critical part of the program to use the computational force efficiently.

For heterogeneous parallel programming, directive-based accelerators are used. Compared to low-level programming (PThreads, CUDA, and OpenCL) using directives-based language such as OpenMP and OpenACC. For parallel programming, OpenACC is the new programming standard, where application program interface characterizes a gathering of compiler orders to depict circles and locales of

C. Navya (✉) · H. A. Sanjay · S. Salvi
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: navya.c@nmit.ac.in

H. A. Sanjay
e-mail: sanjay.ha@nmit.ac.in

S. Salvi
e-mail: sanket.salvi@nmit.ac.in

code in FORTRAN, C, and C++ with the goal that they can be offloaded from CPU to quickening agent, which gives versatility across working frameworks, CPUs, and accelerators. OpenACC is a programming standard for parallel programming developed by PGI, Cray, CAPS, and Nvidia.

The main objective of this work is to utilize the available GPUs efficiently for parallel programming and to compare the time taken for executing scientific applications on CPUs and GPUs using OpenACC and OpenMP. The work focuses on parallelizing benchmark applications like fast Fourier transform, Laplace transform, molecular dynamics, and matrix multiplication using OpenMP and OpenACC directive-based languages. The experiment section demonstrates execution of these parallel applications on various programming environment. The result section demonstrates benchmark application implemented using OpenACC will perform better against to the benchmark application implemented using OpenMP.

The remaining part of this paper is organized as follows: Sect. 2 highlights various national and international efforts, which focuses on understanding the concepts of high performance computing, parallel programming, and directive-based programming models used for parallel programming. Sect. 3 focuses on architectural analysis of programming models. Section 4 discusses the proposed parallel implementation of various benchmark applications. Section 5 describes the performance evaluation, the experimental setup, and the results of the experiment along with the graphs. Finally, Sect. 6 provides conclusion and future work to enhance the proposed system.

2 Related Work

This section describes efforts made by several researchers to deal with issues and challenges in applying the methods for real-time applications. In Li, [1] has shown a performance examination among CUDA and OpenACC. The exhibition investigation manages programming models and fundamental compilers. The examination of execution holes has appeared in nineteen parts of ten benchmarks. They will in general use piece execution time and information affectability as primary principles reported once ends were made. A comparison of knowledge sensitivity may be a new index to explore an easily ignored downside that, however, each programming model is sensitive to changes of knowledge sizes. The vibes of the PRoDS equation brings the USA. A target examination rather than an emotional correlation. The OpenACC programming model is significantly more delicate to information than CUDA with advancements, while CUDA is much touchier than OpenACC to improvements. Generally speaking, the OpenACC execution is practically equivalent to CUDA under a decent correlation, and OpenACC might be a fair contrast to CUDA especially for amateurs in elevated-level equal programming.

The work [2] shows that OpenACC and OpenMP each lacks the power to completely generate a tailor-made multidimensional grid and threads for GPUs. Whereas this downside is often overcome by flattening the loop via the collapse

clause, their experiments have shown that the performance of such associate improvement would possibly still be slower than the CUDA 2-dimensional grid, wherever we have a tendency to achieve our greatest performance result.

In Ledur et al. [3] and Memeti et al. [4], authors show the characteristics of OpenCL, OpenMP, OpenACC, and CUDA with respect to programming productivity, enforcement, and energy. In Wang et al. [5], author shows how realistic it is to use a single OpenACC source code for a set of hardware's with different underlying micro-architecture, Nvidia Kepler, and Intel Knight Corner. In this project, we are considering Nvidia Tesla and Nvidia Quadro. Performance portability of OpenACC is related to the arithmetic intensity, and a big performance gap still exists in specific benchmarks between platforms.

The work [6] demonstrated that GPU-based parallel computer architecture has been showing extended notoriety as a building piece for first class handling and for future Exascale enrolling. They have assessed existing request-based models by porting application pieces from particular authentic spaces to utilize CUDA GPUs, which, along these lines, permits us to perceive fundamental issues in the supportiveness, adaptability, sensibility, and investigate limit of the current models.

3 Architectural Analysis of Parallel Programming Models

Generic directive-based programming frameworks comprise of directives, library routines, and designated compilers. In the request-based GPU programming models, a game plan of solicitations is utilized to stretch out data open to the selected compilers, for example, heading on arranging of circles onto GPU and information sharing rules. The most essential great circumstance of utilizing request-based GPU programming models is that they give bizarre state thought on GPU programming, following the consigned compiler covers the majority of the eccentric subtle parts explicit to the principal GPU structures. Another ideal position is that the solicitation frameworks make it simple to do steady parallelization of occupations, as OpenMP, with the ultimate objective that a client can choose domains of a host undertaking to be offloaded to a GPU gadget in a steady manner and a short time later the compiler, thus, makes related host gadget programs. There exists a couple of order-based GPU programming models. These models give unmistakable degrees of consultation and programming attempts expected to follow their models and smooth out the execution furthermore move.

3.1 *OpenACC Programming Model*

The OpenACC programming model adopts a renowned approach. The first application is explained with orders and calls to a runtime application program interface. The compiler is coordinated to create pieces that execute on the connected GPU or GPUs.

The OpenACC execution model is like that of CUDA; a fundamental program runs on the CPU and starts errands (computational pieces or information moves) on the GPU. The principle program handles synchronization, either through unequivocal client control or certainly. OpenACC embraces the natural feeble memory model utilized in most GPU programming models: The GPU and CPU memory spaces are particular. OpenACC is an API that gives a lot of array orders, runtime libraries, and condition factors that can be used to create equal projects in Fortran, C, and C + to run on quickening agents, including GPUs. Engineers can begin composing their calculations consecutively and introduce directives OpenACC in the algorithm. It resembles giving indications for the compiler to turn the code parallel.

3.2 OpenMP Programming Model

The OpenMP standard was created, and it is kept up by the gathering OpenMP architecture review board shaped from some big organizations, for example, Intel, SGI, SUN Microsystems, IBM, and others, that toward the finish of 1997, assembled power to make a conventional equal programming for shared memory models. The OpenMP API and spotlights on a lot of orders that underpins the making of equal projects with shared memory through the execution of a programmed and improved arrangement of strings. Its highlights would now be able to be utilized in dialects FORTRAN 77, FORTRAN 90, C, and C ++. The benefits of utilizing OpenMP can be shown on straightforwardness and little change in the codes, the powerful help for equal programming, simplicity of comprehension, and utilization of mandates, one help settled parallelism, and the chance of dynamic alteration of the quantity of strings utilized.

3.3 Difference Between OpenACC and OpenMP

The OpenMP program will contain pragma omp directives. When the compiler encounters a pragma directive, it will start executing the program parallel. It will divide the tasks among multiple cores of the CPU, and the program will start executing all the tasks on different CPUs simultaneously. And hence, the parallel programming of a given application can be achieved.

OpenACC will contain pragma acc directives. When the compiler encounters a pragma directive, it will start executing the program parallelly. It will divide the program tasks among multiple cores of GPUs. GPU will contain thousands of cores which can be used to execute the program faster, and time taken for executing a program will be less. And, all GPU cores will be used to execute the program simultaneously to achieve parallel programming.

4 Proposed Methodology

The implementation part describes the parallel implementation of benchmark application by considering available GPU cores efficiently using OpenACC and available CPU cores using OpenMP. Performance of OpenACC programs running on GPU is compared with that of OpenMP programs running on CPU. The runtime of each core is evaluated on the test dataset. The benchmark HPC applications that are used to do the comparison are matrix multiplication, fast Fourier transforms, Laplace transforms, and molecular dynamics.

Parallelizing process may incorporate a few or all of the following:

- Analyzing segments of the code that can be performed simultaneously.
- Outlining the simultaneous bits of code onto multiple processes running in parallel.
- Distributing the input, output, and intermediate data information related with the program.
- Overseeing access to information shared by various processors.
- Integrating the processors at different phases of the parallel program execution.

4.1 *Parallel Implementation of Matrix Multiplication:*

In math, framework augmentation is a parallel operation that takes a couple of grids and produces another network. Numbers, for example, the genuine or complex numbers can be duplicated by number juggling. Then again, networks are varieties of numbers, so there is no one of a kind approach to characterize “the” increase of grids. Accordingly, when all is said in done, the expression “lattice duplication” alludes to various diverse approaches to increase lattices. The key components of any network duplication include: the quantity of columns and segments the first grids have (called the “size”, “request,” or “measurement”) and determining how the passages of the networks create the new grid. Like vectors, networks of any size can be duplicated by scalars, which add up to reproducing each section of the grid by the same number. Like the entry-wise meaning of including or subtracting lattices, augmentation of two grids of the same size can be characterized by reproducing the relating passages, and this is known as the Hadamard item. Another definition is the Kronecker result of two frameworks, to acquire a square network. One can shape numerous different definitions. On the other hand, the most helpful definition can be inspired by straight mathematical statements and direct changes on vectors, which have various applications in connected arithmetic, material science, and designing. This definition is regularly called the lattice product [2, 3]. In words, if A_n is a $n \times m$ grid and B is a $m \times p$ network, their framework item AB is a $n \times p$ framework, in which the m passages

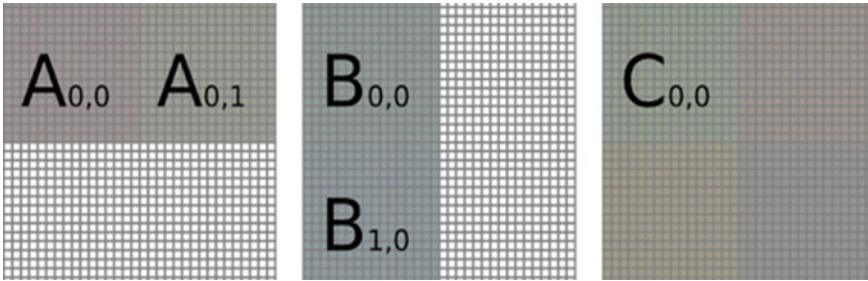


Fig. 1 Shows the parallelization of matrix multiplication

over the columns of A_n are reproduced with the m sections down the segments of B (the exact definition is underneath).

Algorithm

1. Input: matrices $A[m,p]$ and $B[p,n]$
2. Let $C[m,n]$ be a new matrix
3. for i from 1 to m :
4. for j from 1 to p :
5. Let $sum = 0$
6. For k from 1 to n :
7. Set $sum \leftarrow sum + A[i,k] \times B[k,j]$
8. Set $C_{ij} \leftarrow sum$
9. Return C

A local execution does out one string to process one component of grid C . Each string loads one line of network A_n and one segment of framework B from worldwide memory, does the internal item, and stores the outcome back to lattice C in the worldwide memory. To build the “calculation-to-memory proportion,” the tiled grid duplication can be applied. One string square processes one tile of lattice C (Fig. 1).

4.2 Parallel Implementation of Fast Fourier Transforms:

A FFT figures the DFT and creates precisely the identical result as assessing the DFT definition straightforwardly; practically, essential contrast is that a FFT is much quicker. Let x_0, \dots, x_{N-1} be complex numbers. The DFT is defined by the formula.

Evaluating this definition particularly obliges $O(N^2)$ operations: There are N yields X_k , and each yield obliges a total of N terms. A FFT is any framework to enroll the same results in $O(N \log N)$ operations. More conclusively, all known FFT

computations oblige ($N \log N$) operations, but there is no known confirmation that a lower disperse quality score is impossible.

To layout the venture trusts of a FFT, consider the count of complex growths and increments. Surveying the DFT's aggregates particularly incorporates N^2 complex increments and $N(N-1)$ complex additions. The undoubtedly comprehended radix 2 Cooley-Turkey count, for N a power of 2, can enlist the same result with just $(N/2) \log_2(N)$ psyche boggling duplications and $N \log_2(N)$ complex additions. For all intents and purposes, genuine execution on front line PCs is by and large controlled by components other than the rate of calculating operations, and the examination is a confounded subject, yet the general change from $O(N^2)$ to $O(N \log N)$ remains.

Algorithm

1. For $k=0$ to l // where $l=\log(N)/\log(2)$
2. Divide the set into intervals
3. Get the corresponding twiddle factors, W
4. in each interval
5. for each pair of points $J, J+\text{Half-Size}$
6. Analyze using butterfly representation

The portion of work that can be parallelized in this problem is finding out twiddle factors and FFT computation. Mapping of concurrent portions of work can be achieved by using the directive pragma omp for which splits parallel iteration spaces across threads in OpenMP and by using the pragma acc kernels directive in OpenACC to execute the portion of code on GPU. The data samples and the twiddle factors, in form of a complex number, are shared (global) across all the threads. Intermediate results generated during processing are stored in variables that are private (local) to the threads.

Synchronization is needed while moving back and forth in recursive task division, which is achieved through barrier directive in OpenMP and barrier directive in OpenACC.

4.3 Parallel Implementation of Laplace Transforms:

The Laplace change is a by and large used fundamental change as a piece of number juggling and electrical structure named after Pierre-Simon Laplace that changes a portion of time into a segment of complex rehash. The retrogressive Laplace change takes a flighty recurrent region breaking point and yields a cutoff depicted in the time space. The Laplace change is related to the Fourier change, yet while the Fourier change imparts a limit or banner as a superposition of sinusoids, the Laplace change conveys a limit, even more generally, as a superposition of minutes. Given a direct logical or utilitarian delineation of an information or respect a system, the Laplace

change gives an alternative functional depiction that consistently revamps the technique of examining the lead of the structure or in mixing another system taking into account a course of action of determinations. Along these lines, for example, Laplace change from the time locale to the recurrent space changes differential connections into arithmetical numerical enunciations and convolution into steady.

Algorithm

1. Set particle positions.
2. Assign particle velocities.
3. Repeat
 - Calculate force on each particle.
 - Update particle positions and velocity.
 - Measure properties ,Store results.
 - until the preset time steps
5. Analyze properties, print results

The portion of work that can be parallelized in this problem is taking the size of matrix and calculating the time taken for each matrix. Mapping of concurrent portions of work can be achieved by using the directive `pragma omp` for which splits parallel iteration spaces across threads in OpenMP and by using the `pragma acc kernels` directive in OpenACC to execute the portion of code on GPU. The input data, viz. taking the size of matrix and taking their values, these values are shared (global) across all the threads. Intermediate results generated during processing are stored in variables that are private (local) to the threads. Synchronization is needed while calculating the time taken for each matrix. These calculations can be done within the `pragma omp master` directive in OpenMP and `pragma acc master` directive in OpenACC.

4.4 Parallel Implementation of Molecular Dynamics

Molecular dynamics (MD) is a PC reenactment of physical improvements of molecules and particles in the association of N-body diversion. The particles are allowed to interface for a period of time, giving a point of view of the development of the atoms. In the most generally perceived adjustment, the direction of particles and iotas are constrained by numerically handling Newton's examinations of development for a game plan of imparting particles, where qualities between the particles and potential imperativeness are described by interatomic conceivable outcomes or subatomic mechanics power fields. The strategy was at first envisioned inside speculative material science in the late 1950s yet is associated today generally in compound material science, materials science, and the showing of biomolecules.

Since atomic frameworks comprise countless, it is hard to find the properties of such complex systems legitimately; MD reenactment circumvents this issue by

using numerical methodologies. In any case, long MD propagations are deductively seriously adjusted, creating consolidated mix-ups in numerical joining that can be limited with genuine selection of computations and boundaries, yet not cleared out totally.

For structures which agree to the ergodic theory, the advancement of a singular nuclear stream reenactment might be used to concentrate doubtlessly noticeable thermodynamic properties of the system: the time midpoints of an ergodic structure contrast with micro-canonical gathering midpoints. MD has moreover been named “quantifiable mechanics by numbers” and “Laplace’s vision of Newtonian mechanics” of foreseeing the future by vivifying nature’s powers and allowing understanding into sub-nuclear development on an atomic scale.

Algorithm

1. Set particle positions.
2. Assign particle velocities.
3. repeat
 1. Calculate force on each particle.
 2. Update particle positions and velocity.
 3. Measure properties, Store results.
4. until the preset time steps
5. Analyze properties, print results

The portion of work that can be parallelized in this problem is calculating the force on each particle. Mapping of concurrent portions of work can be achieved by using the directive pragma omp for which splits parallel iteration spaces across threads in OpenMP and by using the pragma acc kernels directive in OpenACC to execute the portion of code on GPU. The input data, viz. number of particles, their positions, and initial velocities, are shared (global) across all the threads. Intermediate results generated during processing are stored in variables that are private (local) to the threads. Synchronization is needed while measuring properties such as total kinetic energy and total potential energy. These calculations can be done within the pragma omp master directive in OpenMP and pragma acc master directive in OpenACC.

5 Experimental Setup and Results

This section is aimed to give a brief description of the experimental setup that is required for this project work to obtain the required results. First, the experimental setup is based on the system specifications to meet the requirements. Different applications are used to determine the performance of CPU and GPU and conclude with results and screenshots. Once the framework necessities are settled, then we need to figure out if a specific programming bundle fits framework prerequisites or not. The

Table 1 Tesla GPU configuration

	CPU configuration	GPU configuration
System type	HP Pro 3330 NT PC	PowerEdge R270
Processor	Intel Core i3-322Q CPU@3.30GHZx4	Intel xeon® CPU E5-26,200 @2.00GHZ\1S
RAM	2 GB RAM	32 GB RAM
Operating system type	64-bit	6 4-bit
Hard disk	500 GB	500GBx3
Graphics card	NA	Nvidia Tesla M2075 dual slot graphics card

Table 2 Quadro GPU configuration

	CPU configuration	GPU configuration
System type	HP Pro 3330 NT PC	DELL Precision R5500
Processor	Intel Core i3-3220 CPU @3.30GHZx4	Intel xeon® CPU E5620 2.40 GHz
RAM	2 GB RAM	32 GB RAM
Operating system type	64-bit	64-bit
Hard disk	500 GB	500GBx3
Graphics card	NA	Nvidia Qua dip K2000 dual slot graphics card

software bundle includes Nvidia drivers, PGI compiler, CUDA toolkit (Version 6.5), and Fedora operating system.

The following is the configuration of the system which is used in our experiment setup to run the applications on CPU and GPU (Tables 1 and 2).

PGI compiler is the compiler required to compile and run the OpenACC and OpenMP programs. The following graphs show the comparison of OpenMP and OpenACC on CPUs and GPUs for scientific applications on Tesla and Quadro graphics cards, in which x-axis shows the time in seconds and y-axis shows the number of particles in which blue color shows the results of OpenMP and red color shows the results of OpenACC. Performance is measured by varying the problem size of the benchmark applications.

Fig 2 shows the comparison OpenMP and OpenACC for matrix multiplication. The results on Tesla graphics card shows better performance because it contains more number of CPU cores (24) and more number of GPU cores (448) when compared to Quadro graphics card which contains less CPU cores (8) and less GPU cores (240). As the size of the matrix increases, OpenACC implementation will perform better (Fig. 3).

The figure shows the comparison OpenMP and OpenACC for FFT. On both the devices, we are able to observe large gap in the performance between OpenACC and OpenMP implementation as the problem size increases.

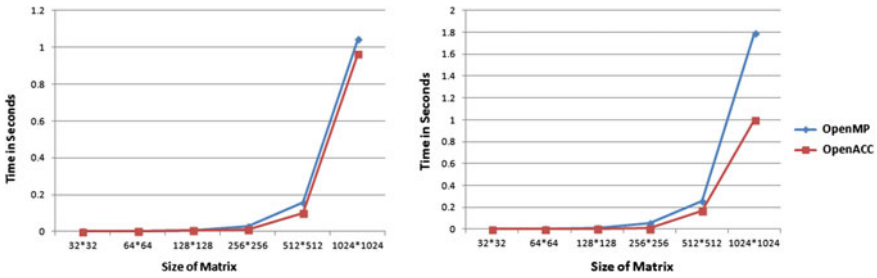


Fig. 2 Matrix multiplication performance on Tesla and Quadro graphics cards

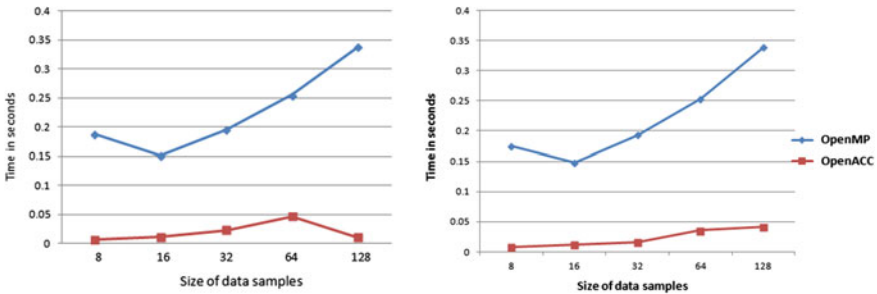


Fig. 3 Fast Fourier transform (FFT) performance on Tesla and Quadro graphics cards

Fig. 4 shows the comparison OpenMP and OpenACC implementation for Laplace transform. The results demonstrate performance improvement with OpenACC implementation as the size of the data samples increases.

(240) (Fig. 5).

The figure shows the comparison OpenMP and OpenACC for molecular dynamics. The results on Tesla graphics card shows better performance because it contains more number of CPU cores (24) and more number of GPU cores (448) when compared to Quadro graphics card which contains less CPU cores (8) and less

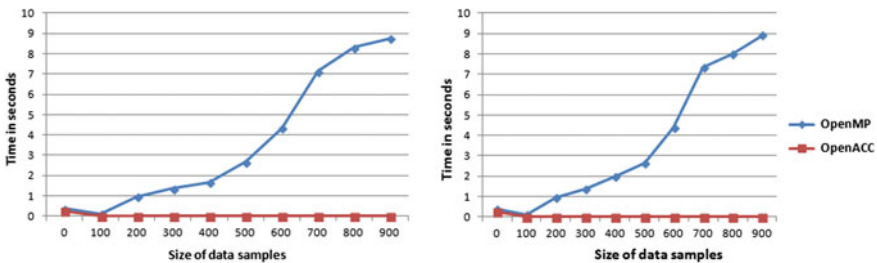


Fig. 4 Laplace transform performance on Tesla and Quadro graphics cards

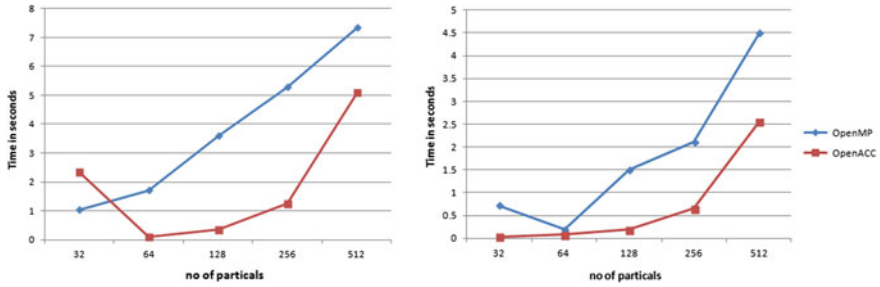


Fig. 5 Molecular dynamics performance on Tesla and Quadro graphics cards

GPU cores (240). Also, results demonstrate improvement in the performance as the number of particles increases.

6 Conclusion

Parallel programming is increasing in the near future, not only in massive computing software, but also in systems of small and medium businesses to generate more speed and providing the programmer more options to exploit the hardware resources. In this project, the performance of parallel program is better than the performance of serial program. When performance measurement was done, it was observed that time taken by OpenACC is better for smaller size data as well as when the size of data increases, when compared with OpenMP. Hence, from our work, we can conclude that OpenACC is a good option for data parallel task on GPUs. OpenMP is less complex and can give nearly equivalent performance to OpenACC on CPUs. Developers who need to utilize the parallelism must change their programming paradigms to address the issues that emerge as speed and better execution programming applications so as to build the limit computations and handling conceivable.

References

1. Li X, Shih PC (2018) An early performance comparison of CUDA and openACC. MATEC Web Conf 208:05002. <https://doi.org/10.1051/mateconf/201820805002>
2. Gayatri R, Yang C, Kurth T, Deslippe J (2018) A case study for performance portability using openmp 4.5. In: WACCPD@SC
3. Ledur CL, Zeve CM, dos Anjos JC (2013) Comparative analysis of openACC, openMP and CUDA using sequential and parallel algorithms
4. Memeti S, Li L, Pillana S, Kolodziej J, Kessler C (2017) Benchmarking openCL, openACC, openMP, and CUDA: programming productivity, performance, and energy consumption. In: Proceedings of the 2017 workshop on adaptive resource management and scheduling for cloud computing. ARMS-CC '17, Association for Computing Machinery, New York, NY, USA, pp 1–6. <https://doi.org/10.1145/3110355.3110356>

5. Wang Y, Qin Q, SEE SCW, Lin J (2013) Performance portability evaluation for openACC on intel knights corner and nvidia kepler
6. Lee S, Vetter JS (2013) Early evaluation of directive-based GPU programming models for productive exascale computing. In: SC '12: proceedings of the international conference on high performance computing, networking, storage and analysis, pp 1–11

Performance Evaluation of HPC Application in Containerized and Virtualized Environment



G. Manoj Kumar, Rohit Danti, Odso Amit, R. Guru Raghavendra, B. R. Kiran , and H. A. Sanjay 

1 Introduction

Cloud computing is a distributed computing model that leverages computing resources and services over the Internet with pay as you go model. Cloud computing offers three basic services such as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). Platform as a service (PaaS) is one of the prominent cloud computing services that provides a platform allowing customers to develop, run, and manage applications. Platform as a service enables customers to change their requirements based on use case with large choices of computational and storage resource that are accessible anytime. Application performance and optimized resource utilization have the focus on virtualizing the hardware resources. Hypervisor-based virtualization involves the creation of virtual machines over the physical machines each of which has its own operating system and the underlying hardware divided virtually. The other method of virtualization is the container-based virtualization in which the container bundles the application together with libraries and other dependencies, providing isolated environments for running the software services.

Container virtualization uses native OS kernel to run multiple instances and allow bundling of set of applications with libraries for providing isolated environment which makes deployment easy and consistent regardless of the target environment. Containers are an abstraction that packages code and dependencies of an application together. Multiple containers can run on the same physical machine or virtual machine and share the same OS kernel with other containers, each of which is running

G. Manoj Kumar (✉) · R. Danti (✉) · O. Amit (✉) · R. Guru Raghavendra (✉) · B. R. Kiran · H. A. Sanjay

Department of Information Science, Nitte Meenakshi Institute of Technology, Bangalore, India

B. R. Kiran

e-mail: kiran.br@nmit.ac.in

as an isolated process in the user space. Containers consume less space than VMs and can handle more applications and require fewer VMs and operating systems. The advantage of containerizing over virtualizing the hardware stack allows running multiple containers on top of the OS kernel directly since they are lightweight, share the same OS kernel, start much faster, and use a fraction of the memory compared to booting an entire OS [1]. The different containerization tools available are Docker, OpenVZ, Mesos Containerizer, CoreOS rkt, and LXC Linux Containers [1]. Docker is a popular, open-source container format that is supported on Google Cloud Platform and by Google. Kubernetes Engine [2]. Docker gives the flexibility for developers to easily pack, ship, and run application as a lightweight, portable, self-sufficient container, which runs virtually, and along with this, Docker also provides security to applications running in a shared environment.

Kubernetes is an open-source container-orchestration system for automating computer application deployment, scaling, and management. The Kubernetes cluster consists of pod, node, master node, and cluster. A pod holds one or more container. Nodes are the hardware components which are likely to be a virtual machine hosted by a cloud provider or a physical machine in a data center. Master nodes control the deployment of pods and the worker nodes. A cluster contains multiple worker nodes and at least one master node.

Metric server tool is a tool which gathers data about utilized assets (memory and CPU) of nodes and pods which is scalable and efficient. The Kubernetes metric server has an in-built autoscaling feature. The Kubernetes metrics server is an aggregator of resource usage data in the cluster. The tlp-stress is used as a workload generator. Tlp-stress is a workload-centric tool for stress testing and benchmarking Cassandra application. Metrics server collects resource metrics from Kubelets and exposes them in Kubernetes API server through metrics API for use by horizontal pod and vertical pod. Metrics server can be used for CPU/memory-based horizontal autoscaling and automatically adjusting/suggesting resources needed by containers.

This paper focuses on providing a detailed performance comparison running an HPC application on a VMware cluster of virtual machines and on a Docker cluster. As a HPC application, Cassandra an open-source NoSQL distributed database widely is selected since it is adopted by companies that use Docker and VMware technologies for processing big data applications. Our experimental results show that the container-based virtual environment had low overhead compared to the VMware virtualized environment.

The presented work is organized as follows: In Sect. 2, we discussed Related Work, Section 3 describes Hypervisor-Based Virtualization, Section 4 describes Container-Based Virtualization, Section 5 describes Performance Metrics and Workload, Section 6 presents Experimental Setup and Results, and we conclude our results in Sect. 7.

2 Related Work

Most of the existing efforts are on either container or hypervisor-based solutions. Couple of efforts exists on performance comparison of container and hypervisor-based solutions. In Singh and Singh [3], author focuses on Docker container-based virtualization which uses native OS kernel to run multiple instances and allow bundling of set of applications with libraries for providing isolated environment. In this type of virtualization, the kernel of OS runs on the hardware node with different isolated guest virtual machines (VMs) called containers. The hypervisor-based virtualization allows sharing of a single hardware host among the multiple virtual machine instances and a hypervisor. Docker is a container technology that offers platform as a service (PaaS) with the operating system level virtualization that makes it easy to bundle and distribute software along with its dependencies. In this paper, the author has discussed the importance of adopting container-based virtualization and Docker in cloud computing. In future, majority of microservice applications will be built on the containers. The author failed to address the security issues which are present in the containers.

Shirinbab et al. [4] evaluate the performance of a HPC application (Cassandra) in a container and virtual machine environment. Cassandra is NoSQL, scalable, and distributed databases that are popularly used in big data platforms that typically run on a virtualized infrastructure built using virtual machine or containers. The author has also highlighted that container-based virtualization provides isolation and less resource overhead whereas the security issues and I/O performance are a concern. In this paper, it has been highlighted that containers virtualize the operating system which eliminates the burden of resource usage monitoring and providing isolation among the applications. The authors also compare the performance between VMware and Docker container, with Apache Cassandra as a workload. The experimental study shows that Docker lowers the overhead in running the workload compared to the VMware. The workload generator selected in the paper does not provide flexibility to select the different workloads and uses simple key value table. Hence, the results of performance computation will not be effective.

Younge et al. [5] focus on the use of containers with advance supercomputing and HPC system software to analyze the overhead and the scalability features of deploying container-based cloud environment. HPCG and IMB benchmarks are used to investigate the overhead and scalability features of a container on a Cray XC30 test bed system. The paper also presents performance evaluation of deploying the containers with the Docker on Amazon's EC2 and Cray XC30 system. The investigation was done to understand the potential advantage and drawbacks of the container technologies, within an HPC environment which provide a model for DevOps. The comparison highlights a possible path for advanced testing and high-performance production which runs by transforming the environment of HPC applications from Amazon EC2 to a Cray supercomputer resulting in a new methodology that maximizes both DevOps efficiency and Cray supercomputer utilization.

Choi et al. [6], propose the framework which highlights the importance of resource management in the container-based cloud environment. The paper presents the general-purpose scheduling framework (GPSF) which provides the resource management functions, and a feature to customize the scheduling policy in a local environment using which the cloud providers and researchers can optimize the resources in the cloud environment. The paper also focuses on implementing the GPSF prototype and evaluates it by applying customized scheduling, and the experimental results show that the proposed framework has great potential for optimizing resources. The proposed work failed to address the challenges related to security in containers.

Madhumathi [7], in this paper, focuses around the incorporation of Docker compartments alongside an instrument for observing called “Sysdig” provides security to the Docker container using which the Docker compartments are sent safely. The creator features that Docker is an open-source containerization enablement stage which encourages the facilitating and execution of numerous sort of programming applications, stages, and any exclusively constructed programming. Holders have risen as the most route for light-footed programming advancement, bundling, delivering, and organization across various IT situations. The quicker provisioning, the constant versatility, and fast transmission of utilization holders are the principle key differentiators of the containerization worldview. The DevOps idea gets forties with the quicker development and dependability of the Docker stage. As a result, IT can dispatch quicker and run a similar application, unaltered, on any gadget. This combination improves and smooths out the checking of compartments, estimation, and the board. Docker holders wrap any exceptionally manufactured, bundled programming in a document framework which has all the fundamental devices for running the wrapped programming. This sort of bundling empowers the product to run wherever with no challenges. At the end of the day, this component ensures that the product will consistently run the equivalent, notwithstanding hidden working situations.

Madhumathi [7], in this paper, mainly focuses on highlighting the differences between Docker containers and hypervisors. Both containers and hypervisor technologies support isolation and multitenancy. The applications built on containers share the underlying operating system resulting in the smaller size deployments compared to the deployments made on the hypervisor; thus, a large number of containers can be stored on a single physical host. Docker is a popular, open-source container format and provides a systematic way to automate the faster deployment of applications inside the portable containers. Containers can also run on VMs if a cloud has the native container. In the case of hypervisor-based VMs, the entire application, container, and OS stack are placed on a VM and run the application.

3 Hypervisor-Based Virtualization

Hypervisor-based virtualization involves the creation of virtual machines over the physical machines each of which has its own operating system and the underlying hardware divided virtually. Since each virtual machine has its own OS and a virtual

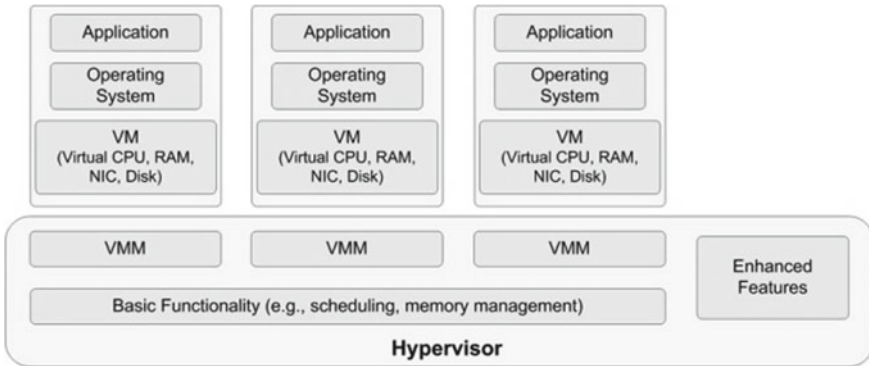


Fig. 1 Hypervisor-based virtualization

copy of all the hardware resources, VM requires significant RAM and CPU resources which cause the overhead (Fig. 1).

4 Container-Based Virtualization

Container virtualization uses native OS kernel to run multiple instances and allow bundling of set of applications with libraries for providing isolated environment which makes deployment easy and consistent regardless of the target environment which makes deployment easy and consistent. Multiple containers can run on the same physical or virtual machine and share the same OS kernel with other containers, each running as isolated processes which eliminates the need of dedicating an entire server to a single application which leads to less overhead (Fig. 2).

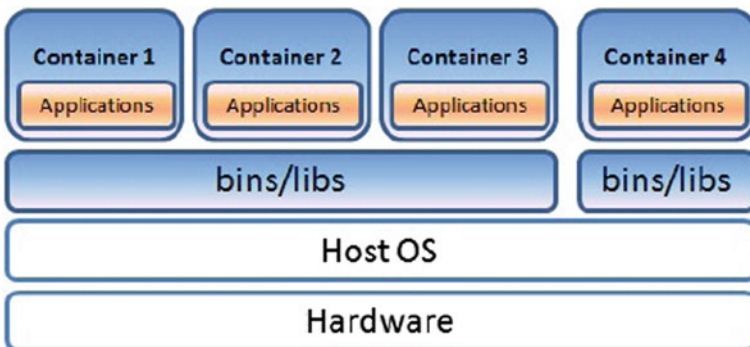


Fig. 2 Container-based virtualization

5 Performance Metrics and Workload

Cassandra tlp-stress tool [8] is an emulator for generating realistic workloads which can be specified via command line parameters. Tlp-stress tool has list of predesigned workloads such as Time Series Key/Value, Materialized Views, Collections (maps), and Counters which can be opted by the users directly as per their requirements. The tool allows the user to design workloads which allows to create thousands of database tables. Drop wizard metrics library will allow the user to monitor the workload statistics.

The performance metrics used to measure the performance of Docker container and VMware are:

- CPU utilization,
- Memory utilization,
- Latency.

CPU utilization is reported as the average usage, in CPU cores, over a period of time. The value is derived by taking a rate over a cumulative CPU counter provided by the kernel. Memory utilization is reported as the amount of memory (in bytes) in-use.

Metric server tool is a tool which gathers data about utilized assets (memory and CPU) of nodes and pods which is scalable and efficient. The Kubernetes metric server has an in-built autoscaling feature. The Kubernetes metrics server is an aggregator of resource usage data in the cluster. Metric server provides an API that is used to retrieve current resource usage. Metric server is scalable and has inbuilt autoscaling property. Metrics server offers single deployment that works on most clusters, scalable support up to 5,000 node clusters, and resource efficiency. Metrics server consumes 0.5 m core of CPU and 4 MB of memory per node (Fig. 3).

The command used to check the metric server-API service is

\$ kubectl get apiservices | egrep metrics.

The command used to check the metric server deployment and service details.

\$ kubectl get deploy, svc -n kube-system | egrep metrics-server.

The command used to retrieve the metrics of the pods:

\$ kubectl top pod.

The command used to retrieve the metrics of the nodes:

\$ kubectl top node.

6 Experimental Setup and Results

Objective of the experiment was to compare the performance of HPC application on VMware virtual machine and Docker container. The HPC application used is Cassandra—an open-source NoSQL distributed database which is widely adopted by companies that uses Docker and VMware for production and in big data applications.

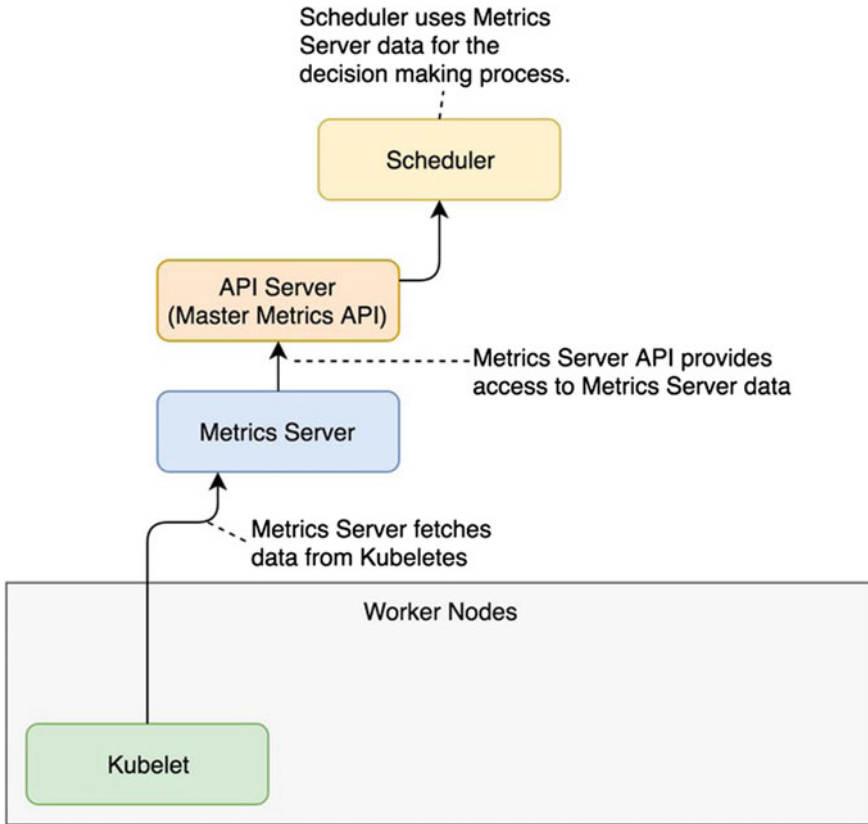


Fig. 3 Basic flow of data to and from metric server

All our experiments were performed on two DELL servers with a total of 16 core processors, 24 GB of RAM, and disk size of 1TB. Cassandra 3.0 was used as an identical workload in both the test cases—Virtualized and Dockerized. In total, 3 Cassandra nodes were configured in cluster with default settings.

Test cases

1. **Cassandra Virtualized**—In this case, one virtual machine is created on each host resulting in three virtual machines and all the three virtual machines are connected to form a Cassandra cluster.
2. **Cassandra Dockerized**—In this case, one container is created on each node resulting in three containers. In each container, we run Cassandra application and all the three nodes are connected to form a Cassandra cluster. In each test case the Cassandra is experimented with the different workload scenarios of read and write.

In this section, we present the results of the experiments. From Figs. 4 and 5, we can conclude that the write and read latency in Docker container is low compared to VMware since the hypervisor-based virtualization involves CPU and memory overhead. With the increase in number of write and read operations, container-based virtualization is able to perform better. In terms of CPU and memory utilization, the Docker shows the lowest CPU and memory utilization compared to the VMware and the same is shown in Figs. 6 and 7. With the increase in workload, the utilization of CPU and memory increases significantly in the VMware environment leading to performance and operational bottleneck which can be clearly visualized in Table 1.

Fig. 4 Comparison of write latency in Docker container and VMware

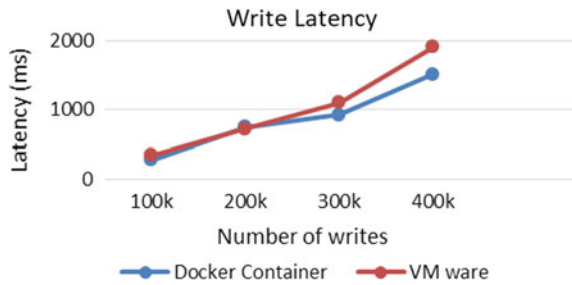


Fig. 5 Comparison of read latency in Docker container and VMware

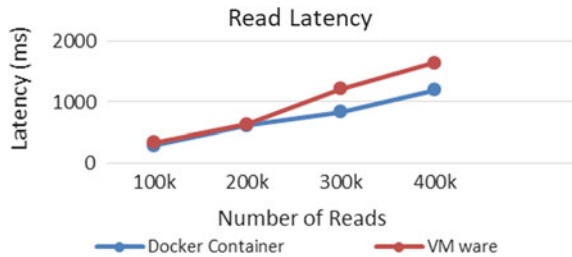


Fig. 6 Comparison of CPU utilization in Docker container and VMware

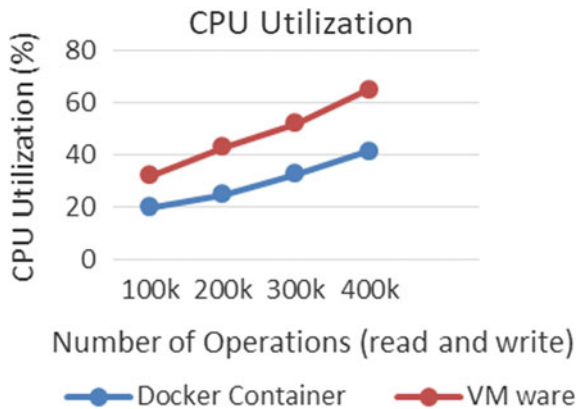
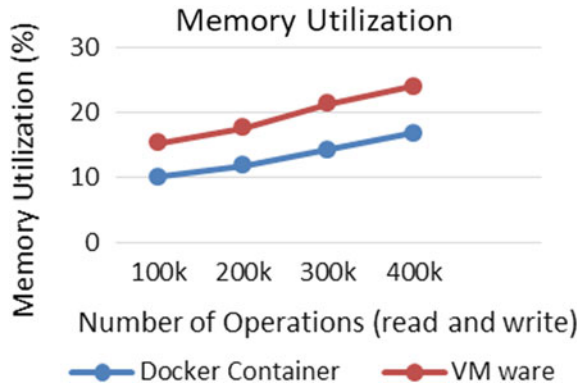


Fig. 7 Comparison of memory utilization in Docker container and VMware



7 Conclusion

In this paper, we try to evaluate the performance of HPC application (Cassandra) on containerized and virtualized environments. The overall results showed that the virtualized environment has a significant resource and operational overhead compared to containerized environment which causes a fall in the performance of the application. From the experimental results summary, it is evident that Cassandra Dockerized seems to address the challenges of virtualization by packing the application and their dependencies into lightweight containers. As a result of this, Cassandra Dockerized consumed fewer resources, and the operational overhead is low which results in overall better performance. Our comparison also justifies the selection of tlp-stress as workload generator because of following features. 1. tlp-stress uses time series data model. 2. Tlp-stress provides extensive list of workloads. 3. Tlp-stress performs one million requests with a 9:1 write-read thread ratio. As a result of this, the experimental results obtained in our paper are more accurate compared to the existing work [4].

Table 1 Summary of the performance results

Workload (k)	Operations	Docker container				VM ware			
		CPU utilization (%)	Memory utilization (%)	Request per second	Latency (ms)	CPU utilization (%)	Memory utilization (%)	Request per second	Latency (ms)
100	Read	20.1	10	215	286	32	15.2	202	340
	Write			217	280			205	340
200	Read	24.8	11.8	265	610	43	17.6	252	630
	Write			258	745			248	720
300	Read	32.6	14.2	325	830	52	21.3	305	1217
	Write			325	920			305	1100
400	Read	41.4	16.8	277	1196	65	24	246	1643
	Write			275	1512			235	1903

References

1. Google, “t1p-stress”. Webpage: <https://thelastpickle.com/t1p-stress/>.
2. Google, “Solve more with Google Cloud”. Webpage: <https://cloud.google.com/>
3. Singh S, Singh N (2016) Containers & docker: emerging roles & future of cloud technology. In: 2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT). IEEE, pp 804–807
4. Shirinbab S, Lundberg L, Casalicchio E (2017) Performance evaluation of container and virtual machine running cassandra workload. In: 2017 3rd International conference of cloud computing technologies and applications (CloudTech). IEEE, pp 1–8
5. Younge AJ, Pedretti K, Grant RE, Brightwell R (2017) A tale of two systems: using containers to deploy hpc applications on supercomputers and clouds. In: 2017 IEEE international conference on cloud computing technology and science (CloudCom). IEEE, pp 74–81
6. Choi S, Myung R, Choi H, Chung K, Gil J, Yu H (2016) Gpsf: general-purpose scheduling framework for container based on cloud environment. In: 2016 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData). IEEE, pp 769–772
7. Madhumathi R (2018) The relevance of container monitoring towards container intelligence. In: 2018 9th international conference on computing, communication and networking technologies (ICCCNT). IEEE, pp 1–5
8. Hemmert KS, Glass MW, Hammond SD, Hoekstra R, Rajan M, Dawson S, Vigil M, Grunau D, Lujan J, Morton D et al. Trinity: architecture and early experience, in Cray Users
9. Bernstein D (2014) Containers and cloud: From lxc to docker to kubernetes. IEEE Cloud Comput 1(3):81–84
10. Merkel D (2014) Docker: lightweight linux containers for consistent development and deployment. Linux J 2014(239):2
11. Google. Google container engine. Webpage: <https://cloud.google.com/container-engine>
12. Messina P (2017) Update on the Exascale Computing Project (ECP). HPC User Forum
13. Shirinbab S, Lundberg L, Casalicchio E (2020) Performance evaluation of containers and virtual machines when running Cassandra workload concurrently. J Concurr Comput: Pract Experience :Pract Experience

Design Improvement of Blade-Less Wind Generator



D. Tejas and Shefali Jagwani 

1 Introduction

The wind turbine has evolved from the time that they were just used to lift water, beat husk or show direction. Power generation capacities have reached scales of megawatts at present day technology [1]. Although the peak power production of commercial onshore and offshore wind turbine generator is in megawatt scales the cost of manufacturing, transportation, installation and commissioning is rather cumbersome and financially straining. Also the mechanism of converting kinetic energy from gusts of wind to rotation of shaft through partial abstraction and to electrical power involves expensive gear mechanism and also labor intensive maintenance. As the research of better structural components are being discovered materials with ability to withstand flexure without symptom of deformation are being employed to manufacture newer designs of wind turbine generator. One such design is the vortex blade-less wind generator (BWG) [2]. This design is inspired from the evolution of a golf ball, which was initially spherical, smooth surfaced. During the evolution of golf ball, it was established that the new ball would not fly the distance that a used ball would. Further study led us to know that considerably entities of mass and volume where flying through air, a region of slow velocity air would cling to the surface of the ball and result in causing friction would slow the ball down. This region of air sticking to the surface of the flying object was established as boundary layer [3]. This presence of boundary layer affects the motion of objects moving through air in various other ways like swaying it off course and creating excessive wear on the surface of the objects. In order to reduce the effects of the boundary layer, studies have been

D. Tejas (✉) · S. Jagwani

Department of Electrical and Electronics Engineering, Nitte Meenakshi Institute of Technology, Bengaluru 560064, India

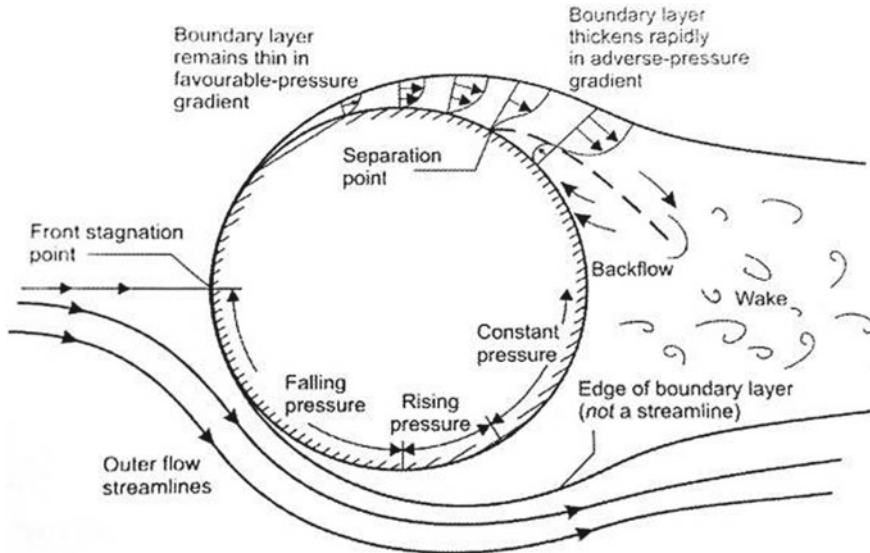


Fig. 1 Presence of boundary layer about a spherical object when moving through air [3]

conducted around the world. In this design of BWG with design inputs, the limitations being caused by this phenomenon will be tried to be overcome and better operational life cycle will be aimed to achieve (Fig. 1).

As the blade-less generator is comparatively a new concept, there are very few papers available in the literature. Though many blogs and articles are present in the literature but the number of technical papers are less in this domain. In this paper, we have studied the work done in this field and proposed the improvement in design of BWG.

A company based in Spain, named **vortex blade-less** [2], a tech start-up, founded in 2012, is working on an environmental-friendly aero-generator. It is a new type of wind energy generation technology for on site generation in residential areas, on-grid/off-grid installation and in conjunction with solar PV systems. They are developing a system which works on aero-elastic resonance which was inspired after observing the Tacoma Narrows bridge oscillating due to cross wind. Its innovation helps it generate energy from its unusual shape by oscillating at a fixed place. Its mast structure is a composite of fiber glass and carbon fiber. It takes the advantage of Von Karman's emission to vibrate in a flowing fluid (air) because of its slender structure. In the lower portion of this design, a carbon fiber rod connects the mast with the alternator, and it generates electricity without any moving parts.

Paper [4] explains the model of the blade-less wind turbine where they have created a design and got linearly varying voltage vs wind speeds. This was tested at increments of 1 to 4 V at every 2 m/s wind speeds till 12 ms^{-1} . Authors in paper [5] convey that the purposed design will reduce the finances involved in production operation and repair as the model minimized number of parts and the need for periodic

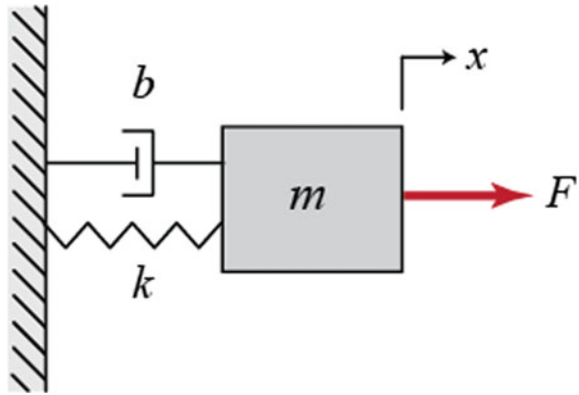
lubrication and changing of working components will not arise. It also mentions that there model is environmentally friendlier to the bird and animal species. Articles [6, 7] have explained the motion of golf ball with dimples through air. Aerodynamic optimization is achieved through dimple pattern design which is a critical part of overall golf ball development. They mention that the indentation on the golf ball create a thin boundary layer of turbulent air that clings to the surface. This reduces the wake around the golf ball thus allowing it to fly farther after striking. Thus, this concept is adopted for our work which includes golf ball along with dimples for the overall design improvement in BWG.

2 Blade-Less Wind Generator

The wind turbine simply works inverse of a ceiling fan. When there is wind blowing in the atmosphere, it possess kinetic energy. This energy is imparted to the wind caused by changing weather patterns. The blades of the wind turbine face the wind in normally with a nominal pitch angle. As the wind exerts pressure on the blades, due to the pitch angle, it get converted into rotational motion through the hub, gear, and generator mechanism. It requires complex operational automation to control the pitch angle, tip speed ratio, stall control and braking. Adding to the complexity, the wind mill consists of aerodynamically shaped blades which have to be precisely manufactured and transported to the installation site. The site for installation also has to be sizeable enough to allow the complete sweeping of the blades. All these aspects make the conventional wind turbine a thing of past. The BWG basically works on the principle of vortex shedding under the influence of force exerted by the wind [8]. Upon placing the BWG in the location of power requirement the breeze is partially obstructed in the area of installation. The motion of wind after striking the mast structure of the BWG exerts force in normal direction to the incident surface. The incident force will not be uniform throughout the surface of the mast. The shape of the mast is conically tapered form a larger diameter at the top to a reduced diameter in the lower portion. Due to this exertion of force, the mast sways in a periodic manner creating alternate regions of high and low-pressure zones around it. This action is referred to as vortex shedding [9]. The effect of vortex shedding experienced in the upper portion of the mast is more than the lower portion of it. The swaying motion is then transmitted to the piezo-electric generator through the stabilizer rod [4]. The entire structure of the BGW is firmly anchored to the ground. The structure of the BWG after attaining natural frequency continues to oscillate till effects of resonance takes over. Here, the frequency f is related to air velocity V , characteristic length L and strouhl number St as mentioned in Eq. 1 [2]. The inverse of length L is the sum of mean diameter D of mast and the product of amplitude a and adjustment factor X dependent on Reynolds Number as mentioned in Eq. 2 [2].

$$f = St.V.L \quad (1)$$

Fig. 2 Spring mass damper system [10]



$$L^{-1} = D + a.X \tag{2}$$

This design of BWG overcomes the limitations of the conventional wind turbine being simple in operation, reduced space requirement, minimal maintenance, and less complex. The flora and fauna of the geographic location are also minimally affected. Bird strikes and disturbances of migration routes will be minimal (Fig. 2).

3 Components of a BWG

The components of the BWG consist of a mast, stabilizer rod, electricity generator, and an heavy base structure.

3.1 Mast

It is an inverted cone shaped, solid filled structure with a endoskeleton support. Its function is to gather the kinetic energy of the wind and convert in to usable form of motion which will be further transmitted to derive work and generate power. The materials used for this component must possess physical and chemical properties which will be able endure the extreme environmental conditions in any geographic location on the earth. It should also effortlessly bond with other materials to form a seamless cohesion. The resonant velocity of mast V_r will be equal to product of natural frequency N of body and crosswind width R by the strouhl number St which is product of frequency f and mean diameter D by fluid stream velocity U [11].

$$V_r = N.R/St \tag{3}$$

$$St = f.D/U \quad (4)$$

3.2 Stabilizer

The stabilizer rod is an critical component of the entire design of BGW. The complete operational life of the BWG is depended on this particular component. This component should possess high-fatigue resistance property which is of at most importance. Its work is to convert oscillating motion into usable form and transmit it to electricity generator. The fatigue strength S_f of the stabilizer is equal to product of a and number of cycles to failure N to the power b where a and b are 10^3 and 10^6 , respectively, for low cycle operations [12].

$$S_f = a.N^b \quad (5)$$

It should also possess properties like corrosion inhibition, malleability, and cohesive property with other materials to form a rigid structure.

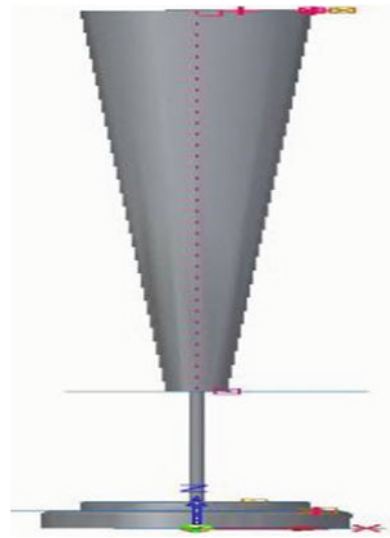
3.3 Electricity Generator

This component can be employed using two principles of operation. One is by using piezo-electric materials. They are materials which can produce electric potential voltage when subjected to mechanical stresses. The variation of stress with voltage is mostly proportional to each other. Contrary to this, the inverse is also true that is if electricity is induced on them, and they undergo a mechanical deformation. Second method is to utilize the use of neodymium magnets and copper coils through a complete electrical circuit. The relative movement of coils in the vicinity of magnets generates a current in the copper coils. Thus, employing this principle, electricity is generated (Fig. 3).

3.4 Base

The base of the entire structure also has multiple functions. Dampening the unwanted vibrations, providing a solid support and an anchor point, counterbalancing the weight of the entire structure are some examples. Further study can lead to a installation of a thrust bearing and introduction of single-axis degree of rotation.

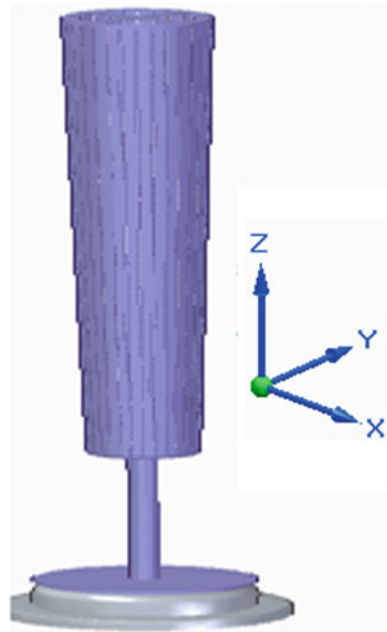
Fig. 3 CAD Model of BWG of mentioned dimensions without perforations designed with Solid EdgeV20 [13]



4 Proposed Design

The design improvement in this segment of study is implementing the perforations. The idea of introducing perforations has been borrowed from the design of golf ball. The gauge of the perforation varies widely on the purpose of use. Golf ball, while flying through air with the help of these perforations disturbs the boundary layer around it which causes the low-pressure region trailing behind it. The low-pressure region tends to pull the ball in direction opposite to direction of motion. By avoiding the low-pressure region, the golf ball moves farther distance upon the strike from the golfer. This is being implemented to the BWG for it to attain damped resonance and to attain max power conditions even at low-wind speeds. Resonance is the phenomenon when frequency of external excitation is equal to the natural frequency of the body. At resonance, the amplitude of oscillation is excessively large. Example of an uncontrolled resonance is the collapse of Tacoma Narrows Bridge in Tacoma, Kitsap Peninsula, USA. Here, the wind accelerated the amplitude of natural frequency to attain a state of resonance. Using the similar principle to this study, when the wind strikes the surface of the BWG where vortex shedding formation is disturbed. Thus, the rate of increase in amplitude for lower cut-in speed can be achieved to attain max power conditions that is resonant conditions more effectively. This can be observed specifically in the virtual wind tunnel simulation test performed using Autodesk Flow Design Software [14]. It clearly visible that the colored eddy formation is more prominent in the CAD model without perforations and less prominent in models with perforations at a rated wind speed. Here in the case of BWG, the maximum power generation condition damping should be tuned to achieve a state of resonant oscillation.

Fig. 4 CAD model of BWG of mentioned dimensions of proposed design with perforations designed using solid edgeV20 [13]



5 Dimensions of CAD Model

The dimensions for the CAD model were adopted from the technical paper mentioned in the reference as [9]. The followings are the dimensions for the CAD model (Fig. 4).

- Taper Ratio, $R_t = 9.34: 1 = \frac{\text{Length of mast}}{D_{\text{max}} + D_{\text{min}}}$.
- Length of the mast = 2.148 m.
- Diameter of mast at the top = 0.19 m (Max).
- Diameter of mast at bottom = 0.04 m (Min).
- Length of stabilizer = 0.776 m.
- Diameter of base = 0.2 m.
- Size of indentation on mast = 0.01 m.

6 Results and Discussions

The results obtained after and before implementing the design is portrayed in the Figs. 5 and 6, respectively, of virtual wind tunnel simulation. It is clearly visible that the swirling action of air is taking place on the trailing side of BWG without the perforations. These swirling of air know as eddies must be avoided in order for quick rate of achieving maximum power generation condition. It is also visible in Fig. 5 that the BWG whose mast has calculated perforations, and the eddy formation

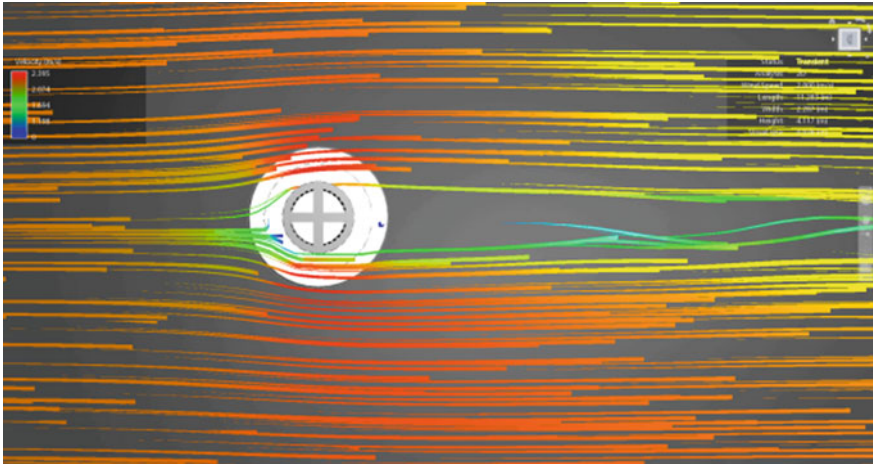


Fig. 5 BWG with perforations on mast [14]

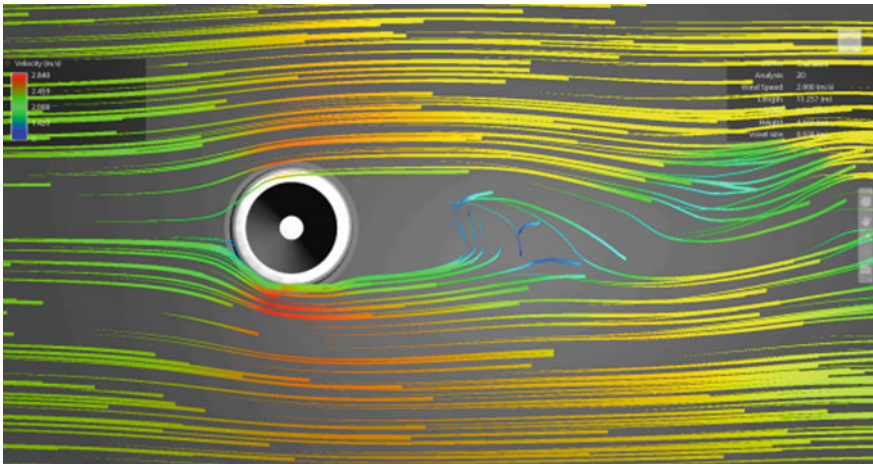


Fig. 6 BWG without perforations on the mast [14]

is drastically reduced. So the improvement of design is evident when observed from figures and arriving at a conclusive understanding about it.

7 Conclusions

The improvement of the present design, for the BWG, is just one dimension of the infinite possibility this design opens to the world. The implementation of this

possible at all circumstances where there is relative motion of air around a entity of mass. With designs like these, renewable energy being a far fetched for many can become reality. The socioeconomic being of many humans habitations around the world would improve and see development. Most of all, the need for petroleum derivatives would reduce simultaneously reducing the carbon footprint to produce the energy. The problems created globally as a result of using them would also reduce to nurture a holistic support to flora and fauna.

References

1. Power Technology. <https://www.power-technology.com>
2. Villarreal DJY, VIV resonant wind generators. Vortex Blade-less S.L. <https://vortexbladeless.com>
3. Technitel. <https://techintel.co.za/blog/boundary-layer-separation-from-a-circular-cylinder/>
4. Dhanusha PB, Shyamraj R, Raj A (2018) Harnessing wind energy using bladeless windmills. In: International conference on circuits and systems in digital enterprise technology (ICCSDET). Kottayam, India, pp 1–6. <https://doi.org/10.1109/ICCSDET.2018.8821085>
5. Goryachev SV, Kharchenko PA (2019) Use of bladeless generator in wind power. In: International conference on industrial engineering, applications and manufacturing (ICIEAM). Sochi, Russia, pp 1–6. <https://doi.org/10.1109/ICIEAM.2019.8742784>
6. Business Insider. <https://www.businessinsider.in>
7. Scientific American. <https://www.scientificamerican.com/article/how-do-dimples-in-golf-ba/#>
8. Akshay C, Mahendra C, Rushewshar C, Dhamone SP (2019) Blade-less windmill power generation. *Int Res J Eng Technol* 6(1):1775–1778
9. Kshirsagar OD, Gaikwad AB (2019) Design and analysis of vortex blade-less windmill for composite materials. *MAT J J Ind Mech* 4(2):15–24
10. Computing the damping value for two masses in a harmonic oscillator. <https://physics.stackexchange.com>
11. Das JBK, Srinivasmurthy PL (2015) Mechanical vibration. ISBN-9788128003714
12. Das JBK, Srinivasmurthy PL (2015) Design of machine elements 1. ISBN-9788128002373
13. Seimens. ® Solid Edge. <https://solidedge.siemens.com>
14. Autodesk. ® Flow Design. <https://www.autodesk.com>

Prediction of Leaf Blotch in Turmeric Plant



Manjula R. Chougala and A. C. Ramachandra

1 Introduction

In India, most of the population is indulged in agriculture for their livelihood directly or indirectly. Indian economy depends on agricultural products. Adopting the advanced techniques in agriculture aims to increase the productivity, quality and quantity with a less expenditure. Emerging trends are hitting almost all areas in agriculture. India is a leading producer, consumer and exporter of turmeric and turmeric products. Turmeric is known as 'Haridra' that symbolise its special bond with Hari or Lord Vishnu. It is believed that the clothes worn by the gods of Hindu mythology were dyed in turmeric. In India, it is customary to use turmeric on social and religious occasions like pooja and marriages. Depending on the size and colour and aroma, there are several varieties of turmeric. Most popular of them are Kasturi, Mundaga, Yalachaga and Balaga. It is used in conventional medicines like Ayurveda and Unani as it promotes the antioxidant capacity of the body and helps in boosting the immune system. Turmeric is a golden spice which is rich in antibacterial, antiviral, anti-tumour, antiseptic, antioxidant, hepato-protective, cardio-protective, radio-protective, nephron-protective and anti-inflammatory properties so widely used in medicines. The curcumin content of turmeric is proved to cure many deadly diseases including cancer. Cosmetic industries use turmeric as it heals and prevents the dry skin problems. Its anti-ageing property made it as the main ingredient in cosmetics. It gives the rich colour, aroma and taste to food thus used in culinary. Dyeing industries also use turmeric for its natural colour.

M. R. Chougala (✉)
VTU, Belagavi, India

A. C. Ramachandra
Department of Electronics and Communication Engineering, NMIT, Bangalore, India
e-mail: ramachandra.ac@nmit.co.in

Turmeric is a flowering plant that has long, oblong leaves with the average height of one metre. It is a ten-month crop which needs moderate rainfall of 1500 mm or irrigation and temperature range of 20–35 °C. Turmeric thrives best in clay loam and well-drained sandy soils. Turmeric leaves are used in selected dishes, and they are purported to improve digestion. Rhizomes, in the powdered form, are used in culinary for colour and its warm flavour. Various foliar diseases affect the turmeric plantation during cultivation that results in low quality and poor yield. The nutritional deficiency in soil and rhizome diseases affects the growth causing big loss to farmers. Different institutes of agriculture and spice research in India are analysing these diseases affecting to turmeric plants. We propose an algorithm for early detection of leaf blotch in turmeric, so that disease may be controlled in the initial stage.

Turmeric can be grown in diverse tropical conditions from sea level to 1500 m above sea level, at a temperature range of 20–35 °C with an annual rainfall of 1500 mm or more, under rain fed or irrigated conditions. Though it can be grown on different types of soils, it thrives best in well-drained sandy or clay loam soils with a pH range of 4.5–7.5 with good organic status.

1.1 Leaf Blotch (Taphrina Maculans)

Anthrachnose or leaf blotch is a fungal disease caused by *Taphrina Maculans*. This fungus is developed in soil, seeds and survives in soil on infected plant debris. Disease symptoms appear on the leaves during the month of October and November. The yellow spots of oval, rectangular or irregular in shape start developing on both the surface of leaf but the upper surface is infected more. These spots are gradually turn wider and to dirty yellow or dark brown in colour. In severe cases, plants appear to be scorched and rhizome yield is reduced. Sample images of turmeric leaves affected with Leaf Blotch are as shown in Fig. 1.

2 Literature Survey

Sumathi and Senthil Kumar developed a DIP technique for automatic detection and classification of plant leaves diseases. The digital images are acquired, and RGB components are separated from collected colour images [1]. If the pixel intensity values of the green component are less than the threshold value, then the red and blue components of the pixel values are assigned to zero because green colour pixels mostly represent the healthy areas of the leaf. The images are converted into a binary image and stored in the system and are given to the backpropagation neural network to detect the infection in the leaf. The review explained the different image processing techniques and algorithms that are used for identification of disease in different plant leaves, vegetables and fruits. Anand and Ashwin Patil developed segmentation-based image processing technique for analysing the plant leaf diseases automatically



Fig. 1 Sample images of leaf blotch infected leaves

and classify them. High-resolution JPEG format images are stored in the database and are converted into YCbCr, HSI and CIELAB colour space models [2]. Colour-transformed images are passed through median filtering technique for removing unnecessary spots, and CIELAB colour space method is adopted for accurately detecting the disease spots. Arivazhagan et al. used image analysis technology to identify the diseases from the symptoms that appear on the plant leaves. The RGB colour images of plant leaves are converted into HSI colour image [3]. The green pixels are masked and removed using specific threshold value followed by the segmentation process, and the texture statistics are computed. Support vector machine (SVM) is used as a classifier for the identification of disease infection in plant leaves. Fang and Huijie used artificial neural network (ANN) for detecting diseases early and accurately in pomegranate plants. The digital images of infected pomegranate leaves are collected, and Gabor filter techniques are used for pre-processing and segmentation [4]. The colour and texture features are extracted, and backpropagation neural network techniques are used to train the feature values to differentiate and categorise the healthy and disease-infected plants.

Manoj et al. developed a modern histogram-based technique to identify bacterial leaf blight diseases in paddy leaves. Digital images of healthy and infected paddy leaves are captured and stored [5]. The RGB colour images are converted into a greyscale image, and the histograms are plotted both healthy and infected greyscale paddy leaf images. Finally, the results are compared to recognise of the infected region in the paddy leaves. Sanjay and Nitin [6] developed a colour transform-based technique for recognition of agricultural plant leaf diseases RGB colour images are converted into HSI images. Texture parameters are calculated and compared for both healthy and infected leaves. The presence of diseases on the plant leaves is evaluated. Bindu and Toran developed a modern machine vision and fuzzy logic system to recognise position of bacterial blight pomegranate plant leaves. The colour image segmentation techniques and K-means segmentation algorithm are used to identify

the infected portion [7]. The positions of infection on leaves are calculated using fuzzy inference systems. The results were compared to manually recognise the position of the bacterial blight disease that infected pomegranate plant leaves. Basavaraj and Rajesh developed a classifier using a modern multi-layered backpropagation neural network, and it uses the combination of texture- and colour-based recognition techniques to identify uninfected and infected vegetables and fruits. The mean, median, standard deviation, hue, saturation and intensity values are extracted and trained to the neural network classifier to categorise healthy and infected images of fruits and vegetables [8]. Qinghai et al. analysed the colour and image enhancement technique to identify and analyse the diseases and pests of cotton leaves. The RGB colour images are converted into greyscale, HIS and YCbCr images [9]. Histogram is plotted for infected and uninfected cotton leaves images. Results are compared for extracting the disease-infected cotton leaves. Sanjay [10] studied the sugarcane leaves and developed an image segmentation technique to identify disease severity in them. He used triangle threshold segmentation techniques to calculate infected region of the sugarcane leaves. Sannakki et al. developed fuzzy logic systems and modern machine vision to recognise the position of the disease infected in plant leaves. K-means image segmentation algorithms are used to identify the disease-infected portion [11]. Fuzzy inference systems are used to calculate the infection in plant leaves. Ajit Danti and Suresha M proposed the methodology for segmentation and classification of areca nuts using the three-sigma control limits. They converted into RGB image into YCbCr colour space and extracted chromatic components of all RGB but suppressed red and blue components [12]. The Gaussian curve is plotted to depict the three-sigma control limits in classification of different variety of areca nuts. Pradnya et al. proposed the system for recognising the grape leaf disease by extracting the leaf features [13]. Different analytical and discriminating algorithms are used to classify the diseases. Spatial grey-level dependence matrices (SGDM method is used for analysing). Nandini et al. worked on identification and classification of leaf diseases in turmeric plant [14] the extraction of colour and texture features is done with grey-level co-occurrence matrix. Second-order statistical features such as energy, entropy, contrast, variance, homogeneity, correlation are derived. Shape features such as solidity, eccentricity, perimeter and centroid are defined. They used the machine learning algorithms support vector machine, decision tree and naïve Bayes are used in classification. Basavaraj et al. proposed an approach for identification and classification of Indian medicinal plants [14]. They extracted colour, texture and edge features to classify the plants as shrubs, herbs and trees.

Tarun kumar and Karun Verma have discussed the theory of converting the RGB image to grey image for feature extraction [15]. They proposed the method to split the R, G, B components using MATLAB and transformation equation to calculate the intensity of grey image, compared the results with Braum method. Anant Bhardwaj et al. discussed the computationally good method in recognising plant species using texture analysis and moment invariants [16]. They discussed various parametric equations recognised different leaf samples using grey level co-occurrence matrix method. Wasim khan et al., proposed the image retrieval system using the histogram and

texture descriptor analysis. The texture of an image is analysed by entropy, standard deviation and local range [17]. The classification is done by calculating the Euclidean distance of test image and the stored image properties in the database. K. R Gokulakrishnan and kapilya have reviewed various techniques in plant disease identification and issues in broadcasting [18]. Aerial videography technique is implemented, and they proposed the partial classification and real-time monitoring of diseased plants. Arti Rathod et al. have studied the different techniques in identifying the plant diseases by analysing the plant leaves. Morphological characteristics extraction, classification using the Otsu threshold and K-means clustering all edge detection algorithms are reviewed [19]. Different colour spaces and clustering techniques are also reviewed. Nitin S Tijare et al. surveyed the image recognition-based identification system for crop diseases [20]. They surveyed on extraction of colour features, shape features and texture features. This survey helps in estimation of severity of the diseases classification. S R Kodituwakku et al. proposed the system for automatic indexing and searching the images in database. They used content-based image retrieval system that represents the feature vector to determine the similarity measures [21]. Colour coherent vectors, histograms and colour moments are used to represent the visual features of an image. Savita N et al., reviewed the different image processing techniques to detect and classify the plant diseases [22]. Different classification techniques such as support vector machine, k-nearest neighbour, probabilistic neural networks, self-organisation graph and artificial neural networks are discussed.

3 Proposed Model

A statistical method to detect the leaf blotch disease of turmeric leaves is done considering the spatial relationship between the pixels. The proposed system works in five phases as shown in Fig. 2.

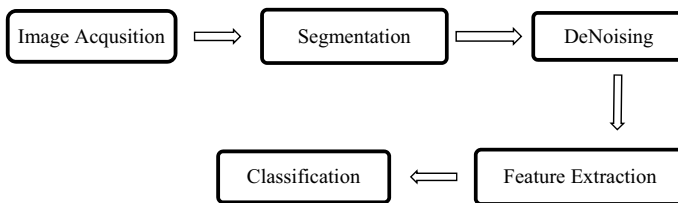


Fig. 2 Proposed Model for detecting Leaf Blotch in Turmeric plant

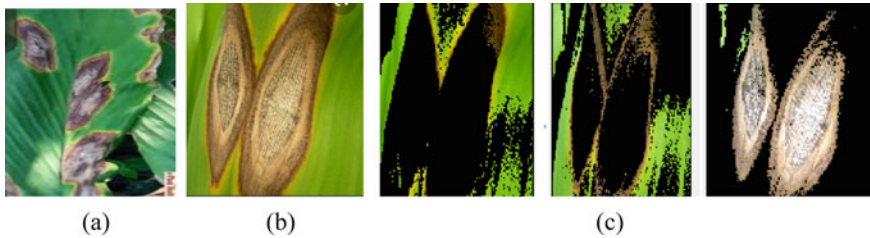


Fig. 3 Images of leaves affected with. **a** Leaf blotch. **b** RoI. **c** Segmented images

3.1 Image Acquisition and Pre-processing

In recent years, the rapid growth in technology, computer-aided processing has evolved in agriculture field. The high-quality images are captured using high-resolution digital camera, stored, and transferred easily and fast at a very low cost using smartphones. The images of healthy and infected leaves are selected from turmeric plantation during fourth month of cultivation. These leaves are washed thoroughly to remove dust particles so as to avoid noise in the digital image. The noise-free image helps in better diagnosis and to get better end results. The images are captured, and background is cleared thus perception of image is improved so as to provide reliable input to the system. For acquiring the related image, authors have visited the turmeric plantation of Sangli variety in Kolhapur District of Maharashtra, India.

3.2 Image Segmentation

This phase includes partitioning the image into regions and selecting the region of interest (RoI) for analysing. K-means segmentation is applied on selected image to get the clear representation of boundaries, lines and curves in the RoI. Segmentation process helps to assign label to each pixel of an image such that the pixels with similar label share some characteristic of the image. The filtering of the image gives sharp contrast and highlights the contour areas as shown in Fig. 3. Median filter is used for enhancing the image. The morphological operations are done on the image to categorize the leaves. The segmented images which are extracted for further analyzing as shown in Fig. 3.

3.3 De-noising

In this phase, the unwanted signals or the disturbances from the image are removed to get the clear image from the selected RoI image. The median filter is applied to remove

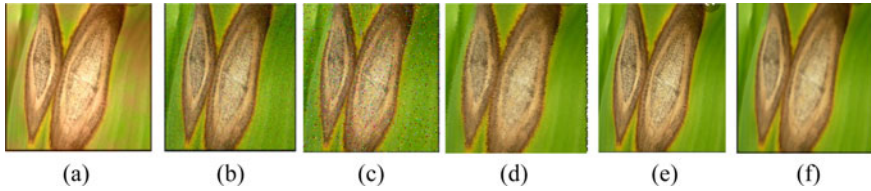


Fig. 4 Images of infected leaf. **a** RoI. **b–c** Images with noise. **d–e** Images after applying median filter. **f** Final image with high contrast

the noise and improve the appearance as shown in Fig. 4. Noise removal makes the image more sharp, contrast and highlights the contour areas. The morphological operations can be done accurately on the image after de-noising.

3.4 *Extracting the RGB Colour Components*

The images of healthy and infected leaves are pre-processed and segmented. The pictures of healthy and infected leaves are collected in standard format of 256×256 and stored. These stored images are trained, and RGB components are extracted. The extracted statistical values are tabulated in database. The colour moments are calculated and compared.

1. Healthy leaf phase
2. Infected leaf phase.

For identifying the leaf blotch disease of turmeric, the colour feature is focussed and the colour moments such as mean, median and the standard deviations are derived using Formulae (1), (2) and (3). The colour features are extracted for all RGB colours using the histogram. Healthy leaf images are collected, and corresponding RoI is extracted and is converted into greyscale and histogram equalised image. The histogram and equalised histogram are plotted to extract the colour components. The resultant images of healthy leaf phase are as shown in Fig. 5. The histogram of a digital image gives the numeric information about the distribution of the number of

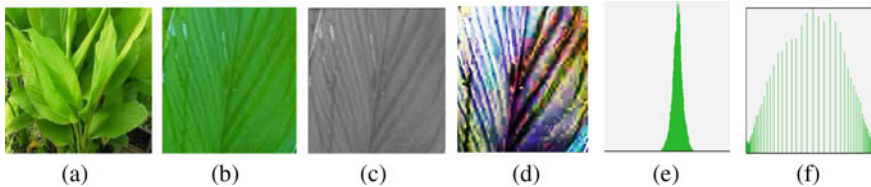


Fig. 5 Sample images. **a** Healthy leaf. **b** RoI. **c** Grey image. **d** Equalised image. **e** Histogram. **f** Histogram equalisation

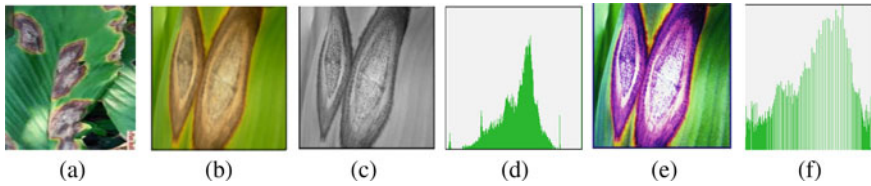


Fig. 6 Image quality enhancement using histogram equalisation: (a) original image; (b) enhanced ROI image of (a); (c) greyscale image; (d) histogram of the image in (c); (e) resulting image obtained from image (c) by histogram equalisation; (f) histogram of the image in (e)

pixels per grey-level value. The histogram of a digital image is a discrete component. Histograms provide the basis various spatial domain processing techniques like image enhancement, image compression and segmentation.

Histogram provides the distribution of univariate data graphically. The histogram of green component is as shown in Fig. 5d. The histogram equalised image is as shown in Fig. 5e, and corresponding histogram is as shown in Fig. 5f to get the clear contrast of the image. The most frequent intensity values in the image are stretched to get the uniform distribution and help in extracting the as the green is the major component in detecting the disease the mean, median and the standard deviation values are considered for disease identification [12].

$$\text{Mean of Green Component} = \mu_G = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N C_G(i, j) \tag{1}$$

$$\text{Median of Green Component} = \mu^{1/2} = \frac{n+1}{n} \text{th term} \tag{2}$$

$$\begin{aligned} &\text{Standard Deviation of Green Component} \\ &= \sigma_G \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (C_G(i, j) - \mu_G)^2} \end{aligned} \tag{3}$$

i and j are pixel coordinates of an image in row M and column N . μ_G is mean, $\mu^{1/2}$ is median and σ_G standard deviation, C_G is chromatic green component in YCbCr colour space (Fig. 6).

3.5 Classification

Leaf images are analysed, and the statistical values of RGB colour components are calculated in YCbCr colour space. Classification is done for infected leaves on the basis of statistical values of green component. The statistical values of healthy

leaves images are collected and stored in dataset. The comparison is made manually for similarity measures with the healthy leaf features vector and infected leaf features vector, and the distance is calculated. The distance of values predicts the infection severity of leaf blotch disease in turmeric plant.

4 Experimental Results

The Red–Green–Blue (RGB) components are extracted from the leaf images and tabulated in database. The red and blue components suppressed as green is the major colour in identifying the leaves. So the green colour moments are considered for further classification and compared. The Sangli Variety of turmeric plantation is visited in Kolhapur District of Maharashtra State, and 200 leaf images are collected and stored. 50 images are considered for training and remaining for testing. The images are resized to 256×256 standard size. The resizing and pixel resolution provides better computation speed and accuracy in analysing and detecting the disease.

The readings obtained from the experimental results are as shown in Table 1. It depicts that the values of mean, median and the standard deviation of green component are decreasing as the severity of the disease increases. These values are decreasing slightly in the healthy leaves as the green component gets deeper in aged leaves as compared to tender or new leaves. The rate of change of values is plotted to predict the disease and its severity. Mean, median and the standard deviation values against ten samples of infected leaves with the red line and the healthy leaves with the blue line are as shown in Figs. 7, 8 and 9, respectively.

Table 1 Samples experimental results of healthy (H) and infected (I) turmeric leaves

Samples	Extraction of green component in healthy leaves			Extraction of green component in infected leaves		
	Mean (H)	Median (H)	Std. Dev (H)	Mean (I)	Median (I)	Std. Dev (I)
1	146	152	19.04	99.65	110	15.84
2	112	120	16.73	101.83	109	16.79
3	118	127	19.56	123.58	130	13.09
4	115	131	20.49	99.68	119	15.83
5	120	129	13.77	100.34	124	15.11
6	104	113	21.21	99.12	115	14.56
7	144	149	36.86	98.70	117	13.56
8	155	162	21.76	99.56	123	14.52
9	117	121	20.55	97.12	121	15.11
10	118	130	17.67	100.29	118	13.75

Fig. 7 Mean for green component

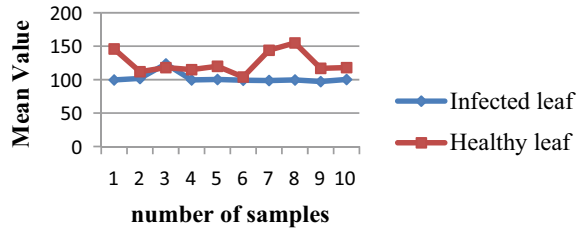


Fig. 8 Median for green component

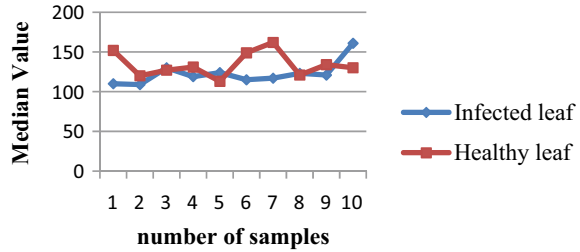
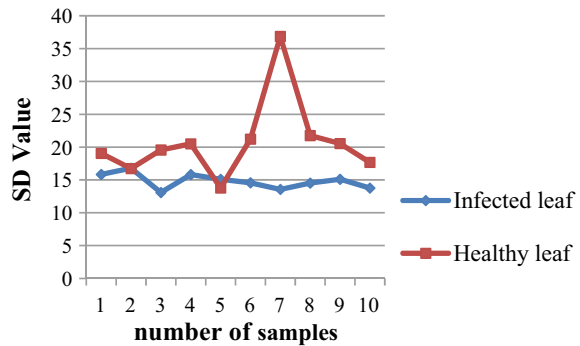


Fig. 9 Standard deviation for green component



The above experimental results show that in healthy turmeric leaves the average mean value of green component ranges from 100 to 150 as compared to 80–100 in case of infected leaves, median value of green component ranges from 120 to 170 as compared to 90–130 in case of infected leaves and standard deviation value of green component ranges from 15 to 30 as compared to 10–20 in case of infected leaves.

5 Conclusion

The proposed method conveyed that leaf blotch disease in turmeric plantation can be identified at the early stage by analysing the infected leaves. The results help the plant pathologists, so that the preventive measures can be suggested to farmers and avoid

spreading of this deadly fungal disease. It is fast, cost-effective and non-destructive method of detecting the plant disease using digital images. Due to the advances in mobile technology, the digital images are easily captured, stored and shared. With the growing trends in the area of image processing, various accurate methods can be adopted to detect plant diseases. Periodic inspection of fields is necessary by the farmers to get good yield and better quality of products. The proposed method can also be used to detect other plant diseases along with edge detection, texture analysis, etc.


References

1. Sumathi C, SenthilKumar AV (2014) Enhancing accuracy of plant leaf classification techniques. *Int J Eng Res Appl* 4(3):40–46
2. Anand H, Ashwin Patil RK (2012) Applying image processing Technique to detect plant diseases. *Int J Mod Eng Res* 2(5):3661–3664
3. Arivazhagan S, Newlin S, Ananthi S, VishnuVarthini S (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agric Eng Int CIGR J* 15(1):211–217
4. Fang H, Huijie L (2014) Plant leaves recognition and classification model based on image features and neural network. *IJCSI Int J Comput Sci* 11(1):100–104
5. Manoj M, Titan P, Debabrata S (2012) Damaged paddy leaf detection using image processing. *J Glob Res Comput Sci* 3(10):7–10
6. Sanjay B, Nitin P (2013) Agricultural plant leaf disease detection using image processing. *Int J Adv Res Electric Electron Instrum Eng* 1(2):599–602
7. Bindu T, Toran V (2013) Identification and classification of normal and infected apples using neural network. *Int J Sci Res* 2(6):160–163
8. Basvaraj S, Rajesh Y (2011) Identification and classification of normal and affected agriculture/horticulture produce based on combined colour and texture feature extraction. *Int J Comput Appl* 3(1):356360
9. Qinghai H, Benxue M, Zhang Q, Jing Z (2013) Cotton pests and diseases detection based on image processing. *Telkonnika Indonesian J Electr Eng* 11(6):3445–3450
10. Sannakki S, Vijay S, Arun Kumar R (2011) Leaf disease grading by machine vision and fuzzy logic. *Int J Comput Technol Appl* 2(5):1709–1716
11. Sannakki SS, Rajpurohit VS, Nargundand Pallavi Kulkarni VB (2013) Diagnosis and classification of grape leaf diseases using neural networks. In: International conference on computing, communications and networking technologies IEEE
12. Danti A, Suresha M (2012) Segmentation and classification of raw arecanuts based on three sigma control limits. Elsevier
13. Narvekar P, Kumbar M, Patil SN (2014) Grape leaf diseases detection & analysis using SGDM matrix method. *IJIRCCE* 2(3)
14. Basavaraj S, Suvarna A, Govardhan A (2010) A combined colour, texture and edge features based approach for identification and classification of Indian medicinal plants. *Int J Comput Appl* 6(12):45–51
15. Kumar T, Verma K (2010) A theory based on conversion of RGB image to gray image. *IJCA* 7:7–10
16. Bharwaj A, Kaur M, Kumar A (2013) Recognition of plants by leaf image using moment invariant and texture analysis. *IJIAS* 3:237–248
17. Khan W, Kumar S, Gupta N, Khan N (2011) A proposed method for image retrieval using histogram values and texture descriptor analysis. *Int J Soft Comput Eng (IJSCE)* 1(2). ISSN: 2231-2307

18. Gokulakrishnan KR (2014) Kapilya "Detecting the plant Diseases and Issues by Image Processing Technique and Broadcasting." IJSR 3(5):1016–1018
19. Rathod AN, Tanawal B, Shah V (2013) Image processing techniques for detection of leaf disease. IJARCSSE 3
20. Tijare N, Bednerkar SS (2014) Image recognition based crop disease identification System: a survey. IJCSMC 3(4):868–873
21. Kodituwakku SR, Selvarajah S, Comparison of color feates for Image Retrieval, IJCSE 1. ISSN: 0976-5166
22. Ghaiwat SN, Arora P, Detection and classification of plant diseases using image processing techniques: a review. ISSN: 2327-2812

Voice-Based Gender and Emotion Prediction Using Convolutional Neural Network



Arshiya Firdos, T. R. Amrutha, Chaithra, N. Tejaswini,
and K. M. Deepika 

1 Introduction

Speech emotion recognition can altogether increase the nature of spoken exchange frameworks. Although VER was read for a long time, machines despite everything experience issues in perceiving speaker's emotions. In numerous studies, gender contrasts are seen in emotional speech samples, recommending that sex data would get definite favorable circumstances in VER [1]. Methods for fusing sex data into VER can be outlined up as two techniques. One procedure is to form a distinctive feeling show for each sexual orientation, which is implied to as Sep-System for the distinctive show. The other procedure is to recognize sex data as a broadened component, which is intimated to as Aug-System.

In accordance with Sep-System, problem is modeled as gender identification followed by emotion. Gender data is not represented in this model. However, differences in the individual gender's utterances are utilized as attributes to prepare male and female feeling classifiers independently [2]. Other than, considering sex introduction freely extends the bumble of sexual orientation recognizable proof and it takes longer term in feeling acknowledgment. Concerning to Aug-System, emotion classifier is trained with all the utterances in training data, so that makes the difficulties in representing gender information. Traditional techniques for encoding sex utilize a novel number.

The paper is organized as follows. Section 2 contains related work. Section 3 describes proposed system. Section 4 defines audio feature extraction and analysis.

A. Firdos · T. R. Amrutha · Chaithra · N. Tejaswini
Bangalore, India

K. M. Deepika (✉)
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore, India
e-mail: deepika.km@nmit.ac.in

Section 5 describes system model; Sect. 6 describes results; and Sect. 7 concludes the paper.

2 Related Work

Souza and Souza [3] utilization of spectral characteristic mel-frequency cepstral coefficients (MFCC) is performed using ANN in which MFCC functions were used as speech parameters, and five unique emotional states have been considered for analysis. Pitch is the most primarily-based feature in the emotion of speech. Gupta et al. [4] voice frames are extracted from speech signal using time period strength (STE) and zero-crossing price (ZCR). The speech signal is transformed from time domain to frequency domain after which cepstral analysis is done to split the records. Cepstral evaluation technique converts data as linear combination in cepstral domain. Bisio et al. [5] providing a gender-driven emotion system is to individuate gender of speakers later classifies emotion characterizing speech indicators. Regarding emotion recognition method uses two classifiers which include two convolutional neural networks (CNN) the one used for male speakers and other for female speakers. Jadav [6] a system of both gender and emotion recognition is developed using CNN classifier which takes gender as an input for emotion recognition. To train CNN classifier, emotional database referred to as Berlin emotional speech database (BESD) is used.

3 Proposed System

We endorse a trendy convolutional neural community (CNN) framework for designing real-time CNNs. We validate our model through developing a real-time vision device which accomplishes the responsibilities of gender and emotion detection in a single mixed step with the usage of our proposed CNN system. We report accuracies of 96% within the gender dataset and 66% within emotion dataset. In conjunction with this, we also brought the very recent actual time enabled backpropagation visualization technique. Backpropagation uncovers the dynamics of the weight adjustments and evaluates learned capabilities. The careful implementation of contemporary CNN architectures, the use of modern regularization techniques, and the visualization of previously hidden features are vital to reduce distance among gradual performances and real-time architectures. Advantages: high accuracy than the existing approach and less time consumption.

4 Audio Feature Extraction and Analysis

4.1 Feature Selection Criterion

Feature extraction is the gathering of elements, called a part vector, from some other game plan of variables (e.g., a chose talk signal time game plan). Feature assurance is on a very basic level the change of those watched vectors to trademark vectors. The fundamental goal of feature assurance is to find a trademark vector to a hugely low-dimensional part region that holds the information suitable to the item while permitting basic connections to be worked on using basic extents of likeness [2].

Despite the way that it might be tempting toward the start to pick all the evacuated limits, the “scourge of dimensionality” go to the nearness amazingly brisk. As we add extra features to the vector, estimations of the features extend, which achieves an outrageous need of count and memory in both getting ready and testing. The enthusiasm for a great deal of planning records to address sound characteristics grows exponentially with the component of the limit space [7].

That grants us to achieve extraordinary precision for sound class, it is miles significant to pick significant features. The trademark extraction techniques can be arranged as common assessment and ghost examination approach. Transient evaluation uses the waveform of the sound sign itself for assessment. While apparition examination uses absurd depiction of the sound sign for appraisal. A specific plan of limits should be picked which would high abstractly set up the character and henceforth the class of the sound sign. These features pick the dimensionality of a specific records factor in the segment district. A selected feature should have the following properties:

Discriminative power: Whole inspiration driving picking a component is to obtain differentially among sound sign that has a spot with different classes. Thus, a trademark needs to take similar characteristics inside the indistinct class anyway specific characteristics across indisputable classes.

Uncorrelated to various limits: In solicitation to show up at higher count execution, cost and time smoothing out, each picked feature should be standout in information.

Invariance to unessential issues: A part needs to show security to immaterial issues including commotion, bandwidth or the ampleness scaling of the sign. One of these points of interest is fundamental in growing a solid model.

4.2 Frame-Level Features

The hidden presumption in most extreme discourse preparing plans is that the attributes of the discourse sign change exceptionally with time. This suspicion closes in a sort of fast time preparing techniques in which brief portions of the discourse markers are extricated and handled as though they had been brief sections from continued sounds with consistent properties. That is rehashed (occasionally) varying. As often as possible these short fragments which can be alluded to as

assessment outlines cover each other. The final product of preparing on each casing might be either in a solitary numeric worth, or a lot of numeric worth. Thusly, such preparing produces a spic and span time subordinate arrangement that may fill in as representation of the sound sign.

4.3 Multi-dimensional Frame-Level Features of Speech

The most noticeably used multi-dimensional component vectors in the region of talk assessment and mix, coding, and affirmation are spectrogram, linear predictive coefficients (LPC), line spectral pairs (LSP), and mel-frequency cepstral coefficients (MFCC)

4.4 Spectrogram

A standard development in trademark extraction of a talk is repeated (apparition) examination. Human talk can be really fixed over the evaluation time period 25 msec. Accordingly, the sign is inspected in dynamic thin timeframes of 20–25 MS width. The ghost evaluation of the talk signal is finished by strategy for finding discrete Fourier change (DFT) of the models inside the packaging. Brisk Fourier change figuring is used to process discrete Fourier change.

The time-changing spooky features of the talk sign can be graphically demonstrated by methods for using sound spectrograph. This makes a two-dimensional model considered a spectrogram in which the vertical rotate depicts repeat and level estimation to time. As the obscurity of the model forms, essentialness of the sign moreover augments. Thus, the resonation repeat of the vocal tract suggests up as darkish gatherings in the spectrogram [2]. Voiced districts are portrayed by strategy for striated look considering the periodicity of the waveform while unvoiced stretches are solid diminish.

4.5 Mel-Frequency Cepstrum Coefficients (MFCC)

The inspiration for the use of MFCC is on the grounds that the sound-related response to the human ear can settle frequencies non-directly. MFCC has been widely used in discourse handling programs. MFCC has the advantage that it has the ability of catching the basic qualities of discourse [2]. A mel is a unit of measure essentially based on the human ear's apparent recurrence. MFCCs depend on the perceived variety of the human ear's pivotal data transmissions with recurrence; channels separated straight at low frequencies and logarithmically at high frequencies had been utilized to extricate the crucial qualities of discourse. This is communicated inside

the mel-recurrence scale, which is direct recurrence separating under 1000Hz and logarithmic recurrence dividing more than 1000Hz.

5 System Model

5.1 *Extraction of Audio Features*

To describe the capacity of the approached technique in recognizing feeling and sexual orientation, to begin with it is divided into three stages, the limits were characterized by the ringer interferences inborn to the assignment. To framework the sound sign inside each stage, we figured the accompanying highlights for every voice epoch: voice age length, zero-crossing rate (ZCR) of the sound sign, mel emphasize cepstral coefficients(MFCC), winning emphasize, hard-hearted emphasize, perceptual shocking centroid(PSC), dazzling levelness, incline and kurtosis of the control unearthly density(PSD), ZCR of the Z-score of PSD (ZCR zPSD) for all ages, the basic, the collaborator, and the tertiary formants, and the rate of sign essentialness over 200, 500, 700, 1000, and 2000 HZ. We in extension cleared the callous, center, standard deviation, most fundamental, and scarcest ZCR zPSD from sliding windows inside parts time and rehash spaces insides each age, and these estimations were dealt with for each component insides each orchestrate [7] (Fig. 1).

5.2 *Gender-Driven Feature for VER*

The motivations to join the gender-driven component to spectrographic information. The fundamental step is checking sexual presentation data aid move forward the accuracy of VER [2]. The gender-driven highlight encodes sex information best with variable qualities. As the gender-driven component is expelled through acoustic highlights, it shows disdain toward of everything that incorporates a few acoustic data (for case F0) that is comparing to spectrogram. In highlight status and combination organizes, the CNN extricates 32-dimensional gender-driven highlights from 384-dimensional acoustic highlights. At that point, the spectrogram and gender-driven component are set as compositional element (F).

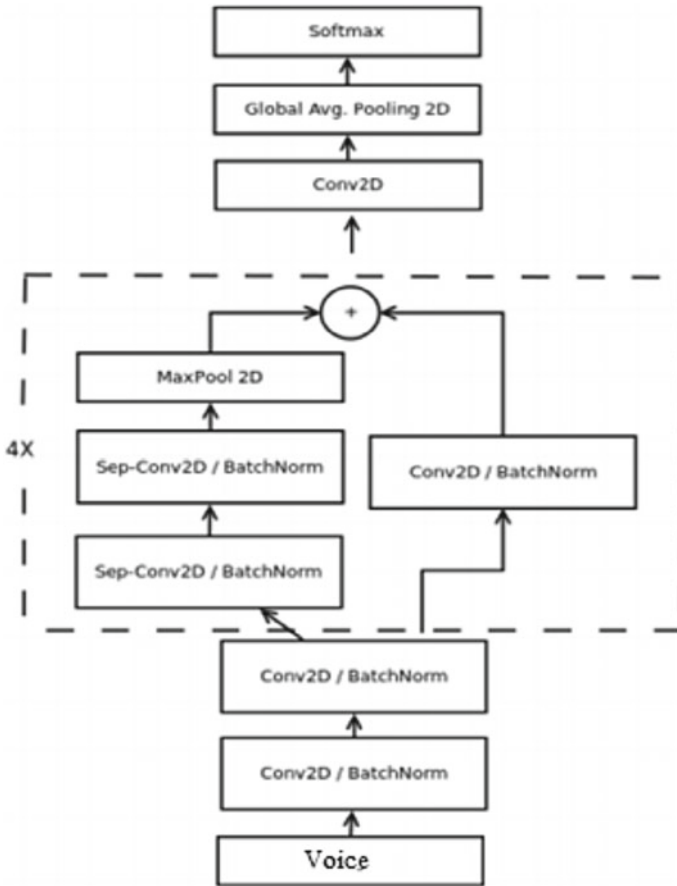


Fig. 1 Architecture diagram

6 Results and Analysis

6.1 Prediction of Gender and Emotion

Prediction of sexual orientation and feeling is done at same time, when compared to the past conspire the overall exactness taken for concurrent sexual orientation and feeling in our framework has been increased.

Figure 2 shows variation of accuracy with epoch which is the number of passes of the entire training dataset.

Figure 3 shows variation of loss with epoch for the training dataset along with validation.

Figure 4 shows an interface which helps user interact with machine and helps the user to upload necessary files required by the model to predict the desired output.

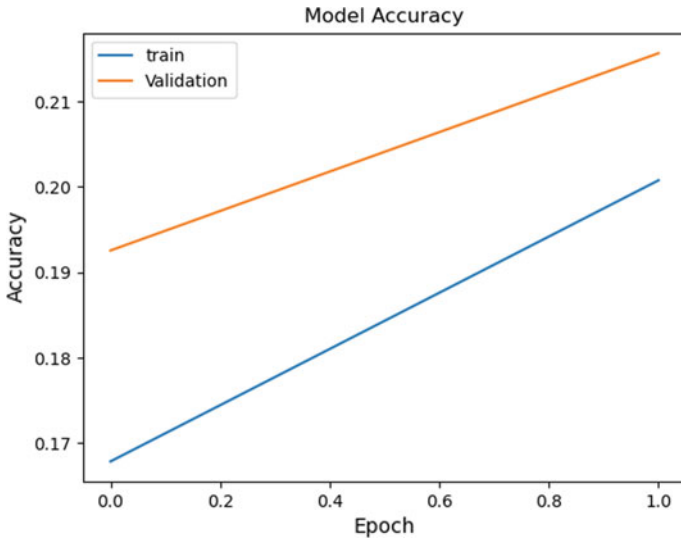


Fig. 2 CNN accuracy

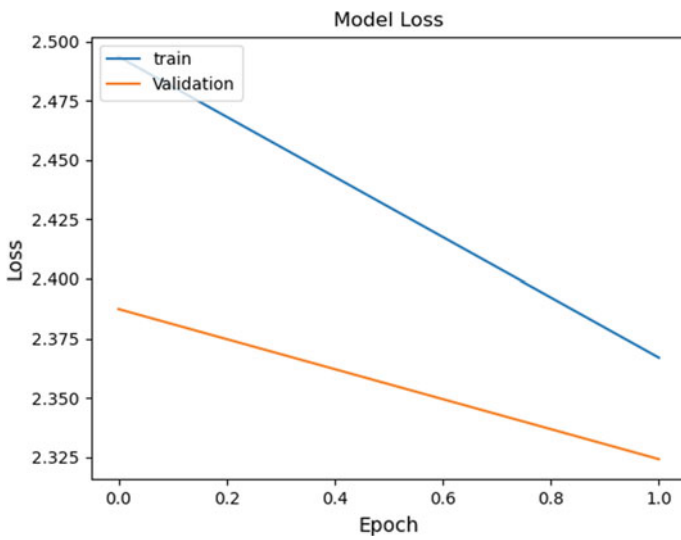


Fig. 3 CNN loss

Figure 5 shows waves whose characteristics show the frequency distribution for gender and emotion prediction.

Figure 6 shows a graph detecting the emotion and gender of the test data passed by user.

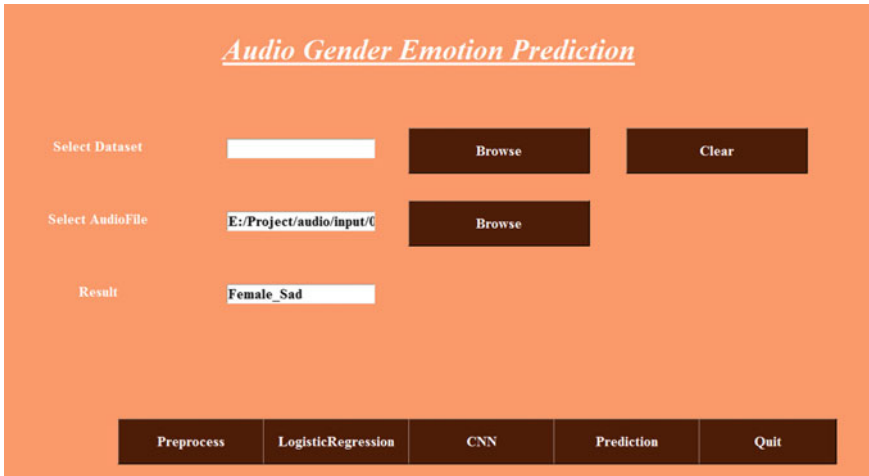


Fig. 4 Prediction (application result)

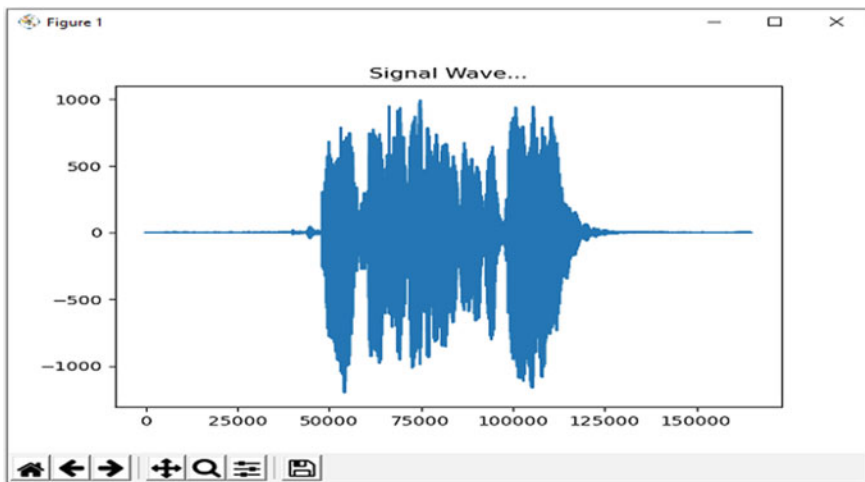


Fig. 5 Signal wave

7 Conclusion

The gender-aware CNN is proposed for discourse feeling acknowledgment. We originally proposed the circulated sex highlight and sexual orientation-driven component. So, the two novel highlights with sexual orientation data were exclusively increased into spectrogram as extra factors. At last, the CNN is utilized to coordinate the final classification. The results of appraisals illustrated that our proposed highlights can misuse sexual introduction information palatably and perform best on VER tasks.

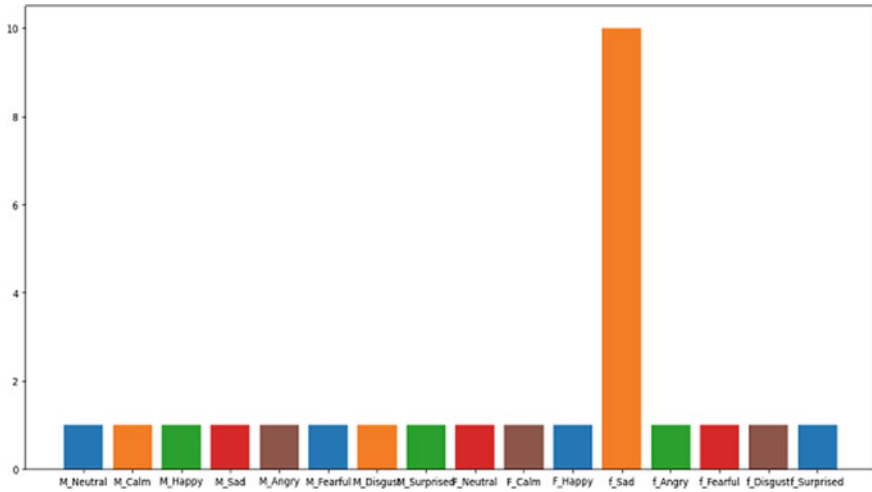


Fig. 6 Gender-emotion chart

References

1. Sidorov M, Ultes S, Schmitt A (2014) Emotions are a personal thing: towards speaker adaptive emotion recognition. In: IEEE International conference on acoustics, speech and signal processing, pp 4803–4807
2. Vogt T, Andre E. (2006) Improving Automatic Emotion Recognition from speech via gender differentiation In: Language resources and evaluation conference. Genona
3. Souza AD, Souza RD (2017) A review on emotion recognition using various methods and classification techniques
4. Gupta M, Bharti SS, Agarwal S (2016) Support vector machine based gender identification using voiced speech frames
5. Bisio I, Delfino A, Lavagetto F, Marchese M, Sciarrone A (2014) Gender-driven emotion recognition through speech signals for ambient intelligence applications
6. Jadav S (2018) Voice-based gender identification using machine learning
7. Schuller B, Steidl S, Batliner A (2009) The INTERSPEECH 2009 emotion challenge. In: Tenth annual conference of the international speech communication association

Architectural Design and Analysis for Economical Operation of Photovoltaic DC Microgrid for Rural Electrification



CH. Venkata Ramesh , K. C. Ravi Kiran , T. C. Balachandra ,
B. L. Namratha , and B. N. Lakshmi Narayana 

1 Introduction

Power performs a critical role in economic development, and the energy region is an compulsory infrastructure in any financial system. Supplying adequate and low cost electric strength is crucial for economic improvement, socio-financial and better popular of dwelling. Demand for strength in a growing united states of America like India is full-size and is growing steadily. The demand for energy in India is vast and is developing steadily. The improvement of numerous sectors of the financial system is distinctly difficult without the improvement of the strength area. In truth, it has ended up an essential ingredient for enhancing the exceptional of existence and its absence is related to poverty and negative exceptional of lifestyles [1]. Presently, all the villages in India had been electrified; it does not imply that all households are electrified. As of August 2018, most effective 91% of the overall households are electrified in India. As the cost of PV modules goes on decreases the deployment of solar primarily based microgrids is becoming popular in growing countries for rural electrification.

Further, using incorporated microgrids containing nearby technology, distribution and intake have its benefits as compared to unconnected structures [2]. The DC microgrids with each generation and masses are of DC could be greater optimized due to efficient technology, distribution, and storage of electrical power because of the absence of conversion degrees. The overall performance of DC microgrids is 80% compared to the AC microgrids, i.e. of round 60% only [3]. This evaluation is solely based totally on DC microgrid of low voltage direct contemporary; it will mitigate or removes the pricey up-conversion and down conversions for the reason

CH. V. Ramesh (✉) · K. C. R. Kiran · T. C. Balachandra · B. L. Namratha · B. N. L. Narayana
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: venkataramesh.ch@nmit.ac.in

that distribution distances are shorter in village degree, and it is far important to layout viable architecture design.

On this paper, a DC microgrid is modelled to evaluate its working efficiency, electricity losses and levelized cost of electricity (LCOE) for the proposed and existing architectures, i.e. cluster architecture, C-kind architecture I, C-kind architecture II, and O-kind architecture. In this work a rural far-flung village in which sixteen houses and water drive of 5 Horse Power is used. It has been shown that proposed (cluster) architecture has more advantages over the existing architectures due to the comparative analysis of its lower power loss and unit electricity cost.

2 Different Architecture Used In Rural Electrification

2.1 C-kind Architecture I

The birds-eye view of the layout for village containing sixteen houses using C-kind architecture I [4] can be seen in Fig. 1. It indicates the generation spatial plan of homes and energy distribution structure together with the placement storage devices. The conductors are placed in linear way, generation and energy processing and storage unit are positioned at one end (left side of the layout) of the road/layout, it is in

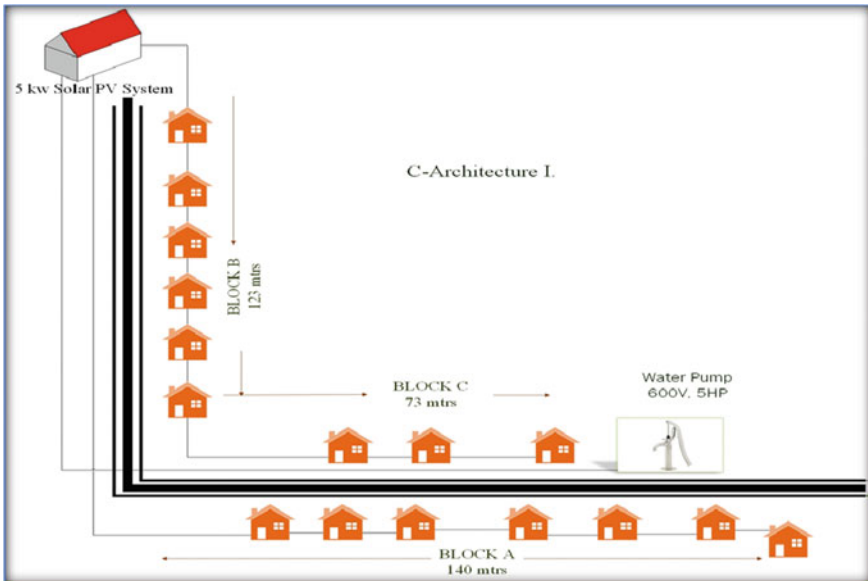


Fig. 1 Spatial plan of homes and energy distribution structure for C-kind architecture I with photovoltaic power

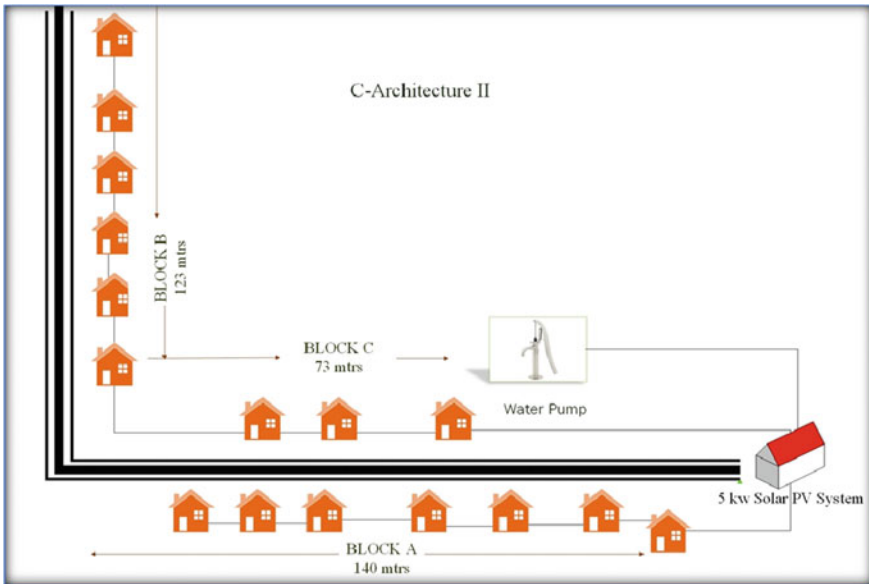


Fig. 2 Spatial plan of homes and energy distribution structure for C-kind architecture II with photovoltaic power

the form of C-like shape and is termed as linearly dispensed C-architecture I DC microgrid. This architecture is a simple version of some linear villages (i.e., home hundreds going for walks parallel to the road linearly) determined in India.

2.2 C-kind Architecture II

The birds-eye view of the layout for village containing sixteen houses using C-kind architecture I can be seen in Fig. 2. It indicates the generation spatial plan of homes and energy distribution structure together with the placement storage devices. The conductors are placed in linear way while generation and energy processing and storage unit are positioned at one end (Right side of the layout) of the road/layout, it is in the form of C-like shape and is termed as linearly dispensed C-architecture-II DC microgrid.

2.3 O-kind Architecture

The birds-eye view of the layout for village containing sixteen houses using O-kind architecture can be seen in Fig. 3. The conductors are placed in linear way while

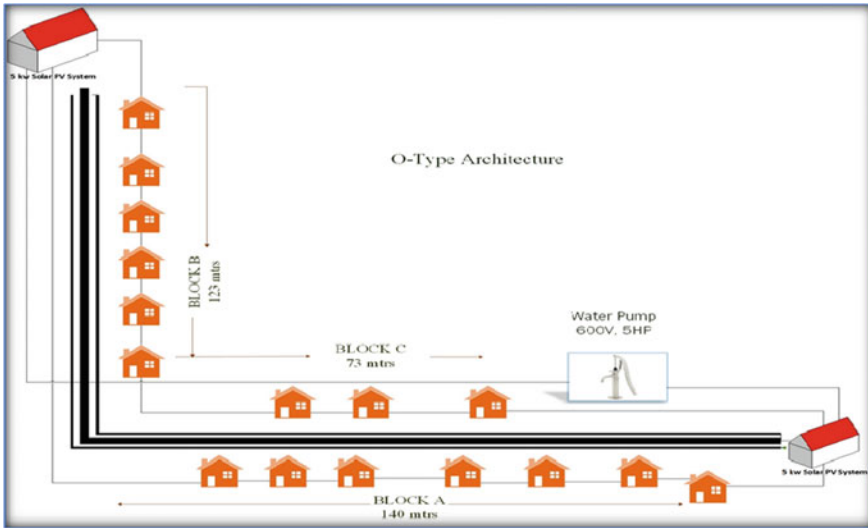


Fig. 3 Spatial plan of homes and energy distribution structure for o-kind architecture with photovoltaic power

generation and energy processing and storage unit are positioned at both the ends of the road/layout.

2.4 Cluster Architecture

Cluster architecture outfit to villages, there may be a clustered of homes or if houses are allotted in a random way. Length of every cluster and range of homes in every cluster may additionally vary relying upon the shape of the village. Furthermore, the gap amongst numerous houses in conjunction with the space of every house from the central generation and storage unit might also vary. Usually, this type of architecture is seen more since the arrangement of houses may not be unique in rural remote areas [5]. Birds-eye view of cluster architecture can be seen in Fig. 4.

3 Dc Microgrid Model

The version proposed in this paper depends on the renewable power supply being used. The proposed version facilitates generation and utilization via solar PV panels during daylight hours and battery is used when solar power is not available. In the proposed version of agricultural village, every house manages to devour the energy

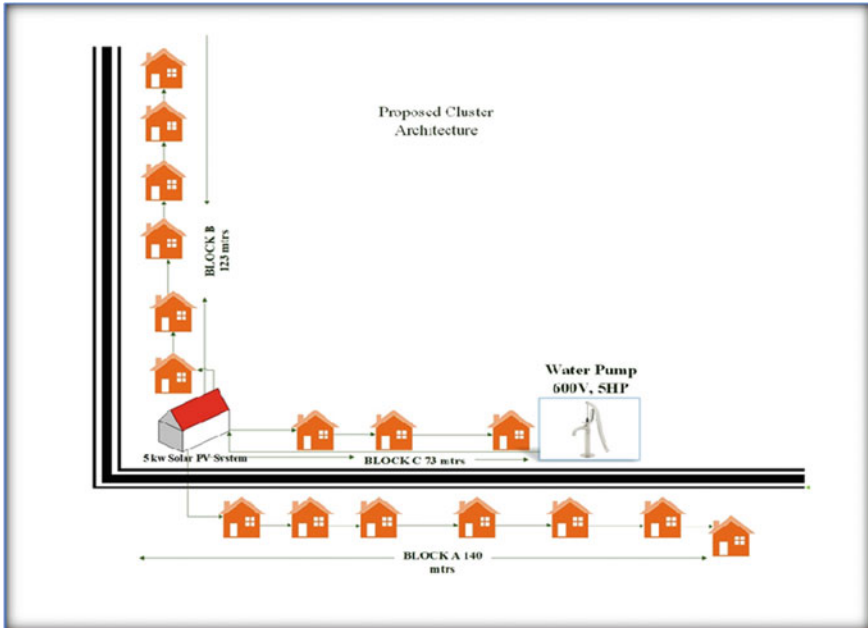


Fig. 4 Spatial plan of homes and energy distribution structure for cluster architecture with photovoltaic power

generated by centralized supply and the storage unit [6]. The power necessary to pump the water is generated centrally.

4 Matlab Modelling

The special sorts of Architectures discussed are simulated in MATLAB/SIMULINK (Fig. 5).

a. Model of solar photovoltaic system

- Number of panels connected in series in each array = 4
- Number of arrays connected in parallel
- Irradiation = 1000 W/m²
- Voltage rating PV = 145 V
- Current rating of PV = 35 V

Manual exchange switch is integrated to make all of the panel in series to achieve 600 V to feed the water drive (Figs. 6, 7 and 8).

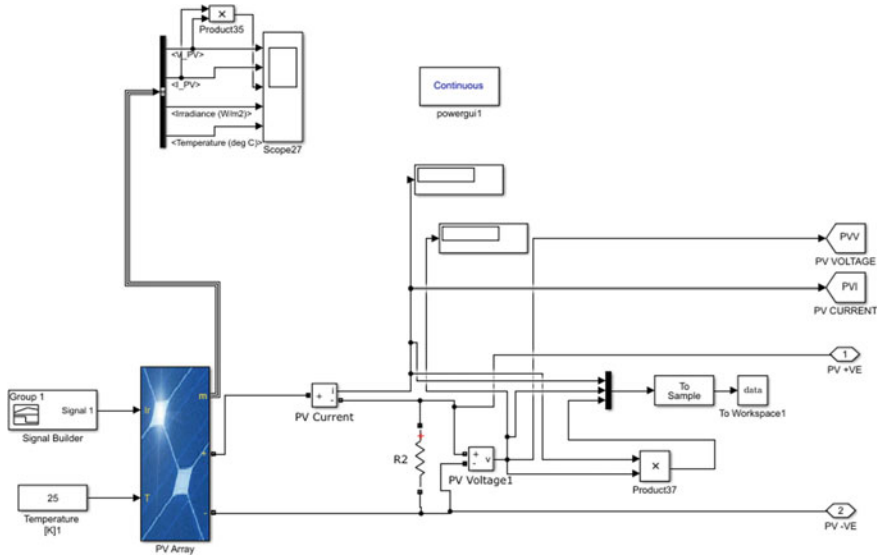


Fig. 5 MATLAB Simulink model for PV Array with Variable Irradiance

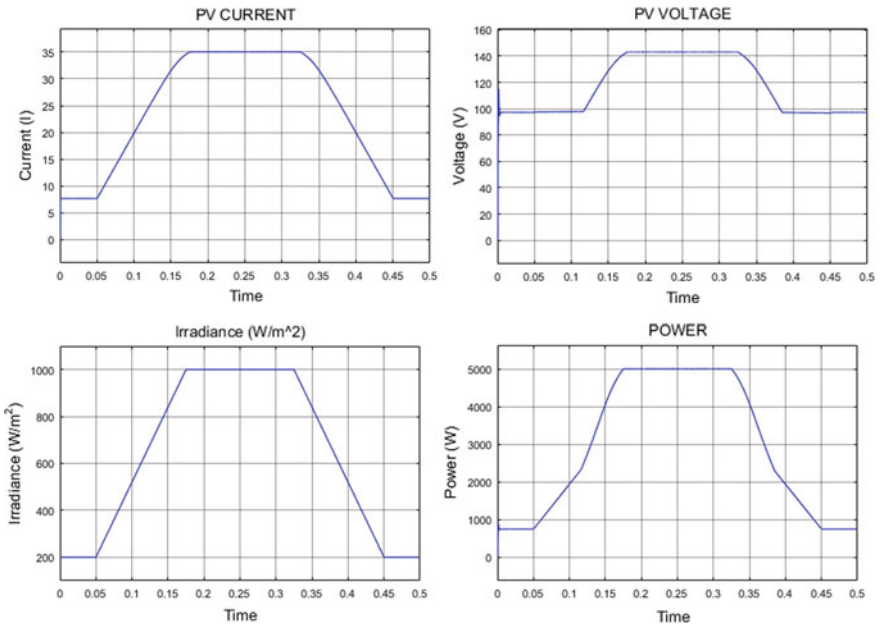


Fig. 6 Simulated results showing PV Current, PV Voltage, Irradiance and Power variation with time.

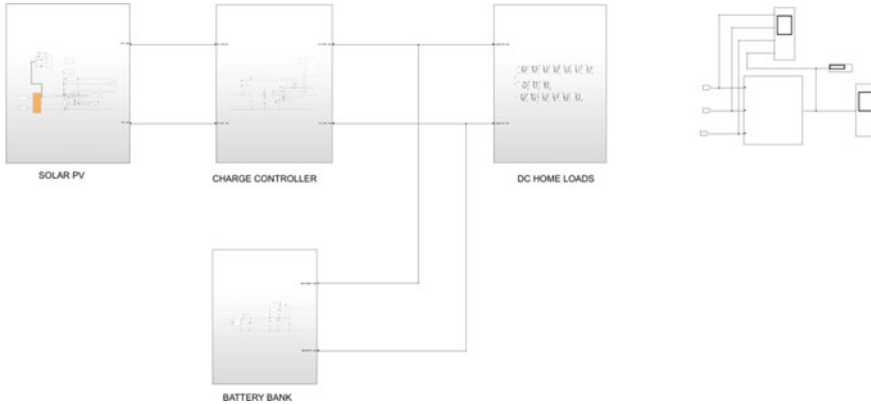


Fig. 7 Overview of the microgrid model

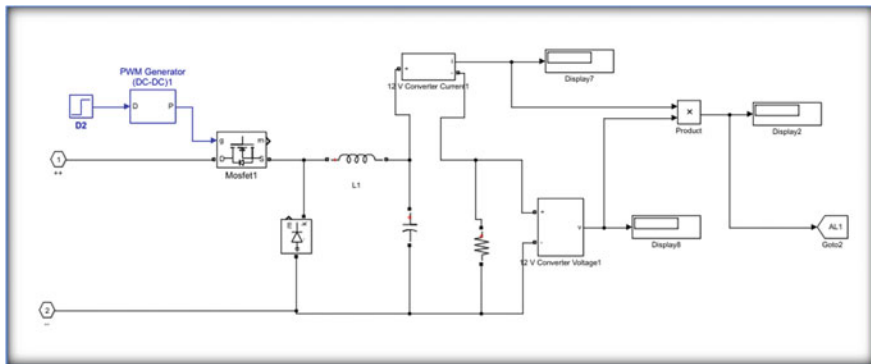


Fig. 8 Simulink model for Buck converter

b. *Home Loads and Buck Converter Model*

In the proposed architecture, the position of homes is not uniform in phrases of distance, the energy generation and storage element are positioned in centre of the layout is simulated. Figure 7 represents the overview of the microgrid model. Figure 8 represents the overview of the Simulink model of buck converter.

c. *Outline of DC MicroGrid Model with buck converter*

5 Economic Analysis of Dc Microgrid

Price evaluation of a task is a very essential factor to recognize the realistic viability of that undertaking. It has a scientific method which allows us to calculate the system unit strength price [7]. The existing worthiness of system can be calculated by means of the subsequent equations.

$$C_{pw} = C_{comp} \left(\frac{1+i}{1+d} \right)^N \tag{1}$$

$$C_{mpw} = \left(\frac{m}{yr} \right) * \left(\frac{1+i}{1+d} \right) * \left(\frac{1 - \left(\frac{1+i}{1+d} \right)^N}{1 - \left(\frac{1+i}{1+d} \right)} \right) \tag{2}$$

- No of years = N
- Bank Inflation = i
- Discount rate = d
- Component cost = C_{comp} .

Further, the existing well worth of operation and renovation value can be evaluated by way of the following equations.

The full life-cycle cost (LCC) of the undertaking has to be calculated by using summing individual present worth price of each component inclusive of O&M price [8]. Then, annualized LCC may be calculated through Eq. (4). Life cycle costing is computed for 25yrs

$$LCC = \sum (\text{presentwell worthpriceof eachcomponent}) \tag{3}$$

$$ALCC = LCC \left(\frac{1 - \left(\frac{1+i}{1+d} \right)}{1 - \left(\frac{1+i}{1+d} \right)^N} \right) \tag{4}$$

In the end, unit energy price is calculated with ALCC and overall power generated.

$$\text{Unit Electrical Cost} = ALCC/365E_L \tag{5}$$

6 Results and Discussion

An average village of 16 houses with 40 W intake is considered for all 4 sorts of spatial distributions in a village. Distance amongst two consecutive homes is varied in simulation. The power losses, performance and worst case voltage drop of the distribution system is designed for each of the proposed configurations. For cluster course of village, randomly allotted cluster architecture is considered with three Blocks/clusters.

The voltage drop and power loss can be decided by using following formulae.

$$E_b = a * \left(\frac{b * \rho * R * L}{S} \right) * I^2 \tag{6}$$

$$V_D = \frac{b * \rho * I * L}{S} \tag{7}$$

MATLAB Results for Different Architectures

- a. C-Architecture I
- b. C-Architecture II
- c. O-Architecture
- d. Cluster Architecture (Figs. 9, 10, 11 and 12).

The power loss in O-Type architecture is less 4.61, and the proposed cluster architecture is 5.269w/h. The difference between two architectures power loss is 0.654w/h which is negligible/comparatively less (Graph 1).

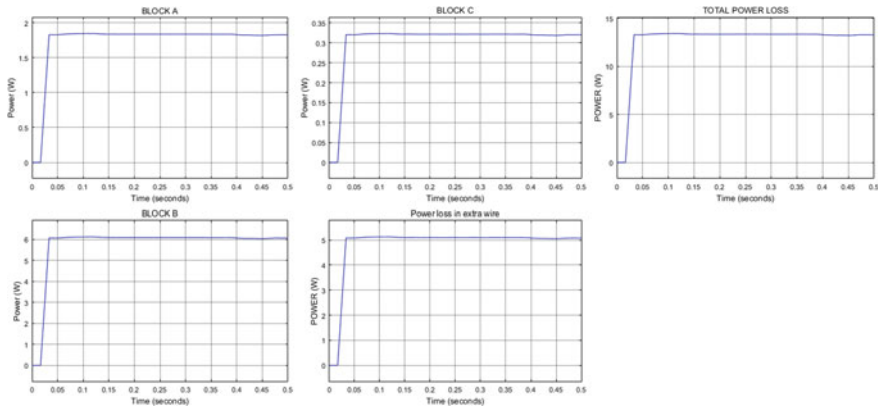


Fig. 9 Power losses in different blocks, extra wire and total power loss as a function of time in C-Architecture I

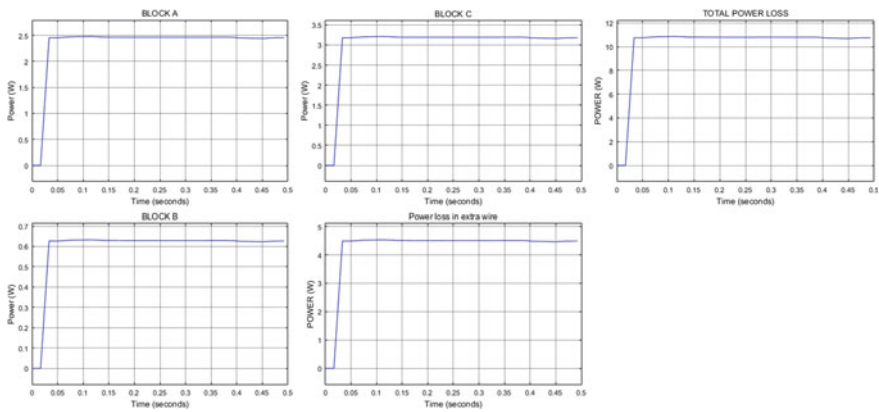


Fig. 10 Power losses in different blocks, extra wire and total power loss as a function of time in C-Architecture II

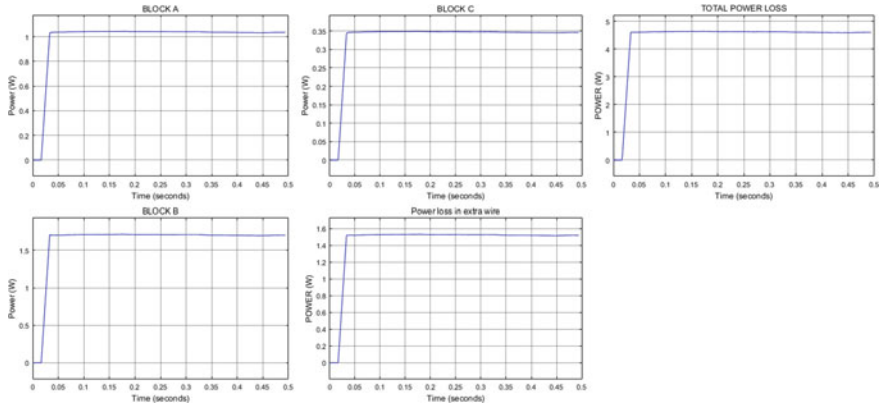


Fig. 11 Power losses in different blocks, extra wire and total power loss as a function of time in O-Architecture

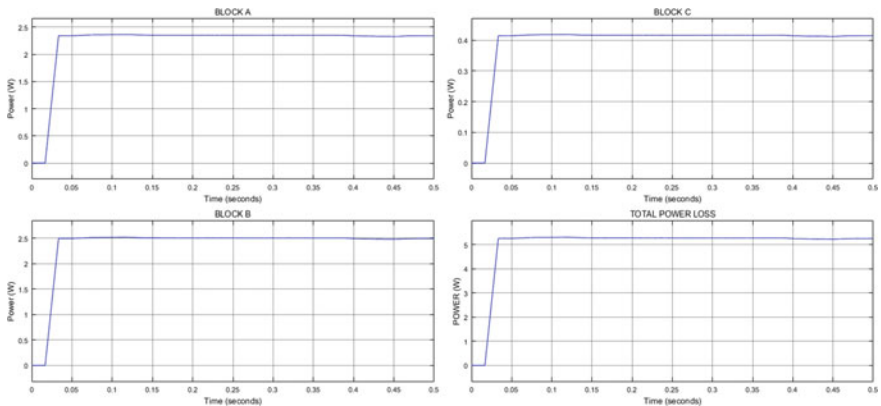
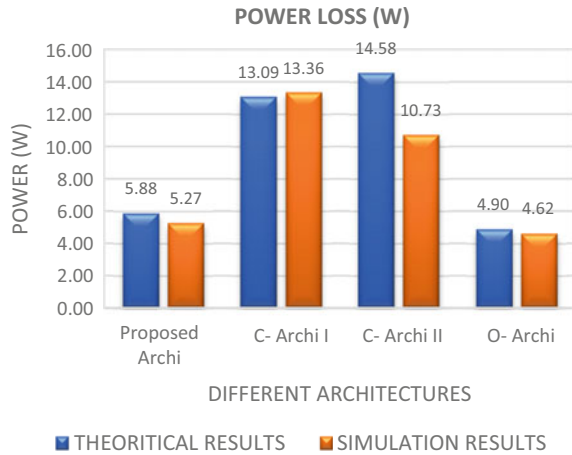


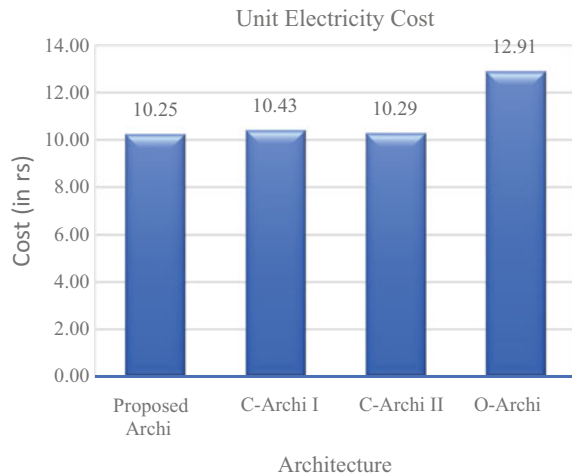
Fig. 12 Power losses in different blocks, extra wire and total power loss as a function of time in proposed cluster-architecture

The comparative analysis of Unit Electrical Cost is presented in the above Graph 2, the cost analysis clearly pictures that the UEC is much lesser in the proposed architecture, i.e.10.25 INR and 10.43INR for C-Architecture I, 10.29 INR for C-Architecture II and 12.91 INR for O-Architecture which is higher. Centralized battery banks must be kept on both sides of the source.

Graph 1 Comparison of theoretical and simulation results



Graph 2 Comparison of unit electricity cost



7 Conclusion

A method for deciding on an efficient distribution structure of sun PV-based DC microgrid with applicability to village orientation has been presented in this work. The targeted loss analysis indicates that O-architecture is the maximum efficient structure for rural electrification. Based totally on power loss evaluation, it is proved that O-structure is efficient. It requires more maintenance since the two sources have to be interconnected, and it requires more safety equipment, to interconnect both the sources extra cable has to be used, if one of the sources fails, then the loss gradually increases and in addition, it increases the distance between source and the load connected at the end of the line. The simulation results indicate that the

Table 1 Cost analysis of different architecture designs

	Proposed cluster archi	C-Archi I	C-Archi II	O-Archi
Total project cost	1,465,085	1,499,592.4	1,473,332	1,959,815
Maintenance cost	240,000	240,000	240,000	240,000
Replacement cost	297,991.5	297,991.5	297,991.5	322,909.3
LCC	2,003,076.5	2,037,583.9	2,011,323.5	2,522,724.3
ALCC	112,252	114,186	112,715	141,373
UEC (in INR)	10.25	10.43	10.29	12.91

Table 2 Architectures and corresponding theoretical and simulation results of power loss

Architectures	Theoretical results of power loss (w/h)	Simulation results of power loss (w/h)
Proposed cluster architecture	5.88	5.269
C-architecture I (Left)	13.08	13.3
C- architecture II (Right)	14.58	10.73
O-architecture	4.90	4.615

distinction between the proposed cluster architecture and the O-type architecture is comparatively less, i.e.0.654w/h. Economic analysis carried out in this work is presented in Table 1 which shows that the difference between total project cost required for the proposed work and O-type architecture is approximately 500000INR (i.e., of 25.3% of the total project cost) and the proposed system’s unit electricity cost is 2.66INR/Unit which is lesser compared to the O-type architecture. It can be concluded that the proposed cluster architecture is advanced (Table 2).

- Initial investment is high in O-architecture since it has dual source, maintenance cost & structure cost increases.
- Since the difference between total power loss in the proposed and O-type architecture is not so high, huge investment is not required.
- While constructing the O-type system cable length will increase and one of the main concerns is the battery maintenance (replacement and enclosure for the battery)
- Battery management requires 30% of the total project cost

Deployment of solar PV-based microgrid can be further enhanced with the smart technology so, that the safety issues, voltage dip, and the power loss can be mitigated. Solar PV-based microgrid can also be integrated into the grid in future if generation exceeded the demand. Based on geographical conditions wind power plant can also be implemented in future and it can be integrated into the proposed system.

References

1. <https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=IN>
2. Rajasekar N, Bilakanti N, Miyatake M (2017) Energy management technique for home micro grid system. In: TENCON 2017 - 2017 IEEE region 10 conference, Penang, pp 1433–1438. <https://doi.org/10.1109/TENCON.2017.8228083>
3. Hamza M, Shehroz M, Fazal S, Nasir M, Khan HA (2017) Design and analysis of solar PV based low-power low-voltage DC microgrid architectures for rural electrification. In: 2017 IEEE Power and energy society general meeting, Chicago, IL, pp 1–5. <https://doi.org/10.1109/PESGM.2017.8274134>
4. Nasir M, Iqbal S, Khan HA (2018) Optimal planning and design of low-voltage low-power solar DC microgrids. *IEEE Trans Power Syst* 33(3):2919–2928. <https://doi.org/10.1109/TPWRS.2017.2757150>
5. Nasir M, Zaffar NA, Khan HA (2016) Analysis on central and distributed architectures of solar powered DC microgrids. In: 2016 Clemson university power systems conference (PSC), Clemson, SC, pp 1–6. <https://doi.org/10.1109/PSC.2016.7462817>
6. Nasir M, Khan HA, Hussain A, Mateen L, Zaffar NA (2018) Solar PV-based scalable DC microgrid for rural electrification in developing regions. *IEEE Trans Sustain Energy* 9(1):390–399. <https://doi.org/10.1109/TSTE.2017.2736160>
7. Haque M (2012) thesis titled “Design and economic analysis of a standalone PV system to electrify the remote area of Bangladesh” in the 2011–2012 session at Institute of Energy, University of Dhaka
8. Ahamed MHF, Dissanayake UDSD, De Silva HMP, Pradeep HRCGP, Lidula NWA (2016) Modelling and simulation of a solar PV and battery based DC microgrid system. In: 2016 International conference on electrical, electronics, and optimization techniques (ICEEOT), Chennai, pp 1706–1711. <https://doi.org/10.1109/ICEEOT.2016.7754977>

Effects of Variable Gravity Field on the Onset of Ferroconvection in an Anisotropic Porous Layer



S. Kiran , Y. H. Gangadharaiah , H. Nagarathnamma, and R. Padmavathi

1 Introduction

Ferrofluid or also known as magnetic fluid is a non-electric carrier fluid that includes small particles of strong ferromagnetic materials [1]. Kaiser and Miskolczy [2] state that ferrofluid has a special feature which is it can maintain its fluid properties in the existence of magnetic field, and the magnetic properties of ferrofluid can be affected by the composition, distribution and also volume concentration. Previously, ferrofluid is known in the rocket fuel by NASA, and currently, ferrofluid has been used in a various field such as in electric devices, mechanical engineering, medical applications and optic. There have been tremendous studies on the convection of ferrofluid. Stiles and Kagan [3] found the instability of ferroconvection in a strong magnetic field. The impact of the vertical magnetic field in ferrofluid was researched by Hennenberg et al. [4]. Mokhtar and Arifin [5] had the impacts of feedback control in the ferrofluid layer system. Laroze. Recently, the impact of magnetic field dependent viscosity in a ferrofluid was proved by Prakash et al. [6].

The ideas of gravity are well known in the theoretical investigation, a lot of studies related to variable gravity had been done before to justify a convection phenomenon in a large scale such as in the ocean and mantle. If the height of the gravity field varies, then the fluid encounters various forces of buoyancy at different stages. So it becomes

S. Kiran (✉) · R. Padmavathi

Department of Mathematics, Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: kiran.s@nmit.ac.in

R. Padmavathi

e-mail: padmavathi.r@nmit.ac.in

Y. H. Gangadharaiah

Department of Mathematics, RV Institute of Technology and Management, Bangalore, India

H. Nagarathnamma

Department of Mathematics, Dr. Ambedkar Institute of Technology, Bangalore, India

important to research fluid convection with variable gravity. Pradhan and Samal [7] have carried out the pioneering study in this area. The authors consider a fluid layer that under the presence of a gravity field is heated from above or below varies with height. The effect variable gravity and internal heating with additionally inclined temperature gradient in a porous medium had been done by Alex et al. [8]. Bala and Chand [9] employed Brinkman porous medium with variable gravity of ferrofluid. The effect of variable gravity and throughflow on the onset of thermal convective motion in a porous layer reported by Suma et al. [10] and Gangadharaiyah et al. [11] using regular perturbation technique. Dhananjay Yadav [12] and Nagarathnamma et al. [13] investigated variable gravity field effects on porous layer by using Gelerkin technique.

The object of this investigation is to study the effect of variable gravity on the onset of ferroconvection in an anisotropic porous medium. In this paper, therefore, we analyze the effect of vertical magnetic field on system with anisotropic porous layer using three different cases of gravity field variations.

2 Conceptual Model

Figure 1 illustrates the physical configuration of the present study. The physical model under consideration is a horizontal anisotropic porous bed bounded between planes at $z = 0$ & $z = d$ and changeable gravity $g(z)$. We assume that the gravity vector \vec{g} is, $\vec{g} = -g_0(1 + \lambda H(z))\hat{k}$.

The surface tension, σ and density of the fluid, ρ are in the form of

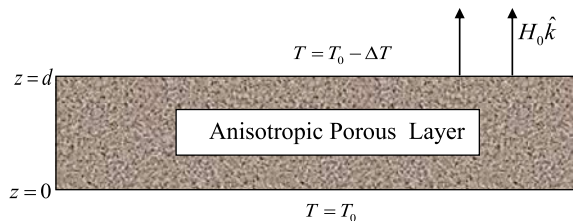
$$\rho = \rho_0[1 - \alpha(T - T_0)] \tag{1}$$

$$\sigma = \sigma_0 - \sigma_T(T - T_0) \tag{2}$$

The basic governing equations for the anisotropic porous layer are:

$$\nabla \cdot \vec{V} = 0 \tag{3}$$

Fig. 1 Physical configuration



$$\frac{\rho_0}{\phi} \left(\frac{\partial \vec{V}}{\partial t} + (\vec{V} \cdot \nabla) \vec{V} \right) = -\nabla p - \frac{\mu}{K} \vec{V} + \rho \vec{g} + \mu_0 (\vec{M} \cdot \nabla) \vec{H} \tag{4}$$

$$\begin{aligned} \varepsilon \left[\rho_0 c_{v,H} - \mu_0 \vec{H} \cdot \left(\frac{\partial \vec{M}}{\partial t} \right)_{v,H} \right] \times \frac{DT}{Dt} + (1 - \varepsilon) (\rho_0 c) \frac{\partial T}{\partial t} + \mu_0 T \left(\frac{\partial \vec{M}}{\partial t} \right)_{v,H} \\ \times \frac{D\vec{H}}{Dt} = k \nabla^2 T \end{aligned} \tag{5}$$

$$\nabla \cdot \vec{B} = 0, \quad \nabla \cdot \vec{H} = 0 \quad OR \quad \vec{H} = \nabla \varphi \tag{6}$$

$$\vec{B} = \mu_0 (\vec{M} + \vec{H}), \quad \vec{M} = [M_0 + \chi(H - H_0) - K(T - T_0)] \frac{\vec{H}}{H} \tag{7}$$

The solutions for the quiescent basic state are as follows

$$T_b(z) = T_0 - \beta z \tag{8}$$

$$\vec{H}_b(z) = \left[H_0 - \frac{K\beta z}{1 + \chi} \right] \hat{k} \tag{9}$$

$$\vec{M}_b(z) = \left[M_0 + \frac{K\beta z}{1 + \chi} \right] \hat{k} \tag{10}$$

The basic state is perturbed in the following form

$$\begin{aligned} \vec{V} = \vec{V}', \quad p = p_b(z) + p', \quad T = T_b(z) + T', \\ \vec{H} = \vec{H}_b(z) + \vec{H}', \quad \vec{M} = \vec{M}_b(z) + \vec{M}' \end{aligned} \tag{11}$$

Substituting Eq. (11) into Eq. (7) and by using equation the basic state, yields

$$H_x + M_x = \left(1 + \frac{M_0}{H_0} \right) H_x \tag{12}$$

$$H_y + M_y = \left(1 + \frac{M_0}{H_0} \right) H_y \tag{13}$$

$$H_z + M_z = (1 + \chi) H_z - KT \tag{14}$$

The normal mode expansion is assumed in the form:

$$(w, T, \varphi) = [W(z), \Theta(z), \Phi(z)] \exp[i(lx + my)] \tag{15}$$

Substituted Eq. (15) into momentum equation, energy equation and also Maxwell equation. After that we performing the linearization. We obtain the following ordinary differential equations

$$\left[(D^2 - a^2)^2 - \frac{1}{Da} \left(\frac{1}{\xi} D^2 - a^2 \right) \right] W - R a^2 (1 + \lambda G(z))(D\phi - \Theta) - R a^2 (1 + \lambda G(z))\Theta = 0 \tag{16}$$

$$(D^2 - \eta a^2)\Theta = (1 - M_2) W \tag{17}$$

$$(D^2 - M_3 a^2)\Phi = D\Theta. \tag{18}$$

where $Da = k/d^2$ is the Darcy number, $M_2 = \mu_0 T_0 K^2 / (1 + \chi)$ is the magnetic parameter, $R_T = \alpha_t g \beta d^4 / \nu \kappa$ is the thermal Rayleigh number, $\xi = k_h / k_v$ is an anisotropic permeability, M_1 is the magnetic number, $R_m = M_1 R_T = \mu_0 K^2 \beta / (1 + \chi) \alpha_t \rho_0 g$ is the magnetic Rayleigh number, $M_3 = \left(1 + \frac{M_0}{H_0} \right) / (1 + \chi)$ is the nonlinearity of the ferrofluid, and $\eta = k_h / k_v$ is an anisotropic effective thermal diffusivity.

The boundary conditions take the form

$$W = DW = D\Theta = \Phi = 0 \quad \text{at } z = 0 \tag{19}$$

$$W = D\Theta = D^2 W = D\Phi = 0 \quad \text{at } z = 1 \tag{20}$$

3 Solution by Regular Perturbation Technique

Accordingly, the variables W, Θ and Φ are expanded in powers of a^2 as

$$(W, \Theta, \Phi) = \sum_{i=0}^N (a^2)^i (W_i, \Theta_i, \Phi_i) \tag{21}$$

$$D^4 W_0 - \frac{1}{Da \xi} D^2 W_0 = 0 \tag{22}$$

$$D^2 \theta_0 = W_0 \tag{23}$$

$$D^2 \Phi_0 - D\theta_0 = 0 \tag{24}$$

$$W_0 = DW_0 = D\Theta_0 = 0 = \Phi_0 \text{ at } z = 0 \tag{25}$$

$$W_0 = D^2W_0 = D\Theta_0 = 0 = D\Phi_0 \text{ at } z = 1 \tag{26}$$

The solution to the zeroth order Eqs. (22) till (24) by using boundary conditions in Eq. (25) and (26) are as follow:

$$W_0 = 0, \quad \Phi_0 = 0 \text{ and } \Theta_0 = 1 \tag{27}$$

First-order equations are

$$D^4W_1 - \frac{1}{Da\xi} D^2W_1 = (1 + \lambda G(z))R_T - (1 + \lambda G(z))R_m \tag{28}$$

$$D^2\Theta_1 - \eta = W_1 \tag{29}$$

$$D^2\Phi_1 = D\Theta_1 \tag{30}$$

$$W_1 = DW_1 = D\Theta_1 = 0 = \Phi_1 \text{ at } z = 0 \tag{31}$$

$$W_1 = D^2W_1 = D\Theta_1 = 0 = D\Phi_1 \text{ at } z = 1 \tag{32}$$

Eqs. (28) to (29) will be solved for three cases of gravity field variations for the three cases: (i) $H(z) = -z$ (ii) $H(z) = -z^2$ and (iii) $H(z) = -(e^z - 1)$ by using MAPLE. The equation of R_{TC} will be generate in term of M_1, λ, ξ, η and Da . We assumed $M_2 = 0$ by referring to Finlayson [14], M_2 will not affect the convection since the value of it will be approximated to zero because the value too small which is 10^{-6} .

4 Results and Discussion

In this paper, with the presence of variable gravity, the resulting eigenvalue problem of ferroconvective motion in an anisotropic porous layer was analytically solved using regular perturbation technique. The boundaries are regarded rigid-free and insulating with a linear stability assessment. The outcomes collected are described graphically in Fig. 2 to show the effect of different parameters on the critical thermal Rayleigh number, R_{TC} .

Figure 2 depicts the vertical velocity W for different values of λ in the case of linear gravity variance ($G(z) = -z$). The effect of gravity parameter λ has no noticeable influence on W in the case of linear gravity variance and for case of exponential variation it shows accelerate W which is presented in Fig. 3. Figures 4, 5

Fig. 2 Vertical velocity eigen functions W for different values of λ with $G(z) = -z$

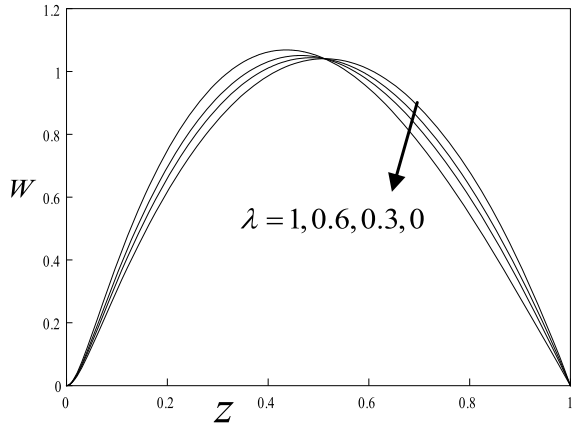


Fig. 3 Vertical velocity Eigen functions W for different values of λ with $G(z) = -(e^z - 1)$

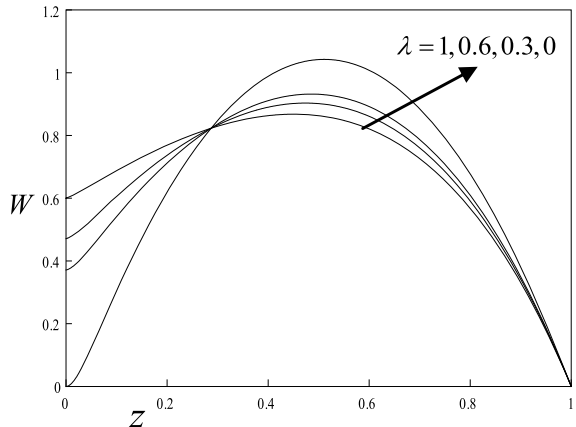


Fig. 4 R_{TC} versus λ in the case of gravity field $G(z) = -z^2$

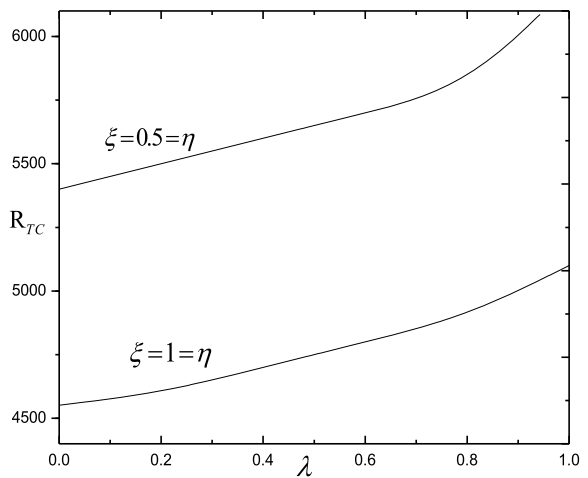
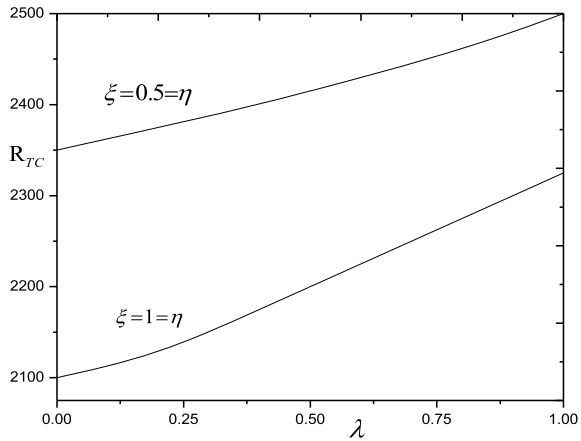


Fig. 5 R_{TC} versus λ in the case of gravity field $G(z) = -z$



and 6 illustrate the influence of the anisotropy ($\xi = 0.5 = \eta$) against isotropic case ($\xi = 1 = \eta$) with λ for $Da = 0.001$ and $M_1 = 1$. As expected, it is observed from Figs. 4, 5 and 6 that system shows more stabilizing in the case anisotropy porous system.

Figures 7, 8 and 9 show the effects of mechanical anisotropic parameter, ξ , on the onset of ferroconvection for different values of M_1 with different gravity variance field for $\lambda = 1$, $Da = 0.001$ and $\eta = 0.5$. On increasing the value of ξ , we found that the value of R_{TC} decrease. The value of ξ is directly proportional to the horizontal permeability, K_x of the porous and inversely proportional to the vertical permeability, K_z of the porous medium. This is due to the facts that increasing the value of K_x will cause the size of the cell to become larger while decreasing the value of K_z will result in larger temperature difference between the lower and upper plate as reported by Degan et al. [15]. Furthermore, it is noticed that the system more unstable for case (ii), while for case (iii) it is found to be more stable.

Fig. 6 Critical R_{TC} versus λ in the case of gravity field $G(z) = -(e^z - 1)$

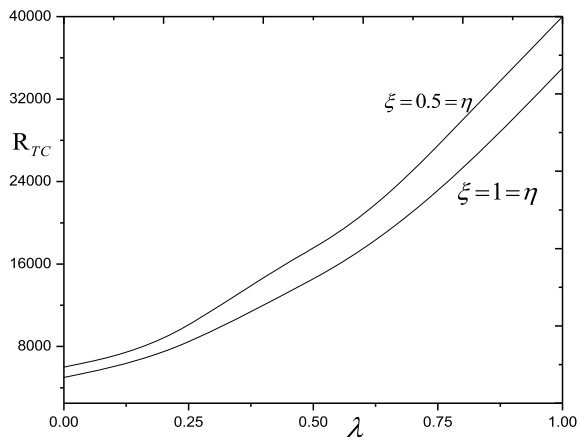


Fig. 7 R_{TC} versus ξ for different values of M_1 for the gravity field $G(z) = -z$

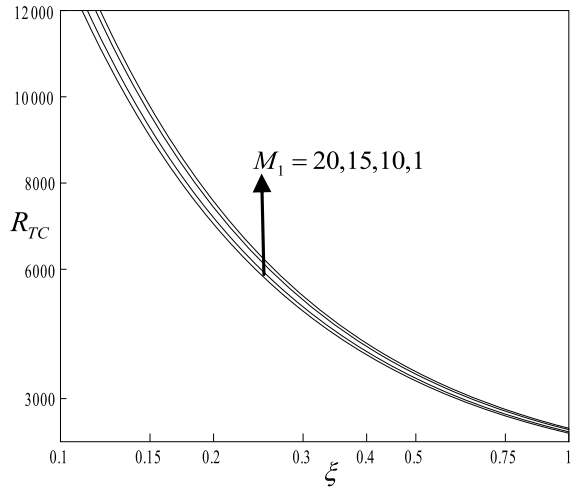


Fig. 8 R_{TC} versus ξ for different values of M_1 for the gravity field $G(z) = -z^2$ with $\lambda = 1$

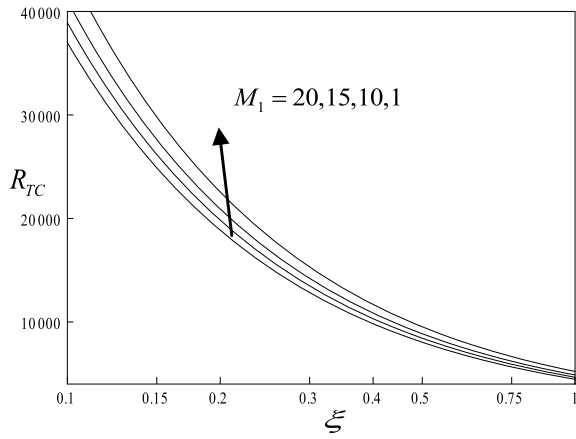


Fig. 9 R_{TC} versus ξ for different values of M_1 for the gravity field $G(z) = -(e^z - 1)$

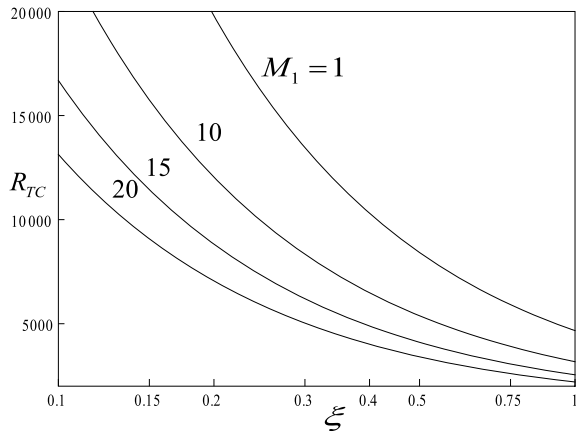


Fig.10 R_{TC} versus η for different values of λ for the gravity field $G(z) = -(e^z - 1)$

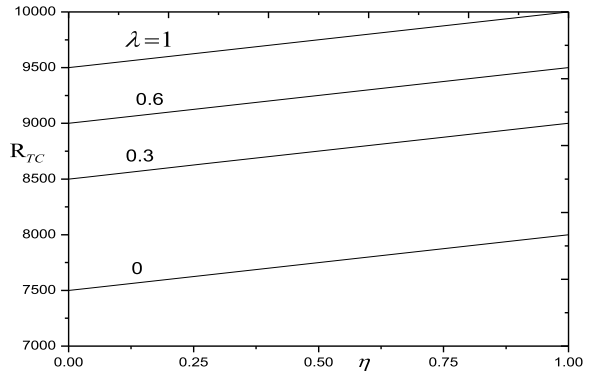
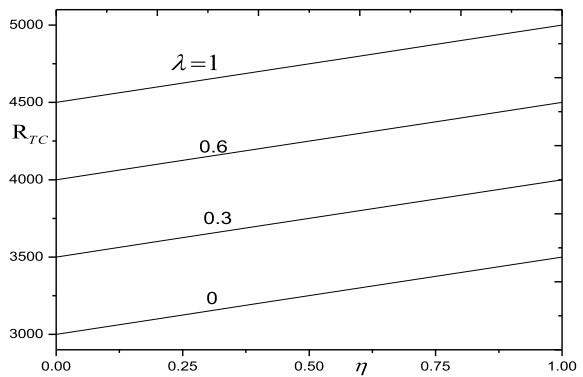


Fig. 11 R_{TC} versus η for different values of λ for the case of gravity field $G(z) = -z$

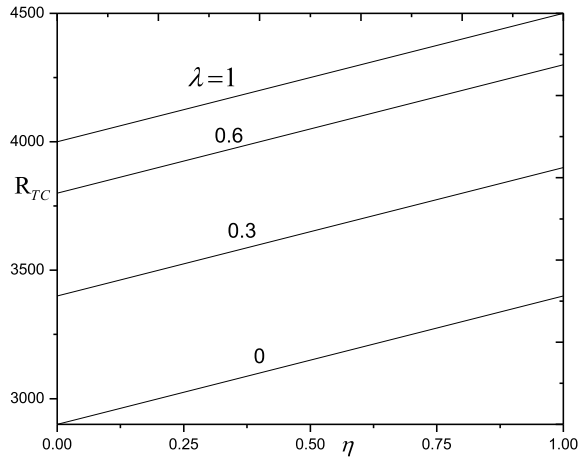


The critical thermal Rayleigh number R_{TC} with thermal anisotropic parameter, η for different for different gravity variance field with for different values of λ are presented in Figs. 10, 11 and 12. It is noted that with increasing the value of η the value of R_{TC} also increase. Increasing η is corresponds to decrease in vertical thermal diffusivity κ_{Tz} which will slow down the heat flow vertically through it. Furthermore, it is noticed that the system more unstable for case (ii), while for case (iii) it is found to be more stable.

5 Conclusions

The stationary thermal ferroconvection in a fluid saturated an anisotropic porous medium in the presence of variable gravity field is investigated analytically using linear stability analysis. The resulting eigenvalue problem obtained from the governing equations is solved analytically by regular perturbation method. The study was conducted in three different cases of variance of the gravity field: (i) $G(z) = -z$ (ii) $G(z) = -z^2$ and (iii) $G(z) = -(e^z - 1)$. The function of thermal Rayleigh

Fig. 12 R_{TC} versus η for different values of λ for the case of gravity field $G(z) = -z^2$



number is obtained with the purpose of investigating the effect of magnetic Rayleigh number, magnetic parameter, nonlinearity of the ferrofluid, mechanical anisotropy, ξ and thermal anisotropy parameter, η with different values of gravity parameter on the onset of thermal ferroconvection. Results show that the effects of increasing the increasing values of M_1 , Da and will enhance the convection of a ferrofluid layer system while the increasing of η will help to stabilize the system. It is noted that for case (iii) the system is more stable, while for case (ii) the system is more unstable.

References

1. Raju K (2018) Effect of temperature dependent viscosity on ferrothermohaline convection saturating an anisotropic porous medium with Soret effect using the Galerkin technique. *Int J Heat Technol* 36(2):439–446. <https://doi.org/10.18280/ijht.360208>
2. Robert K, Miskolczy G (1970) Some applications of ferrofluid magnetic colloids. *IEEE Trans Magnet* 6(3):694–698. <https://doi.org/10.1109/TMAG.1970.1066834>
3. Stiles PJ, Kagan M (1990) Thermoconvective instability of a ferrofluid in a strong magnetic field. *J Colloid Interface Sci* 134(2):435–448. [https://doi.org/10.1016/0021-9797\(90\)90154-G](https://doi.org/10.1016/0021-9797(90)90154-G)
4. Hennenberg M, Weyssow B, Slavtchev S, Alexandrov V, Desaive T (2005) Rayleigh–Marangoni– Bénard instability of a ferrofluid layer in a vertical magnetic field. *J Magnetism Magnet Mater* 289:268–271. <https://doi.org/10.1016/j.jmmm.2004.11.076>
5. Mokhtar NFM, Norihan MA (2012) Benard-Marangoni ferroconvection with feedback control. *Int J Modern Phys: Conf Ser* 9:552–559. World Scientific Publishing Company. <https://doi.org/10.1142/S201019451200565X>
6. Prakash J, Kumar R, Kumari K (2017) Thermal convection in a ferromagnetic fluid layer with magnetic field dependent viscosity: a correction applied. *Studia Geotechnica et Mechanica* 39(3):39–46. <https://doi.org/10.1515/sgem-2017-0028>
7. Pradhan G, Samal P (1987) Thermal stability of a fluid layer under variable body forces. *J Math Anal Appl* 122:487–495
8. Alex SM, Patil PR, Venkatakrishnan KS (2001) Variable gravity effects on thermal instability in a porous medium with internal heat source and inclined temperature gradient. *Fluid Dynam Res* 29(1):1. [https://doi.org/10.1016/S0169-5983\(01\)00016-8](https://doi.org/10.1016/S0169-5983(01)00016-8)

9. Ankuj B, Chand R (2015) Variable gravity effect on the thermal instability of Ferrofluid in a Brinkman porous medium. *Int J Astronomy Astrophys Space Sci* 2(5):39
10. Suma SP, Gangadharaiah YH, Indira R (2011) Effect of throughflow and variable gravity field on thermal convection in a porous layer. *Int J Eng Sci Tech* 3:7657–7668
11. Gangadharaiah YH, Suma SP, Ananda K (2013) Variable gravity field and throughflow effects on penetrative convection in a porous layer. *Int J Comput Technol* 5:172–191. <https://doi.org/10.24297/ijct.v5i3.3519>
12. Yadav D (2019) Numerical investigation of the combined impact of variable gravity field and throughflow on the onset of convective motion in a porous medium layer. In: *International communications in heat and mass transfer* 108:104274. <https://doi.org/10.1016/j.icheatmasstransfer.2019.104274>
13. Nagarathnamma H, Gangadharaiah YH, Ananda K (2020) Effects of variable internal heat source and variable gravity field on convection in a porous layer. *Malaya J Matematik* 8:915–919. <https://doi.org/10.26637/MJM0803/0031>
14. Finlayso BA (1970) Convective instability of ferromagnetic fluids. *J Fluid Mech* 40(4):753–767. <https://doi.org/10.1017/S0022112070000423>
15. Degan G, Vasseur P (2003) Influence of anisotropy on convection in porous media with nonuniform thermal gradient. *Int J Heat Mass Transf* 46:781–789. [https://doi.org/10.1016/S0017-9310\(02\)00352-6](https://doi.org/10.1016/S0017-9310(02)00352-6)

Graph-Based Keyword Extraction for Twitter Data



S. Vijaya Shetty, S. Akshay, B. S. Shritej Reddy, Hemant Rakesh, M. Mihir, and Jyothi Shetty

1 Introduction

Keyword extraction is a technique that sorts unstructured text and extracts the most important words in a text. Keywords are representations of the most relevant information present in a document. Natural language processing, text mining, text analysis, information retrieval, search engine optimization make use of keyword extraction techniques to obtain better results. Due to the increased growth in the quantity of text records on the internet and in digital libraries, manual analysis of those documents is no longer feasible. Various methods like TextRank, word2vec, YAKE, HITS have been introduced that handle different kinds of data with high accuracy. Keyword extraction has been a flourishing exploration field for quite a whilst, covering different applications in text mining, text analysis, search engine optimization, information retrieval and natural language processing. Be that as it may, it is anything but an area of exploration that is totally bound together. Regardless of the presence of the numerous methodologies inside the field, there is no standard technique for keyword extraction that has been tried and working on all kinds of data.

Methods for automatic keyword extraction are often supervised as they require a label for each training example, whereas unsupervised learning algorithms use the textual properties from unlabelled training data. Semi-supervised learning algorithms use both labelled and unlabelled data to extract keywords. Unsupervised methods are often further divided into other concentrated study domains and use the individual

S. Vijaya Shetty (✉) · S. Akshay · B. S. Shritej Reddy · H. Rakesh · M. Mihir
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: vijayashetty.s@nmit.ac.in

J. Shetty
NMAM Institute of Technology, Nitte, India
e-mail: jyothi_shetty@nitte.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_68

863

domains or a combination of them as an ensemble. The summarization of the key contribution of this paper is as follows:

- 1 We leverage the topical information in knowledge bases to enhance the performance of keyword extraction.
- 2 We model the keyword extraction as an optimization problem and supply the corresponding solution also as a pruning approach to scale back the complexity.

2 Related Work

Based on a large amount of text data obtained from a microblogging platform in China called “Sina Weibo”, the paper [1] uses a “word2vec” model that extracts semantic features to analyze short text in Chinese. It performs TF-IDF on datasets with emotional tags that result in a word embedding. Weibo datasets are trained in the word2vec model to result in the word embedding which then is used to train classifiers to perform sentiment analysis. Word2vec and TextRank are popular methods when dealing with social media short text to understand the semantic features between words in a selection of text. As shown in Yu et al. [2], word2vec is used to get the relation between the words whilst TextRank makes use of the word frequency, semantic and direction relation to rank the words also shown in Gopan et al. [3] to extract keywords. On the other hand, graph-based methods as shown in Marina and Mark [4], Yu and Vincent [5], Bougouin et al. [6], Page et al. [7], Wang et al. [8] have vertex representation and edge representation for the concept, weight, direction, and label classifications with sub-classifications for each. Graph-based methods for keyword extraction is a developing research topic and paper [9] discusses several methods and approaches providing guidelines for future research and development in the field. In natural language processing and as mentioned in Chatterjee and Goyal [10], Hossny and Mitchell [11], Latent Dirichlet Allocation (LDA) is a statistical model that segregates a set of documents into a set of topics and assigns each word under a certain group by clustering similar words. Each document is represented as a collection of topics and the model attributes each word to a topic which explains the similarity between the words. Various state-of-the-art studies perform keyword detection in a document by making use of an existing collection of keywords as a reference but these existing methods are not feasible for social media data as its corpus is ever increasing which makes unsupervised methods necessary. Various unsupervised methods are used to understand dynamic data by understanding multiple local features as shown in the “YAKE” algorithm implemented in Campos et al [12]. We observe drawbacks in Sarna and Bhatia [13] due to lack of datasets since social media data is dynamic and changing, its dictionaries change as well, and the possibility of overlap of domains. The method involved in Dastanwala and Patel [14] to segment followers and describe a group of high-value social audience members to generate seed words by word frequency and perform fuzzy keyword match, SVM, Twitter LDA creates a comparative analysis. Generation of an associated keyword space with distance measure, uniformization of the distribution in associated keyword

space is the usual workflow. There is another approach as mentioned in Wada et al [15] with incremental clustering to build the ASKS and visualize a temporal variation of keyword frequency and its resulting trends. As mentioned in Chatterjee and Goyal [10], Belkaroui [16], Lee and Kim [17], “TF-IDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus”. It is frequently utilized as a weighting factor in data recovery, text mining and client displaying. The TF-IDF value increments with the word recurrence in the record and is counterbalanced by the quantity of documents in the corpus that contain the word, which helps to adjust for the higher frequency of some words compared to others.

3 Proposed Method

In this section, we describe the dataset, preprocessing and algorithm used in the implementation. Since the objective of this work is to find pervasive patterns emerging across graphs in twitter data, we will make use of twitter data.

3.1 Preprocessing

The data involves tweets pertaining to a certain search query that includes all tweets with retweets filtered that relate to the search query. The data extracted contains a lot of noisy data that do not have any value to the keyword extractions aspect and thus need to be cleaned. The tweets are extracted and formed into sentences using POS tagging which are further tokenized using the spacy tokenizer. We clean these sentences by removing semantic barriers such as stop words, punctuations, emojis, digits and normalizing whitespaces. An initial complete graph with the cleaned text using the first w terms, iterate over the remaining terms and generate a set of candidate edges. For every candidate edge, if it is not a self-edge and has already been seen, we update its weight and if not seen, create it and assign it a unit weight. We now form the weighted and unweighted graphs by adding edges and removing self-loops, generate and order a list of keywords.

3.2 *K-truss Algorithm for Tweets*

The proposed method makes use of an improved truss decomposition algorithm as shown in Table 1, to construct the graph of keywords. Consider V as a finite site of vertices, v or actor and E as a finite set of edges, e or ties/relation between each vertex/edge.

The algorithm begins by initializing variables that compute the support for every edge e in unweighted graph G . The $\text{sup}()$ of an edge, $e = (u, v) \in EG$ in G ,

Table 1 K-Truss Decomposition algorithm [18]

<p>Input: Graph, $G = (V_G, E_G)$</p> <p>Output: the k-class, φ_k, for $2 \leq k \leq k_{max}$</p> <ol style="list-style-type: none"> 1. $k \leftarrow 2, \varphi_k \leftarrow \emptyset$; 2. calculate $\text{sup}(e)$ for each edge $e \in E_G$; 3. sort all of the edges in ascending order of their $\text{sup}()$; 4. while ($\exists e$ such that $\text{sup}(e) \leq (k - 2)$) 5. let $e = (u, v)$ be the edge with lowest support; 6. assume without loss of generality., $\text{deg}(u) < \text{deg}(v)$; 7. for each $w \in \text{nb}(u)$ do 8. if ($(v, w) \in E_G$) 9. $\text{sup}((u, w)) \leftarrow (\text{sup}((u, w)) - 1)$; $\text{sup}((v, w)) \leftarrow (\text{sup}((v, w)) - 1)$; 10. reorder (u, w) and (v, w) according to their new support; 11. $\varphi_k \leftarrow (\varphi_k \cup \{e\})$; 12. remove and edge, e from Graph, G; 13. if (all edges in G are not removed) 14. $k \leftarrow k + 1$; 15. goto Step 4; 16. return φ_j, for $2 \leq j \leq k$;

denoted by $\text{sup}(e, G)$, is defined as $|\{\Delta uvw : \Delta uvw \in \Delta G\}|$. Support of an edge is determined by the number of triangles in graph that contain that edge. The edges are then sorted in ascending order of their computed support which are then stored in an array A . On completion of initial steps, the algorithm proceeds to iteratively remove the lowest support edge, i.e. the edge at the beginning of the sorted array. These steps are repeated until support of the edge is not less than $k - 2$ with each iteration starting at $k = 2$.

The edge e , which is removed, is added to φ_k (the k -class of G) since $\text{sup}(e) \leq (k - 2)$ and therefore cannot be in φ_{k+1} . Once the edge is removed, the support value of all edges that form a triangle with that edge is decremented as well which will result in the updation of positions in the sorted array. At step 8, we test whether $(v, w) \in E_G$, by efficiently keeping E_G in a hash table where the addresses of the

edges are stored. The updates in the arrays are completed in constant time similar to the methods in Batagelj and Zaversnik [19]. The pointer positions are then updated by increment in order for it to point to the consecutive edge with the lowest support value instead of explicitly removing the edge from the graph. This process is repeated until the iterations and conditions mentioned above are satisfied.

Using the K -truss decomposition algorithm, we find k -connected components [20] of a graph. The k -connected components are generated from the list of keywords which are ranked and used to generate subgraphs. These subgraphs are the result of maximal trusses depending on the value of k , the truss size. Since the degree of a vertex is the number of edges incident on it, multigraph shows loops counted twice as a result which must be removed in the process of retrieving the components. The neighbourhood of a vertex v in a graph G is the subgraph of G induced by all vertices adjacent to v .

4 Experiments

This section showcases the experimental results of the proposed method using the tweets extracted from the Twitter API based on a search query. The extracted tweets are related to the topic mentioned in the search query and contain tweet data that would give a brief overview of the topic as seen on social media. The dataset varies in word size depending on the number of tweets there are available for the given search query. The retrieved tweets are cleaned, pre-processed, and used to create weighted and unweighted graphs to produce subgraphs which would then generate the keyword list.

As shown in Table 2, each topic has a set of 7 top-ranked keywords that are extracted from the tweets using a combination of the subgraphs and scores generated by the closeness centrality measure. The initial graph construction contains a high number of candidate words and extracting the essential words that are of relevance to the tweets involves understanding how important other nodes are to a certain node in the graph.

Table 2 Topics and their respective top-ranked keywords

Black lives matter	Oneplus	Lockdown extension	Wonder woman
Black	Oneplus	Extension	Wonder
Lives	Nord	Lockdown	Woman
Matter	Price	Covid	Like
Holding	Phone	Cases	DC
People	PeteLau	July	Spectre
Powerful	Like	Government	GalGadot
Support	Compatible	Economy	New

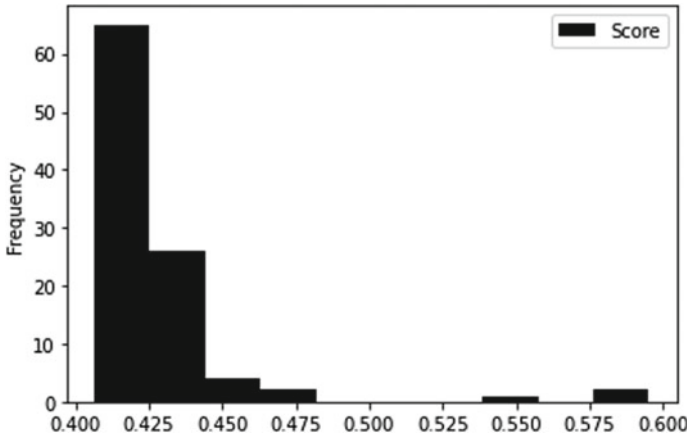


Fig. 1 Frequency distribution of closeness centrality of the top 100 keywords

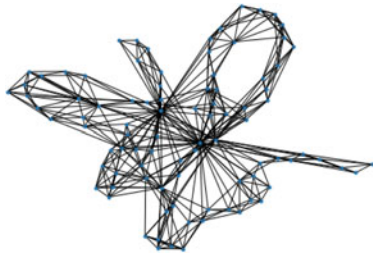
Most of the words have relatively low importance when compared to the keywords as shown in Fig. 1 which is a frequency distribution plot of the top 100 candidate words and their centrality scores. There are a very low number of nodes that have a high level of importance to the graph and thus would come to be the keywords.

5 Results and Discussion

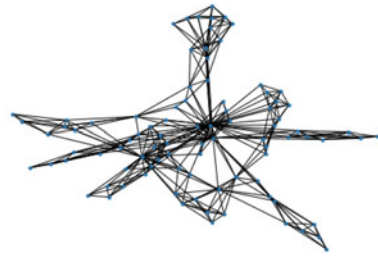
The results of the proposed method are graphs constructed from the tweets that can be reduced to get the *k*-core subgraphs. These subgraphs contain a set of words represented by nodes and the relations between the words represented as edges. The graphs and subgraphs for the tweets generated by the search query “oneplus nord” are shown below.

Figure 2a and b represents weighted and unweighted graphs for a collection of tweets by search query “oneplus nord”. Passing this list of tweets through the *K*-truss algorithm for keyword ranking generates subgraphs as shown below in Fig. 3. Each of the subgraphs is dependent on the sorted edges list and until any more edges cannot be removed. The subgraphs represent how keywords are related in the context of the tweets.

Figure 3a–d describes the subsequent subgraphs for the search query “oneplus nord”. The subgraphs decrease in density with each pass, whilst filtering keywords that are ranked lower in the dictionary used to build the graphs. It can be seen in the subgraphs that reducing the self-loops and low-ranked edges make the graph more relevant to the topic and give a better understanding of the way the keywords are related. From Fig. 3c and d, it is easy to conclude that certain keywords have more in common with the topic compared to others.

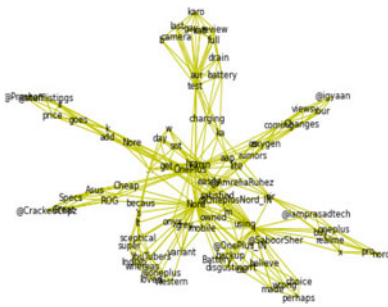


(a) Weighted graph of tweet search



(b) Unweighted graph of tweet search

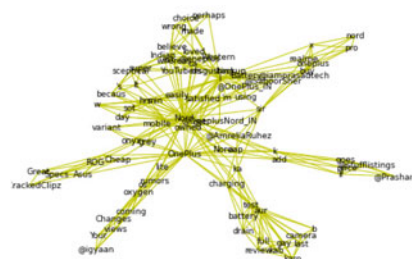
Fig. 2 Initial constructed graphs for tweet search



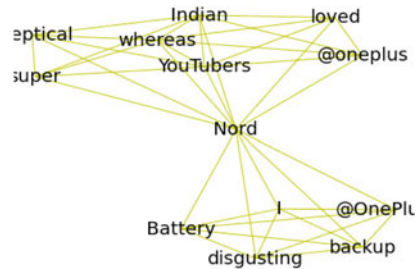
(a) Subgraph 1



(b) Subgraph 2



(c) Subgraph 3



(d) Subgraph 4

Fig. 3 Subgraphs for search query “oneplus nord”

The analysis of accuracy of keyword extraction can be done by calculating precision, recall and F-score values. This measure is evaluated by using a gold standard set of keywords for the tweets as true values and the extracted keywords as predicted values is shown in Table 3.

Table 3 F-1 score for N = 15

Method	F1—score
TF-IDF	47.8
TextRank	53.3
K-Truss decomposition	55.45

6 Conclusion

The proposed system is an alternative to other existing keyword extraction works and extracts keywords from social media data to understand the context of the data. By doing so, the resulting system will be able to improve artificial intelligence systems, do fraud detection better, and more. Through the survey, it is clear that better-performing text mining systems are essential for better AI and ML products and we have also understood the current trends and level of NLP technologies. This paper showcases keyword extraction from social media using a graph-based method called k-core truss decomposition on data extracted using the twitter API. Graph-based methods extract keywords in the form of pairs which can be represented in a graph to define how those keywords are related to each other in the context of the topic. Furthermore, centrality of a keyword graph also helps us understand what keywords contribute more to the topic and separate them from the less significant candidates that do not contribute to the overall understanding of the tweet data.

References

1. Shi B, Zhao J, Xu K (2019) A Word2vec model for sentiment analysis of weibo. In: 2019 16th international conference on service systems and service management (ICSSSM), pp 1–6. <https://doi.org/10.1109/ICSSSM.2019.8887652>
2. Yu Z, et al (2019) Hot event detection for social media based on keyword semantic information. In: 2019 IEEE fourth international conference on data science in cyberspace (DSC). pp 410–415. <https://doi.org/10.1109/DSC.2019.00068>
3. Gopan E, et al (2020) Comparative study on different approaches in keyword extraction. In: 2020 fourth international conference on computing methodologies and communication (ICCMC). <https://doi.org/10.1109/iccmc48092.2020.iccmc-00013>
4. Marina L, Mark L (2008) Graph-based keyword extraction for single document summarization. In: Proceedings of the workshop on multi-source multilingual information extraction and summarization. Association for Computational Linguistics, pp 17–24
5. Yu Y, Vincent N (2018) Wikirank: improving keyphrase extraction based on background knowledge. arXiv preprint [arXiv:1803.09000](https://arxiv.org/abs/1803.09000)
6. Bougouin A, Florian B, Béatrice D (2013) Topicrank: graph-based topic ranking for keyphrase extraction
7. Page L, et al. (1999) The page rank citation ranking: bringing order to the web. Stanford InfoLab
8. Wang R, Wei L, Chris MD (2014) Corpus-independent generic keyphrase extraction using word embedding vectors. *Softw Eng Res Conf* 39
9. Slobodan B, Ana M, Sanda M (2015) An overview of graph-based keyword extraction methods and approaches. *J Inf Organ Sci* 39(1):1–20

10. Chatterjee R, Goyal M (2015) Tactics of twitter data extraction for opinion mining. In: 2015 2nd international conference on computing for sustainable global development (INDIA Com), pp 761–766
11. Hossny AH, Mitchell L (2018) Event detection in twitter: a keyword volume approach. In: 2018 IEEE international conference on data mining workshops (ICDMW), pp 1200–1208 <https://doi.org/10.1109/ICDMW.2018.00172>
12. Campos R et al (2020) YAKE! keyword extraction from single documents using multiple local features. *Inf Sci* 509:257–289. <https://doi.org/10.1016/j.ins.2019.09.013>
13. Sarna G, Bhatia MPS (2016) A probabilistic approach to automatically extract new words from social media. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 719–725. <https://doi.org/10.1109/ASONAM.2016.7752316>
14. Dastanwala PB, Patel V (2016) A review on social audience identification on twitter using text mining methods. In: 2016 international conference on wireless communications, signal processing and networking (WiSPNET), pp 1917–1920. <https://doi.org/10.1109/WiSPNET.2016.7566476>
15. Wada S et al (2013) Associated keyword analysis for temporal data with spatial visualization. In: 2013 international joint conference on awareness science and technology Ubi-Media computing (iCAST 2013 UMEDIA 2013), pp 243–249. <https://doi.org/10.1109/ICAwST.2013.6765441>
16. Belkaroui R, Faiz R, Elkhelifi A (2014) Conversation analysis on social networking sites. In: 2014 tenth international conference on signal-image technology and internet-based systems, pp 172–178. <https://doi.org/10.1109/SITIS.2014.80>
17. Lee S, Kim H (2008) News keyword extraction for topic tracking. In: 2008 fourth international conference on networked computing and advanced information management, vol 2. pp 554–559. <https://doi.org/10.1109/NCM.2008.199>
18. Wang J, James C (2012) Truss decomposition in massive networks. arXiv preprint [arXiv:1205.6693](https://arxiv.org/abs/1205.6693)
19. Batagelj V, Zaversnik M (2003) An $o(m)$ algorithm for cores decomposition of networks. *CoRR*, cs.DS/0310049
20. Matula DW (2020) k -Components, Clusters, and Slicings in Graphs. *SIAM Journal on Applied Mathematics*, vol 22, no 3, 1972. JSTOR, pp 459–480. www.jstor.org/stable/2099755. Accessed 11 July 2020

DSEORA: Integration of Deep Learning and Metaheuristics for Web Page Recommendation Based on Search Engine Optimization Ranking



Ketan Vaish, Gerard Deepak, and A. Santhanavijayan

1 Introduction

With hundreds of millions of active URLs on the World Wide Web (WWW), the process of information retrieval (IR) has become more crucial than ever. Hence, such a large amount of information on the internet makes it difficult for the user to retrieve the relevant information. Due to the low precision of the results produced by search engines, optimization techniques are in the discussion for a long time. An information retrieval (IR) system captures all the URLs which are relevant to the user's query term and then ranks these URLs based on certain important measures. To make IR experience more personalized, several approaches have been put forward [1–3].

Motivation: In this study, a novel deep learning approach for web search incorporating Search Engine Optimization Ranking has been discussed. A plethora of information available on the internet is unorganized and unlabeled but for training the model, a large corpus of the labelled dataset is obtained from the web. The techniques incorporated in this study are LSTM based classification using WWW index terms, the keywords extracted from webpages, and Ant Colony Optimization technique for ranking the most relevant webpages using Jaccard's index for measuring the similarity between a webpage and user query.

Contribution:

This proposed approach is devised to improve the relevance of the web search by the users. The proposed architecture yields a very reliable and accurate deep learning inspired classification-based model. The ranking algorithm incorporated

K. Vaish (✉) · G. Deepak

Department of Information Science and Engineering, RV College of Engineering, Bengaluru, India

A. Santhanavijayan

National Institute of Technology, Tiruchirappalli, Tamil Nadu, India

in this methodology is one of the most efficient and effective algorithms capable of finding solutions to very complex and difficult problems, making this method even more accurate and precise. The integration of Deep Learning with Ant Colony Optimization has yielded satisfactory results. Experimental evaluations confirm the superiority of proposed methods over cluster-based methods.

Organization: Related Works are depicted in Sect. 2. Section 3 gives a brief introduction to incorporated techniques of this proposed method. Section 4 provides a detailed working of the proposed Deep SEO Ranking Algorithm (DSEORA). In Sect. 5, results and performance evaluation are discussed and the conclusions are given in Sect. 6.

2 Related Work

Harish Kumar et al. [1] have put forward an alternative to K-means clustering. They have used Group Average Agglomerative Hierarchical Clustering (GAAHC) incorporating Modified Levenshtein Distance (MLD) and ranks the page using a larger amount of time and the frequently visited page for IR. Suruchi Chawla proposed a method of information retrieval based on clustering with optimal ranked clicked URL with the help of a Genetic Algorithm Personalized Web Search technique (GAPWS) [2]. Information scent metrics and K-means clustering method have been used for ranking of webpages. Debajyoti Mukhopadhyay et al. [4], have used the syntactical classification technique for grouping. In their method, some properties of the web pages and documents are selected based on user's demand, measured and assigned some weights when ranking is done. Then, in syntactic classification, on a set of webpages, they carried out fuzzy c-means clustering algorithm and classification using neural network. But, in Syntactic Classification, the meaning of the webpage's text, pictures, audios or videos is not taken into account. Liang Pang et al. [5], have proposed a strategy for detection and extracting the relevant contents and building a measuring network by utilizing CNN or 2-D GRU. A global relevance score is produced for final ranking. Zhenzhen Li et al [6], have given a deep learning model based on LSTM RNN for answer ranking given the condition that questions, as well as answers, should be limited to one sentence only. Lian-Wang Lee et al. [7], have put forward an IR method involving user query, in which a set of training queries similar to a testing query is determined then ranking is done with help of a weighing mechanism. Several semantic-based ontologies or ontology application are discussed in [8–20], which yield confidence in approaches of a similar problem class.

In the proposed method, a labelled dataset trains the LSTM model, which is later used for label prediction and Ant Colony Optimization algorithm involves artificial ants which work like natural ants and find the optimum solution for difficult problems by their simple natural process of pheromone release and attracting other ants to the same path. This gives ACO an advantage over genetic algorithms.

3 Background

3.1 Long Short-Term Memory (LSTM)

Recurrent Neural Network (RNN) has proved to be highly significant in various domains such as sentence prediction, speech recognition, etc. RNNs face the problem of long dependencies and LSTMs is a solution to this problem. There are mainly three gates viz. input (i_t), forget (f_t) and output (o_t), given by Eqs. (2), (3) and (4) respectively in the LSTM architecture [21], vector c denotes the cell memory unit, $a^{(t)}$ is a hidden vector, which passes the information of previous time steps at any time step t . W and b display network parameters for respective processes. Let the input sequence be denoted by x , where $x = [x(1), x(2), \dots x(t)]$, then following Equations wrap the working of LSTMs.

$$\tilde{c}^{(t-1)} = \tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c) \quad (1)$$

$$i_t = \sigma(W_i[a^{(t-1)}, x^{(t)}] + b_i) \quad (2)$$

$$f_t = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f) \quad (3)$$

$$o_t = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o) \quad (4)$$

$$c^{(t)} = i_t * \tilde{c}^{(t-1)} + f_t * c^{(t-1)} \quad (5)$$

$$a^{(t)} = o_t * \tanh c^{(t)} \quad (6)$$

σ denotes sigmoid function

It is evident from the Eqs. (2), (3) and (4) that gates control the cell memory state. Equation (1) represents old cell memory which is updated using gates into new cell state as depicted in Eq. (5). Input gate decides which input sequence should alter the state to what extent. Forget gate allows the network to forget unnecessary data while output gate controls the hidden vector a , which in turn controls next time step calculations as shown in Eq. (6). In this study, there are four different sources of input data for more accurate results and hence the classification part is done using LSTM.

3.2 Ant Colony Optimization (ACO) Algorithm

ACO algorithm is a heuristic method applied for optimization of a wide range of difficult problems [5]. Stigmergy is the main inspiration of the ACO algorithm. Stigmergy refers to how the organisms in nature interact and coordinate with one another by temporarily modifying the environmental factors around them. In ACO, artificial ants search for an optimal solution. Ants release a pheromone in their pathway for other ants to follow up. Similarly, we use a weighted graph of required nodes and pheromone model for other ants to walk through nodes and update these parameters.

Mathematical Model: Consider two nodes (i, j) on the graph. Let τ denote level of pheromone and k be the k th ant. L denotes the length of the path. Then:

$$\Delta\tau_{i,j}^k = \left\{ \frac{1}{L_k}; \text{ when } k\text{th ant travels on path between nodes } i, j \right. \quad (7)$$

$$\Delta\tau_{i,j}^k = 0; \text{ when ants do not move} \quad (8)$$

$\Delta\tau_{i,j}^k$ denotes the pheromone amount released by the k th ant travelling on the path between nodes i, j as given by Eqs. (7) and (8).

Thus, the net level of pheromone present between nodes i, j after k th ant travels on it, is given by Eq. (9):

$$\tau_{i,j} = (1 - \rho) * \tau_{i,j} + \sum_{k=0}^m \Delta\tau_{i,j}^k \quad (9)$$

where ρ denotes evaporation constant

Thus, the probability of any ant choosing path i, j is given by Eq. (10):

$$P_{i,j} = \frac{(\tau_{i,j})^\alpha (\eta_{i,j})^\beta}{\sum (\tau_{i,j})^\alpha (\eta_{i,j})^\beta} \quad (10)$$

where $\eta_{i,j}$ is defined as $\frac{1}{L_{i,j}}$ and parameters α and β manipulate influence of $\tau_{i,j}$ and $\eta_{i,j}$ respectively, on $P_{i,j}$.

Thus, the optimization takes place according to the size of the ant population involved. The next section discusses the proposed work.

4 Proposed Work

In this paper, we propose a technique of finding the most relevant URL set and most relevant URL using Recurrent Neural Network Long-Short Term Memory (RNN LSTM) for effective personalized web search. This method requires an input of

data from four different sources which makes it more versatile. The standard data sources include user's web usage data, the index words harnessed from the WWW, the Query Subject Ontology, and the Knowledge Bases. At the initial stage, data is extracted from these data sources for further processing. The dataset used for this study comprises of web usage data from different domains along with labels. Therefore, as the URLs are fetched from the dataset, we tend to select frequently clicked URLs for a better result from this method. The second source is index words, which are extracted from the WWW using the search query entered by the user. The other two sources i.e. Knowledge Base and Query Subject Ontology, are already present and easily accessible for use.

4.1 Data Preparation

Data preparation involves pre-processing of data before feeding to the algorithm. Pre-processing is implemented using existing techniques for preparing the data. The primary step of preparing and pre-processing the data include the Canonicalization of URLs, to a common URL format. Since a large set of URLs is extracted from the dataset, there is a chance that two different URLs for the same page might be present. If the algorithm finds the same content on two different URLs then it may get confused in the process of ranking the two URLs. In such a case, the best URL is chosen. This is followed by Tokenization to break down the query text into tokens. Tokenization is achieved using a `WhitespaceTokenizer` from natural language toolkit python library.

Followed by Tokenization is Lemmatization, which realizes the morphological analysis of the words. The main aim is to remove inflections from the endings of the words and get the dictionary form of it, which is termed as the lemma. In this study, the technique is applied to the index words from the web and to user's search query. This allows user to use any of the inflectional forms of a word in the query and get relevant results. The `LemmaProcessor`, a class of lemmatizing processor offered by Stanford coreNLP, has been used. Figure 1 shows the architecture of the methodology proposed in the paper. It gives a sequential understanding of the flow of the data during classification as well as ranking.

4.2 Classification of URLs Using LSTM

This is the phase I of the proposed methodology and is carried out offline. Classification is a supervised machine learning technique that works on the basic principle of, first, training the model with the labelled dataset and then using test dataset for predicting the labels and measuring the performance of the trained model. Thus, in this study, the core idea of classifying the URLs revolves around two steps. First, the LSTM network is trained by feeding it with a labelled URL set, which has its

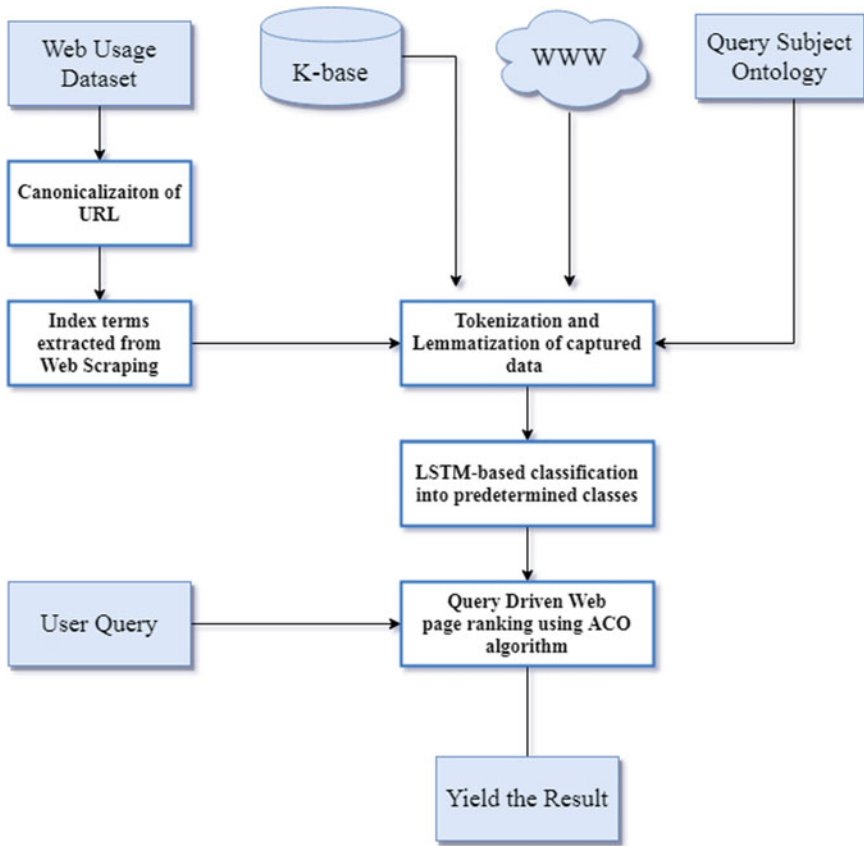


Fig. 1 Proposed system architecture

respective categories labelled on them. After this, a trained model is obtained which can be used to fetch the most relevant URL from classified URL set, for a specific user query. This is done using Ant Colony Optimization Algorithm. To proceed with the classification process, a labelled dataset and the index terms are obtained, which are extracted from the URLs. Meta tags also provide some information regarding the content as well as the very purpose of a web page. The model developed in this paper uses meta tags for classifying webpages. The metadata, tags, title, description, keywords, et cetera are extracted, which becomes key information for mapping the URLs to respective categories. For instance, there be three URLs: url_1, url_2 and url_3, the data extracted from the webpages of respective URLs be: text_1, text_2 and text_3 and the labels attached to these URLs be label_1, label_2 and label_3, then at the first step, all the above data is fed to the network where the text data or index terms extracted from the webpages are mapped with the labels already provided, ultimately mapping URLs with their labels. Thus, in the above process, the LSTM network learns to classify the URLs to predetermined classes based on the index

terms extracted from their respective webpages. An LSTM based URL classifier is ready.

4.3 Ranking Relevant URL Using Ant Colony Optimization Algorithm

Query driven online ranking and yielding of the most relevant webpage with the help of classified model prepared by an offline process (LSTM-based) and using Ant Colony Optimization Algorithm is phase II of the proposed methodology. Here, artificial ants are used for finding the most relevant webpage in accordance with the user query. Based on the similarity between query and webpage contents, artificial ants release pheromone for other artificial ants to follow and release more pheromone. Thus, this mechanism ensures that the pheromone level for the most relevant webpage is highest. Deciding the amount of pheromone to be released is the most crucial task of the algorithm. Jaccard's similarity index is used to measure the similarity between the webpage and user query. This is done with the help of the extracted keywords from different sources. If A and B be sets of keywords, then Jaccard's similarity index is given by Eq. (11).

$$J(A, B) = \frac{(A \cap B)}{(A \cup B)} \quad (11)$$

5 Implementation, Results and Performance Evaluation

5.1 Dataset

The dataset used for this study is an open-source Website Classification Dataset from UK Selective Web Archive. UKWA claims that their entire selective archive has been collected and managed manually. They have classified the URLs into a two-tiered hierarchical system and provided labels on them. It provides Primary category, Secondary category, Title and URL. For each URL, titles of every page and a set of keywords that summarizes the site is obtained. These keywords are used to find the similarity index when the ranking of the webpages, in accordance with the user query, is done.

Table 1 Proposed ACO driven DSEORA algorithm for ranking web pages

```

begin
  Initialize parameters
  while the most relevant page is not found do
    each ant is positioned at the starting node
    repeat
      for each ant do
        Choose the next point on basis of Jaccard's similarity index calculated.
        Update local pheromone level
      end for
    until each ant finds a solution for one trip
    Update the optimal and most relevant solution
    Update pheromone level for every path
  end while
end

```

5.2 Algorithm for Webpage Ranking

ACO algorithm has been used for webpage ranking, as discussed in sub-Sect. 3.2. In the first step, parameters are initialized and then ants are positioned at the starting node. Ants choose their next node based on Jaccard's similarity index between the user query and the index terms extracted from the webpage. The path with highest amount of pheromone is yielded after no node is left unvisited. The ACO driven DSEORA Algorithm for Ranking Web Pages is given in Table 1.

5.3 Results and Performance Evaluation

The proposed method has been successfully implemented in Google Collab environment on a system with Intel Core i7 9th generation processor and 16 GB of RAM. The URL corpus was taken from UKWA. Performance metrics of the proposed method includes percentages of Precision, Recall, F-measure, Accuracy and the False Discovery Rate (FDR). This study also compares the results of the proposed method, Deep Search Engine Optimization Ranking Algorithm, with those of three cluster-based information retrieval methods, namely GAPWS, GAAHC_MLD and SLAHC_MLD.

Table 2 shows the performance metrics of the proposed technique on 5 arbitrary queries. The average percentages of Precision, Recall, F-measure and Accuracy were found to be 88.69, 85.10, 86.86, and 86.90% respectively. A mean FDR that was achieved was 0.116. The average accuracy of DSEORA (proposed work) was found to be higher than GAPWS, GAAHC_MLD and SLAHC_MLD by 13.93, 9.37 and 12.08% respectively. Results justify the fact that clustering does not work well if initial partitions are misjudged. In classification, we take advantage of known

knowledge to solve unknown problems. Therefore, classification is a better choice for categorization when known topics or labelled dataset is available. The graph in Fig. 2 shows a comparison of Accuracy in the percentage of four different methods of information retrieval. It is evident from the graphs in Figs. 2 and 3 that classification-based proposed methodology produces, by far, better results than other methods. The relevance of URLs is not assured in the cluster-based grouping as it is unsupervised learning and hence, the classification-based grouping is done, which matches the keywords extracted from different sources, in the proposed method. False Discovery Rate (FDR) is also compared. Figure 3 shows a comparison of FDR among the methods.

Table 2 Performance measures for DSEORA

Query	Precision%	Recall%	F-Measure %	Accuracy %	FDR
Q1	89.41	86.17	87.76	87.79	0.11
Q2	87.62	83.25	85.38	85.44	0.13
Q3	88.12	85.73	86.91	86.93	0.12
Q4	92.14	88.18	90.12	90.16	0.08
Q5	86.17	82.17	84.12	84.17	0.14
Average	88.69	85.10	86.86	86.90	0.116

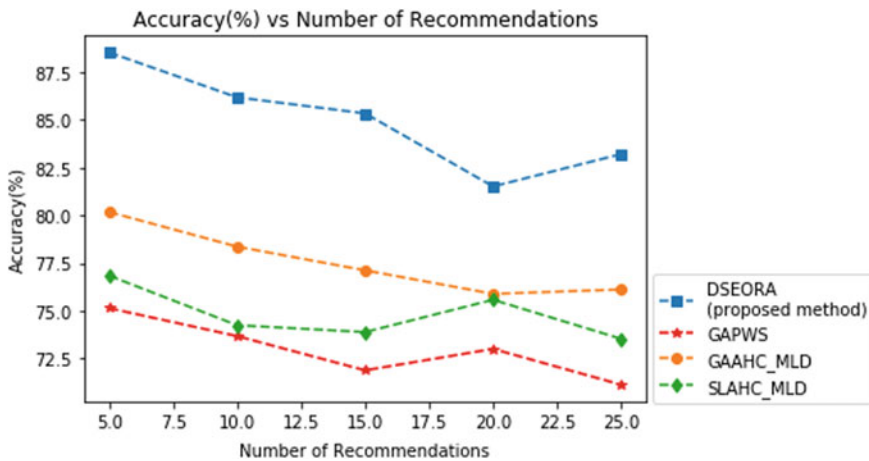


Fig. 2 Accuracy comparison of DSEORA among several approaches

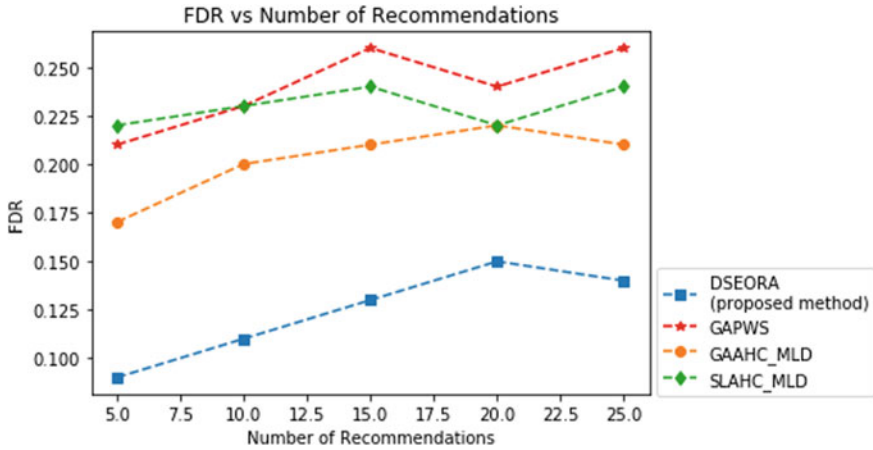


Fig. 3 FDR comparison of DSEORA among several approaches

6 Conclusions

The proposed method, Deep Search Engine Optimization Ranking Algorithm (DSEORA), is evidently more precise and accurate than other approaches. The techniques used in the study were LSTM and ACO. A large corpus of URLs is classified into pre-determined classes using RNN LSTM. This ensures that each webpage of a website is classified correctly by matching the index terms extracted from webpages and the labels provided by UKWA. A trained model is achieved after the classification process. In the second phase, ACO is used for picking up the most relevant URL in response to the user query. Experimental results show an average accuracy and f-measure of 86.90 and 86.86% respectively, which is higher than all the clustering-based approaches.

References

1. Harish Kumar B, Vibha L, Venugopal KR (2018) Personalized recommendation of web pages using group average agglomerative hierarchical clustering (GAAHC)
2. Chawla S (2016) A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective personalized web search
3. Harish Kumar BT, Vibha L, Venugopal KR (2016) Web page access prediction using pal K R, web page access prediction using hierarchical clustering based on modified levenshtein distance and higher order markov model. In: IEEE International conference organized by IEEE region 10 symposium (TENSYPMP-2016), Bali, Indonesia on 9th to 11th May 2016
4. Mukhopadhyay D, Biswas P, Kim Y-C (2011) A syntactic classification based web page ranking algorithm
5. Pang L, Lan Y, Guo J, Xu J, Xu J, Cheng X (2017) DeepRank: A new deep architecture for relevance ranking in information retrieval

6. Li Z, Huang J, Zhou Z, Zhang H, Chang S, Huang Z (2016) LSTM-based deep learning models for answer ranking
7. Lee L-W, Jiang J-Y, Wu CD, Lee S-J (2009) A query-dependent ranking approach for search engines
8. Gerard D et al. (2019) OntoQuest: an ontological strategy for automatic question generation for e-assessment using static and dynamic knowledge. In: 2019 fifteenth international conference on information processing (ICINPRO). IEEE
9. Deepak G, Santhanavijayan A (2020) Onto best fit: a best-fit occurrence estimation strategy for RDF driven faceted semantic search. *Comput Commun*
10. Deepak G, Shwetha BN, Pushpa CN, Thriveni J, Venugopal KR (2018) A hybridised semantic trust-based framework for personalised web page recommendation. *Int J Comput Appl* 1–11
11. Deepak G, Kumar AA, Santhanavijayan A, Prakash N (2019, November). Design and evaluation of conceptual ontologies for electrochemistry as a domain. In: 2019 IEEE international WIE conference on electrical and computer engineering, pp 1–4
12. Santhanavijayan A, Naresh Kumar D, Deepak G (2019) A novel hybridized strategy for machine translation of indian languages. *Soft Comput Signal Process* 363
13. Santhanavijayan A, Naresh Kumar D, Deepak G (2021) A semantic-aware strategy for automatic speech recognition incorporating deep learning models. In: *Intelligent system design*. Springer, Singapore, pp 247–254
14. Deepak G, Priyadarshini JS (2018) A hybrid semantic algorithm for web image retrieval incorporating ontology classification and user-driven query expansion. In: *Advances in big data and cloud computing*, pp 41
15. Sairam H et al. (2019) A novel approach for ontology focused inter-domain personalized search based on semantic set expansion. In: 2019 fifteenth international conference on information processing (ICINPRO). IEEE
16. Deepak G, Gulzar Z (2017) Ontoepds: enhanced and personalised differential semantic algorithm incorporating ontology driven query enrichment. *J Adv Res Dynam Control Syst* 9(Specia):567–582
17. Deepak G, Teja V, Santhanavijayan A (2020) A novel firefly driven scheme for resume parsing and matching based on entity linking paradigm. *J Discrete Mathemat Sci Cryptograp* 23(1):157–165
18. Kumar N, Deepak G, Santhanavijayan A (2020) A novel semantic approach for intelligent response generation using emotion detection incorporating NPMI measure. *Proc Comput Sci* 167:571–579
19. Deepak G, Priyadarshini S (2016) A hybrid framework for social tag recommendation using context driven social information. *Int J Soc Comput Cyber-Phys Syst* 1(4):312–325
20. Deepak G, Kumar N, Santhanavijayan A (2020) A Semantic approach for entity linking by diverse knowledge integration incorporating role-based chunking. *Proc Comput Sci* 167:737–746
21. Kaushik IS, Deepak G, Santhanavijayan A (2020) QuantQueryEXP: a novel strategic approach for query expansion based on quantum computing principles. *J Discrete Mathem Sci Cryptograp* 23(2):573–584

Traffic Symbol Detection and Recognition System



K. Md. Zakir Hussain, Komal Nagaraj Kattigenahally, S. Nikitha, Pabitra Priyadarshini Jena, and Y. Harshalatha

1 Introduction

Developing intelligent vehicles for better and safer driving practices with maximum efficiency is the necessity for the rapidly growing generation in this era. The technological development of Advanced Driver-Assistance System (ADAS) provides a framework to build smarter vehicles that can make driving much more comfortable and easier. Traffic Symbol Recognition system (TSDR) is one of the driver support systems that automatically detects traffic sign boards on roads and assists the driver recognizing it in advance.

In recent years, there has been an increase in the number of road accidents. To avoid road disasters, following the traffic signs and abiding by traffic rules is necessary. The road hazards can be prevented with the help of TSDR. It is a system designed to operate in a real-time environment that aims at enhanced safety and assists the driver in controlling and performing scrutinized operations. It can also be implemented as a part of unmanned vehicles that are yet to come up, for self-regulatory controls while in motion.

The traffic sign boards are a vital part of the roadways for the safe movement of the vehicle users and pedestrians. As per the Indian Road Congress (IRC), the road signs are categorized into three types: (i) Mandatory / Regulatory Signs (Fig. 1), (ii) Cautionary/Warning Signs (Fig. 2), and (iii) Informatory Signs (Fig. 3). Mandatory Signs are obligatory signs that indicate what one must do, rather than must not do. These are generally round in shape with a red border (there are exceptions). Violation of these signs is a punishable offense. Cautionary signs warn the driver of the potential danger on the road ahead so that the driver can take necessary action. Usually, these signs are triangular in shape with a red border. Informatory signs are

K. Md. Zakir Hussain · K. N. Kattigenahally · S. Nikitha · P. P. Jena · Y. Harshalatha (✉)
Siddaganga Institute of Technology, Tumakuru, Karnataka 572103, India
e-mail: latha_rupesh2002@sit.ac.in

Fig. 1 Mandatory signs: U turn prohibited and no parking sign



Fig. 2 Cautionary sign: hump ahead, men at work and left hand curve

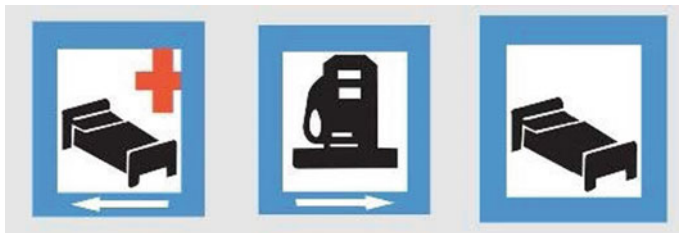


Fig. 3 Informatory signs: hospital, petrol pump and resting place

meant for providing information to the driver on direction, destination, and roadside facilities. These signs help in saving time, and also make it easy to reach destinations. These are known as facilitators to the driver.

Traffic sign detection involves color-based segmentation and shape-based detection, followed by recognition system which involves classification of traffic signs using different classifiers. In [1], for sign detection, region of interest (ROI) is extracted using a color segmentation algorithm. The color space used for the process is RGB. In [2], the author uses RGB color space for color-based segmentation, and results obtained are 98.25% accurate. Thresholding method is used to get binary image and Hough Transform for shape and contour detection. In [3], HSV color space is used and Douglas-Peucker algorithm and border following algorithm are used for shape and contour detection, respectively. Feature extraction is performed using HOG (Histogram Oriented Gradient) and SVM classifier and convolution neural networks (CNN) are used for classification and recognition of the road signs. Two color spaces are compared-YUV and RGB, and it was concluded that RGB produces better results [4]. A neural network-LWN++ is used as classifier. In [5], the author has proposed

an improved version of object-based detection system called You Only Look Once (YOLO). YOLO is based on CNN, produces more accurate and faster results than Fast R-CNN method. In this paper, YOLOv2 is used for detection and recognition of traffic symbols. The detection and recognition system is explained in Section II. Distance calculation to acquire the traffic symbol is explained in Section III. Implementation details are discussed in Section IV. Results are discussed in Section V followed by conclusion in Section VI.

2 TSDR System

An automatic TSDR system consists of image acquisition, traffic signboard detection, and recognition of detected signboard as shown in Fig. 4. The algorithm used for implementation of traffic symbol detection and recognition is YOLOv2 that performs both detection and recognition. After the image acquisition, YOLOv2 divides the entire image into, say $M \times M$ grid, where each grid cell predicts only one object and also predicts “n” boundary boxes. Each boundary box has a confidence score which contains information regarding the likeliness of an object being present in that boundary box and the accuracy of the boundary box. Each grid also uses the concept of conditional class probability which tells about the class it belongs to. The root concept used in YOLOv2 is Convolutional Neural Networks (CNN) and linear regression of boundary boxes. Finally toward the end all the boundary boxes with very less confidence scores are removed and only the ones with objects in them remain.

Working of YOLO Algorithm is explained as follows:

Entire image is divided into $S \times S$ grid cells depending upon the size of the object being detected. For each grid cell, the label Y will be a $(5 + N)D$ vector. YOLO predicts the bounding boxes and their corresponding class probabilities for objects. After the image is divided into $S \times S$ grids, the target vector (Y) is calculated for each grid as shown in Fig. 5. In Fig. 6, the vector Y is shown for 3 class problem where, P_C defines whether the object is present in the grid or not. bx , by , bh and bw specify the bounding box if there is an object. These will be calculated relative to the corresponding grid cell. bh and bw are obtained by normalizing the bounding box height h and width w by the image height and width. x and y are offsets to the corresponding cell. Hence, x , y , h and w are all between 0 and 1. c_1 , c_2 , c_3 represent the classes. The value of class probabilities will be high if the object belonging to that class is present in a particular grid. The value of S depends on the size of



Fig. 4 Traffic symbol detection and recognition system

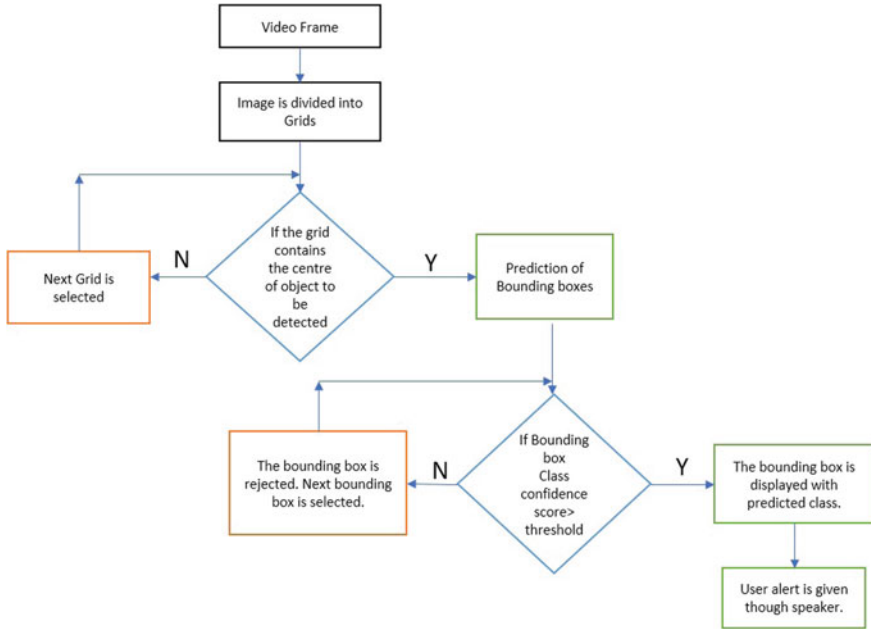


Fig. 5 Detection and recognition of sign board

Fig. 6 Label vector

Y=	P _c
	bx
	by
	bh
	bw
	c1
	c2
	c3

object being detected. The image dataset and their corresponding target vector are fed into convolution neural network (CNN) (Fig. 7) and the model is trained using both forward and backward propagation.

In general for a $\times S$ grid image and N class, the label vector $\text{dim} = S \times S \times (5 + N)$.

Intersection over Union (IoU) & Non-Max Suppression:

IoU:

$$\text{IoU} = \frac{\text{Areaoftheintersection}}{\text{Areaoftheunion}}, \text{ Greater IoU implies good localization of object.}$$

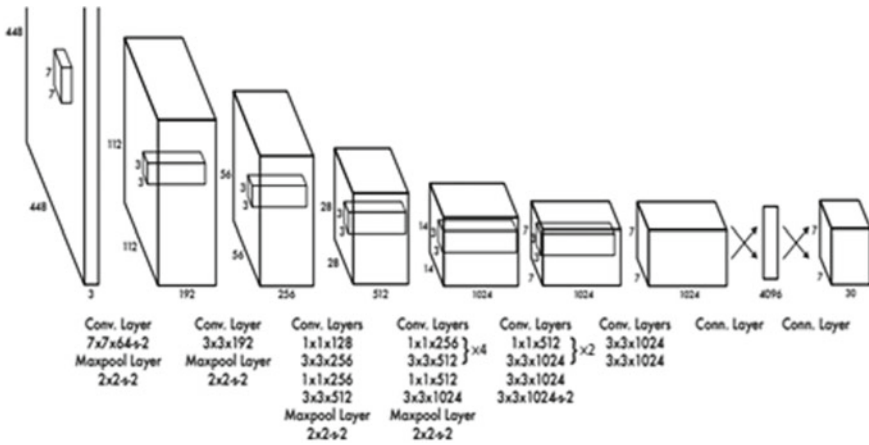


Fig. 7 YOLO architecture

3 Distance Calculation

In this section, the minimum distance from which our system should be able to capture, detect and recognize the traffic symbol is calculated. The situation is considered in which traffic symbol is on one side of the road and the car is traveling on the opposite extreme side of the road as shown in Fig 8. In this situation, the camera should capture the image before the car reaches the point “z.” If it fails to capture, then the traffic sign board goes out of the range and the system fails to detect the traffic symbol.

Two cases are considered in which car moves at different speed.

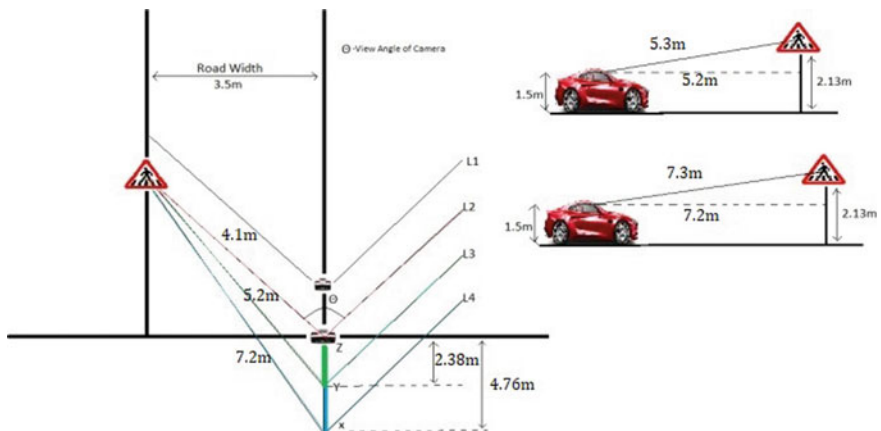


Fig. 8 Recognition distance calculation

Case 1: Car moves at a speed of 60km/hr or 16.66m/s

The TSDR system designed can process image at rate of 7fps which implies that the system can perform detection at every 2.38m interval.

In the Fig 8, the system should capture and process one image between point Y and Z (2.38m interval). Otherwise, the sign board goes out of range. In the worst case, if the image is captured at point Y, the distance from the car to the signboard is 5.3m (considering the height of the car and the sign board). Therefore, the system should be able to detect and recognize a traffic symbol at a minimum distance of 5.3m, if the car travels at a speed less than or equal to 60km/hr.

Case 2: When the car moves at a speed of 120km/hr or 33.33m/s, a similar approach made as in case 1. It can be calculated and shown that the system should be able to detect and recognize a traffic symbol at a minimum distance of 7.3m if the car travels at a speed less than or equal to 120km/hr

4 Implementation Details

This section provides the implementation details of traffic symbol detection and recognition System. All the steps are implemented on Python IDE using python support libraries. The input video is captured using mobile camera (Infinix zero 5) at 30 frames /second having the image dimension 640x480.

Training the model: The input for training the model will be images and their corresponding Y labels. During the training phase, images are given as inputs and their corresponding label matrix to the model. Both forward and backward propagation will train the model.

Training Parameters:

- Grid size: 13×13 (each grid of 32×32 px)
- Anchor boxes: 5
- No. of classes: 10
- Model: tiny-yolo-voc

After training the model, a metafile of size 62 MB is obtained which is used for testing.

Testing the model: The testing image will be divided into the same number of grids as chosen during the training period. For each grid, the model will predict an output of shape $13 \times 13 \times 75$. Non-Max suppression technique will be applied on the predicted boxes to obtain a single prediction per object.

Detection and Recognition of sign boards using the trained model:

Darkflow framework is created in raspberry pi and meta file generated after training the model is loaded. Python code is written to call the model, pass input image to it, display the predicted results, and give audio alert to the user.

The input image obtained through image acquisition process is divided into *sxs* grids as shown in Fig. 9a, where the *s* value depends on the size of the object being detected. Each grid cell predicts only one object. The grid containing the center of



Fig. 9 a Grids formed on input image, b Bounding boxes formed for the grid cells

the object being detected is responsible for the detection of the object. Bounding boxes are drawn around those grids as shown in Fig. 9b and class confidence score is calculated for each bounding box formed.

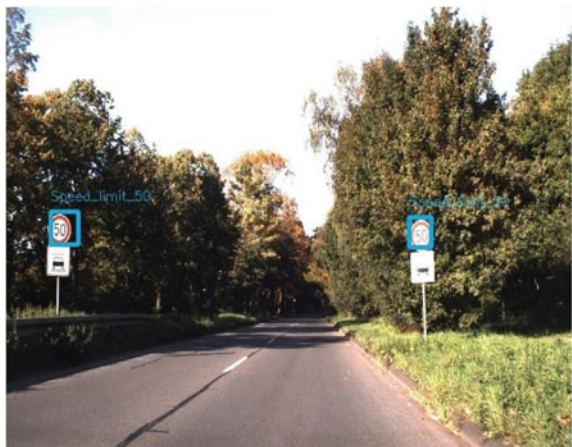
The class confidence score for each prediction box is computed as:

$$\text{Class confidence score} = \text{box confidence score} \times \text{conditional class probability}$$

where the confidence score reflects how likely the box contains an object and the conditional class probability is the probability that the detected object belongs to a particular class.

If the score is greater than the threshold set, the bounding box is displayed with the predicted class as shown in Fig. 10 and at the same time an audio alert is given to the user. The process continues for all the bounding boxes formed.

Fig. 10 Detected and recognized sign board



5 Results

German Traffic Sign Detection Benchmark (GTSDDB) dataset is used. The GTSDDB dataset contains 900 high-resolution natural scene images of the traffic signs as shown in the Fig 11, each image is 1360×800 pixels. German Traffic Sign Recognition Benchmark (GTSRB) dataset contains the traffic signs varying from 16×16 pixels to 128×128 pixels. The GTSRB dataset contains 39,209 images for training and 12,630 images for testing. The signs are divided into three main categories, namely, mandatory, cautionary, and informatory including 43 subclasses. The signboards which are similar to Indian traffic signboards are used in the experiment.

Labeling is a graphical image annotation tool and is used for creating annotations for traffic sign boards. Darkflow framework is used to run the YOLOv2 model used. The model is coded in python 3.7, and python compiler is used for training the model. Real-time testing on Indian traffic symbols and traffic symbol is recognized. In this experiment, a test video is recorded using a mobile camera (Infinix zero 5) fixed on the vehicle, riding at the speed of 40 km/hr on BH road, Tumakuru. The camera has a resolution of 640×480 at 30 frames /second. The recorded video is fed as input to detection model, and the video is processed at 10frames/second. The model recognizes the sign board and detects type as give way and speed limit 30 as shown in Figs. 12 and 13, respectively.

In the experiment conducted, for German test images some of the recognized images are shown in Figs. 14 and 15, and for china test data, some of the recognized images are shown in Figs. 16 and 17. Some of the misclassified images are shown in Figs. 18, 19 and 20, and misclassification is due to (i) our model has been trained to speed limit (10 categories), stop and give way sign boards, (ii) China traffic sign boards are kept at greater heights compared to other countries, (iii) due to the different number of training images in different categories. So, the model is able to detect the



Fig. 11 German traffic dataset images



Fig. 12 Give way sign board detection



Fig. 13 Speed limit 30 sign board detection

sign boards with the detection accuracy of 95% and with recognition accuracy of 97.36% in real time.

6 Conclusion

In this paper, the automatic detection and recognition of traffic symbols is briefly explained. The algorithm used for detection and recognition is You Only Look Once (YOLO). The model is trained on German dataset, with the images under different illumination conditions. The experiments performed shows that YOLO takes a better stand in terms of performance (with an accuracy of 97.3% in recognition) as compared to other existing algorithms for traffic sign board detection and recognition system.



Fig. 14 Speed limit 50 sign board detection



Fig. 15 Speed limit 30 sign board detection



Fig. 16 Speed limit 40 sign board detection



Fig. 17 Give way sign board detection



Fig. 18 No right turn detected as speed limit



Fig. 19 3.5 m height limit detected as speed limit



Fig. 20 Speed limit 100 detected as speed limit 50

References

1. Sajjad Hossain M, Mahmudul Hasan M, Ameer Ali M, Humayun Kabir M, Shawkat Ali ABM (2010) Automatic detection and recognition of traffic signs. Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh, East West University, Mohakhali, Dhaka-1212, Bangladesh School of Computing Science Central Queensland University, Rockhampton, Australia
2. Malik Z, Siddiqi I (2014) Detection and recognition of traffic signs from road scene images. In: 12th international conference on frontiers of information technology. Bahria University, Islamabad, Pakistan
3. Swathi M, Suresh KV (2017) Automatic traffic sign detection and recognition. In: International conference on algorithms, methodology models and applications in engineering technologies, 16 Feb 2017
4. Broggi A, Cerri P, Medici P, Porta PP, Ghisio G (2007) Real time road signs recognition. In: 2007 IEEE intelligent vehicles symposium Istanbul, Turkey, June 13–15
5. Redmon J, Farhadi A (2017) “YOLO9000: better, faster, stronger. University of Washington, Allen Institute for AI

Investigation of Gain and SAR in Dual Band Monopole Antenna Using Slotted Electromagnetic band Gap Structures



Prasanna G. Paga, H. C. Nagaraj, V. P. SaiKarthik, and B. R. Bhuvan

1 Introduction

Right now monopole integrated with EBG antenna, technology has been improving at a rapid pace which is mainly due to its technological benefits of gain improvements and SAR but comes out with circuit complexity. To cope up with increased circuit complexity, the size of the components containing the circuitry should be made as small as possible. Especially, the semiconductor technology growth is vast and is proved with the increased use of wireless technology. Today, there are many devices that operate in multiple bands of frequencies [1]. Frequency bands like GSM, Wi-Fi, Wi-Max, LTE have become more popular. Dual-band antennas provide the users to use the above frequency bands of frequencies for various applications in daily life. One such application is mobile phones. Other than these bands, there are other bands of frequencies like high band and ultra wide band (UWB) that are most helpful in Satellite applications. In [1], a dual-band monopole antenna is proposed for GSM, UMTS, LTE, Wi-Fi, Wi-Max and UWB applications. The proposed antenna in [1] uses two parallel sickle shaped patch and is small in size so that it can be mounted on any mobile devices. Monopole antennas have better advantages compared to dipole antenna as they exhibit omni-directional radiation pattern which is because of their length. In [2], a tri-band monopole band antenna is designed using FR4 substrate which operates in 2.4/3.5/5.2 GHz bands of frequencies. E-type monopoles are employed in the design of the antenna because of which better return loss and

P. G. Paga (✉) · H. C. Nagaraj · V. P. SaiKarthik · B. R. Bhuvan
Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of
Technology, Bangalore, India
e-mail: prasanna.paga@nmit.ac.in

H. C. Nagaraj
e-mail: principal@nmit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication
and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_71

899

gain were obtained for the operating bands of frequencies. Dual-band antenna generally operates at 2.4–3.5 GHz frequency bands since most of the applications cover these bands. In [3], one such antenna is proposed which is intended for Wi-Max and WLAN applications which operate at 3.5 GHz and 5.8 GHz frequency bands. Rectangular patch is designed in [3] and the gain of 3.93 dB for 3.5 GHz and 4.91 dB for 5.8 GHz frequencies were obtained. In [4], MIMO antenna is designed using square shaped patch with defective ground structure (DGS) for same applications. DGS was employed to get isolation between the multiple bands. In [5], tri-band monopole antenna is designed to operate in *C*-band and *X*-bands. Other than these applications, antennas find its uses in the field of medicines. In [6], wearable antenna is proposed which is desired to operate in ISM band. Here, micro strip patch is modified to *E*-shaped textile antenna, and hence, the performance parameters are increased.

The proposed antenna in this paper also deals with EBG structures which enhances the performance parameters of the antenna. Different EBG structures can be used to increase the performance parameters of wearable antennas that serve itself in medical field as proposed in [7]. The proposed antenna, in [7], makes use of EBG structure and the antenna gives a maximum gain of -15.18 dBi at 2.4 GHz operating frequency band. EBG structures fall under the class of meta materials. EBG also helps in the reduction of SAR (specific absorption rate) as proposed in [8]. Dual-band antennas also find its applications in one of blooming fields of today, i.e. IoT (Internet of Things). One such patch antenna using two via slot EBG structure for mutual coupling reduction were presented for IoT applications [9]. In [10], a dual-band antenna for 5G operating in the Ka band has been presented resonating at 28 GHz and 38 GHz frequency bands. The structure has been placed on a Rogers's substrate of permittivity of 2.2 and height of 0.254 mm. The peak gains reported were 7.71 dB and 7.73 dB for the lower and the upper frequency bands, respectively along with a SAR of 0.37 W/kg and 1.34 W/kg.

2 Proposed Antenna

The proposed antenna design is as shown in Figs. 1 and 2. Figure 1 shows the top view of the dual-band monopole antenna without EBG structure, whereas Fig. 2 shows the ground plane view of the same. The long patch, i.e. $L_{0.9}$ is calculated to operate at 0.9 GHz frequency, and short patch, i.e. $L_{2.4}$ is calculated to operate at 2.4 GHz frequency (Fig. 3).

The slotted EBG structure is incorporated around the patch to improve the performance parameters of the antenna like return loss, gain and bandwidth coverage of GSM 900 and Wi-Fi bands.

Dual-band monopole antenna with slotted EBG structures is shown in Figs. 4 and 5. EBG structures surround the patch of the antenna.

The complete specifications of the proposed antenna with EBG structure is shown in Table 1.

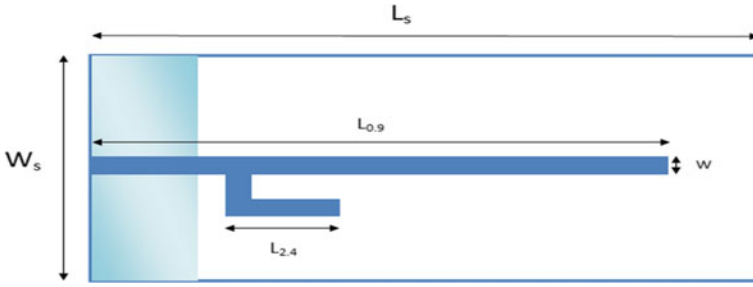
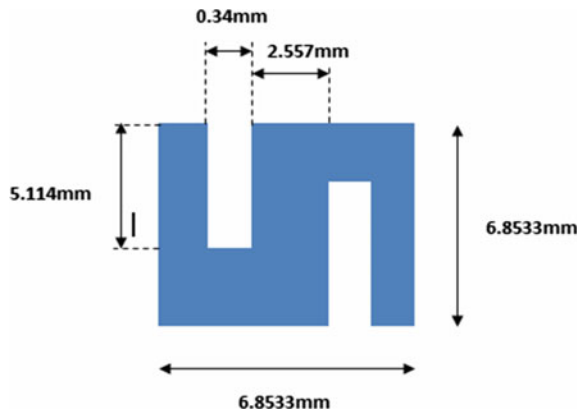


Fig. 1 Top view of proposed dual-band monopole antenna



Fig. 2 Ground plane view of proposed dual-band monopole antenna

Fig. 3 Slotted EBG unit cell of size $5.114 \times 6.8533 \text{ mm}^2$



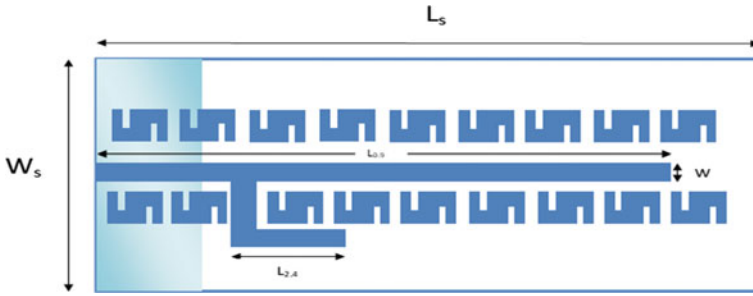


Fig. 4 Top view of dual-band antenna with slotted EBG structure



Fig. 5 Ground plane view of dual-band antenna with slotted EBG structure

Table 1 Antenna design parameters

Dimensions	Value (mm)
Substrate length, L_S	100
Substrate width, W_S	40
Ground plane length, L_G	11.5
Ground plane width, W_G	40
Feed width, W	3
Length of the patch for 0.9 GHz, $L_{0.9}$	88.9
Length of the patch for 2.4 GHz, $L_{2.4}$	12
Slotted EBG	6.8533
Slot length	5.114
Slot width	0.34
Gap between patch and EBG	0.9066
Gap between EBG	0.5

3 Antenna Design Consideration

3.1 Methodology

In the proposed work, a dual-band antenna has been placed on a FR4 substrate of permittivity 4.4 and height of 1.6 mm to tune between GSM 900 and Wi-Fi (2.4 GHz) frequency band. The longer conductor is designed to resonate at GSM 900 frequency band that is at 0.9 GHz, whilst the shorter strip is coupled to resonate at 2.4 GHz frequency band that happens to be the Wi-Fi application's working frequency. To improve the performance parameters slotted, EBG structure has been printed along the length of the conductors. Here, monopole class of antenna structures is investigated that helps in overcoming the space constraint characterized by a partial ground plane.

3.2 Design Equations

For 0.9 GHz operating frequency,

$$L_{0.9} = \lambda_g/4 \quad (1)$$

For 2.4 GHz operating frequency,

$$L_{2.4} = \lambda_g/4 \quad (2)$$

$$\lambda_g = \lambda_0/\sqrt{\epsilon_{\text{reff}}} \quad (3)$$

where λ_g = guide wavelength corresponding to GSM band, i.e. 0.9 GHz.

ϵ_{reff} = Effective permittivity of the substrate, λ_0 = Free space wavelength.

The effective permittivity is calculated using the expression,

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12 \frac{h}{W} \right]^{-1/2} \quad (4)$$

where ϵ_r = relative permittivity, h = thickness of the substrate, W = width of the monopole strip.

The length of Ground plane is given by, $L_G = 6h + L_{2.4}$

The width of Ground plane is given by, $W_G = 6h + W_{2.4}$

4 Simulated Results

The detailed simulated results of the dual-band monopole antenna with and without slotted EBG structure are discussed.

Figure 6 shows the return loss plot of dual-band monopole antenna without EBG structures. The return loss of -17.9963 dB is obtained for GSM 900 band, and -12.05881 dB is obtained for Wi-Fi band. With the introduction of EBG structures, the resonating frequency and also the return loss are improved. The return loss comparison plot of both without and with EBG is shown in Fig. 7. The simulated values of return loss of antenna with EBG structures are -17.75147 dB for GSM band and -19.98459 dB for Wi-Fi band.

The angular axis represents angle in degrees and the vertical axis represents gain in dBi. The peak gain reported was 6.29939 dBi.

Fig. 6 Return loss plot of dual-band monopole antenna without EBG structures. Horizontal axis represents frequency (GHz) and vertical axis represents S_{11} (dB)

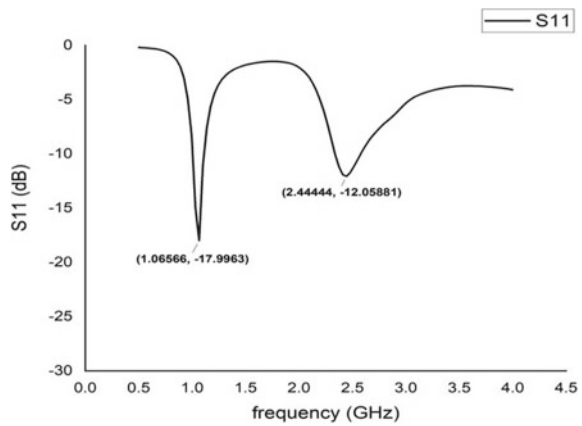
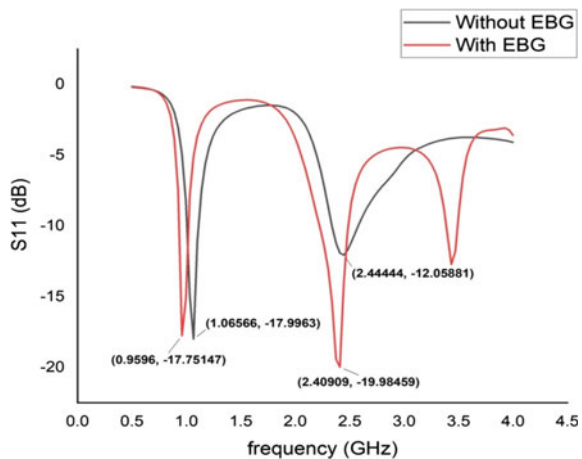


Fig. 7 Return loss plot of dual-band monopole antenna with and without EBG structures. Horizontal axis represents frequency (GHz) and vertical axis represents S_{11} (dB)



The angular axis represents angle in degrees and the vertical axis represents gain in dBi. The peak gain reported was 6.76136 dB.

The peak gain reported was 6.39 dBi. The vertical axis plots the calculated gain in dBi, and the circular axis shows the angle in degrees.

The peak gain reported were 7.17 dBi. The vertical axis represents the gain in dBi, whilst the angular axis represents the angle in degrees.

EBG structures improved the gain of the antenna for both bands of resonance as shown in Figs. 8 and 9.

Slotted EBG structures improved the gain of the antenna at both frequencies as shown in Figs. 10 and 11. The maximum gain of 6.39079 dB and 7.1703 dB for GSM 900 and Wi-Fi 2.4 GHz bands, respectively. By the data provided, we can infer that gain is increased for both the bands (Fig. 12).

The -10 dB impedance bandwidth is extending from 1.0656 to 1.116 GHz in the lower band and 2.33–2.572 GHz in the upper band. The parallel axis represents the frequency in (GHz), whilst the perpendicular axis represents the return loss in dB.

Fig. 8 Radiation pattern of dual-band antenna without EBG structure at GSM band resonating at 0.9 GHz

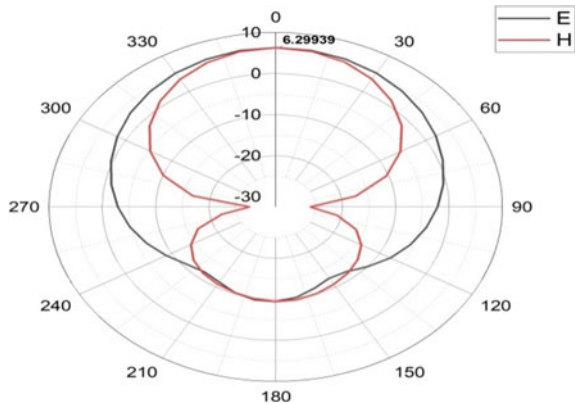


Fig. 9 Radiation pattern of dual-band antenna without EBG structure at Wi-Fi band resonating at 2.4 GHz

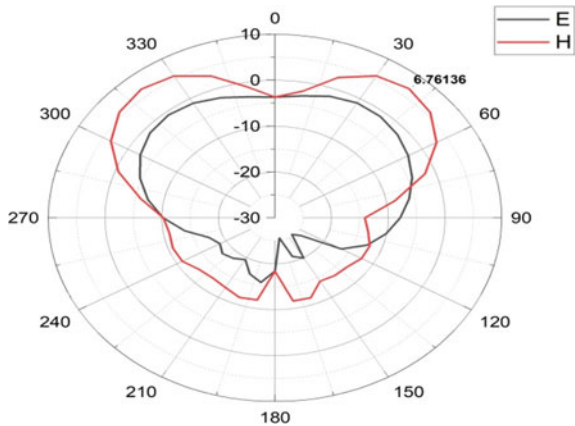


Fig. 10 Radiation pattern of dual-band antenna with Slotted EBG structure at GSM band

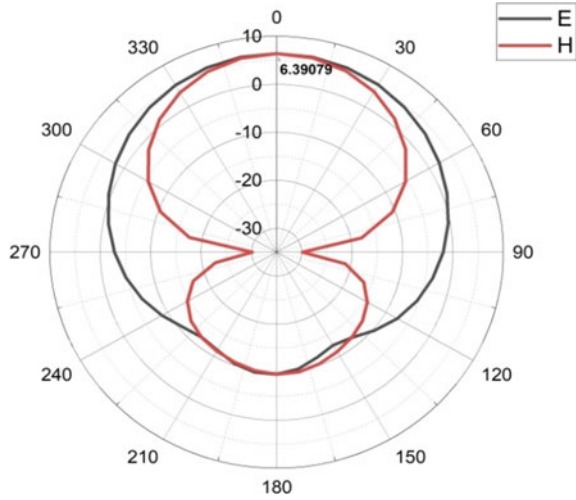
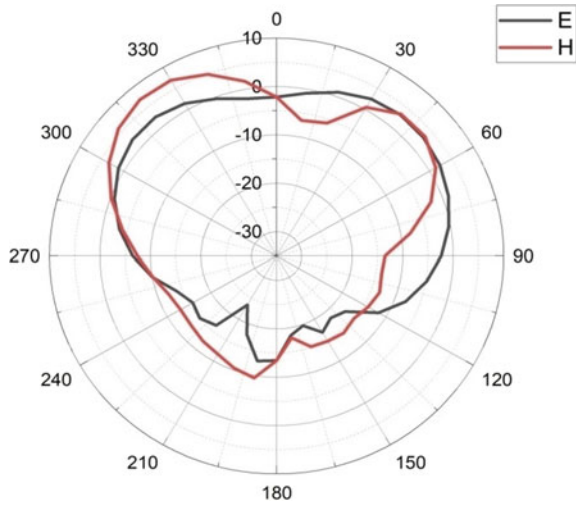


Fig. 11 Radiation pattern of dual-band antenna with Slotted EBG structure resonating at 2.4 GHz



The -10 dB impedance bandwidth is extending from 0.9375 to 1.0243 GHz in the lower band and 2.245–2.487 GHz in the upper band. The parallel axis represents the frequency in (GHz), whilst the perpendicular axis represents the return loss in dB.

The bandwidth coverage of the antenna with EBG structure for both the operating frequencies is shown in Fig. 13. It is observed that the higher and lower cut-off frequencies of GSM band are 1.0243 GHz and 0.9375 GHz, respectively. Whereas the higher and lower cut-off frequencies of Wi-Fi band are 2.4870 GHz and 2.2457 GHz, respectively. Thus, the bandwidth coverage for GSM band is found to be 86.8 MHz, and that of Wi-Fi band is 241.3 MHz. The SAR values of the antenna at 0.9 GHz

Fig. 12 Return loss plot of dual-band antenna without EBG structure

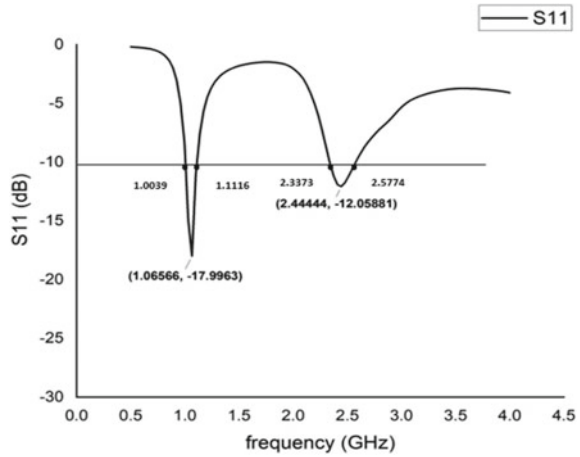
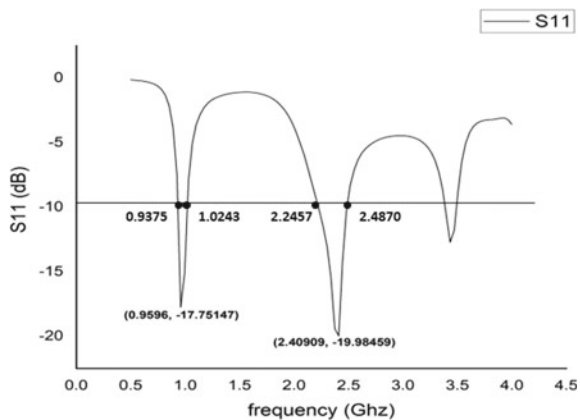


Fig. 13 Return loss plot of dual-band antenna with Slotted EBG



without and with slotted EBG are as shown in Figs. 14 and 15 respectively while the corresponding results of SAR at 2.4 GHz without and with EBG are as shown in Figs. 16 and 17 respectively.

5 Results and Discussion

The proposed dual-band monopole antenna with slotted EBG structure as shown in Fig. 4 can be used for GSM and Wi-Fi applications. The geometrical specifications of the EBG cell and the antenna were reported in Fig. 3 and Table 1, respectively. Simulated results with and without EBG are represented in Tables 2 and 3.

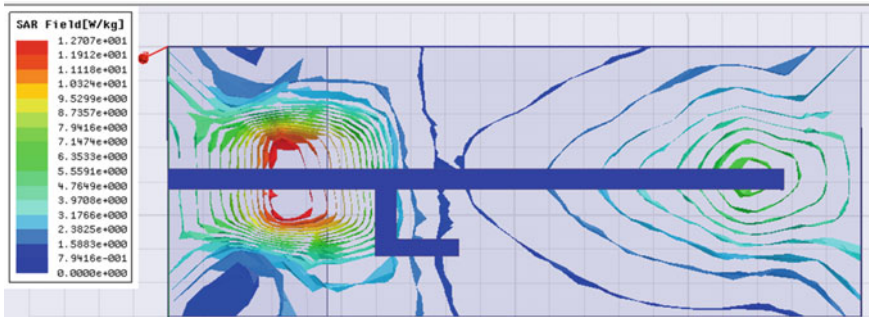


Fig. 14 SAR value of the antenna without EBG at GSM indicating 12.7 W/kg

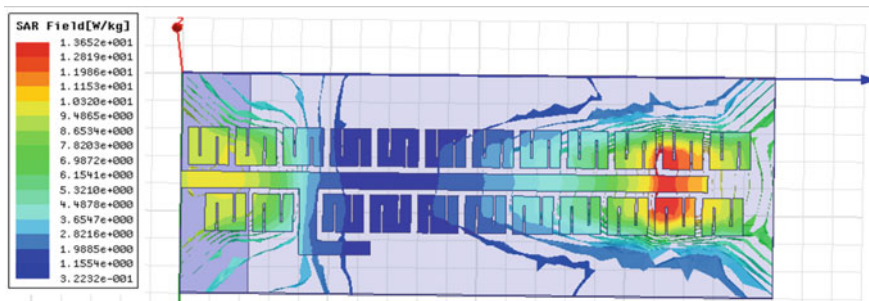


Fig. 15 SAR value of antenna at 0.9 GHz with EBG indicating 13.65 W/kg

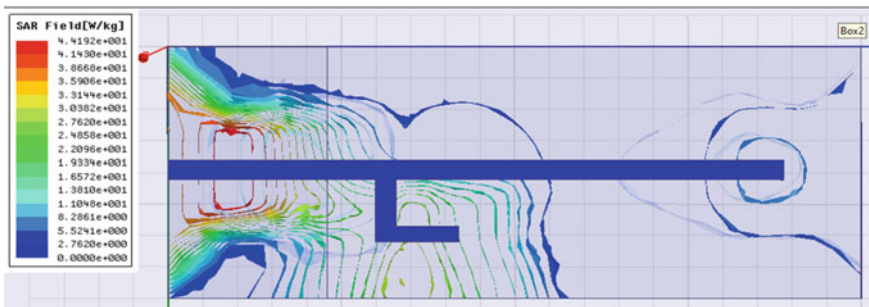


Fig. 16 SAR value of antenna at 2.4 GHz without EBG indicating a value of 44.19 W/kg

6 Conclusion

The above results infer that Slotted EBG structure enhanced the performance parameters of the antenna by improving the return loss, increasing the gain for the bands and improving the bandwidth coverage. Without EBG, the GSM 900 band antenna attained a return loss of -17.9963 dB, gain is 6.29939 dB and bandwidth of

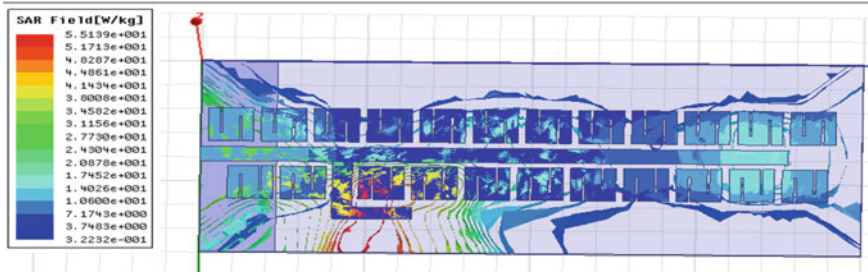


Fig. 17 SAR value of antenna at 2.4 GHz with EBG indicating a value of 55.13 W/kg

Table 2 Simulated results of dual-band monopole antenna without EBG structure

Parameter	GSM—900 MHz	Wi-Fi—2.4 GHz
Return loss (dB)	-17.9963	-12.05881
Gain (dB)	6.29939	6.76136
Bandwidth (MHz)	107.7	244.1

Table 3 Simulated results of dual-band monopole antenna with slotted EBG structure

Parameter	GSM—900 MHz	Wi-Fi—2.4 GHz
Return loss (dB)	-17.75147	-19.98459
Gain (dB)	6.39709	7.1703
Bandwidth (MHz)	86.6	241.3

107.7 MHz from Table 2. As against Wi-Fi band, where the corresponding return loss values were -12.05881, gain of 6.76136 dB and bandwidth of 244.1 MHz. With the integration of slotted EBG surrounding the antenna, there has been a significant improvement in the return loss of -17.75147 dB, gain of 6.39709 dB and bandwidth is 86.6 MHz. The corresponding parameters in the Wi-Fi band were S_{11} equal to -19.98459 dB, gain of 7.1703 dB and bandwidth of 241.3 MHz as reported in Table 3. Corresponding SAR values reported were 12.7 W/kg and 44.19 W/kg at GSM 900 and Wi-Fi without EBG which are comparatively less compared to 13.65 W/kg and 55.13 W/kg for the structure with EBG, respectively, inferring the EBG structures can be used for improving the gain of antenna for GSM and Wi-Fi applications.

References

1. Roshan R, Prajapati S, Tiwari H, Govind G (2018) A dual wideband monopole antenna for GSM/UMTS/LTE/Wi-Fi/and lower UWB application. In: International conference on microwave and photonics (ICMAP). <https://doi.org/10.1109/ICMAP.2018.8354635>
2. Fang Q, Mi D, Yin YZ (2015) A triband MIMO antenna for WLAN/WiMAX applications. Prog Electromagn Res Lett 55:75–80. <https://doi.org/10.2528/PIERL15070202>

3. Thaher RH, Jamel ZS (2018) New design of dual-band microstrip antenna for Wi-Max and WLAN applications. In: International scientific conference of engineering sciences (ISCES). <https://doi.org/10.1109/ISCES.2018.8340541>
4. Khade SS, Badjate SL (2018) Square shape MIMO antenna with defected ground structure. In: International conference on recent advances in information technology (RAIT). <https://doi.org/10.1109/RAIT.2018.8388995>
5. Althubitat Al Amro WH, Abdelazeez MK (2018) monopole tri-band notched characteristics UWB antenna for WiMAX, C-band, WLAN and X-band applications. In: International conference on information and communication systems (ICICS). <https://doi.org/10.1109/IACS.2018.8355466>
6. Oguntala GA, Abd-Alhameed RA, Noras JM (2018) Inverted E-shaped wearable textile antenna for medical applications. IEEE. <https://doi.org/10.1109/ACCESS.2018.2847280>
7. Fan Y, Huang J, Chang T, Liu XY (2018) A miniaturized four-element MIMO antenna with EBG for implantable medical devices. J Electromagnet RF Microwave Med Biol 2(4). <https://doi.org/10.1109/JERM.2018.2871458>
8. Rahmatian P, Movahhedi M, Ghafoorzadeh-Yazdi A (2018) Dual-band dual-mode wearable antenna for on-off body communication based on metamaterial. In: Iranian conference on electrical engineering (ICEE). <https://doi.org/10.1109/ICEE.2018.8472524>
9. Bhavarthe PP, Rathod SS, Reddy KTV (2018) Mutual coupling reduction in patch antenna using electromagnetic band gap (EBG) structure for IoT application. In: International conference on communication, information and computing technology (ICCICT). <https://doi.org/10.1109/ICCICT.2018.8325867>
10. Khan J, Sehrai DA, Ali U (2019) Design of dual band 5G antenna array with SAR analysis for future mobile handsets. J Electr Eng Technol 14. <https://doi.org/10.1007/s42835-018-00059-9>

Profit Driven Blockchain Based Platform for Land Registry



Md. Amaan Ahmad, Pooja Singh, M. Sushmitha, H. A. Sanjay,
and N. Madhu

1 Introduction

Land ownership and maintaining land records are a challenging task and intrigues a lot of issues in India. Most of the institution dealing with financial transaction related to land ownership has become so critical that they ask for guarantee to keep land based properties that can be used against collateral security. It has also become a tedious and a herculean job to function normally in the domain of real estate as there are lot of uncertainties on ownership claims of the land [1].

The real estate transaction or the process of buying and selling the land is done using the sale deeds and other authentication documents. Submitting these documents for property ownership does not always results in successful transfer of the property but it records the transaction. Such records are duly maintained by the govt. owned departments and documented as land records. India being a subcontinent having a population of approximately 1.38 billion and land area of 3.287 million km² it is very difficult to keep track of the ownership (clear titles) due to the outdated method used to maintain land records. This issue also causes lot of discrepancies during land transfer. Over few decades, many schemes and implementation ideas have been put forth to make land ownership transfers clear and robust. Digitization of such records helps in avoiding frauds. The records maintained by the advisories appointed through government in each area are primarily used to collect revenues [2]. Therefore, the registration of area unit is much of presumption than being a conclusive matter [3],

Md. A. Ahmad · P. Singh · M. Sushmitha · H. A. Sanjay · N. Madhu (✉)
Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore 560064, India
e-mail: madhu.n@nmit.ac.in

H. A. Sanjay
e-mail: sanjay.ha@nmit.ac.in

which implies responsibility of verification of the previous ownership depends on the loyalty and faith of customers on the registrar's appointment by the government.

Blockchain technology can be used to address the above mentioned issues. Blockchain is a distributed ledger technology (DLT) that acts as localized computation and knowledge sharing platform that permits multiple reliable domains. Multiple individual records/transactions are stored as block that are coupled along in single list known as chain. It is a shared distributed ledger that helps in processing any digital group actions over the business network and tracks the physical or untouchable assets concerned that facilitates the method of recording transaction and pursuit assets in a scalable business network. Assets are often physical like land, house, cash, automobile or untouchable like material possession, like patents, copyrights, or stigmatization. Effectively, any price is often half-track and listed on a blockchain network, reducing risk and cutting prices for all concerned [4].

Blockchain consists of blocks that are chained along with cryptographic hashes. A data structure is maintained to contain each block that comprises of:

- (1) Encrypted hash of block ($n-1$)
- (2) Transaction data
- (3) Timestamp

Ethereum is one such blockchain that has capability to run good contracts on high of it. Creating the whole blockchain, a world state machine that might execute code that's protected by the properties of blockchain. Ethereum would check for the code which follows the set parameters. This process will avoid advisories between the two parties.

This work involves developing a decentralized framework for land registry. This platform authorizes the originality of the land registration documents which are tamper-proof, immutable, transparent and encrypted hence, provides security. This technology has been chosen because it is definite, tamper-proof and encrypted. Adoption of blockchain technology will remove the problem of fake documents. In this work, the document uploaded is first converted to data structure and the square measure value is added to the network. Computed hash will be placed on the blockchain network. The main aim is to computerize all land records, as well as mutations, up transparency within the land record maintenance system, alter maps and surveys, update all the settlement records and minimization of the scope of land disputes and conjointly to produce clear titles of land possession that would be monitored simply by administration, for facilitating faster transactions. The framework allows the user to transact using fiat currency. Completeness, security and privacy are the three important features that any of the blockchain based solution should comply which is satisfied in the proposed framework.

The work also meets the below mentioned objectives:

- Unique identification for every land record.
- Eliminates advisories as the framework eliminates external verification process.
- Allows fiat currency for transaction.
- Authenticity of the present information toward land possession.

- Eliminating the need for physical archives of contracts and files.
- Greater security against frauds.

The rest of the paper is organized as follows. Section 2 provides the overview of the existing efforts in the domain of blockchain and land record digitalization. Section 3 describes the proposed framework with the internal details. Section 4 describes the implementation of the decentralized application and the tools used to develop the DApp. Section 5 provides the experimentation details and results with the cost effectiveness, followed by the conclusion and the future scope.

2 Related Work

Most of the existing efforts are in the proposal state. Few countries have started planning to deploy the land registration process on blockchain platform. The work, by Kshetri and Voas [3], explains that creating a tamper-proof digital ledger of transactions, and sharing the ledger helps offer transparency and using cryptography allows access to add to the ledger securely. It says that it is extremely difficult to change or remove data recorded on a ledger making it possible to reduce or eliminate integrity violations such as fraud and corruption thus giving an upper hand as compared to other technologies. Ehmke et al. [5], a lightweight and scalable blockchain protocol, explained how one can use Ethereum to keep the state of the system clear in the present block and further include the ongoing system state in new transactions enabling all participants to validate incoming transactions. It further explains how the transactions can be validated without having access to the whole system state, thus enabling users to participate in the network without having to download the blockchain beforehand. Ahram et al. [2], describe how enabling secure trust frameworks, creating agile value chain production, and tighter integration with technologies such as cloud computing, and IoT work in innovations in blockchain technology. All these works and many more helps us understand blockchain, its implementations in today's world and how securely it can manage data with atomicity and consistency. It can also detect supply chains by removing paper-based trails, making businesses able to pinpoint incapability within their supply chains faster, as well as find items in real time. It allows businesses, and possibly even consumers, to view how products performed from a quality-control perspective. Kan et al. [6], elucidate a dynamic network of multi-chain generated for inter-blockchain communication for obtaining an interactive multiple blockchain architecture providing atomicity and consistency for crossing-chain transactions. Work by Nandi et al. [7], a secured land registration framework on blockchain. Land is a stable and non-liquid plus having high worth. The integrity and correct track of ownership/transfer records of land may be extremely difficult task. Because the possession of land will perpetually amendment over time which too typically terribly offtimes, it poses a frightening task of keeping elaborate and long-possession transfer records. The matter any escalates because of the presence

of dishonorable or incomplete registries that are terribly troublesome to trace back through time.

Some nations, like Sweden [4], European country, Honduras [4], India [8] and Dubai [9], have planned to shift the land register system into a blockchain primarily based system. The Sweden project has some documentation associated with their project obsessed by the Chromaway [1] within the year 2016, the Sweden project is in the second stage of development.

3 Proposed Framework

The proposed framework provides the necessary modules to perform land registration on the blockchain platform. The platform is created on top of the Ethereum blockchain and solidity is used for the smart contract development. Stakeholders at different levels can use the platform and browse the services. The framework allows the user to pay the transaction fee using fiat currency which leads to mass adoption. The framework converts fiat currency to cryptocurrency for enabling transactions in blockchain environment.

The complete working process of the system is explained in Fig. 1. Our framework eliminates the concept of advisories like government agencies, but these agencies can monitor the process and intervene as and when required. The step by step process of the complete working of the system is as follows:

- (i) The super user adds the super admin who is responsible for adding the owner of the land.

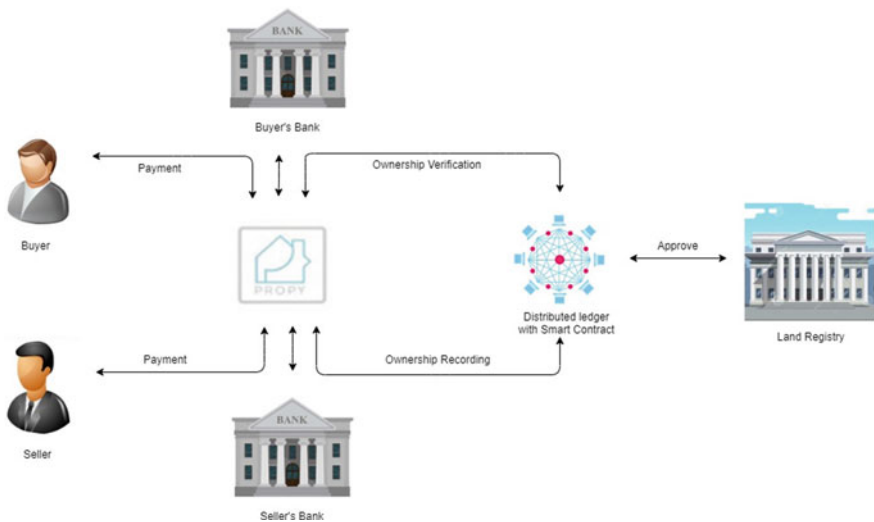


Fig. 1 The architectural design

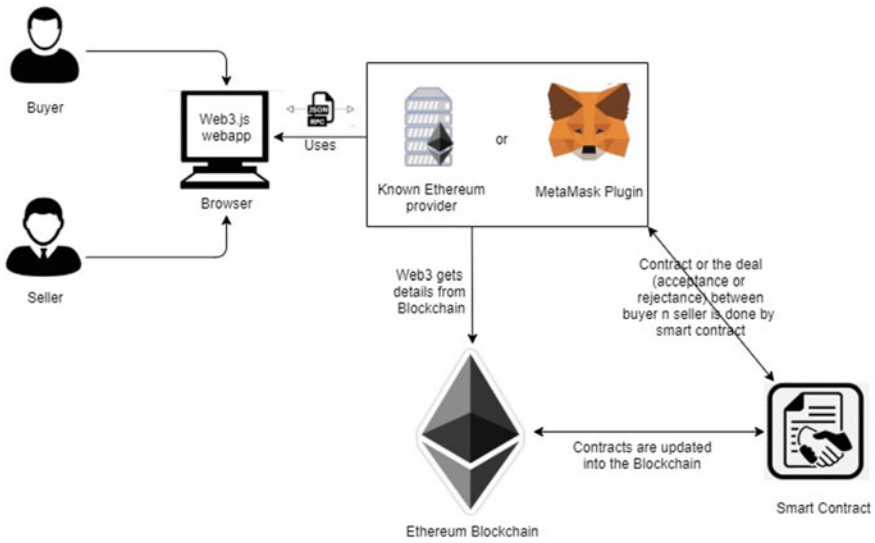


Fig. 2 Proposed framework

- (ii) The owner who wants to sell the land makes it available for other buyers.
- (iii) Buyer who is interested, place a request for the land.
- (iv) The owner of the land either accepts or rejects the request.
- (v) If the request is accepted then the seller makes payment to the buyer and becomes the current buyer of the land.
- (vi) If the request is rejected the land is not available for the seller.

The DApp has been developed to make the land registration and land transaction transparent and decentralized. The framework eliminates the need of advisories. The proposed framework has been shown in Fig. 2. The smart contract consists of mainly two functions:

1. Registration: The detailed documents are submitted to the authority named by the government, registered as super admin. The super admin verifies the details with the existing records and logs the details into the DApp.
2. Transaction: The transaction is for registration of land, selling, buying and adding the super admin for a particular village/area. It ensure that the transactions are atomic in nature and when the transaction is completed the new owner details are updated automatically without intervention from any users.

The proposed blockchain architecture will have land and identities of parties involved as assets, whole registration process will be done through distributed ledger (DLT), and the transactions done will be through smart contracts. Figure 3 depicts the architecture of land registry using blockchain technology.

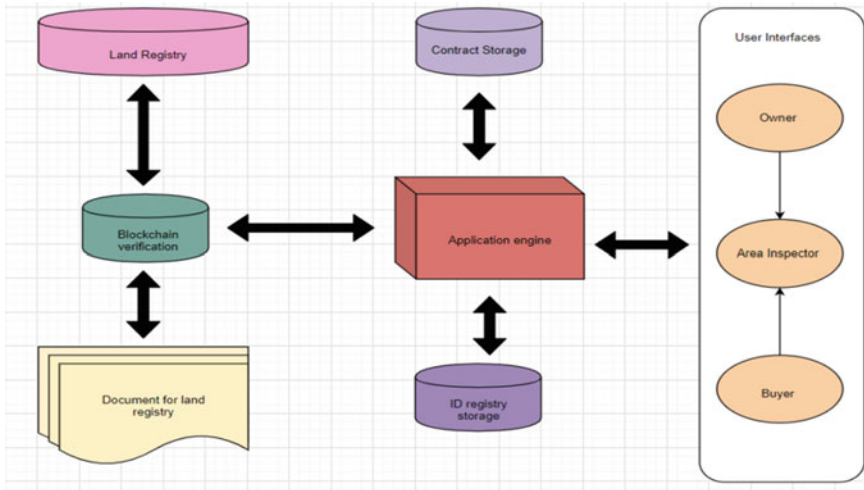


Fig. 3 Land registry blockchain architecture

4 Implementation

The platform is implemented over the Ethereum blockchain. The logic of the application is coded in multiple smart contracts. The implementation has been done in the test driven development (TDD) methodology which provides good security first design pattern. The complete system is decentralized; however, the centralized server doesn't stop from in any way from interacting people to the smart contract. The development has been carried out using the truffle blockchain development framework.

This work uses the following tools that help in development of the decentralized application on Ethereum.

- **Truffle:** The truffle suite for the development of decentralized application on Ethereum. It is a spun off spoke from Consensus LLC. Truffle takes care of smart contract life-cycle management and provides support for user-defined deployments, link-libraries and advance Ethereum applications.
- **Smart Contract Development:** The smart contract development in the platform is done using solidity language. It has the extension of “.sol”. The bytecode generated from the solidity code is the input to the Ethereum virtual machine.
- **Angular JS:** It is a JavaScript framework which enables embedding into HTML page using a <script> tag. This is possible because Angular JS provides extension which binds data in expressions with the help of attributes with Directives.
- **Ganache:** Ganache is used to deploy contracts and run tests on the DApp as it is a private blockchain used in development on Ethereum. It provides more features as compared to other platforms.
- **Meta-mask:** Meta-Mask is a Google Chrome extension that is seen as a bridge between Internet browsers, Ethereum, DApps built on Ethereum such as My Ether

Wallet. So basically, meta-mask is a wallet for the browser. It allows users to execute Ethereum DApps on internet browser directly not requiring to run on a full Ethereum node. Meta-mask enables users to store, send, receive and facilitate interactions with the Ethereum network.

- **Web3js:** For the DApp to interact with the user, it is required for the app to interact with the Ethereum and create smart contracts. The Web3js library allow the user to transfer Ether from one account to another, perform read and write from smart contracts.

This decentralized application makes the process of land registration clear and distributed. It consists of an angular project taking care of the front-end. It is much easier to build a single-page application using Textscript language of angular framework. Code repetition is reduced to a great extent by making use of the modules in angular framework. Back-end components are mainly a local blockchain, called Ganache which keeps running on the local host and is connected to the browser through meta-mask, a web3js provider. Most of the functionalities are implemented using the smart contract that is written using solidity programming language. This smart contract is deployed on the blockchain and invoked automatically on the being deployed. The DApp flow of execution is shown in Fig. 4.

The smart contract consists of primarily two functions.

1. Registration

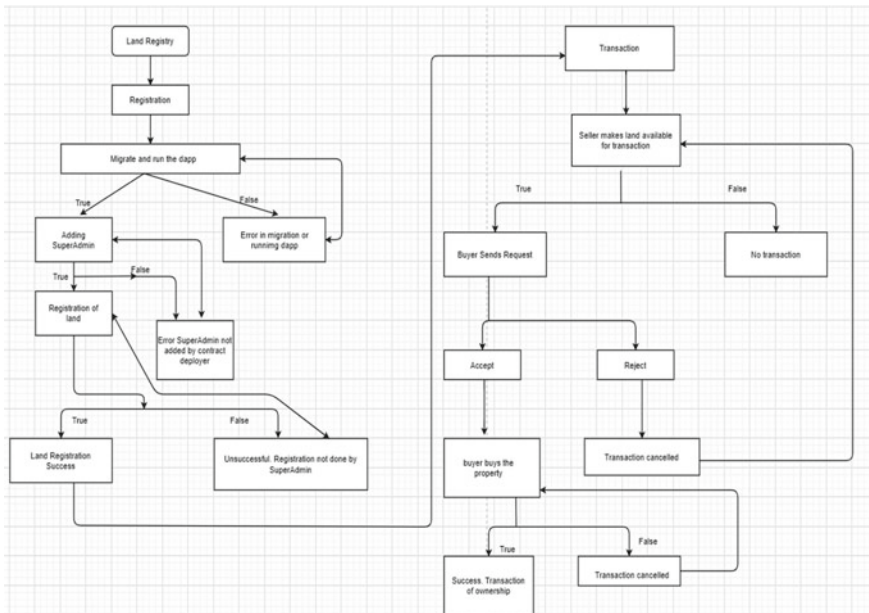


Fig. 4 Execution flow of DApp

Land details are stored in the structure created in the smart contract and contains field for state, district, village, survey number, owner's address. Profiles of the clients are stored as a structure and an array for all assets are included in the structure. A mapping is provided for each address to the profiles structure, so that all the assets for a client can be listed. A function is dedicated in the smart contract for the purpose of calculating unique id for every land piece based on its details. This id is used for sorting out the Land as well. "computeId" function uses keccak256() to calculate id. Super admin (Contract creator) can create village admins by invoking the function in the smart contract which takes the village and address of the village admin as input. Both super admin as well as the village admin has the permission to add a new land record by making call to the function specifically defined in the same smart contract to add land.

2. **Transaction**

Buying and selling of the land consist of two transactions one for each purpose, requiring gas charges and land costs both in Ethers (cryptocurrency). Also, the transactions are automatically reverted in case of transaction failures. The status of every new land is unavailable and owner of the land can make it available for sale. Once the status is changed to available then any client can search for the land and request for it, which again triggers a function in the smart contract and notifies the owner about the request. The owner can choose to accept or reject it, once accepted the buyer can buy the land by completing the transaction using the cryptocurrency (Ethers). Rejected request turns the request status again to rejected and can be requested again by another buyer.

Blockchain is getting used to make distinctive digital units for assets like land property, to verify digital file believability, identity, order in time and places and to own a secured management on all the processes. Because of the dearth of maintenance of rationalized land records, there are legal actions, scams and property disputes over land possession. Hence, a digital and secured department need to be established, for higher maintenance of land records.

The projected blockchain design can have land and identities of parties concerned as assets, whole registration method is done through distributed ledger (DLT), and also the transactions done are through sensible contracts.

The following is the description of the different modules developed in the framework:

1. **Registration:**

Super admin or the village admin can add land details in the form and register the land details with the DApp. Addition of land requires some gas charges (transaction fee) which is paid through a transaction. On receiving the land details, a function in the smart contract creates a property id for that land for future retrieval of the data.

2. **Village Admin Registration:**

The super admin has the right to allot village admin to a village. Village admin has the jurisdiction for that village and can register land for that village. Super

admin invokes the function in the smart contract to add the village admin by providing the village as string and village admin's address.

3. **Registering land for sale.**

For every new piece of land added, it is not available for sale, and the owner has the right to change the land status to available so that it can be searched by the buyer. Owner has to pay some gas charges to make it happen.

4. **Request for buying:**

The buyer can search for the land using land details and if available can make a request to the owner for buying that land. For this purpose the buyer has to pay some gas charges to invoke a function in the smart contract which is responsible to send the owner of the land request notification, and changing the status of the request from default to pending.

5. **Request Viewing:**

On the home page of the owner will pop up a notification for the request describing the address of the requester.

6. **Processing the requests:**

After a request is made to the owner of the land, the owner has the right to accept or decline the request. If the owner finds the request to be legit and then owner can accept the request by invoking an internal function in the smart contract and thus turning the status of the request was approved.

7. **Buying:**

After the request is approved by the owner, the sender of the request can search for the land and buying option is enabled. By clicking on the button, the buyer can initiate the land transfer. After successfully completion of the transaction, the ownership is transferred to the buyer. Internally, there are functions which are invoked for removing the older ownership and adding new owner to the DApp.

5 Experiment and Results

Experiments were conducted on the machine running Ubuntu 19.04 as the host operating system with 8 GB of RAM. The local blockchain environment during the development was Ganache. Using Ganache in this work 10 user accounts were created, and each account was credited with 100 ethers. The first account created a genesis block which had the super-admin details. Figure 5 shows the user interface of the DApp.

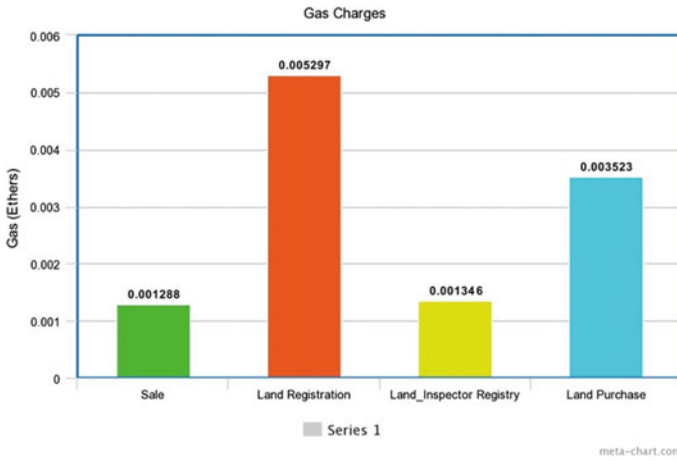
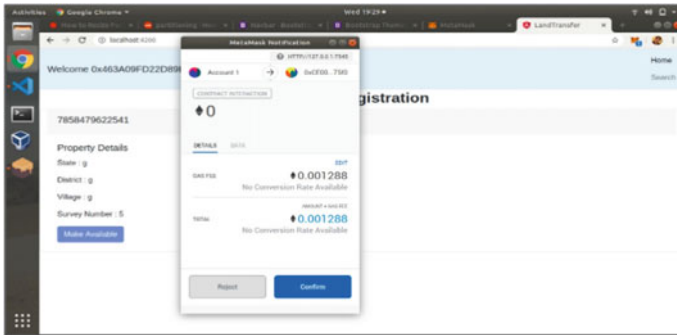


Fig. 5 Transaction fee in ethers



For any changes made to the state of the blockchain, a transaction is initiated which cost some gas for the successful completion. These transactions create a log about the states of the blocks in the blockchain and can be verified individually by all the nodes, since all the nodes replicate a copy of the main blockchain network, on the host machine. Also it can be noted from the graphs that the solution is very much cost effective in comparison with the traditional land registration process which includes large charges as stamp duty. Table 1 provides a gist of transaction fee in INR for traditional land registration in India (as per the current rates), and the same

Table 1 .

Property value (INR)	2,000,000
Stamp duty (INR)	112,000
Land registration cost (INR)	20,000
BBMP and BMRDA charges (INR)	168
Total cost (INR)	2,132,168

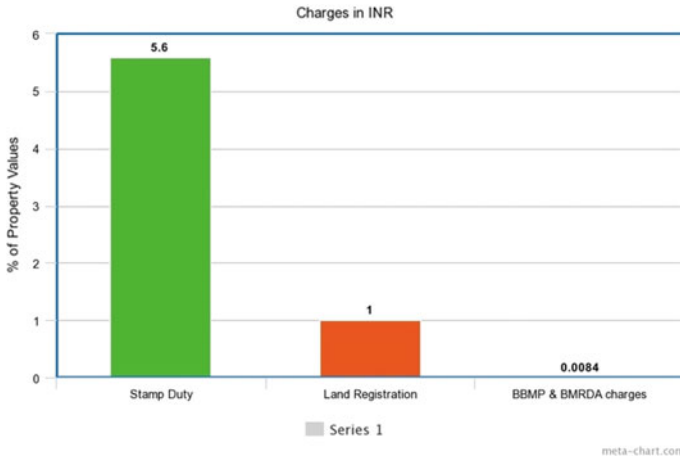


Fig. 6 Transaction fee in a traditional land registry

is depicted in Fig. 6. Figure 5 represents the transaction fee for various operations in terms of ether.

Stamp duty can be effectively reduced by using blockchain based land registry, where all of the registration process doesn't need physical record keeping mechanism, and thus, the process of registration becomes quite affordable for the buyer and seller.

The transaction fee in the proposed framework is based on the gas price fixed by the Ethereum network. As per the implementation for the amount of ether consumed for a land transaction is 0.01145 Ether which is equivalent to INR 206 (1 Ether equals 17,972.90 INR). This shows the cost effectiveness of the proposed framework. The framework accepts the transaction fee from the consumers through fiat currency which allows easy adoption by the Indian government and public.

6 Conclusion

In this paper, we have explicated the idea of improved and digitized land registry process using blockchain technology as a platform to build the underlying trust infrastructure. blockchain in land registry will bring the transparency in process and not only detect the fraud but also it will prevent it too. This solution will also increase citizens' confidence in the government and make the overall customer experience less cumbersome. Most importantly, it will enhance the data security and ensure authenticity of land records.

When considering the future scope tokenization of land and assets can be made possible using cryptocurrency for a blockchain platform, which will provide enhancement to the liquidity of land and assets. Customers can buy tokens from the ICO held at the start of a land or property sale. Fragmentation of large assets such as malls,

commercial complex, etc., can be made, so that people can invest even small amount of money and get a share in that asset.

References

1. Nandi M, Bhattacharjee RK, Jha A, Barbhuiya FA (2020) A secured land registration framework on blockchain. In: 2020 third ISEA conference on security and privacy (ISEA-ISAP), Guwahati, India, pp 130–138. <https://doi.org/10.1109/ISEA-ISAP49340.2020.235011>
2. Ahram T, Sargolzaei A, Sargolzaei S, Daniels J, Amaba B (2017) Blockchain technology innovations. In: 2017 IEEE technology and engineering management conference (TEMSCON), Santa Clara, CA, pp 137–141. <https://doi.org/10.1109/TEMSCON.2017.7998367>
3. Kshetri N, Voas J (2018) Blockchain in evolving countries. 20(2)
4. Xu M, Chen X, Kou G (2019) A systematic review of blockchain. *Financ Innov* 5:27. <https://doi.org/10.1186/s40854-019-0147-z>
5. Ehmke C, Wessling F, Friedrich CM (2018) Proof-of-property—a lightweight and scalable blockchain protocol. In: 2018 IEEE/ACM 1st international workshop on emerging trends in software engineering for blockchain (WETSEB), Gothenburg, Sweden, pp 48–51
6. Kan L, Wei Y, Hafiz Muhammad A, Siyuan W, Gao LC, Kai H (2018) A multiple blockchains architecture on inter-blockchain communication. In: 2018 IEEE international conference on software quality, reliability and security companion (QRS-C), Lisbon, pp 139–145. <https://doi.org/10.1109/QRS-C.2018.00037>
7. <https://lisk.io/academy/blockchain-basics/how-doesblockchain-work/what-is-hashing>
8. Mertz L (2018) (Block) chain reaction: a blockchain revolution sweeps into health care, offering the possibility for a much-needed data solution. *IEEE Pulse* 9(3):4–7. <https://doi.org/10.1109/MPUL.2018.2814879>
9. Abhishek G, Shukla S (2019) Property registration and land record management via blockchains. M. Tech. Thesis, IIT Kanpur
10. Sun H, Mao H, Bai X, Chen Z, Hu K, Yu W (2017) Multi-blockchain model for central bank digital currency. In: 2017 18th international conference on parallel and distributed computing, applications and technologies (PDCAT), Taipei, pp 360–367. <https://doi.org/10.1109/PDCAT.2017.00066>
11. <https://committee.iso.org/files/live/users/fh/aj/aj/tc211contributor%40iso.org/files/ISO%20TC%20211%2044th>
12. <https://blockgeeks.com/guides/blockchain-applications/>

Bio-Modification—An Emerging Ground Improvement Technique



Divya Viswanath  and M. N. Asha 

1 Introduction

The in-situ soil at project sites is very often found to be incapable of bearing loads from the superstructure above. This situation occurs when the field soil possesses very low-shear strength and bearing capacity. Any structure constructed on such soils may fail in shear or may be subjected to settlements. As developing countries are focusing on infrastructure development at every nook and corner, the engineers often face the challenges of proceeding with construction activities in soils that exhibit poor engineering characteristics. This problem was overcome in the past by mixing the in-situ soil with locally available materials and thereby preparing the blended soil ready for construction. With the progress in science and technology, many research works were reported wherein existing engineering properties of any soil are modified with the help of stabilizing agents and was termed as ground improvement or soil stabilization. The types of stabilizing agent used for any project would largely depend upon ease of availability as well as the economic budget of the project. The different stabilizing agents used in the past for improving the properties of soils were that of lime, cement, rice husk ash, construction and demolition waste, bagasse ash, fly ash, etc. The role of lime and curing temperature on the unconfined compressive strength of expansive soil was explored by [1]. Based on their studies, they reported that the effect of curing temperature was more significant in improving the unconfined compressive strength (UCS) of stabilized soil. Harichane et al. [2] studied the effect of natural pozzolana and lime stabilization on consistency limits and shear strength of cohesive soils. They concluded that improvement in shear strength of cohesive

D. Viswanath (✉) · M. N. Asha
Department of Civil Engineering (VTU RRC), CMR Institute of Technology, Bangalore, India
e-mail: divya.v@cmrit.ac.in

M. N. Asha
e-mail: asha.n@cmrit.ac.in

soils were observed when treated with natural pozzolana-lime combination rather than the individual effect. The effect of rice husk ash stabilization on laterite soil treated with cement content was explored by [3]. They inferred from their studies that the curing periods played a significant role in increasing the UCS and California bearing ratio (CBR) of stabilized soil. Yadu and Tripathi [4] explored the potential of non-grounded granulated blast furnace slag (GBS) in stabilizing soft soils and reported significant reduction in the swelling characteristic of the treated soft soil. The improvement in engineering and strength properties of lithomargic clay/soil blended with sand and coir mat were studied by [5]. They inferred from their experimental results that CBR values increased for coir and sand blended soil, while UCS values were nearly same for soil treated with coir alone and for the combination of soil, sand and coir. Rashid et al. [6] studied the feasibility of Xanthan gum as an eco-friendly stabilizer for improving laterite soil. Based on their experimental studies, they concluded that Xanthan gum has the potential to increase the shear strength of laterite soil at an optimum content of 1.5%. Onyelowe [7] explored the behavior of lateritic soil when stabilized with varying proportions of nanosized waste paper ash (NWPA) and inferred that 12% is the optimum content of NWPA which yields higher strength properties. They also concluded from their studies that NWPA can be used as a partial replacement in pavement subgrade or sub base.

In recent years, a new method of ground improvement, termed as bio-modification, has been gaining significance and attention of researchers. Bio-modification is the technique which makes use of microorganisms such as bacteria or enzymes in modifying the soil properties. The advantage of this method when compared with the conventional materials is its eco-friendly and sustainable characteristic. According to [8], the different types of bacteria that can be used for bio-modification are oligotrophic bacteria, aerobic gram-negative bacteria, cellulose degrading bacteria, gram-positive facultative aerobic and anaerobic bacteria, etc. The effect of Terrazyme treatment on strength properties of BC and red soil was studied by [9]. They inferred from their experimental studies that air drying is best suited for Terrazyme treatment and concluded that 200 ml per 3 cubic meter of soil is the optimum dosage. They also observed from their results that the swelling characteristics of BC soil improved with Terrazyme treatment. Viswanath and Asha [10] conducted experimental investigations to evaluate the potential of enzymatic stabilization in BC and lateritic soil. Based on their results, they reported improvement in mechanical properties of enzyme blended soils and concluded that the effect was maximum for BC soil especially at lower dosages. Shankar et al. [11] studied the effect on bio-enzymatic stabilization on strength and permeability characteristics of laterite soil and inferred that enzyme is ineffective in improving properties of cohesion-less soil. The potential of enzymatic cement stabilization in improving the strength of soft clay was explored by [12] and concluded that the combination of soft clay with 1% of cement and 0.06 ml/kg of enzyme was the best suited for achieving higher UCS. Ganapathy et al. [13] discovered that mountain soil with poor engineering parameters when treated with bio-enzyme can be used efficiently in the construction of light traffic earth roads, pedestrian walkways and bicycle tracks. The modification of clayey soil using enzymatic lime as stabilizer was experimentally investigated by [14]. They

reported that this method was more effective in improving the characteristics of soils with minimum 20% clay content. Khan et al. [15] studied the changes in strength and swelling properties of clayey soils when treated with three different types of bio-enzymes and concluded that bio-enzymes are effective in reducing the moisture absorption capacity of clayey soils.

In the present work, it is proposed to understand the feasibility of bio-modification on laterite soil and compare the results with the available literature. It is proposed to evaluate the strength parameters of enzyme blended laterite soil and compare the outcomes for different dosages and curing periods. Further, efforts have also been made to predict the type of soil which is most suitable for bio-modification.

2 Methodology

The methodology adopted in the present study are as follows:

- (a) Collection of laterite soil and characterization of the same for index and engineering properties.
- (b) Identification of trial dosages and curing periods for Terrazyme based on literature.
- (c) Experimental studies to evaluate the performance of Terrazyme on strength parameters.
- (d) Comparison of results from the current study with those available from literature and prediction of the type of soil most feasible for Terrazyme treatment.

3 Characterization of Materials

The laterite soil used in the present study was obtained from Mangalore district and was red in color. The specific gravity of the soil was found to be 2.68 and was classified as CI as per IS soil classification system. The natural laterite soil achieved a maximum dry density of 1.76 g/cc at an optimum moisture content of 18.75%. The strength parameters of soil are significant for any construction activity. In the present study, the engineering parameters of UCS and CBR of untreated laterite soil were evaluated and obtained as 250.38 kPa and 9.23%, respectively.

The enzyme used in the present study, Terrazyme, was obtained from a commercial supplier. The enzyme was dark brown in color and has specific gravity almost equal to that of water. It was formulated from vegetable extracts and was of non-toxic grade. The different dosages of Terrazyme used in the current study are 150 ml per cubic meter and 250 ml per cubic meter of soil.

4 Experimental Studies

The effect of bio-modification on lateritic soil is compared with reference to strength indices viz., unconfined compressive strength and California bearing ratio and is illustrated in the following sections.

4.1 Effect on Unconfined Compressive Strength

The natural laterite soil obtained at a depth of 2 m from top soil is pulverized and oven dried in the laboratory for 24 hours before the testing process. Such oven dried soil is first tested for its index and engineering properties before treating with Terrazyme. The soil is then blended with pre-fixed Terrazyme dosages of 150 ml/m³ and 250 ml/m³ and subjected to curing periods of 7 and 14 days. The samples were then analyzed for change in compressive strength parameters, and results are indicated in Fig. 1. It can be observed from Fig. 1 that varying dosages produce considerable enhancement of compressive strength along with increase in the axial strain capacity of enzyme blended soil. This could be due to the expulsion of moisture which results in a denser packing of soil particles. It can also be inferred from the results that the significance of curing period is noteworthy in improving the UCS of treated soil. The increase in UCS was of the order of 41% and 67.5% for 150 ml/m³ dosage after 7 days and 14 days of curing, respectively. This increase was observed to be at a higher range of 68.5% and 97.2% for 250 ml/m³ dosage after 7 days and 14 days of curing period, respectively.

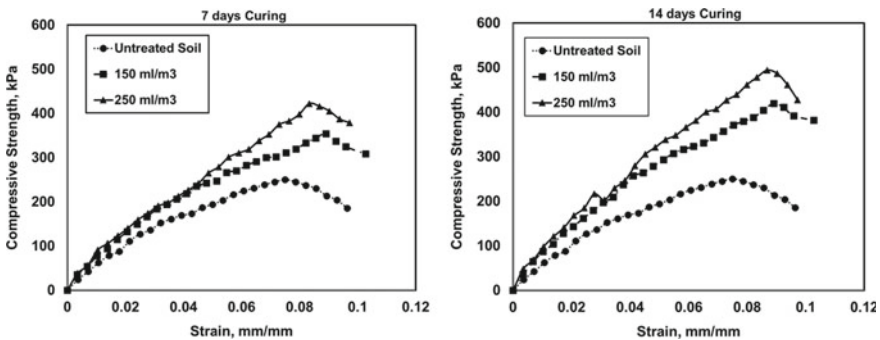


Fig. 1 Variation in UCS of treated and untreated lateritic soil

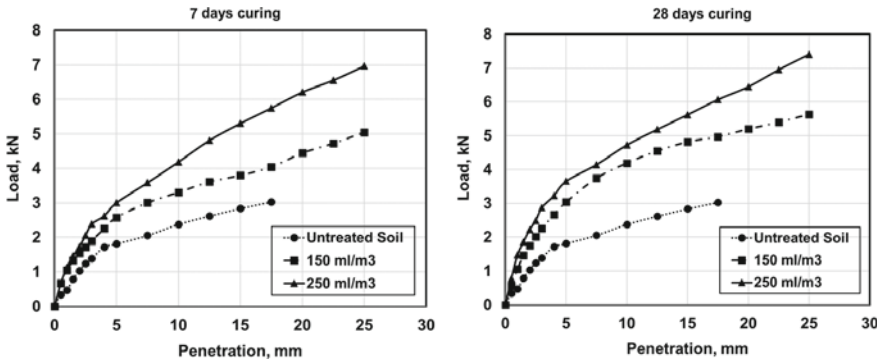


Fig. 2 Load penetration analysis of treated and untreated lateritic soil

4.2 Effect on Bearing Ratio

Figure 2 represents the results of load penetration tests carried out on natural and Terrazyme blended laterite soil. The results show larger load carrying capacity and thereby indicates a higher bearing ratio values for treated soil. With increasing dosages of Terrazyme, the moisture absorption phenomenon becomes expeditious and thus contributes to higher strength. As per IRC 37-2012, bearing ratio results are very important in the design and construction of flexible pavements. The curing period is also found to play an active role in improving the bearing ratio of Terrazyme blended soil. Hence, the higher range of CBR obtained during bio-modification implies higher probability of application for subgrade improvement.

5 Comparative Studies

In this session, comparative studies have been performed to collate the effect of different types of stabilization on unconfined compressive strength of laterite soil. Table 1 shows the general properties of the various laterite soil, and the details of literature taken for comparative analysis. Based on the results from present work and literature, variation in UCS at 7 days curing period is represented in Fig. 3.

The analysis indicates that strength of laterite soil can be best improved when coir fiber/mat alone or in combination with sand is added as a stabilizer. This is because sand being cohesion-less enables uniform mixing of the coir fibers. However, since sand is an expensive ingredient, alternative solutions need to be explored. The second-best stabilizing agent is observed as Terrazyme of dosage 250 ml per cubic meter which is used in the present study. Depending on the availability, cost and ease of logistics of securing the stabilizing agent, selection can be made between sand-coir combination and Terrazyme.

Table 1 Literature details for comparative study

General properties of laterite soil (LS)	Stabilizer used	Literature
Specific gravity—2.45–2.7 Plasticity index—10–34 Maximum dry density—1.3–1.9 g/cc Optimum moisture content—13–19%	^a Sand and coir fiber of 175 mm length	[5]
	Xanthan gum	[6]
	Cement and rice hush ash (RHA)	[3]
	Nanosized waste paper ash (NWPA) taken for curing period of 7 days	[7]
	Terrazyme	[10] and present work

^aNot subjected to curing period as there is no pozzolanic activity

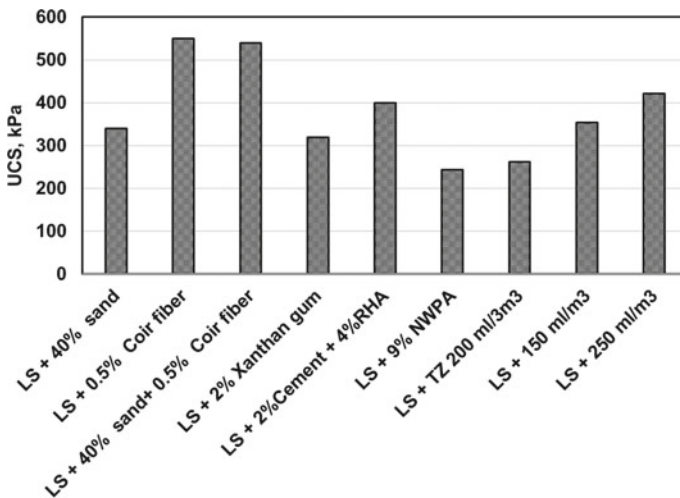


Fig. 3 Comparative study of UCS variation with literature

6 Conclusions

1. The effect of bio-modification in laterite soil was carried out at varying dosages and curing periods in the current study, and following conclusions were arrived at:
2. Terrazyme treatment is effective in improving the unconfined compressive strength and CBR characteristics of laterite soil.
3. Curing period plays a vital role in improving the engineering characteristics of enzyme treated soil.
4. Comparative studies indicate that desirable properties can be achieved at lower enzyme dosages.

References

1. Nasrizar AA, Muttharam M, Illamparuthi K (2010) Role of lime content on soil-lime reaction under thermal curing. In: Proceeding Indian geotechnical conference—2010, GEOTrendz, 16–18 595–598
2. Harichane K, Ghrici M, Kenai S, Grine K (2011) Use of natural pozzolana and lime for stabilization of cohesive soils. *Geotech Geol Eng* 29:759–769. <https://doi.org/10.1007/s10706-011-9415-z>
3. Alhassan M, Mustapha A (2007) Effect of rice husk ash on cement stabilized laterite. *Leonardo Electron J* 47–58
4. Yadu L, Tripathi RK (2013) Effects of granulated blast furnace slag in the engineering behaviour of stabilized soft soil. *Proc Eng* 51:125–131. <https://doi.org/10.1016/j.proeng.2013.01.019>
5. Shivashankar R, Ravi Shankar AU, Jayamohan J (2015) Some studies on engineering properties, problems, stabilization and ground improvement of lithomargic clays. *Geotech Eng* 46:68–80
6. Rashid ASA, Latifi N, Meehan CL, Manahiloh KN (2017) Sustainable improvement of tropical residual soil using an environmentally friendly additive. *Geotech Geol Eng* 35:2613–2623. <https://doi.org/10.1007/s10706-017-0265-1>
7. Onyelowe K (2017) Nanostructured waste paper ash stabilization of lateritic soils for pavement base construction purposes. *Electron J Geotech Eng* 22:3633–3647
8. Ivanov V, Chu J (2008) Applications of microorganisms to geotechnical engineering for bioclogging and biocementation of soil in situ. *Rev Environ Sci Biotechnol* 7:139–153. <https://doi.org/10.1007/s11157-007-9126-3>
9. Ramesh HN, Sagar SR (2015) Effect of drying on the strength properties of terrazyme treated expansive and non-expansive soils. In: 50th Indian geotechnical conference Pune, Maharashtra, India
10. Viswanath D, Asha MN (2020) Experimental investigations on bio-modified soil. *Geotechnical Special Publication*
11. Shankar AUR, Rai HK, Mithanthaya RI (2009) Bio-enzyme stabilised laterite soil as a highway material. *J Indian Roads Congr Pap* No 553:143–151
12. Thomas AG, Rangaswamy BK (2019) Strength behavior of enzymatic cement treated clay. *Int J Geotech Eng* 00:1–14. <https://doi.org/10.1080/19386362.2019.1622854>
13. Ganapathy GP, Gobinath R, Akinwumi II et al (2017) Bio-enzymatic stabilization of a soil having poor engineering properties. *Int J Civ Eng* 15:401–409. <https://doi.org/10.1007/s4099-016-0056-8>
14. Eujine GN, Chandrakaran S, Sankar N (2017) Accelerated subgrade stabilization using enzymatic lime technique. *J Mater Civ Eng* 29:04017085. [https://doi.org/10.1061/\(asce\)mt.1943-5533.0001923](https://doi.org/10.1061/(asce)mt.1943-5533.0001923)
15. Khan TA, Taha MR, Khan MM, et al (2020) Strength and volume change characteristics of clayey soils: performance evaluation of enzymes. *Minerals* 10. <https://doi.org/10.3390/min10052>

Cropping Pattern Decision for a Canal Distributary for Varied Discharge Using Linear Programming Approach



B. R. Ramesh, S. B. Ganesh Kumar, and H. J. Surendra

1 Introduction

Agriculture has a major contribution towards the Indian economy as it contributes 17% of total GDP [1]. By improving the canal irrigation efficiency between 30 and 65% based on project location and management policy, the agricultural production was substantially increased in India [2]. However, the canals under some south Indian command area were found to be working under low efficiency [3]. Poor utilization of the irrigation facility results in a low overall efficiency of irrigation projects [4]. Therefore, to increase the overall efficiency of canal, structural and non-structural interventions are necessary [5]. Structural interventions involve large expenditures and time for implementation [6]. Therefore, non-structural intervention methods are economical, easy to adapt. Optimization of cropping pattern along with the constraints of land and water availability is one of the non-structural intervention techniques [7]. Hence, this paper discusses the optimal cropping pattern design to improve crop production with available water resources.

An efficient irrigation system ensures to meet the pre-determined goal of serving all the canal network receives pre-determined and designed water distribution. However, in a real-time scenario, tail-enders are always deprived of sufficient water due to various reasons [8]. An optimization model is developed for cropping pattern prediction in the canal command area, which includes equitable water allocation for

B. R. Ramesh

Department of Civil Engineering, NMIT, Bangalore, Karnataka, India
e-mail: ramesh.br@nmit.ac.in

S. B. G. Kumar

Department of Civil Engineering, EWCE, Bangalore, Karnataka, India

H. J. Surendra (✉)

Department of Civil Engineering, ATRIA Institute of Technology, Bangalore, Karnataka, India
e-mail: surendra@atria.edu

the head, middle and the tail portion which intern minimizes surplus water supply to non-essential reaches and maximizes cultivation and net returns. The paper concludes linear programming is one of the best tools for optimal allocation of land and water resources [9].

The linear model for the optimization and prediction of cropping pattern is quite popular because of its simplicity [10]. Developed a linear optimization model (LOM) which predicts a mixed-mode of vegetables and cereals to maximize the profit. Artificial neural network (ANN) is used [11] to solve the LOM problems with objective function related to cropping, and it performs better as compared to the analytical model [12] used a genetic algorithm (GA) along with LOM for the prediction of optimal cropping pattern. The finding revealed that GA is suitable as a nonlinear optimization technique for crop pattern planning; however, this poses an additional computational complexity as compared to the LOM. A mixed optimal fuzzy linear programming model (MOFLP model) was developed to determine the cropping pattern with the objective function of maximization of the net profit under the conflicting context and constraints of unique parameters of labour, water, manure and production cost [12].

The use of the binary particle swarm optimization algorithm is found in the work for the study on irrigation lateral canals to arrive an efficient water resource utilization and to maintain water distribution equity [13]. A study concluded that the GenSRT technique suggests an optimal approach of the resource allocation for the higher net profit as compared to other methods used for the cropping pattern prediction [14].

Developed the optimal cropping pattern model, where results were compared with existing adopted cropping patterns in the command area, and it is analysed that the farmers are getting less profit by existing cropping patterns [15]. A suitable model for the semi-arid region and surface irrigation scheme with rotational water supply is proposed optimized AWA model that takes constraints of the crop, soil, geographic conditions, and the irrigation system and suggest the allocation of the water supply with the balance of the land allocation for the matching crops [16]. This type of approach is very much useful for the irrigation and agricultural sector to improve agricultural activity to obtain optimal productivity [17]. A similar type of work is carried out by different authors for different localities to provide information regarding land suitability for different types of crops [18–23].

Hence, this paper presents a methodology of use of linear programming technique for suggesting cropping pattern with an objective of maximization of the yield of the command area through with the available water for the season to a distributary canal and constraints thereon for a selected distributary of Harihar Branch canal of Bhadra command area of Karnataka state-India.

2 Materials and Methods

2.1 Study Area

The Bhadra command area lies between $14^{\circ}21' - 14^{\circ}30' N$ latitude and $75^{\circ}47' - 75^{\circ}57' E$ longitude and falls under the Land of Agroclimatic reach-4 of Karnataka (India). The Canal irrigation system is having the Right and Left bank canal. The Right Bank Canal runs 103 km with a culturable command area of 75,311 ha. Among the four-branch canal, from the right branch canal, the Davangere branch canal takes off at 103 km. The canal runs for a distance of 99 km with a command area of 45,278 ha. It has a discharging capacity of 33.4 cumecs with 27 distributaries. The Harihar Branch Canal takes off at 65 km of the Davangere Branch Canal. This runs for a total distance of 22 km with 18 distributaries whose total length is 80.40 km with a discharging capacity of 10.307 cumecs to the Command Area of 14,996 ha. The study area from the present study is under the 8th distributary of the head portion of the Harihar branch canal.

The soil characteristics of the right bank canal comprise 8.5% Black Soil and 93.5% Red soil, whereas in the left bank canal 7% black soil and 93% Red Soil. The command area receives 58% southwest monsoon and 22% northeast monsoon and 20% of rainfall during summer. The average rainfall of Davangere station is 715 mm, and of Harihar station is 687 mm. The area is identified as dry Argo reach-4 and is categorized as drought-prone [24].

The Harihar branch command area falls under a plane region, covered with red sandy soil and black soil with a composition of red loams, medium black, and red sandy soils [25]. The depth of groundwater in the command area is about 35 to a maximum of 200 m below ground level. Artificial neural network implementations with linear programming model provides a better solutions [26].

2.2 Deficit–Surplus Analysis in the Distributary Based on the Operation Policy

In the Bhadra command area, the number of days the main canal has to run is decided by the irrigation department after ascertaining water available in the Bhadra dam at the beginning of the Kharif and Rabi season. The water is led to the canals coming under the main canal and is run in rotation (generally 10 days ON and 10 days OFF). The crops to be grown in the command area will be notified by the department after knowing the days of the canal to be run during a particular season.

The water requirement of the crops notified during the study period is computed using the Penman Montaitth method considering the various meteorological data for 2001–2014. The main canal under this system is working for 110 days with 12 rotation. The distributary under the present study is working under operation during 2, 4, 6, 8, 10 of each 10 days and 12th rotation for 5 days. The water to be supplied

for the crops shall match for successive two rotation, as one of the rotations will be OFF. Hence, the water requirement for the notified crops were presented in the Table 1.

The water supplied in the distributary during the above rotation must match the water requirement of the crop. However, it is observed that there was a deficit supply in all rotation as shown in the Table 2. Hence, there is a need for the development of the model for cropping pattern.

To properly distribute the water till tail end, 8th distributary (which is taken for the present study) is divided into Head reach, Middle reach, a Tail reach. There are of the 68 outlets in the distributary, based on outlets configuration and the area served by each outlet, the canal is trifurcated in to head reach (total discharge supplied in

Table 1 Crop water requirement for different rotations based on the operation policy

Water required in rotation no	Paddy	Sugarcane	Plantation	Semidry
2	0.299	0.240	0.178	0.185
4	0.338	0.277	0.205	0.243
6	0.338	0.285	0.209	0.249
8	0.309	0.282	0.202	0.212
10	0.220	0.150	0.135	0.094
12	0.071	0.040	0.042	0.024
Total	1.574	1.273	0.973	1.008

Depth of water requirement in m.

Table 2 Surplus/Deficit supply in the distributaries for notified crops during the rotation operation

Rotation operation	Irrigation water requirement for notified crops in ha-m					IWR as per operation policy	Surplus/Deficit
	Paddy	Sugarcane	Plantation	Maize	Total		
Rotation-1	0	84.70	20.12	47.73	152.55	OFF	
Rotation-2	0	94.02	22.35	63.62	179.99	377.03	-217.97
Rotation-3	0	94.02	22.35	80.67	197.04	OFF	
Rotation-4	0	104.73	24.87	91.87	221.48	457.98	-298.92
Rotation-5	0	111.87	26.55	98.08	236.51	OFF	
Rotation-6	0	111.87	26.55	98.08	236.51	470.45	-311.39
Rotation-7	0	111.12	25.95	96.88	233.94	OFF	
Rotation-8	0	110.37	25.35	95.67	231.38	437.22	-278.15
Rotation-9	0	110.37	25.35	70.12	205.83	OFF	
Rotation-10	0	86.34	23.31	54.16	163.82	224.71	-65.65
Rotation-11	0	31.16	10.64	19.10	60.90	OFF	
Rotation-12	0	31.16	10.64	19.10	60.90	60.90	18.63

this reach = 0.582 cumecs. Total command area served in this reach = 544 Ha), middle reach (total discharge supplied in this reach = 0.483 cumecs. Total command area served in this reach = 439 Ha), tail reach (total discharge supplied in this reach = 0.782 cumecs. Total command area served in this reach = 863 Ha).

3 Linear Model for Cropping Pattern Optimization

Proper cropping patterns should be designed to bridge the gap of deficit in the water supply so that the distributary functions at optimal conditions.

For optimal cropping pattern in the command area of the distributary, the main objective is to maximize the yield or production in the command area with the available resources in the crop duration. Mathematically, it is represented as:

3.1 Objective Function

$$\text{Max}Y = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K Y_i A_{ijk} \tag{1}$$

where Y_i = Production per hectare of crop area of i th Crop, A_i = Area of the i th crop.

I = number of crops grown in season (1 = Paddy, 2 = Sugarcane, 3 = Plantation (Areca nut), 4 = Maize, 5 = Aerobic rice, 6 & 7 is Baby corn (double crop)).

J = Reach of the irrigation crop (1 = Head reach, 2 = Middle reach, 3 = Tail reach).

K = Irrigation crop water requirement for required for different rotation.

(1 = First rotation, 2 = second rotation, 3 = third rotation, 4 = fourth rotation, 5 = fifth rotation, 6 = sixth rotation).

A_{ijk} = area of i th crop under j th reach for k th rotation in (ha)

3.2 Constraints of Water Availability

The total water availability for command area (A) in the respective distributary is fixed, and it is divided into three reaches, J = Head(H), Middle(M), Tail(T), the mathematical representation of area constraint is given in equation

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K A_{ijk} IWR_{ik} \leq WAR_{jk}$$

where IWR_{ik} is the depth irrigation water requirement in m.

WAR_{jk} is the volume of water available at the distributary head for rotation at different reach in ha-m.

3.3 Constraints for Area Availability

The total area under various crops during any period of irrigation shall be less than or equal to the total notified cropping pattern from the department (A) in the distributary. The command area is fixed and is divided into three reaches, j is Head (H), Middle (M), and Tail (T) reach, and therefore, the mathematical representation of area constraint is given in equation

$$\sum_{i=1}^I \sum_{j=1}^J A_{ij} \leq A \tag{3}$$

where A = total notified command area from the department of the all the crop in the command area in ha.

3.4 Crop Priority by the Farmer as Per Soil Health Card Report Constraints

Farmer prefers to grow certain crops in their land by applying extra fertilizer required for his preferred crop. The farmer consults in regards to extra fertilizer application for his preferred crop from the agricultural department. The total area under each preferred crop shall be less than the notified crop area (Generally, farmers prefer either wet crop or plantation crop). The constrain can be written as

$$\sum_{i=1}^I APF_i \leq \sum_{i=1}^I A_i \tag{4}$$

where APF_i = Area of i th crop preferred by the farmer.

3.5 Crop Area Constraints in Respect to Natural Soil

The command area is suited for a certain crop, which is suggested for a particular crop by the agricultural department for its natural fertility using nominal fertilizer dosage, which is to be used.

The soil fertility and land suitability based on multicriteria analysis are carried out by using GIS(AymenAL-Taani 2020), by considering this constraint equation can be written as

$$\sum_{i=1}^I ASF_i \leq \sum_{i=1}^I A_i \tag{5}$$

where ASF_i = Area of ith crop suggested based on soil fertility with the optimal dosage of the fertilizer.

3.6 Crop Area Constraints

The command area is already having certain plantation crops. In deciding the cropping pattern, we cannot alter the area under plantation. Therefore, in this model, water required for plantation crops during the rotation period cannot be altered. In the command area, there are two sugar factories, hence to a maximum of 60% area is suggested for sugarcane. The constraint for sugarcane for head and middle reach is fixed to a maximum of 25% of the notified sugarcane crop area. As the farmer prefers to grow wet crop and plantation at the head, dry crop like maize is not preferred at the head. However, the maximum area up to 75% of the notified dry crop area is suggested for the middle and tail, reach. Aerobic rice is a better choice for the farmers who prefer to grow the wet crop in place of sugarcane or dry crop, as it yields more and profitable crop with less water requirement. In the present study, minimum of 10% of the notified dry crop to 100% is suggested. If any extra water is available in any rotation baby corn (short duration), crop can be used as a double crop within the season is suggested in the constraint as below

$$\sum_{i=1}^I A_{ij} \leq \sum_{i=1}^I A_i \times CF_i \tag{6}$$

where CF_i = crop percentage factor for ith crop.

3.7 Fertilizer Constraints

The fertilizer availability for the command area is fixed; as the Government allocates fertilizer and the seeds to each district in the state, this constrains is to be used in cropping pattern decision. The fertilizer constraint is written as below

$$\sum_{i=1}^I \sum_{j=1}^J A_{ij}FR_{ij} \leq TFA \tag{7}$$

where TFA = Total Available Fertilizer for the distributary command area.

FR_i = Fertilizer requirement for ith crop for jth reach of the distributary command area.

4 Results and Conclusion

It is identified from the crop water requirement studies that the water supplied to the distributaries during the rotation period cannot cater to the demand of the notified crop in the season. Also, the water availability in the reservoir will not be always 100% varies during years; hence, in the present study, water availability to the distributary is varied between 100–50% for the same rotation period to understand how much of notified cropping can be achieved and the consequent economic benefit.

The optimal cropping areas for different types of crops under LP modelling for a varied percentage of water availability in each reach (head, middle and tail reach) is done separately, and the total crops suggested for the command area is presented in Table 3, and there on the yield that can be achieved is presented in Table 4.

The predicted percentage of cropping pattern at the different reaches for the varied percentage of water from 100 to 50% is presented in Fig. 1, and it can be seen a

Table 3 Predicted cropping pattern through LP model in the 8th B distributary of Harihar branch canal of the Bhadra command area

WA-TOTAL	Paddy	Sugarcane	Plantation	Maize	Aerobic Rice	Babycorn-1	Babycorn-2	Total
100	0.00	302.51	250.74	0.00	78.14	0.00	28.41	659.81
95	0.00	274.50	250.74	0.00	78.14	0.00	27.96	631.34
90	0.00	246.48	250.74	0.00	78.14	0.00	27.52	602.89
85	0.00	218.46	250.74	0.00	78.14	0.00	27.08	574.43
80	0.00	190.44	250.74	0.00	78.14	0.00	26.64	545.97
75	0.00	162.43	250.74	0.00	78.14	0.00	26.20	517.51
70	0.00	134.41	250.74	0.00	78.14	0.00	25.76	489.06
65	0.00	106.39	250.74	0.00	78.14	0.00	25.32	460.60
60	0.00	78.37	250.74	0.00	78.14	0.00	24.88	432.14
55	0.00	50.35	250.74	0.00	78.14	0.00	24.44	403.68
50	0.00	23.07	250.74	0.00	78.14	0.00	20.00	371.95

Area in hectares.

***WA = % of Available Water in the Distributary.

Table 4 Predicted yield from the cropping area with help of linear programming model in the distributary (Yield in Quintal & for Sugarcane in Tonne)

WA-Total	Paddy	Sugarcane	Plantation	Maize	Aerobic rice	Babycorn-1	Babycorn-2	Total
100	0.00	21,478.49	3134.25	0.00	4297.92	0.00	539.81	29,450.46
95	0.00	19,489.24	3134.25	0.00	4297.92	0.00	531.32	27,452.72
90	0.00	17,499.98	3134.25	0.00	4297.92	0.00	522.96	25,455.11
85	0.00	15,510.73	3134.25	0.00	4297.92	0.00	514.60	23,457.49
80	0.00	13,521.48	3134.25	0.00	4297.92	0.00	506.23	21,459.88
75	0.00	11,532.22	3134.25	0.00	4297.92	0.00	497.87	19,462.26
70	0.00	9542.97	3134.25	0.00	4297.92	0.00	489.51	17,464.65
65	0.00	7553.71	3134.25	0.00	4297.92	0.00	481.15	15,467.04
60	0.00	5564.46	3134.25	0.00	4297.92	0.00	472.79	13,469.42
55	0.00	3575.20	3134.25	0.00	4297.92	0.00	464.43	11,471.81
50	0.00	1638.06	3134.25	0.00	4297.92	0.00	379.98	9450.22

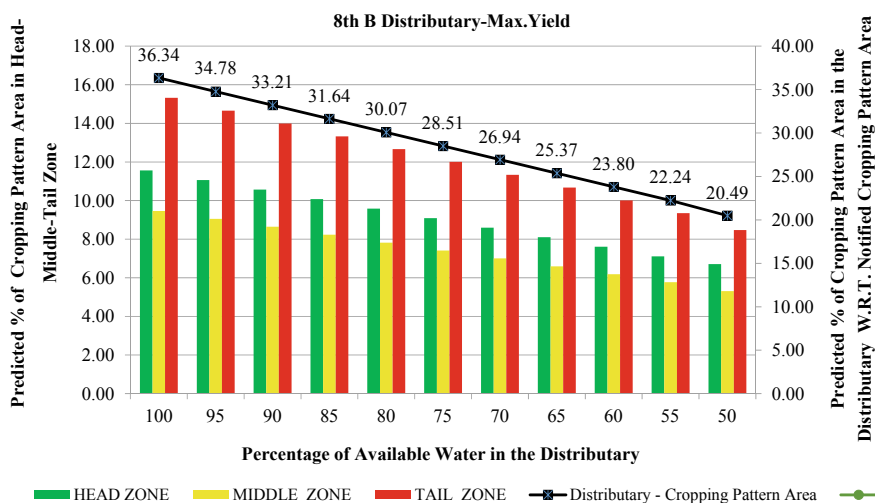


Fig. 1 Predicted cropping pattern area covered in the distributary by linear programming

maximum of 36.34% of the notified area when water availability is 100%–20.49% of the notified area when water availability is 50% can only be achieved.

The predicted economic benefit concerning the notified cropping pattern is for a carried percentage of water availability from 100 to 50% is presented in Fig. 2. The percentage of economic benefit that can be achieved carried from 57.7 to 39.9%.

In conclusion, linear programming can be used as a versatile tool for fixing the cropping pattern for a command area for varied water distribution and cropping patterns.

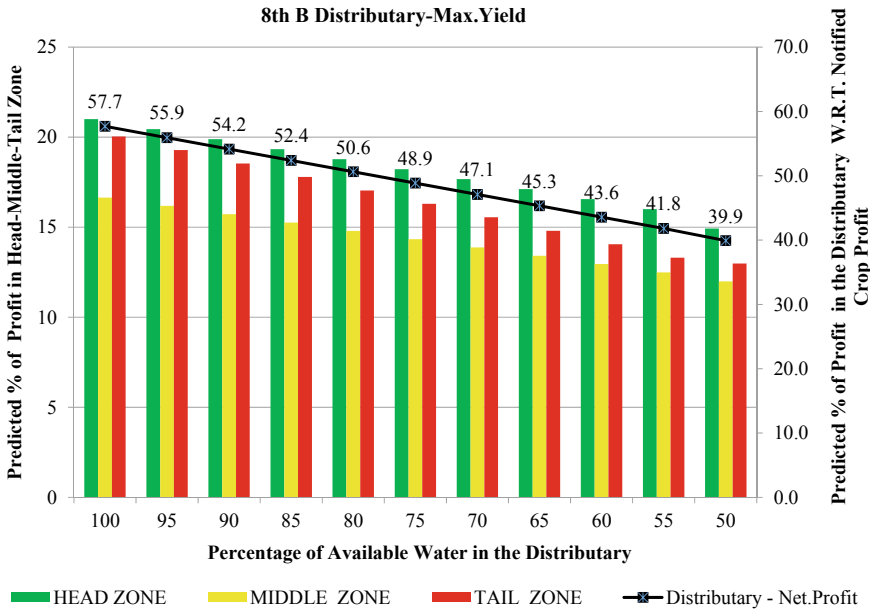


Fig. 2 Predicted net returns from the cropping pattern area covered in the distributary by linear programming

References

1. Hwa E-C (1988) The contribution of agriculture to economic growth: some empirical evidence. *World Dev* 16(11):1329–1339. [https://doi.org/10.1016/0305-750X\(88\)90208-2](https://doi.org/10.1016/0305-750X(88)90208-2)
2. Central Water Commission (2014) Guidelines for improving water use efficiency in irrigation, domestic & industrial sectors. Performance Overview and Management Improvement Organization, Central Water Commission, Govt. of India, R. K. Puram, SewaBhawan, New Delhi-110066. http://www.mowr.gov.in/sites/default/files/Guidelines_for_improving_water_use_efficiency_1.pdf
3. Ramesh BR, Venugopal K, Karunakaran K (2009) Zero-one programming model for daily operation scheduling of irrigation canal. *J Agric Sci* 1(1):13
4. X-P. Deng, L. Shan, H. Zhang, N.C. Turner, Improving agricultural water use efficiency in arid and semiarid areas of China. *Agric. Water Manag.* **80**, 1–3 (2006). <https://doi.org/10.1016/j.agwat.2005.07.021>
5. Rosegrant MW, Binswanger HP (1994) Markets in tradable water rights: potential for efficiency gains in developing country water resource allocation. *World Dev* 22(11):1613–1625. [https://doi.org/10.1016/0305-750X\(94\)00075-1](https://doi.org/10.1016/0305-750X(94)00075-1)
6. Lam WF (1996) Improving the performance of small-scale irrigation systems: the effects of technological investments and governance structure on irrigation performance in Nepal. *World Dev* 24(8):1301–1315. [https://doi.org/10.1016/0305-750X\(96\)00043-5](https://doi.org/10.1016/0305-750X(96)00043-5)
7. Singh A, Panda SudhindraNath (2013) Optimization and simulation modelling for managing the problems of water resources. *Water Resour Manage* 27(9):3421–3431. <https://doi.org/10.1007/s11269-013-0355-7>
8. Kumar D (2017) Canal Based Irrigation Scheduling and Conjunctive Water Use Planning for Optimal Cropping Pattern In Selected Panam Canal Command In Panchmahal District, Gujarat.

- College of Agricultural Engineering And Technology Anand Agricultural University, Godhra, Phd Diss
9. Panda RK, Panigrahy N, Mohanty S, Brahmanand PS, Kumar A, Raju PV, Rao VV (2018) Optimal cropping pattern design for a major distributary of Hirakud canal command in India. *Sustain. Water Resour. Manag.* 4(4):1051–1062. <https://doi.org/10.1007/s40899-018-0241-9>
 10. J Jebelli, B. Paterson, A. Abdel wahab, A linear programming model to optimize cropping pattern in small-scale irrigation schemes: an application to Mekabo Scheme in Tigray, Ethiopia. *Int. J. Environ. Agric. Res.* 2(8), 24–34 (2016)
 11. K.S. Raju, D. Nagesh Kumar, Irrigation planning using genetic algorithms. *Water Res. Manag* 18(2), 163–176 (2004)
 12. D. Khare, M.K. Jat, J. Deva Sunder, Assessment of water resources allocation options: conjunctive use planning in a link canal command. *Resour. Conserv. Recycle.* 51(2), 487–506 (2007). <https://doi.org/10.1016/j.resconrec.2006.09.011>
 13. Y.P. Mathur, R. Kumar, A. Pawde, A binary particle swarm optimisation for generating optimal schedule of lateral canals. *IES J Part A: Civil & Structural Engineering* 3(2), 111–118 (2010). <https://doi.org/10.1080/19373261003619936>
 14. F.-S. Juan, A. Gonzalez-Sanchez, M. Larre, in *A New Method for Optimal Cropping Pattern*. Mexican International Conference on Artificial Intelligence (Springer, Berlin, Heidelberg, 2009), pp. 566–577
 15. A. Rath, S. Samantaray, P.C. Swain, Optimization of the Cropping Pattern Using Cuckoo Search Technique in *Smarter Techniques for a Smarter Planet* (Springer, Cham, 2019), pp. 19–35. https://doi.org/10.1007/978-3-030-03131-2_2
 16. I.K. Smout, S.D. Gorantiwar, Multilevel approach for optimizing land and water resources and irrigation deliveries for tertiary units in large irrigation schemes. I: method. *J. Irrig. Drain. Eng.* 131(3), 254–263 (2005). [https://doi.org/10.1061/\(ASCE\)0733-9437\(2005\)131:3\(254\)](https://doi.org/10.1061/(ASCE)0733-9437(2005)131:3(254))
 17. A.L.-T. Aymen, Y. Al-husban, I. Farhan, Land suitability evaluation for agricultural use using GIS and remote sensing techniques: the case study of Ma'an Governorate, Jordan. *Egyptian J. Remote Sens. Space Sci.* (2020). <https://doi.org/10.1016/j.ejrs.2020.01.001>
 18. Feizizadeh B, Blaschke T (2013) Land suitability analysis for Tabriz County, Iran: a multi-criteria evaluation approach using GIS. *J Environ Planning Manage* 56(1):1–23. <https://doi.org/10.1080/09640568.2011.646964>
 19. Akinci, Halil, Ayşe YavuzÖzalp, and BülentTurgut, Agricultural land use suitability analysis using GIS and AHP technique. *Comput. Electron Agric* 97, 71–82 (2013). <https://doi.org/10.1016/j.compag.2013.07.006>
 20. Steiner F (1983) Resource suitability: methods for analyses. *Environ Manage* 7(5):401–420. <https://doi.org/10.1007/bf01867120>
 21. S.G. Yalew, A. van Griensven, M.L. Mul, P. van der Zaag, Land suitability analysis for agriculture in the Abbay basin using remote sensing, GIS and AHP techniques. *Model. Earth Syst. Environ.* 2(2), p. 101 (2016). <https://doi.org/10.1007/s40808-016-0167-x>
 22. N. Walke, G.P.O. Reddy, A.K. Maji, S. Thayalan, GIS-based multicriteria overlay analysis in soil-suitability evaluation for cotton (*Gossypium spp.*): a case study in the black soil region of Central India. *Comput. Geosci.* 41, 108–118 (2012). <https://doi.org/10.1016/j.cageo.2011.08.020>
 23. A. Rahman, A.E. Mohamed, A. Natarajan, R. Hegde. Assessment of land suitability and capability by integrating remote sensing and GIS for agriculture in Chamarajanagar district, Karnataka, India. *Egyptian J. Remote Sens. Space Sci* 19(1), 125–141 (2016). <https://doi.org/10.1016/j.ejrs.2016.02.001>
 24. G.S. Srinivasareddy, H.S. Shivakumarnaiklal, N.G. Keerthy, P. Garag, E.P. Jothi, O. Challa, Drought vulnerability assessment in Karnataka: through composite climatic index. *MAUSAM* 70(1), 159–170 (2019). <https://metnet.imd.gov.in/mausamdocs/57012.pdf>
 25. Ground water information booklet for Davangere district, Karnataka, http://cgwb.gov.in/District_Profile/karnataka/DAVANAGERE_BROCHURE.pdf
 26. A.K.F. Hassan, F.K. Fathallah, Artificial neural network implementation for solving linear programming models. *J. Kufa Math. Comput.* 2(1), 113–121 (2014)

Energy-Aware Task Scheduling Approach Using DVFS and Particle Swarm Optimization for Heterogeneous Multicore Processors



K. Siddesha and G. V. Jayaramaiah

1 Introduction

Nowadays, we are noticing, the drastic growth in various technologies. This has made the necessity of high speed and better performance computing systems. For high-performance computing, most notably heterogeneous computing environments are used. The proper utilization of the resources and energy consumption is always a challenge in these systems. The development of task scheduling algorithm that allocate each task to its suitable machine, such that their execution and precedence requirements are satisfied, is an option to achieve the objectives like improved system performance and minimized energy consumption. Usually, attaining best solutions for the task assignment problem in any computing system is assumed to be NP-hard problem [1]. In this regard, there is need of developing certain precise algorithms to solve this kind of problems.

In particular, the real-time embedded systems have gained importance in the field of multicore platform due to the present demand and high-speed computations in various applications [2]. The multicore processors provide significant system performance by accomplishing the tasks efficiently. The overall power consumption in processors is addition of dynamic power and static power. Switching activity of the processor leads to dynamic power consumption and leakage current contributes to static power consumption [3]. However, the dynamic power consumption is the main component in the total power consumption [4].

K. Siddesha (✉)

Department of Electronics and Communication Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India
e-mail: siddesha.ec@drait.edu.in

G. V. Jayaramaiah

Department of Electrical and Electronics Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India

Multiprocessor systems are grouped into homogeneous and heterogeneous. For the performance requirement always, heterogeneous systems are considered [5]. Minimizing the power consumption of any high-performance computing system is always a challenging and can be considered as a hot topic of research. Many methods have been identified to tackle the issue of power consumption, most of these techniques target only for homogeneous processors. Meanwhile, the techniques that are suitable for homogeneous domains may not be suitable for heterogeneous domains.

In most of the embedded applications, dynamic voltage and frequency scaling (DVFS) is a known used technique to optimize the power consumption issue [6]. In DVFS method, attention is given to processor's voltage and clock frequency, based on some pre-settings. These voltage and clock frequency are scaled together to attain the performance and power requirements. Overall, the main of DVFS technique is to provide, several discrete voltage levels, and the processor can independently and dynamically adjust its own voltage supply to save energy. But when the number of executing tasks are increased, then the DVFS-based schemes fail to handle the situation and degrades the system performance in terms of resource utilization and energy consumption. So, in this work, we tried to address the negative impact of DVFS-based schemes and proposed a combined energy-aware scheduling approach using DVFS and particle swarm optimization.

In this paper, we discuss a novel method of processor scheduling for DVFS-enabled heterogeneous systems using particle swarm optimization (PSO) algorithm, and the proposed method tries to improve the system performance with respect to resource utilization and energy consumption comparatively. The proposed method attains the aim of solving the problem of processor scheduling in a heterogeneous computing environment, under the constraints of meeting all task requirements with reduced energy consumption.

The main objectives of this work are as follows.

- Study about existing scheduling techniques to minimize the power consumption for processor-based systems.
- Formulating the DVFS system model.
- Proposing a combined scheduling approach using DVFS and particle swarm optimization (PSO) for DVFS-enabled heterogeneous multicore processors.
- Achieving the optimality of the proposed approach with the help of comprehensive experiments and comparisons.

The rest of the manuscript is organized as follows. Section 2 presents related work study of energy-aware task scheduling. Section 3 presents the system models, and an approach to overcome energy consumption issues. Section 4 provides the performance analysis of the proposed model. Lastly, Sect. 5 explains conclusion and also the future work direction.

2 Related Works

The research work on optimizing the energy consumption of heterogeneous multicore embedded systems has made considerable progress. In this section, we highlight some of the research works related to energy-aware task scheduling algorithms.

Lin et al. [7] discussed regarding DVFS-based scheme for power management and reported that the power consumption by processor core is always, assigned task dependent. Due to this, an imbalance occurs in energy consumption and performance. In order to overcome this issue, authors introduced DVFS-based approach for task scheduling with task configurations such as time constraints, task arrival time, task completion time, task execution order and processing speed of the task, etc.

Chengming et al. [8] presented two energy-aware multiprocessor scheduling algorithms. The algorithms are named as reinforcement learning (RL) algorithm and mathematical morphology scheduling (MMS) algorithm. These algorithms are used to schedule, time constrained directed acyclic graph (DAG) tasks to DVFS-enabled embedded multiprocessor system. The performances of these algorithms were analyzed with the genetic algorithm, and the simulation results shown that the proposed algorithms are more energy efficient when compared with genetic algorithm.

Hua et al. [9] developed a scheduling method to allocate set of periodic tasks to heterogeneous processors using ant colony optimization (ACO) algorithm. The scheduling method worked successfully to solve the assigned NP-hard problem, by meeting all timing constraints of the system. In their evaluation work, authors also shown that ACO-based algorithm performs well in getting optimal solutions over genetic algorithm (GA) and LP-based approaches.

Mo et al. [10] presented a concept that, a huge task can be divided into multiple sub-tasks to achieve the desired Quality of Service (QoS). Also, the authors have suggested, combining DVFS with task allocation and proper adjustment helps in improving the power performance. In [11] also, presented a modified ant colony optimization algorithm for task scheduling in a multiprocessor environment. The proposed work has made an attempt to improve the energy consumption of the system.

Amjad et al. [12] formulated the real-time task scheduling problem in DVS-enabled multiprocessor systems. Authors have imposed improved capabilities to genetic algorithm to make it hybrid. An effective study has done to analyze the performance of the proposed approach over existing genetic algorithm and ant colony optimization algorithm. Authors have evaluated the experimental performance of the proposed method using both synthetic and real benchmark data.

Zhang et al. [13] analyzed the energy-aware real-time task scheduling issue in heterogeneous processors and presented a new heuristic algorithm based on particle swarm optimization. The particle swarm optimization (PSO) [14, 15] is a universally accepted optimization algorithm. The proposed algorithm successfully optimized the energy solution with feasible scheduling options. In this work, authors have

made it clear that PSO-based variant algorithms can also be used in DVFS-enabled multiprocessor systems to find feasible energy-aware task scheduling solutions.

Medhat et al. [16] also addressed the energy-aware scheduling issue for heterogeneous multiprocessors. Authors proposed a modified PSO variant algorithm to find optimal solution for partitioning and allocating periodic real-time tasks in heterogeneous multiprocessor environment. Authors also adopted suitable priority assignment methods in their research approach. The proposed method attains the goal of minimizing the overall system energy consumption, by meeting all the system constraints.

In [17], authors presented a population-based co-operative search form called shuffled frog leaping algorithm (SFLA) to solve energy-aware real-time task scheduling problem in heterogeneous multiprocessor systems. The proposed method has been experimented on synthetic data as well as real data from different benchmarks. The experimental results confirm the robust performance of the proposed algorithm over GA and ACO approximation algorithms.

3 Proposed Model

In this paper, we consider the problem of energy-aware task scheduling for DVFS-enabled heterogeneous multicore systems. In this section, we formulated the DVFS modeling for multicore systems and proposed an Adaptive Particle Swarm Optimization-based optimal approach for task scheduling. The proposed method helps to reduce load on the processor by properly allocating the tasks among available cores based on predetermined threshold.

3.1 DVFS Model

The recent revolutions in semiconductor technology leading to a greater number of cores on a single die. This enhancement also leading to high power density on the same chip. Increased power dissipation always influences various adverse effects on cooling costs, reliability, and overall performance of the processors. So, power management issue is more important in embedded computing systems. The techniques like, core layout, gating, etc., are there to address the issue of power management in multicore systems. These techniques work toward power consumption, based on the hardware level approach. Some of the compiler level techniques reduce the power consumption by optimizing the source code with less computations, execution time and memory. On the other hand, the dynamic voltage and frequency scaling (DVFS) [7] and dynamic power management (DPM) are the operating system level techniques, used to optimize the power consumption in embedded computing systems. DVFS performs the scaling of CPU voltage and frequency dynamically,

whereas DPM uses transition devices to minimize the power consumption. DVFS method has been adopted in various real-time applications.

According to DVFS, the maximum frequency of operation is based on the supply voltage, which is given as:

$$f = k \times \frac{(V_{dd} - V_t)^2}{V_{dd}} \quad (1)$$

V_{dd} denotes the supply voltage, V_t is the threshold voltage, f represents the operating frequency, and k is the constant. This can be represented as:

$$f = a \times V_{dd} \quad (2)$$

where a is constant, expression 2 shows that the operating frequency and voltage have linear relationship with each other. Let us consider a processor based on CMOS technology, and the active power consumed by the processor (P_{active}) can be expressed as:

$$P_{\text{active}} = f \times V_{dd}^2 \times C_{\text{ef}} \quad (3)$$

where C_{ef} is the effective capacitance, V_{dd} is the supply voltage, and f is the processor operating frequency. With the help of (2) and (3), the active power of the processor can be given as:

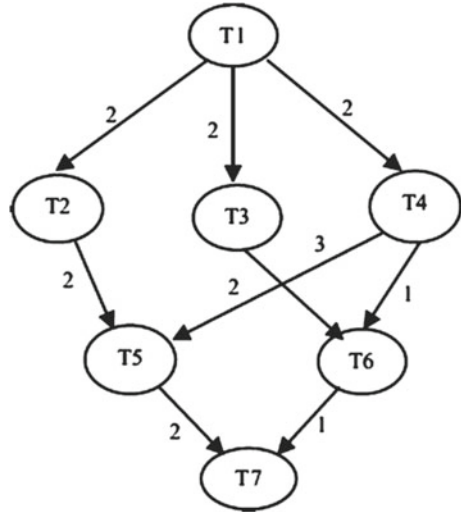
$$P_{\text{active}} = f^3 \times \frac{C_{\text{ef}}}{a^2} \quad (4)$$

The main aim of DVFS-based schemes is to dynamically scale the frequency and voltage to reduce the power consumption. However, these DVFS-based techniques do not focus on the processor scheduling or task scheduling to improve the resource utilization and minimized energy consumption. This particular issue is addressed in this paper and presented a PSO-based scheduling solution to DVFS-enabled heterogeneous multi-core systems, in the next subsection.

3.2 Adaptive Particle Swarm Optimization Algorithm for Energy-Aware Scheduling

Here, we present an optimization approach for dynamic processor scheduling to reduce the power consumption and efficient resource utilization. First of all, we consider the models of system architecture, tasks, and energy, which are vital to formulate the scheduling problem. Let $P = \{p_1, p_2, \dots, p_m\}$ denote DVFS-enabled multiple heterogeneous processors based on CMOS technology, which are fully

Fig. 1 Directed acyclic task graph [18]



interconnected. Each processor $p_j \in P$ is different in terms of task execution time, performance and capacity, due to the heterogeneity. In a heterogeneous multicore embedded system, an application consists of multiple tasks. In the proposed approach, directed acyclic graph (DAG) can also be used to model the multiple tasks, which is denoted as $G = (V, E)$, where V is the set of nodes represents the tasks, as $V = \{T_1, T_2, \dots, T_n\}$, and E denotes edges between the nodes. Each edge $e_{i,j} \in E$ represents precedence and data dependency constraint among the tasks. A task without any predecessor is considered as entry task, and a task without successor is exit task. A simple DAG task graph is shown in Fig. 1.

Matrix $X_{n \times m}$ corresponds to task allocation, where n represents number of tasks and m represents the number of processors in the system. Here we consider that each task has a basic execution time, which may vary for each core. Each w_{ij} is the basic executing time of task T_i on processor p_j . Now the average executing time of task T_i is expressed as:

$$\bar{w}_i = \sum_{j=1}^m \frac{w_{i,j}}{m} \tag{5}$$

For task T_i on processor p_j , we consider two time parameters, earliest start time and earliest finish time of execution and are denoted as $EST(T_i, p_j)$ and $EFT(T_i, p_j)$, respectively. The $EST(T_i, p_j)$ and $EFT(T_i, p_j)$ can be computed as:

$$EST(T_i, p_j) = \begin{cases} 0, & \text{if } T_i = T_{\text{entry}} \\ \max\{t_{\text{av}}(T_i, p_j)\} & \end{cases} \tag{6}$$

where t_{av} is the time at which processor p_j is available for task execution T_i at earliest, Consequently, the EFT(T_i, p_j) of task T_i is given by:

$$\text{EFT}(T_i, p_j) = \text{EST}(T_i, p_j) + w_{i,j} \quad (7)$$

Further, we consider the QoS-related parameters such as dynamic power consumption and makespan to formulate the optimization problem. Here, the considered heterogeneous multicore system is DVFS-enabled. The dynamic energy consumption is always the energy consumed by a task when it is executed, and dynamic energy is the main contribution in the total energy consumption of a processor. As already discussed in Sect. 3.1, energy consumption of DVFS-enabled processor can be calculated as:

$$E = P_{\text{active}} \times w \quad (8)$$

where w denotes the time taken for execution. The energy consumption of a task T_i at voltage v_i and frequency f_i on each cycle at various DVFS levels can be formulated as:

$$E_i = f_i v_i^2 C_{\text{ef}} w_i \quad (9)$$

where C_{ef} is the effective capacitance, w_i is the execution time of T_i . The overall energy consumption of the processor for the execution of all tasks is defined as:

$$E_{\text{active}} = \sum_{i=1}^n f_i v_i^2 C_{\text{ef}} w_i \quad (10)$$

During idle time, the processor can be turned into the sleep mode with the help of DVFS mechanism and relative voltage supply and frequency will be lowered to reduce the energy consumption. So, the energy consumption of processor p_j in idle case p_i can be considered as:

$$E_{\text{idle}} = \sum_{j=1}^m \sum_{\text{idle}_{j,k} \in \text{IDLE}_j} f_{\text{min}} C_{\text{ef}} v_{\text{min}}^2 L_{j,k} \quad (11)$$

where IDLE denotes the group of lists of IDLE processors and $L_{j,k}$ is the idle time duration. Thus, the total energy consumption can be expressed as:

$$E_{\text{Total}} = E_{\text{idle}} + E_{\text{active}} \quad (12)$$

Further in formulating optimization problem, we consider another parameter as timing constraint, where T_{total} is the time from submitting the task to finishing the

task in the processor. In other words, it is computed as the actual finish time of exit task. It can be expressed as:

$$T_{total} = \text{Actual}_{\text{finish}}(T_{\text{exit}}) \tag{13}$$

The main aim of energy-aware scheduling is assigning the tasks of a particular application to heterogeneous processors such that the makespan and total energy consumption is minimum, subject to task’s deadlines and other system constraints are satisfied. Hence, in order to proceed with this, we create a task mapping process M , which schedules the processors and tasks according to the task configurations. Thus, the task scheduling optimization problem can be defined as:

$$\begin{aligned} \text{Makespan} &= \text{minimize } T_{total}(M) \\ \text{Energy} &= \text{Minimize } E_{total}(M) \end{aligned} \tag{14}$$

This problem is assumed to be NP-hard; we solve this problem with the help of optimization scheme and propose a modified PSO scheme to generate the optimal solution for the given problem. The particle swarm optimization (PSO) [19] is one of the most powerful technique for solving optimization problems. PSO can be adopted successfully, in most of the optimization’s scenarios. PSO is utilized in many scientific areas, and there are many deviations of the algorithm. In the PSO, each member is considered as particle, the considered particles move in a given search space. The particles move toward the better search area during the searching process. Each particle is assigned with some fitness value, determined by the function to be optimized to indicate the performance. In our case, the task finish time and energy consumption are considered as the objective functions. Each particle’s position is initialized such that, the particle has attained so far as personal best position ($pBest$) and gained by any particle in the swarm, as global best position ($gBest$). Further the particle updates its velocity and position according to the following equations:

$$\begin{aligned} V_i^{k+1} &= \omega V_i^k + c_1 r_1 (pBest_i - X_i^k) + c_2 r_2 (gBest_i - X_i^k) \\ X_i^{k+1} &= X_i^k + V_i^{k+1} \end{aligned} \tag{15}$$

where V and X represent the velocity and position of the particle, respectively, r_1 and r_2 are the two random real numbers, c_1 and c_2 are positive constant parameters, ω is an inertia factor.

The standard procedure of PSO is as follows:

Initialize the random population of particles.

Do {

For each particle. {

Design a fitness function and evaluate the local and global best positions as $pBest$ and $gBest$.

For each particle. {

Update the position and velocity in each iteration, using (15).

```
}}
```

Repeat the procedure until the desired optimal solution is not obtained.

In this work, we have adopted PSO algorithm to solve energy-aware task scheduling problem. According to the proposed work, if any processor in the network is exceeding the processing capacity (overloaded), then the suitable processor needs to be identified and the removed task needs to be switched to that processor. This is achieved through PSO and termed as adaptive particle swarm optimization scheme suitable for DVFS-enabled systems. Proposed adaptive swarm optimization method performs a search operation to find the suitable processor and its location by communicating with neighbouring swarms. Here, removed task is considered as particle in the swarm, which uses bi-objective function associated with the defined optimization problem, for optimal processor selection and to complete the task in desired time. In this approach, various basic constraints are also considered such as, initial load of any processor should not be greater than the threshold, deadline constraints are considered for multiple processor's scenario where tasks can be migrated from heavily loaded processors to low loaded processors. Here, removed task deadline $T_{deadline}$ provides the information about assigned tasks to the processor such that, if $T_{deadline}$ is higher for removed task then minimum deadline tasks are present, whereas if $T_{deadline}$ is medium for removed task then less number of high deadline tasks are present. We also apply processor grouping, where under-loaded ($P_{Group_{ul}}$) and overloaded ($P_{Group_{ol}}$) processors are grouped. Task is removed from overloaded processors ($P_{Group_{ol}}$) and assigned to under-loaded processors ($P_{Group_{ul}}$) with the help of objective function computation. This complete process of task removal is performed until ($P_{Group_{ol}} = \text{NULL}$). This process helps to reduce the overall energy consumption and also improves resource utilization.

The details of the proposed adaptive PSO algorithm used for processor scheduling are depicted in Table 1. Using this technique of energy saving and task allocation, we can improve the heterogeneous processors performance.

4 Results and Discussion

This section explains the experimental analysis of the proposed scheduling approach. The proposed method has been implemented using MATLAB tool and executed on a PC. In our experiments, the complete simulation parameters are adopted from [20], where the number of processor cores, task set with execution characteristics of each task are considered as inputs to the simulator. The obtained performances of the proposed method have been compared with energy-aware multicore scheduling techniques discussed in [18]. Figures 2 and 3 describe the energy consumption performance of the proposed approach.

According to the experiment presented in Fig. 2, we obtained the average energy consumption as 125.9, 112.5, 101.9, and 88.7 mJ using DVFS [18], PDAM [18], Scheduling [18], and Proposed Scheduling for 3 cores. Similarly, as described in Fig. 3, we conducted another experiment for 6 number of cores and obtained the

Table 1 Proposed PSO-based algorithm

Proposed adaptive particle swarm optimization algorithm for processor scheduling
(1) Initialize simulation with complete desired parameters
(2) $APSO_S_m$ is the m^{th} particle in the available S swarms during simulation
(3) Set $(P_{Group_{ol}}) = NULL$ and $(P_{Group_{ul}}) = NULL$
(4) Do
(5) For each processor p_j in the p^{th} heterogeneous system
(6) If $L_p > Threshold$ then $P_{Group_{ol}} \leftarrow p_j$
(7) Else $P_{Group_{ul}} \leftarrow p_j$
(8) End if
(9) End for
(10) For each m , generate neighbouring, in the swarms
(11) If $L_p = Threshold$
(12) Apply task sorting in ascending order based on the deadline constraints
(13) Do
(14) For each assigned task in overloaded processors ($P_{Group_{ol}}$)
(15) Find best processor from under-loaded processors ($P_{Group_{ul}}$)
(16) End for
(17) Apply task migration or reallocation to the best suitable under-loaded processors ($P_{Group_{ul}}$)
(18) Else move processor to sleep mode // Energy constraints
(19) Update parameters such as task list, load on processors and available processors
(20) End

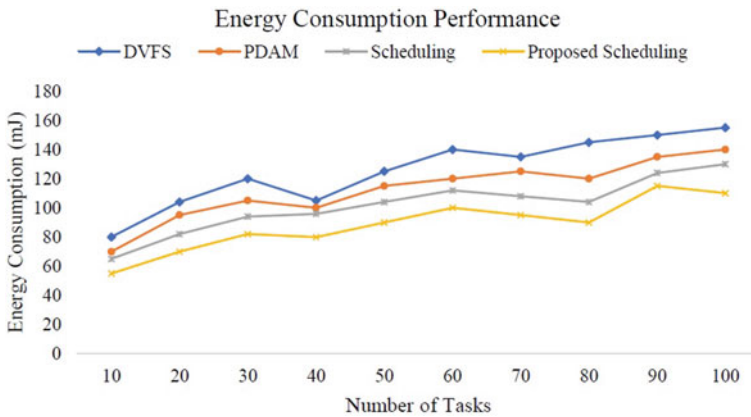


Fig. 2 Energy consumption performance for 3-cores

average energy consumption as 165.4 mJ, 152.1 mJ, 133.7 mJ, and 120.9 mJ using DVFS [18], PDAM [18], Scheduling [18], and Proposed Scheduling techniques, respectively. Experimental study confirms that the proposed scheduling method, due its adaptive dynamic nature of switching the tasks from overloaded processors to under-loaded processors, accomplishes better performance in terms of energy saving, when compared with state-of-art techniques.

Table 2 Energy consumption performance for 3-cores

No. of tasks	DVFS	PDAM	Scheduling	Proposed Scheduling
10	80	70	65	55
20	104	95	82	70
30	120	105	94	82
40	105	100	96	80
50	125	115	104	90
60	140	120	112	100
70	135	125	108	95
80	145	120	104	90
90	150	135	124	115
100	155	140	130	110

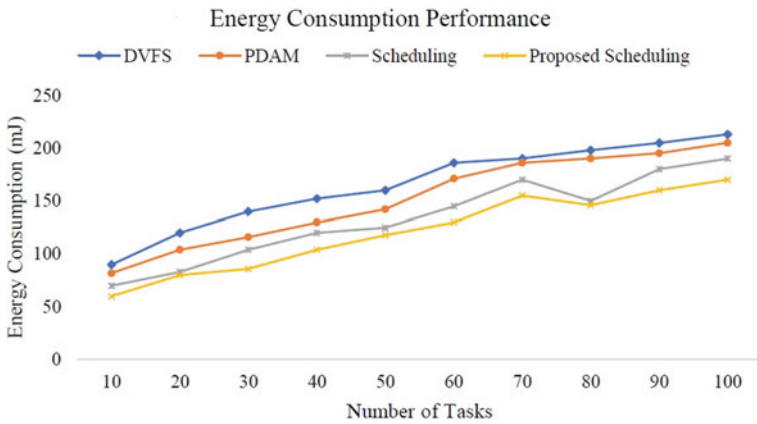


Fig. 3 Energy consumption performance for 6-cores

Table 3 Energy consumption performance for 6-cores

No. of tasks	DVFS	PDAM	Scheduling	Proposed Scheduling
10	90	82	70	60
20	120	104	83	80
30	140	116	104	86
40	152	130	120	104
50	160	142	125	118
60	186	171	145	130
70	190	186	170	155
80	198	190	150	146
90	205	195	180	160
100	213	205	190	170

5 Conclusion

The main aim of this work is to present a combined scheduling approach using DVFS and particle swarm optimization to reduce the energy consumption of the heterogeneous multiprocessor system. Here, to mitigate the negative impact of the conventional DVFS scheme, we presented a novel scheme using particle swarm optimization, which finds an optimum processor to accomplish the task within in the given deadline. The proposed algorithm due its adaptive nature, finds the feasible scheduling solution of minimum makespan and lesser energy consumption. The proposed method has been experimented on different task sets and results indicated that, the proposed PSO-based algorithm performs better when compared with state-of-art techniques. The future work will take into consideration of dynamic load balancing.

References

1. Fernandez-Baca D (1989) Allocating modules to processors in a distributed system. *IEEE Trans Software Eng* 15(11):1427–1436
2. Qin Y, Zeng G, Kurachi R, Li Y, Matsubara Y, Takada H (2019) Energy-efficient intra-task dvfs scheduling using linear programming formulation. *IEEE Access* 7:30536–30547
3. R Jejurikar, C Pereira, R Gupta (2004) Leakage aware dynamic voltage scaling for real-time embedded systems. In: *Proceedings of 41st design automation conference, IEEE, San Diego, CA, USA*, pp 275–280
4. L Niu, G Quan (2004) Reducing both dynamic and leakage energy consumption for hard real-time systems. In: *Proceedings of CASES, ACM, Washington DC, USA*
5. Rakesh K, Dean M Tullsen, Norman P Jouppi (2006) Core architecture optimization for heterogeneous chip multiprocessors. In: *Proceedings of the 15th international conference on parallel architectures and compilation techniques, ACM*, pp 23–32
6. Vartziotis F, Kavousianos X, Chakrabarty AJ, Parekhji R (2015) Time-division multiplexing for testing DVFS based SoCs. *IEEE Trans Comput Aided Des Integr Circuits Syst* 34(4):668–681
7. Lin CC, Syu YC, Chang CJ, Wu JJ, Liu P, Cheng PW, Hsu WT (2015) Energy-efficient task scheduling for multi-core platforms with per-core DVFS. *J Parallel Distrib Comput* 86:71–81
8. Zou C, Liu P, Liu X (2018) Energy-aware task scheduling strategies for multi-core embedded systems. *EasyChair preprints*
9. Chen H, Cheng AMK (2005) Applying ant colony optimization to the partitioned scheduling problem for heterogeneous multiprocessors. *ACM SIGBED Review* 2(2):11–14
10. Mo L, Kritikakou A, Sentieys O (2018) Energy-quality-time optimized task mapping on DVFS-enabled multicores. *IEEE Trans Comput Aided Des Integr Circuits Syst* 37(11):2428–2439
11. Chen H, Cheng AMK, Kuo Y (2011) Assigning real-time tasks to heterogeneous processors by applying ant colony optimization. *J Parallel Distrib Comput* 71(1):132–142
12. Mahmood A, Khan SA, Fawzi A, Noor A (2017) Energy-aware real-time task scheduling in multiprocessor systems using a hybrid genetic algorithm. *Electron J*, 6(2)
13. Zhang W, Xie H, Cao B, Cheng AMK (2014) Energy-aware real-time task scheduling for heterogeneous multiprocessors with particle swarm optimization algorithm. *J Math Probl Eng*. <https://doi.org/10.1155/2014/287475>
14. Y Shi, R Eberhart (1998) A modified particle swarm optimizer. In: *Proceedings of the IEEE international conference on evolutionary computation. IEEE Press, USA*, pp 69–73

15. R Eberhart, Y Shi (1998) Comparison between genetic algorithms and particle swarm optimization. In: Proceedings of the 7th international conference on evolutionary programming. Springer, Berlin, Germany, pp 611– 618
16. Medhat A, Abdullah E (2016) Enhanced PSO approach for real time systems scheduling. *Int J Comput Theory Eng* 8(4):285–289
17. Zhang W, He H, Wang X (2015) Solving energy-aware real-time tasks scheduling problem with shuffled frog leaping algorithm on heterogeneous platform. *Sensors J* 15(6):13778–13804
18. Ranvijay VKK (2018) Enhanced energy aware scheduling in multicore processors. *J Intell Fuzzy Syst*, 35(2), 1375–1385. <https://doi.org/10.3233/JIFS-169680>
19. Zhang Y, Wang S, Ji G (2015) A Comprehensive survey on particle swarm optimization algorithm and its applications. *J Math Probl Eng*. <https://doi.org/10.1155/2015/931256>
20. FMM Ul Islam, Lin M, Yang LT, Choo K-KR (2017) Task aware hybrid DVFS for multi-core real-time systems using machine learning. *Inf Sci* 433(434):315–332. <https://doi.org/10.1016/j.ins.2017.08.042>

Theft Detection and Monitoring System Using Machine Learning



Jatin Arora, Aayush Bangroo, and Shivi Garg 

1 Introduction

Theft is the most common crime committed across the world. According to the National Crime Records Bureau (NCRB), ~ 80% of the criminal cases are related to theft [1] as shown in Fig. 1. Increasing theft rates cause people to suffer both financially and emotionally. Therefore, there is a need to develop a more deterrent surveillance system, which is convenient to use, free from false alarms, minimize human interference, and cost-effective.

Machine learning (ML) techniques prove to be fruitful in developing efficient surveillance systems. This paper aims to design a theft detection and monitoring system, which would be capable to detect theft using a motion-sensing camera using ML and alarm the owner with an alert message along with the captured image of that instance of motion.

The major contributions of this paper are:

- To detect and activate motion in the still place according to requirements.
- To recognize facial expressions and detect people wearing the mask using the ML model.
- To detect the suspiciousness in the surrounding for any kind of weapon and raise alert messages.

After analyzing all this, in the last, the system itself decides to whom it should report the suspicious behavior whether to inform the owner or to cops or both so that the theft can be detected, monitored and a fair possibility to take possible measures according to the situation. The rest of the paper contains four more sections and is described as follows. Section 2 talks about the background and related techniques

J. Arora (✉) · A. Bangroo · S. Garg
Faculty of Informatics and Computing, J.C. Bose University of Science and Technology YMCA,
Faridabad, India

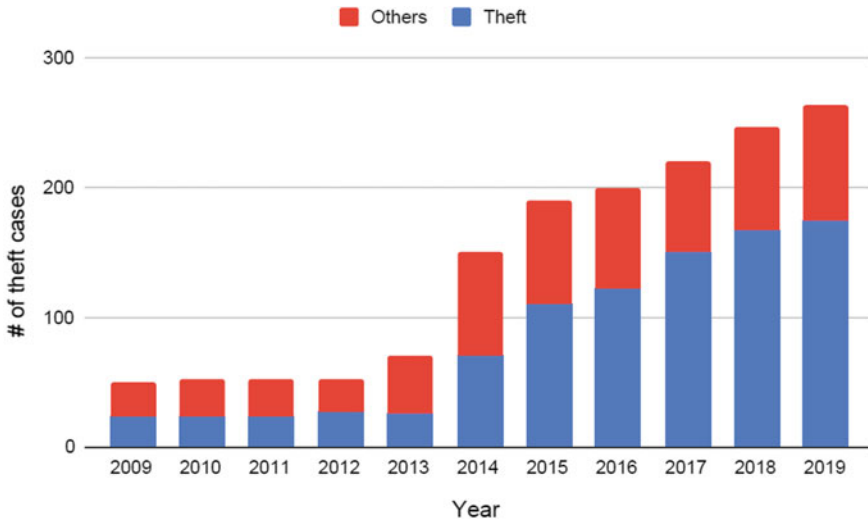


Fig. 1 Contribution of theft in crime

used in theft detection. Section 3 discusses the data proposed methodology. Section 4 presents the results. Finally, the paper is concluded in Sect. 5.

2 Literature Review

The importance and recognition of human motion analysis, mask detection analysis, emotion analysis, weapon detection analysis, pose analysis, activity analysis has led to previous surveys.

Kushwaha et al. [2] examined the burglary location utilizing ML, where they observed theft based on just motion detection.

Bhatia et al. [3] proposed an approach for the extraction of the intransient confront at that point the acknowledgment of four facial expressions. They used the eyebrow and mouth corners since the preeminent ‘anchor’ focuses for confront acknowledgment. There was no manual intercession a touch similar to the beginning manual task of include focuses). Their approach was able to distinguish fractional occlusions also.

Zeng et al. [4] examined a process of face detection and expression acknowledgment. Eight expressions, viz. cheerful, outrage, astonish, disdain, appall, fear, pity, and fear were recognized from the faces. The face detection process was fulfilled by Haar cascade.

Gregg et al. [5] talked about the two particular errands of computerized detection and acknowledgment of dangerous situations. The specificity and affectability of the knife detection algorithm were 94.93 and 81.18%, respectively. The results were essentially superior to the others. Our solution to the knife detection problem deals

with destitute quality and low-resolution images. This can be vital since numerous CCTV frameworks only provide such quality of the footage. It ought to be noted that the algorithm is prepared in real time.

Kocabas et al. [6] examined the pose residual network, which seems to precisely relegate key focuses on individual discoveries outputted by a multi-task learning engineering (MultiPoseNet). Posture estimation strategy accomplished state-of-the-art execution among bottom-up strategies and comparable comes about with top-down methods.

Gupta et al. [7] examined that closed captions are frequently acclimated consequently to train a video action recognizer without requiring any manual labeling of video clips which this activity recognizer may indeed be acclimated to improve the accuracy of caption-based video recovery. Furthermore, training a classifier to spot captions that depict current activities can improve accuracy indeed further. Results on accumulating retrieval rankings from both video and caption classifiers give extra evidence that exploiting the multimodal nature of closed-captioned video can improve the adequacy of activity recognition and video retrieval.

3 Proposed Methodology

The proposed methodology consists of three phases, data collection and acquisition, ML models, and actions taken. The workflow of the proposed methodology is shown in Fig. 2.

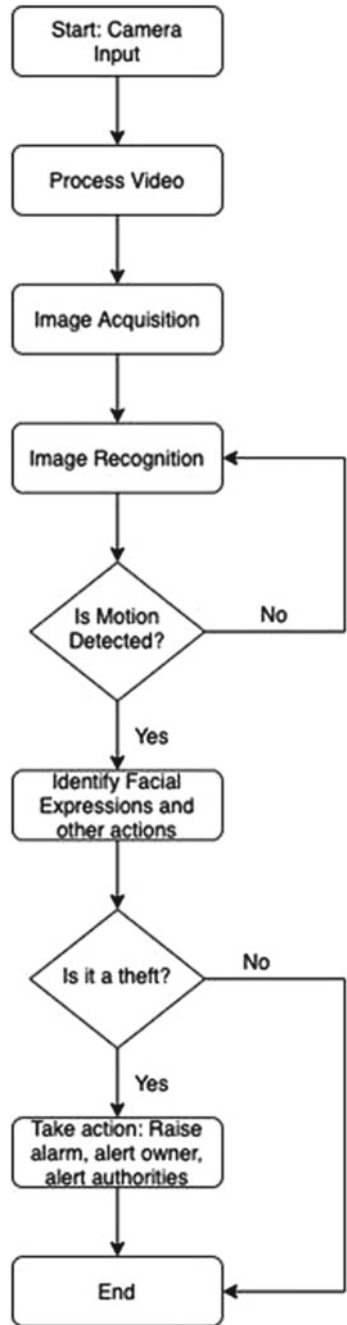
3.1 *Data Collection and Acquisition*

Datasets utilized are taken from Google Pictures, Kaggle, and Flickr 8 k dataset are utilized, which contains 8000 pictures each with five captions. This dataset is fitting sufficient to prepare the demonstration. Bigger the number of pictures, troublesome to prepare the show with the given computational prerequisites.

Video is captured utilizing an IP camera, which is associated with the remote organizer. This is often done by utilizing the OpenCV library. Once the video is captured, it is at that point handled and broken into outlines of pictures. These picture outlines are at that point nourished to ML models to encourage investigation. Information is cleaned and preprocessed sometime recently applying ML models. In information cleaning, a lexicon of one of a kind words shown in all the captions of the dataset pictures is created and spared on the disk. To create the show ended up more strong to outliers, we considered as it were those words, which happen at the slightest 10 times within the whole corpus.

In information preprocessing, the image is changed to a fixed size vector, which is then fed to a convolutional neural network (CNN) based on exchange learning. This demonstration is prepared on ImageNet dataset to perform picture classification on

Fig. 2 Flowchart of the proposed methodology



1000 distinctive classes of pictures, but since we needed to extricate a fixed-length instructive vector for each picture, we evacuated the final softmax layer from the show and extricated a 2048 length vector (bottleneck highlights) for each picture.

3.2 *ML Models*

Image frames are processed and fed to different ML models. Each model performs a task in a particular sequence so as to analyze different evaluation parameters.

3.3 *Actions*

According to the results obtained from the ML models, the actions will be taken such as raising alarms, sending alert messages to the owner only, sending alert messages to cops only, sending an alert message to the owner and cops both.

The system consists of several levels of surveillance; at each level, the activity in each frame of the video will be monitored thoroughly using ML models, which are solely trained to perform their specific job. There are a total of six levels of surveillance, and the system consists of two modes (i.e., day and night), and it will totally depend on the user which mode is required at the moment.

1. **Motion Detection:** To detect the motion, we firstly detected if any human being is present in the frame or not, and to do so, we used CNN. Convolution can be used to achieve the blurring, sharpening, edge detection, noise reduction, which is not easily achieved by other methods. It is represented in mathematical form as follows

$$g(x, y) = f(x, y) * h(x, y)$$

where $g(x, y)$ = output image $f(x, y)$ = input image $h(x, y)$ = filter/kernel/mask. In the convolution method center of the kernel is kept at each element of the image input and the corresponding elements are multiplied and added together. CNN is a type of neural network, made up of a large number of neurons with weights and biases which define the relations between them. Each neuron receives several inputs, takes a weighted sum over them, passes it through an activation function, and responds with an output.

After getting quite normal results from the model trained on a dataset with 2000 images with a human being and 2000 images without a human being, we tried transfer learning for that we used the YOLOv4 neural network and extracted the weights just before the last two layers of the network and then used those pre-trained weights to train the model which can detect the human being in the image and using the transfer learning technique accuracy got improved. After

detecting the human being in the frame, we now can detect the motion and so we used python library 'OpenCV.'

To identify the motion different strategies like frame differencing, background subtraction, optical flow, etc., and got the most excellent comes about with frame differencing. The frame differencing strategy employs the two or three adjacent frames based on time series picture to subtract and gets diverse pictures, its working is exceptionally comparative to background subtraction, and after the subtraction of the picture, it gives moving target data through the edge esteem. This method is straightforward and simple to execute, conjointly it is comparable to the background subtraction. But this strategy is exceedingly versatile to energetic scene changes; in any case, it by and large falls flat in identifying entirely significant pixels of a few sorts of moving objects. Extra strategies that ought to be received in order to identify ceased objects for the victory of the next level are computationally complex and cannot be utilized in real-time without specialized equipment.

2. **Mask Detection:** Thieves usually wear masks to hide their identity while attempting theft. Therefore, the output of this module will help in determining if the theft is real or false. Dataset used in this module is from Google Images and Kaggle. Google images are scraped using the selenium package and Chrome Driver extension. To get better results in a short time the approach, transfer learning is used and the MobileNetV2 architecture is trained on the weights of ImageNet. Amid training, the main focus was on stacking the face cover detection dataset from disk, training a model (utilizing Keras/TensorFlow) on this dataset, and after that serializing the face mask detector to disk. The model was trained on the information set containing 1000 images for a person with a mask and without a mask each.
3. The distinctive steps utilized in this module are:
 - Train mask detector: Accepts the input dataset and fine-tunes MobileNetV2 upon it to make a model.
 - Detect mask image: This module performs face mask detection in static images
 - Detect mask video: Utilizing the camera, this module applies to confront cover discovery to each frame within the stream.

For mask detection, we utilized a crossover model in which we, to begin with, recognized the face utilizing Haar cascade and after that identified the cover on the face utilizing MobilenetV2 prepared on the weights of the Imagenet. Once the confront veil finder was prepared, we at that point moved on to stacking the mask locator, performing face discovery, and after that classifying each face as 'with mask' or ' without mask.'
4. **Facial Expression Detection:** As a preprocessing step, the images are cropped around the faces and intensity normalized. Features are extracted and local descriptors are calculated. This computation is represented using a Vector of Locally Aggregated Descriptors (VLAD). This labeled dataset is trained on the emotion classes using the SVM classifier. Lastly, the face is identified from the

image using a Haar cascade classifier and the image is cropped into 256×256 resolutions.

5. **Weapon Detection:** Weapon detection is another important module for differentiating between false and actual threats [8]. The approach we have used to detect weapons is similar to that of mask detection except that we use a Haar cascade to detect the hand of a person and then detect the weapon in the hand. To train the model, we used the data set containing 4000 images of person with weapon and person without weapon each. The label of the images employed in the dataset is categorical. Therefore, we used a one-hot encoding to convert the categorical labels to binary values. As our dataset is small, we used augmentation technique to achieve good accuracy. Data augmentation could be a technique that may be used to artificially expand the dimensions of a training dataset by creating modified versions of images within the dataset. Kind of like mask detection, transfer learning is used for training the model. The MobileNetV2 architecture is trained on the weights of ImageNet to realize better accuracy during a short time. During training, callbacks like checkpoints and early stopping are used to save the best-trained model.
6. **Pose Detection:** The issue of human posture estimation is frequently characterized since the computer vision strategies that foresee the circumstance of changed human keypoints (joints and points of interest) like elbows, knees, neck, bear, hips, chest, etc. It is a very challenging issue since different components like little and barely unmistakable parts, occlusions, and huge inconsistency in enunciations. The classical approach to enunciated posture estimation is utilizing the pictorial structures system. The fundamental thought here is to speak to a question by a bunch of “parts” orchestrated in an exceedingly deformable setup (not inflexible). A “portion” is an appearance format that is coordinated in a picture. Springs appear in the spatial associations between parts. When parts are parameterized by pixel area and introduction, the coming about structure can demonstrate verbalization which is uncommonly significant in posture estimation. (A organized expectation assignment). This strategy, be that as it may, comes with the impediment of getting a posture demonstrates not looking at picture information. As a result, inquire has centered on improving the representational control of the models. After getting a few unsuitable results from the above-mentioned strategies inside the final, we utilized a neural organization by Google which gave us the desired results. After getting the posture recognized, we prepared a show that portrays whether the postures recognized in outlines ought to concern approximately suspiciousness inside the outline or not [9]. To do so, we made a fake dataset that contains 10,000 pictures labeled with their postures and a word reference which contains data almost the postures and their pertinence to the sort of action conceivable at that point.
7. **Activity Captioning:** Activity captioning is a task similar to image captioning but in real time that involves computer vision and natural language processing. It takes an image and can describe what is going on in the image in plain English. This module is essential for determining the course of actions that are being performed by the person in view. In surveillance systems, it can be used to

determine theft by captioning the frames obtained from the camera and finally alerting the authorities. We again used the approach of transfer learning by utilizing the ResNet50 architecture trained on the weights of ImageNet. After getting the results from the above six modules, those results are combined and then further becomes the input to an ML model which decides to whom it should address for the alert message whether it should be the owner or the cops or both [10].

The main aim of the project is to embed computer vision into a camera. We designed the system in such a way that the camera is connected to an external computer on which the algorithms will run. The client is also connected to this computer through the Internet. Whenever suspicious activity is detected, an alert message will be sent to the client. The message will consist of an image captured through the camera along with the various features extracted by our six modules. We can also use Raspberry Pi instead of a regular computer to reduce the cost of the project.

4 Results

To distinguish the movement, we basically utilized the outline differencing method without identifying the human creatures in it which gave great comes about but inevitably, movement get identified indeed by the development of the creepy crawlies or other undesirable things, to expel that commotion we attempted to set a limit nearly same as the alter in pixels to happen due to nearness of a human within the outline. But in this case, moreover, the clamor was tall so in the end we made an ML show that can detect human creatures within the outline. After identifying, people within the outline movement is recognized utilizing outline differencing techniques.

In the mask detection module where masks are going to be detected using the ML model, the model was first trained using simple CNN a combination of some convolution layers, max-pooling layers, and some dropout layers followed by dense layers but the accuracy was not up to the mark. Then, to improve the accuracy of the model trained on the smaller datasets, we used the transfer learning technique and got great results with that. Figure 3 represents the training loss vs. validation loss and training accuracy vs. validation accuracy. We get 90% of validation accuracy in detecting all types of masks on human faces and we get a loss of 3.5%.

In facial expression detection, we primarily developed a model which detects the frontal face of humans present in the frame and then we build another model to determine the facial features using CNN with activation function 'ReLU' but we got a poor response with that so we replaced 'ReLU' with 'elu' and got great accuracy.

In weapon detection, we detected the weapons in the image and we used the same technique as of mask detector and achieved good accuracy as shown in Fig. 4.

In pose detection, we used a model trained by Google to detect the poses and then getting results from that we processed those results to an ML model which detects

Fig. 3 Training loss and accuracy in mask detection

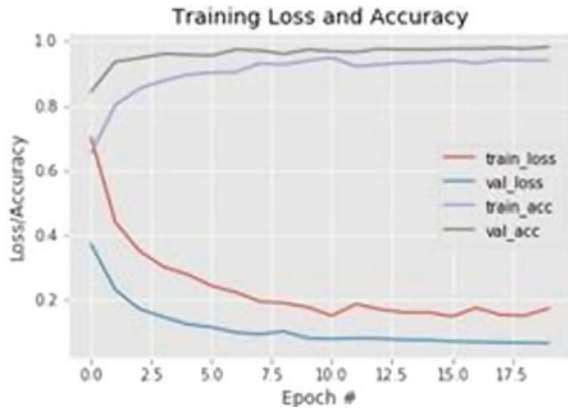
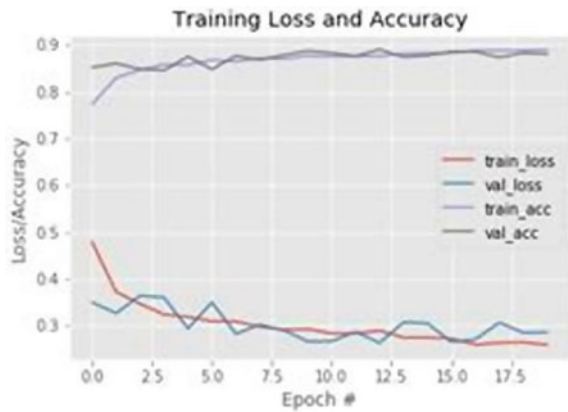


Fig. 4 Training loss and accuracy in weapon detection



the poses in the image relates to suspicion or not and to do so we used classification techniques and got best results with multivariate logistic regression.

In activity captioning, where we designed an ML model which captions the activity is the combination of computer vision and natural language processing and we firstly tried basic image classification and text generation algorithms but did not get good outcomes; then, we used transfer learning for image classification and Bert for text generation, and eventually, we got great results.

Finally after combining the outcomes of these modules, the input for another ML model is created which determines to whom the alert message should be sent, and to design such a model, we used several classification techniques like logistic regression, SVM, etc. But got the best results with decision trees.

5 Conclusion and Future Scope

The work carried out in this paper is basically centered to plan and create effective and helpful observation frameworks to unravel security issues which are able to offer assistance to reduce/stop a theft. Though a significant amount of research has been done in the past to solve such security problems, it still remains challenging due to increased complexity and various theft actions that are taking place daily. The system will capture images only when there is any human being in the frame and motions exceed a certain threshold that is pre-set in the system. It thus reduces the volume of data that needs to be processed. Also, it will help to save data space by not capturing static images which usually do not contain the object of interest. Users using this system need not worry about supervising the cameras all the time instead the system will inform the user about the activities happening and will also suggest that the user should take some action or not. After successfully implementing the project, it can be applied in a smart home security system which would be very helpful in auto theft detection for security purposes. It can also be useful in banks, museums, and streets at midnight.

References

1. Theft's most common crime in Delhi shows NCRB data. <https://ncrb.gov.in/>
2. Kushwaha A, Mishra A, Kamble K, Janbhare R (2018) Theft detection using machine learning. *Int Conf Innov Adv Technol Eng* 8:67–71
3. Sarode N, Bhatia S (2010) Facial expression recognition. *Int J Comput Sci Eng* 2(5):1552–1557
4. Hossain MZ, Soheli F, Shiratuddin MF, Laga H (2019) A comprehensive survey of deep learning for image captioning. *ACM Comput Surv (CSUR)* 51(6):1–36
5. Grega M, Matiołański A, Guzik P, Leszczuk M (2016) Automated detection of firearms and knives in a CCTV image. *Sensors* 16(1):47
6. Zheng W, Liu C (2016) Facial expression recognition based on texture and shape. In: 2016 25th wireless and optical communication conference (WOCC). IEEE, pp 1–5
7. Zhao ZQ, Zheng P, Xu ST, Wu X (2019) Object detection with deep learning: a review. *IEEE Trans Neural Netw Learn Syst* 30(11):3212–3232
8. Pathak AR, Pandey M, Rautaray S (2018) Application of deep learning for object detection. *Procedia Comput Sci* 132:1706–1717
9. Zhu L, Xu Z, Yang Y, Hauptmann AG (2017) Uncovering the temporal context for video question answering. *Int J Comput Vis* 124(3):409–421
10. Verma GK, Dhillon A (2017) A handheld gun detection using faster r-cnn deep learning. In: Proceedings of the 7th international conference on computer and communication technology. pp 84–88

COVID-19 Contact Tracing Using Geolocation



R. Mahanthesh, Surabhi V. Keerthi, and S. Nidhi Gowri

1 Introduction

The biggest threat for us during this pandemic is the current unavailability of a vaccine for this virus. This would imply that we can only take preventative measures. The only defense at our disposal is to follow social distancing, which is followed by maintaining physical distance and avoiding contact with a person that has been infected with this pathogen to slow down the transmission of the virus in a community. While this could be a good precautionary practice, a country like India with such a massive population would find it grueling to adhere to these practices for prolonged durations. This could be implied because a large fragment of population lives in crowded homes, shared by six to eight people. Factually, this insinuates that we need more measures to track the transmission of the virus to help prevent the spread of this virus.

In January of 2020, a novel coronavirus (2019—nCoV) was recognized to cause an outbreak of viral pneumonia in Wuhan, China. The first cases had their symptoms in December 2019. Since then, there have been rapid studies to understand the nature of this virus and the incubation period of the virus. The awareness of the incubation period of the virus becomes extremely important in assessment of the screening and contact tracing. At this stage of the pandemic, there have been various studies to show that the incubation period for the virus is between 2 to 12 days [1]. Based on these results, we have decided to use the period of 12 days as our most inclusive case. This information can help us in tracking data of a patient for the incubation period interval to perform effective contact tracing.

The developments and innovations in modern technology like telecom, cellular network and Internet has paved the way to reach this huge population in every corner

R. Mahanthesh (✉) · S. V. Keerthi · S. Nidhi Gowri
Department of Information Science and Engineering, SJB Institute of Technology, Bengaluru
560060, India

of the world at a very affordable price. Telecom network has been providing a wide range of services in education, healthcare and entertainment fields. In a crisis like this, where the easiest way to avoid the spread of the virus is to detect the spread, the advancements in the telecom network field will help us leverage services to manage communication and coordination required to mitigate the spread of the virus. The application softwares available on today's smartphones perform a variety of services ranging from video conferencing to sharing data files and real-time locations. From the 4G powered smartphones to the 2G connectivity feature phones, mobile networks are available in even the most remote parts of the country making a variety of applications actively available to use. The paper [2] by Ashok Jhunjhunwala, talks about 3G and 4G enabled devices making use of GPS and Bluetooth technologies to help track the whereabouts of a person whereas, 2G enabled devices taking advantage of interactive voice response system (IVRS) to help follow up with an infected individual. While the existing system shows considerable efficiency, the hassles of the app running in the background, GPS and Bluetooth turned on at all times is an issue we want to consider in our paper. Our application makes use of geolocation; data available in an individual's Google account to pinpoint the locations of that individual [3, 4]. Every Google account comes with a default enabled feature called Google timeline that uses geolocation, i.e., MAC address and cellular towers to pinpoint an individual's location based on the trilateration algorithm. This application does not run in the background and needs no Bluetooth or GPS to be turned on at all times. The individual's location history is stored in the user's account since the time the account has been linked to the user's mobile phone, and this data contains every country, state, cities, any restaurant or even a shop the user has visited.

The Global Positioning System is widely used as it provides extremely accurate location information of a person or place than earlier tracking techniques. Almost every mobile device is GPS enabled, and this serves as a huge benefit in terms of the global reach [5, 6]. This technology comes built-in and is accessible worldwide, which makes it easy to integrate it in cellular phones, computers and vehicles even. The GPS tracking system makes use of Global Navigation Satellite System (GNSS) which incorporates a range of satellites that use microwave signals. These signals are transmitted to the GPS device using a wireless or cellular network giving information about the location. Many times, GPS is not available when your line of sight is blocked for the GPS satellites to get a proper fix of your location. With our application, we address this problem by using geolocation. It detects Wi-Fi routers and cellular towers, combines the strength of these signals to determine where they are coming from. With this knowledge, it can calculate how far away from each signal you are, to triangulate and fix your exact position.

With many such location-based services, the number of privacy concerns has also seen a rise. With increasing cases, there is an emergence of numerous contact tracing applications that function by collecting relevant personal data such as name, age, profession, symptoms (if any), underlying health conditions and locations. Concerns have been raised due to COVID-19 situation as the data can be used post-COVID-19 to trace the movements of an individual and their contacts. A contact tracing app requires complex data protection laws because of the sensitivity of its data.

These apps have to follow certain data protection legislation laid out by the country's government. There are many existing apps in several countries such as Trace Together in Singapore, COVID safe in Australia and NHS COVID-19 app in the UK [7, 8]. All of these apps are based on Bluetooth technology and collect limited data but lack in providing either data destruction features or transparency. Studies show that these apps might not be fully compliant with the data protection legislation criteria. Our application tries to overcome these issues by collecting the geolocation of the user with the help of their Google accounts with options to turn off their location tracking whenever needed. The location history and all the analysis are carried out in the user's device; the data does not go past the user's device which makes it more secure than the existing system.

2 Objectives

- The main objective of our project is to help automate the process of contact tracing with a goal to track the spread of the virus.
- When a person tests positive, their location history is published on our application by the concerned authorities.
- Users can upload their location history to compare their data with the positive tested patient's data.
- Finding the distance between the infected person and the concerned user as efficiently as possible.
- We try to achieve maximum efficiency in regards to the search and comparison of locations to find the most vulnerable data points by taking a risk assessment survey.

3 Software Implementation

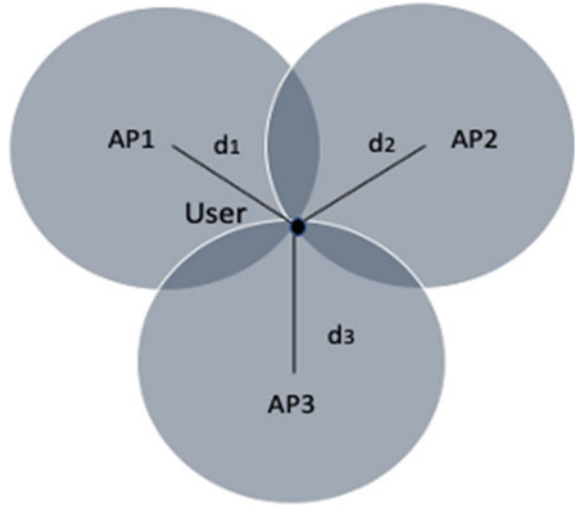
3.1 *How Trilateration Works*

Google gets the geolocation API using the concept of trilateration. The API accepts the address, the signal strengths of cellular towers and Wi-Fi access points near you (Fig. 1).

The Google API scans for access points/cellular towers and returns a unique identifier for those points. It also returns the signal strength along with frequency of these access points/cellular towers [9]. It will calculate how far you are from the cellular tower using the free-space path loss formula which is represented as:

$$\text{FSPL(dB)} = 20\log(d) + 20\log(f) + 32.44 \quad (1)$$

Fig. 1 Trilateration localization representation



where

d = distance of the receiver from the transmitter (km).

f = signal frequency (MHz).

This formula is only used to calculate the distance from one of the access points. If you are around many, then it will scan for multiple access points, and the distance from each of those points is expressed as follows:

$$\begin{aligned}
 d_1^2 &= (x - x_1)^2 + (y - y_1)^2 \\
 d_2^2 &= (x - x_2)^2 + (y - y_2)^2 \\
 d_3^2 &= (x - x_3)^2 + (y - y_3)^2
 \end{aligned}
 \tag{2}$$

We will obtain a linear equation which connects the intersection of two circles as follows:

$$\begin{aligned}
 2(x_1 - x_2)x + 2(y_1 - y_2)y &= x_1^2 - x_2^2 + y_1^2 - y_2^2 - d_1^2 + d_2^2 \\
 2(x_2 - x_3)x + 2(y_2 - y_3)y &= x_2^2 - x_3^2 + y_2^2 - y_3^2 - d_2^2 + d_3^2
 \end{aligned}
 \tag{3}$$

The above equations are represented in the form of a matrix:

$$\begin{aligned}
 A_{2,2}X &= B_{2,1}, \text{ where} \\
 A_{2,2} &= \begin{pmatrix} 2(x_1 - x_2) & 2(y_1 - y_2) \\ 2(x_2 - x_3) & 2(y_2 - y_3) \end{pmatrix}, X = \begin{pmatrix} x \\ y \end{pmatrix} \\
 B_{2,1} &= \begin{pmatrix} (x_1^2 - x_2^2) + (y_1^2 - y_2^2) - (d_1^2 + d_2^2) \\ (x_2^2 - x_3^2) + (y_2^2 - y_3^2) - (d_2^2 + d_3^2) \end{pmatrix}
 \end{aligned}
 \tag{4}$$

We multiply both sides with the inverse of A , to predict the estimated location of the user:

$$X = A_{2,2}^{-1} \cdot B_{2,1} \quad (5)$$

3.2 Procedure to Upload Location History

In this booming era of technology, almost everyone owns a smart phone. Specially in a situation like this created by COVID-19, where everything has gone digital, people have resorted to downloading applications to pay bills, for entertainment and carry on with their work life, and it is necessary to have a Google account for all the above-mentioned tasks. Taking all such scenarios into consideration and assuming the majority of the populations have a Google account, we have based our application in taking advantage of the Google timeline, a default feature in all Google accounts. Our application extensively searches and compares location to find matching time and coordinates of possibly infected persons. Since Google does not have an API to access this data, the user can download their location history. To make this process easy, the user will have to request for their location timeline by going to takeout.google.com on their respective web browsers. Once the page loads, the location history should be exported with the “*export every 2 months for one year*” option selected. It has to be downloaded as *zip file* and *2 GB* export, both default selected options. The relevant month’s data has to be uploaded in our application, of which we will extract past 12 days data from the current date (Fig. 2).

3.3 Data Extraction

The dataset obtained from Google takeout has various details such as duration spent on walking, riding or driving based on speed, etc. [10]; some of these details are trivial for our application and are hence not considered. The data fields that we actually require are latitudes, longitudes and timestamps classified based on when the user has entered a location (start timestamp). The timestamp contains date, hours, minutes and seconds are obtained by converting raw JSON data to CSV for better processing. Another important parameter is accuracy which illustrates how accurate the data is. From reference [11], “Experiments and Terminology,” the location provided by Google is considered accurate if the actual position and the location is less than the accuracy provided by Google, i.e., the actual location should fall inside the circle and radius of the Google location and the provided Google accuracy. Considering the previous mentioned phenomenon, we found that more than 80% of our data to be most accurate when the accuracy value considered is 20.

Fig. 2 Screenshot of the raw data user gets from downloading their Google location history

```
1 {
2   "locations" : [ {
3     "timestampMs" : "1431948762628",
4     "latitudeE7" : 129116549,
5     "longitudeE7" : 775230586,
6     "accuracy" : 69,
7     "activity" : [ {
8       "timestampMs" : "1431948703460",
9       "activity" : [ {
10        "type" : "UNKNOWN",
11        "confidence" : 48
12      }, {
13        "type" : "IN_VEHICLE",
14        "confidence" : 33
15      }, {
16        "type" : "STILL",
17        "confidence" : 17
18      } ], {
19        "type" : "ON_BICYCLE",
20        "confidence" : 2
21      } ]
22    } ]
23 }
```

All of the mentioned data fields in Fig. 3 are extracted with the help of a library called GeoPy. This library is used to convert latitude and longitude coordinates to Geocode which is an address in a human readable form. These parameters help in increasing the efficiency of searching and comparing the data. Our application has two main categories of users, an infected patient and a vulnerable citizen.

3.4 Implementation of User Perspectives

3.4.1 An Infected Patient

In this case, when a person tests positive for the virus, from the time of the test, we collect the patient’s data and split it based on their country and city. The GeoPy library is used to sort the patient’s data to meet the requirements. Here, we sort the time data field for the patient, and store it in descending order, i.e., the last time of the visit to a certain location is displayed first. Upon sorting the data based on the remaining constraints, we publish the necessary information to a remote server which is accessible by the public. During this process, we maintain complete anonymity with respect to the patient’s private information.

Fig. 3 After data extraction, JSON data is converted to CSV with the above fields for 12 days

	column 1	column 2	column 3
1	Time	Latitude	Longitude
2	2020-07-01 00:01:23	12.91171080	77.52296210
3	2020-07-01 00:50:59	12.91169740	77.52295230
4	2020-07-01 00:51:01	12.91169740	77.52295230
5	2020-07-01 00:51:22	12.91169740	77.52295230
6	2020-07-01 00:56:22	12.91169740	77.52295230
7	2020-07-01 01:01:50	12.91169740	77.52295230
8	2020-07-01 01:07:06	12.91169740	77.52295230
9	2020-07-01 01:14:19	12.91169740	77.52295230
10	2020-07-01 01:25:25	12.91169740	77.52295230
11	2020-07-01 01:36:10	12.91169740	77.52295230
12	2020-07-01 01:46:50	12.91169740	77.52295230
13	2020-07-01 01:58:03	12.91169740	77.52295230
14	2020-07-01 02:13:04	12.91169740	77.52295230
15	2020-07-01 02:23:04	12.91169740	77.52295230
16	2020-07-01 02:34:13	12.91169740	77.52295230
17	2020-07-01 02:45:00	12.91169740	77.52295230
18	2020-07-01 02:55:13	12.91169740	77.52295230
19	2020-07-01 03:04:07	12.91169740	77.52295230
20	2020-07-01 03:07:27	12.91169740	77.52295230
21	2020-07-01 03:10:28	12.91169740	77.52295230
22	2020-07-01 03:13:30	12.91169740	77.52295230
23	2020-07-01 03:15:30	12.91169740	77.52295230
24	2020-07-01 03:18:56	12.91168380	77.52295970
25	2020-07-01 03:20:58	12.91168060	77.52295200
26	2020-07-01 03:22:59	12.91167870	77.52295910
27	2020-07-01 03:25:00	12.91169390	77.52294750
28	2020-07-01 03:27:01	12.91169370	77.52294950
29	2020-07-01 03:29:26	12.91169370	77.52294950
30	2020-07-01 03:35:02	12.91169370	77.52294950
31	2020-07-01 03:40:10	12.91169370	77.52294950
32	2020-07-01 03:51:00	12.91169370	77.52294950
33	2020-07-01 04:01:23	12.91169370	77.52294950
34	2020-07-01 04:12:24	12.91169370	77.52294950
35	2020-07-01 04:20:14	12.91169370	77.52294950
36	2020-07-01 04:25:00	12.91169370	77.52294950
37	2020-07-01 04:35:00	12.91169370	77.52294950

3.4.2 A Vulnerable Citizen

In this case, the person obtains a copy of his geolocation from Google with the help of the above-mentioned steps and then uploads the data for the corresponding month. In the background, GeoPy is used to sort their data based on country and city, along with which, we will sort it based on date and time in a descending order, for a period of 12 days. We then check the user’s data points from the sorted data and compare it to the attributes of the infected person on the remote server which is accessible publicly. This helps in optimizing the search and comparison (Fig. 4).

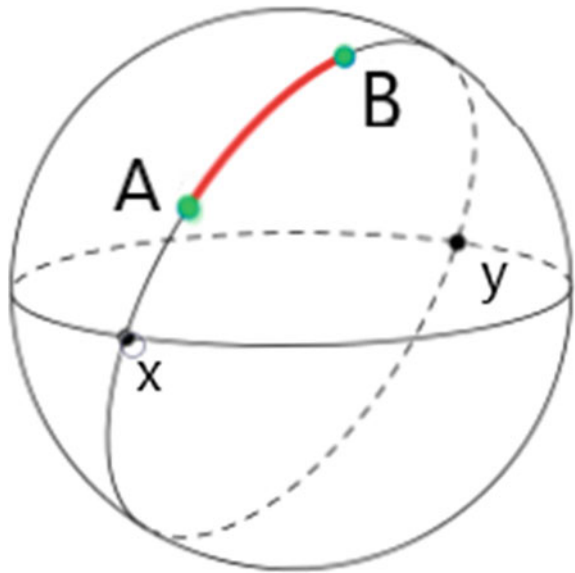
Each individual’s geolocation coordinates will be compared to the data present in the remote server [11]. If the user’s data points or latitudes and longitudes overlap with the infected patient’s data points with a [12] radius of 6 feet or less, we mark that data point as the point at which the vulnerable person has come in close contact with the infected person. We calculate this distance between two people using great circle distance (Fig. 5).

We calculate the distance between the points using great circle distance formula represented below:

Index	Latitude	Longitude	Time
74	13.0031108	77.5991922	2020-03-12 10:21:46
73	13.0031127	77.5991991	2020-03-12 10:21:44
72	13.0024922	77.598908	2020-03-12 10:21:35
71	13.0019178	77.5990151	2020-03-12 10:21:19
70	13.0010606	77.5995061	2020-03-12 10:21:04
69	13.0010329	77.5995144	2020-03-12 10:21:03
68	13.0009056	77.5996474	2020-03-12 10:20:51
67	13.0005043	77.5998742	2020-03-12 10:20:47
66	12.9998018	77.5997344	2020-03-12 10:20:32
65	12.9996291	77.5991088	2020-03-12 10:20:16
64	13.0000909	77.5988	2020-03-12 10:20:01
63	13.0015491	77.5982432	2020-03-12 10:19:33
62	13.0015902	77.5981172	2020-03-12 10:19:32
61	13.001631	77.5981038	2020-03-12 10:19:31
60	13.0016698	77.5980358	2020-03-12 10:19:30
59	13.0020622	77.5985347	2020-03-12 10:19:04
58	13.0016926	77.5951138	2020-03-12 10:18:48
57	13.0015887	77.5945681	2020-03-12 10:18:32
56	13.0011636	77.5945772	2020-03-12 10:18:21
55	13.0008616	77.5946754	2020-03-12 10:18:18
54	12.9979724	77.5955348	2020-03-12 10:17:47
53	12.9966123	77.5958327	2020-03-12 10:17:31
52	12.9965412	77.5958406	2020-03-12 10:17:30
51	12.9956318	77.5959338	2020-03-12 10:17:16
50	12.9945887	77.5959785	2020-03-12 10:17:00
49	12.9936592	77.5962217	2020-03-12 10:16:45
48	12.9932499	77.5965456	2020-03-12 10:16:30
47	12.9932114	77.5965103	2020-03-12 10:16:29
46	12.9924758	77.5948597	2020-03-12 10:16:16
45	12.9917191	77.5942037	2020-03-12 10:16:00
44	12.9915746	77.5941034	2020-03-12 10:15:44
43	12.9912983	77.593958	2020-03-12 10:15:29

Fig. 4 Representation of data, after sorting based on time, in descending order for 12 days

Fig. 5 Great circle distance between points A and B on a sphere; in our case, A: Infected person, B: Cautious Citizen



$$\begin{aligned}
 a &= \sin^2(\Delta\varphi/2) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2(\Delta\lambda/2) \\
 c &= 2 \cdot a \tan 2(\sqrt{a}, \sqrt{1-a}) \\
 d &= R \cdot c
 \end{aligned}
 \tag{6}$$

Where:

φ is latitude, λ is longitude, R is Earth's radius (Mean Radius = 6,371 km).

a = Square of half the chord length between the two points.

c = Angular distance in radians.

This point is used to alert a person if he's at high risk. We do not stop searching at that instant of time. We continue the process for a period of 12 days to check if the vulnerable person has come in contact with an infected patient at multiple data points. If the vulnerable person has come in contact with one or more infected patients at multiple data points, we store all of these points of contact in an array. Once the search is completed, a detailed report is provided to the user, indicating all the places, dates and times on which the user might have come in contact with the infected person.

We then conduct a survey to provide a more accurate risk assessment of the user. We query the user to know multiple things such as: if they remember visiting that location at that particular date and time and if it was crowded. If they were wearing a mask and were disinfecting and sanitizing themselves frequently. If they had any form of physical contact with a person (like shaking hands, hugging, exchanging objects, etc.). If they have any kind of symptoms related to the disease. Based on the obtained answers, we provide a complete risk assessment stating the probability of contracting the virus. This probability is provided in terms of percentage. If the risk is below 75%, then the user must continue to be safe and follow precautionary guidelines such as social distancing. If the risk is between 75 to 90%, the user is at risk of contraction of the virus and must refrain from meeting more people. If the risk is 90% and above then the patient is at high risk of contraction of the virus and must get tested immediately.

4 Results and Discussion

To test our application's credibility at a fundamental level, we downloaded our own location history for the month of March. Our application then extracted the required data to test from the infected patient's perspective from March 1st to March 12th and converted the format from JSON to CSV to create mock/random data. For the vulnerable citizen's perspective, we followed the same procedure and extracted data from the dates March 3rd to March 15th. Furthermore, we classified this mock data into its respective country, state and city for enhanced search optimization. In addition to this, we sorted the data with respect to the time. After the sorting process, we received 150 location instances for the vulnerable citizen and 110 location instances for the infected person. Out of these abundant instances, a majority of the locations were from Bengaluru, Karnataka, India. Hence, we carried out the search for the same. We obtained 78 instances for the vulnerable person and 74 instances for the infected patients which were from Bangalore. With the help of the above scenarios, we encountered 15 points of vulnerabilities, i.e., 15 common points of contact between the infected patient and the vulnerable citizen (Figs. 6 and 7).

```
1 ('13.0031127', '77.5991991') , 2020-03-12 10:21:44
2 ('12.9843827', '77.5876542') , 2020-03-12 10:11:27
3 ('12.9843827', '77.5876542') , 2020-03-12 10:11:26
4 ('12.9843796', '77.5876475') , 2020-03-12 10:10:59
5 ('12.9843827', '77.5876542') , 2020-03-12 10:11:26
6 ('12.9843796', '77.5876475') , 2020-03-12 10:10:59
7 ('12.9843796', '77.5876475') , 2020-03-12 10:10:59
8 ('12.9116776', '77.5239624') , 2020-03-05 11:33:38
9 ('12.9116798', '77.523969') , 2020-03-05 11:33:23
10 ('12.9116884', '77.5239748') , 2020-03-05 11:33:07
11 ('12.9116798', '77.523969') , 2020-03-05 11:33:23
12 ('12.9116884', '77.5239748') , 2020-03-05 11:33:07
13 ('12.9116884', '77.5239748') , 2020-03-05 11:33:07
14 ('12.9059462', '77.5125484') , 2020-03-05 10:04:17
15 ('12.9059243', '77.5125627') , 2020-03-05 10:03:46
```

Fig. 6 Result of search and compare with respective latitudes, longitudes, date and time

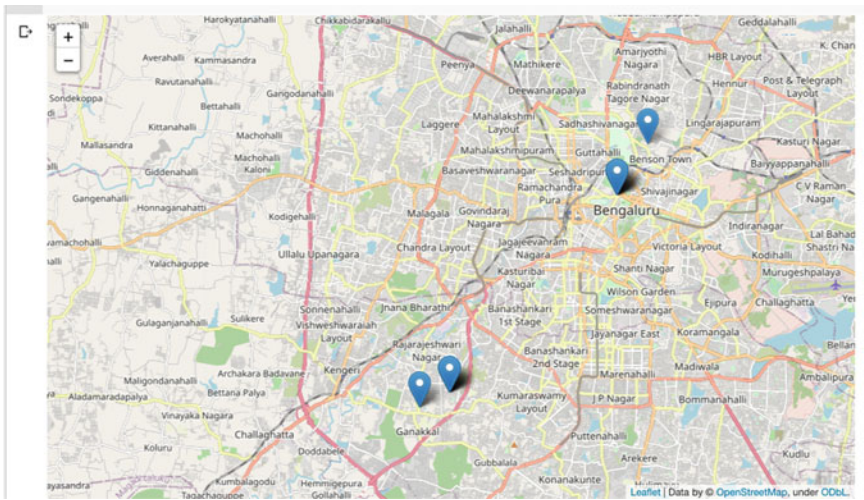


Fig. 7 Map representation of the vulnerable data points

Here, we can observe a detailed report of the place and time at which a vulnerable person came in contact with the infected patient. Further, to narrow down the prospective points of catching the virus and to enhance our contact results, we recommend the user to take the survey (Fig. 8).

Upon obtaining the 15 vulnerabilities, based on whether we have visited them or not, we took a survey to analyze the percentage of risk. From the survey we took, we reduced the number of vulnerabilities to eight relevant locations visited by us. Seven

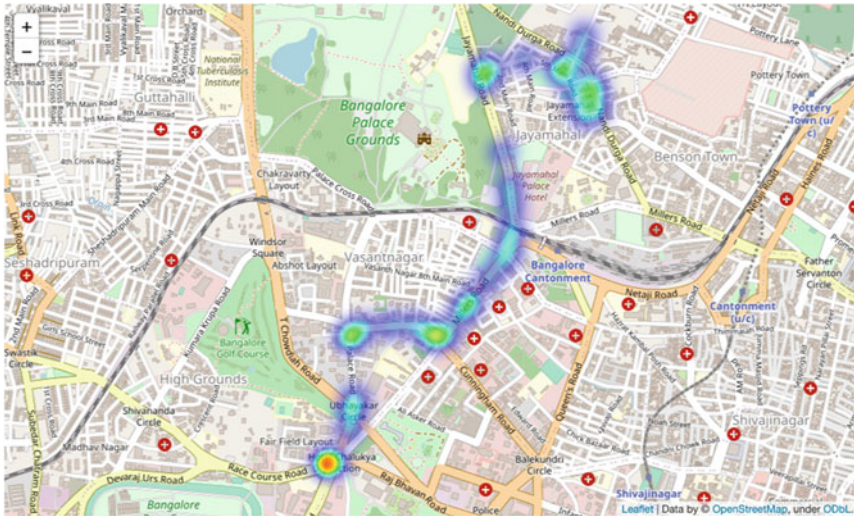
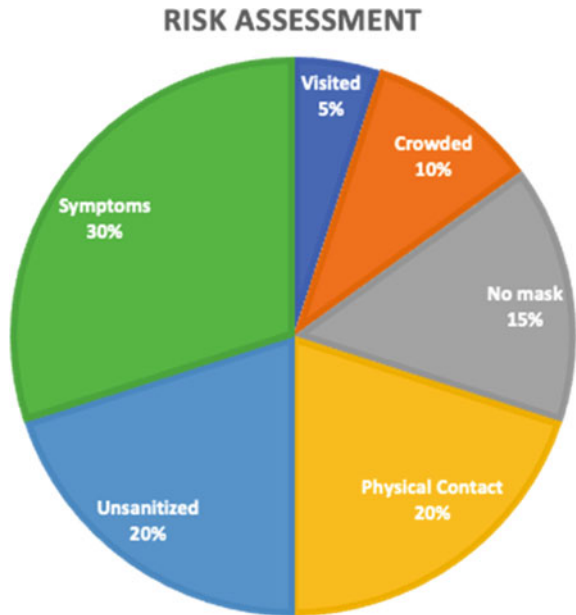


Fig. 8 Heat map representation of the points where user could have come in contact with the infected person

out of the 15 vulnerabilities were not considered, due to the persistent movement of the user, which were detected due to a moving vehicle (Fig. 9).

Fig. 9 Fundamental risk assessment parameters considered using the survey along with their respective percentages of risk



We analyzed the risk percentages for 8 relevant location points out of 15 vulnerabilities and obtained a result which illustrated that we were at a risk of 35–55% of contracting the virus.

5 Challenges and Future Enhancements

- Requirement of high computational capacity—Our application requires a high computational capacity for a few tasks such as the need to reduce a number of irrelevant data fields for multiple users as we need to categorize the data in terms of countries, states and cities. In the future, we intend to implement better algorithms and methods to reduce the time required to sort and clean the data to meet our requirements and to reduce the required computational capacity.
- Unclear data fields and types—While working with Google’s geolocation data, it has been highly meticulous to understand a few of the data fields, their functionality and their data types. Our application can have a great future direction if we can understand some of these fields and attempt to use them to increase the efficiency of the application altogether.
- Vulnerabilities caused due to user movement—Some of the vulnerabilities that were detected were caused due to the movement of the user. This movement could have been caused due to some mode of transportation of the user. These specific vulnerabilities are currently of less significance in terms of the survey involved in our application. In the future, we wish to classify the movement data based on the medium of travel of the user (such as walking, running, traveling in a bike, car, bus, train, flight, etc.) which can be further obtained by the geolocation details. This classification can help increase the efficiency of the survey. More important details can be obtained about whether the user was traveling in their own mode of transport (low risk) or in a public transport (medium to high risk), which can help improve the accuracy of our survey and risk assessment tests.
- Disability of geolocation feature—Android devices have geolocation feature enabled by default. But, some devices may have disabled this feature due to which it is not possible to access the location history. This becomes a limitation for the concerned user while using our application.

6 Conclusion

- We have achieved an automated process for contact tracing of the spread of virus by using geolocation data available in the Google timeline of every Google account.
- The positive person’s location history is extracted, preprocessed and published on a remote server which is made accessible for the public.

- The concerned user can upload their relevant location history to access their risk of contracting the virus. This is carried out by searching and comparing the user's data with the infected patient's data present in the remote database.
- We used great circle distance to find the distance between the infected person and the concerned user. Using this, we were able to accurately locate the most vulnerable data points.
- The risk assessment survey helped narrow down the vulnerabilities to the ones which are more likely to be of higher risk in contracting the virus which helped in avoiding unnecessary panic by the user.

References

1. Backer JA, Klinkenberg D, Wallinga J (2020) Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travelers from Wuhan, China. In: Incubation period distribution. Euro Surveill <https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062>
2. Jhunwala AJ (2020) Role of telecom network to manage COVID-19 in India: Aarogya Setu. <https://www.ncbi.nlm.nih.gov/>. <https://doi.org/10.1007/s41403-020-00109-7>
3. Macarulla Rodriguez A, Tiberius C, Van Bree R, Geradts Z (2018) Google timeline accuracy assessment and error prediction 3(3). <https://doi.org/10.1080/20961790.2018.1509187>
4. Sharma M, Morwal S (2015) Location tracking using Google Geolocation API. Int J Sci Technol Eng 1(11):29–32
5. Bajaj R, Lalinda Ranaweera S, Agrawal D (2002) GPS: location tracking technology. https://www.researchgate.net/publication/220477594_GPS_Location-Tracking_Technology
6. Kharaisat YSH, Al-Khateeb M, Abu-Alreesh Y, Ayyash A (2011) GPS navigation and tracking device. In: Int J Interact Mobile Technol. <https://doi.org/10.3991/ijim.v5i4.1781>
7. Dhar T (2020) Aarogya Setu—carrying your privacy in your hands. In: National academy of legal studies and research academy. <https://dx.doi.org/https://doi.org/10.2139/ssrn.3614506>
8. Garg S, Bhatnagar N, Gangadharan N (2020) A case for participatory disease surveillance of the Covid-19 pandemic in India, vol 6, no 2. JIMR Publications
9. Jeong H, Kim S (2018) Indoor smartphone localization based on LOS and NLOS identification. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6263745/>
10. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith H, Azman AS, Lessler J, Reich NG (2020) Incubation period of corona virus disease 2019(Covid-19). ACP J <https://www.acp-journals.org/doi/https://doi.org/10.7326/M20-0504>
11. Tirkey A (2020) Unchartered territory—emerging world order post Covid-19. Observer Research Foundation
12. Dubov A, Shoptaw S (2020) The values and ethics of using technology to contain the Covid-19 epidemic. Am J Bioethics <https://doi.org/10.1080/15265161.2020.1764136>
13. Lauer SA, Grantz KH, Bi Q et al (2020) Estimated incubated period of Covid-19. American College of Cardiology, Venkatesh Locharla Murthy, MD, PhD, FACC
14. Korpilo S, Virtanen T, Lehvavirta S (2017) Smartphone GPS tracking— inexpensive and efficient data collection on recreational movement, vol 157. Landscape and Urban Planning. <https://doi.org/10.1016/j.landurbplan.2016.08.005>
15. Sardianos C, Varlamis I, Bourlas G (2018) Extracting user habits from Google Maps history logs. In: IEEE: international conference on advances in social networks analysis and mining

16. Mukhopadhyay A, Mallissery A (2018) TELIL: A trilateration and edge learning based indoor localization technique for emergency scenarios. In: International conference on advances in computing, communications and informatics. <https://doi.org/10.1109/ICACCI.2018.8554587>
17. Sajed AN, Amgain K (2020) Corona Virus Disease (Covid-19) Outbreak and the strategy for prevention. Euro J Med Sci

Optimized Two-Dimensional Chaotic Mapping for Enhanced Image Security Using Sea Lion Algorithm



H. R. Latha and A. Ramaprasath

Nomenclature

Abbreviation	Description
T-DES	Triple-Data Encryption Standard
IDEA	International Data Encryption Algorithm
DES	Data Encryption Standard
AES	Advanced Encryption Standard
PTSTFrFT	Phase-Truncated Short-Time Fractional Fourier Transform
EU	Encryption Unit
MGA	Modified Genetic Algorithm
CTM	Chaotic Tent Map
LTM	Logistic-Tent Map
DNA	Deoxyribonucleic Acid Coding
ECC	Elliptic Curve Cryptosystem
KKC-NN	Chaotic Key Controlled Neural Networks
KCFF-NN	Key Controlled Finite Field Neural Network
LDMLNCML	Logistic-Dynamic Mixed Linear-Nonlinear Coupled Map Lattices
2D	Two Dimensional
AF-SL _n O	Average Fitness-Based Sea Lion Optimization Algorithm
SL _n O	Sea Lion Optimization Algorithm
H-CS	Hyper-Chaotic System

H. R. Latha (✉)

Department of Computer Applications, Jindal College for Women, Bangalore, Karnataka, India

Research Scholar, Hindusthan Institute of Technology & Science, Chennai, India

A. Ramaprasath

Department of Computer Applications, Hindusthan Institute of Technology & Science, Chennai, India

e-mail: rprasath@hindustanuniv.ac.in

HS-IEA	High-Sensitivity Image Encryption Algorithm
IEA	Image Encryption Algorithm
PP and CB	Phase Portrait and Chaotic Behaviors
2D-LM	2D- Logistic Map
CS	Chaotic System
LD	Logistic Diffusion
LT	Logistic Transposition
LP	Logistic Permutation
LSG	Logistic sequence generator

1 Introduction

Over the decades, the industrial communications technology is facing a faster pace of expansion owing to the increase in the quantity of the information transmitted over the Internet [1–3]. The data transmitted are not alone in the form of texts, but also audio, image and other multimedia files. Among all these, the images are more commonly utilized and the security of image data transfer is becoming a crucial requirement. With the aid of the computer networks, the authorized persons are permitted to share their information from a distance [4, 5]. In the last few years, the security of the multimedia information is provided by a “class of tool-sets and design insights” with the intention of enhancing its protection against diverse attack scenarios. The security of the information is higher when it is concealed from illegal admittance (privacy), secluded from unauthorized transform (truthfulness) and accessible to an endorsed individual when it is required (accessibility). Further, when there is a shortage in bandwidth for communication or storage, there occurs compression of data. Particularly, while sharing data via the wireless communication network with bandwidth limitations [6–8], it is more vital to deploy the low-bit-rate compression algorithms. During this compression process, one natural question that arises is security, loss of data and confidentiality of the data. Consequently, encryption is performed to enhance user privacy.

In the case of images, the image encryption algorithms are preferred to assure security [9–11]. In literature, a huge count of encryption systems have be formulated to encrypt as well as decrypt the image data. But, till now, there is no single encryption algorithm to be applied for all images. Typically, most of the encryption algorithms available today are utilized for textual data and not for multimedia data. [12–14]. Even though T-DES and IDEA can achieve high security, they possibly will not be appropriate for multimedia applications [15–17].

Even though the traditional encryption algorithms might be good in encrypting the images directly, they fail to hold good under two reasons. The image dimension is greater than the dimension of the text, and the time consumed for performing the encryption is higher. Secondly, the text that is decrypted need to be equivalent to the original one, but this is not compulsory for image data. Diverse encryption systems

have been utilized in different applications like “internet communication, multimedia systems, medical and military imaging systems.” Each of the encryption systems has its own set of characters like high correlation, high redundancy, etc. Thus, the novel procedure should be utilized to guard the privacy of image data from illicit access.

The foremost objective of the current research work is highest below:

- An O2DCM is proposed for image encryption.
- A new optimization algorithm referred to AF-SLnO is introduced to fine-tune the initial chaotic system parameters.
- The security improvement is demonstrated using standard security examinations such as key sensitivity analysis, histogram analysis and adjacent pixel autocorrelation test.

The rest of the paper is organized as: Section 2 addresses about the most interesting literature works done in image encryption. Section 3 describes the proposed image encryption approach: a short overview. Section 4 portrays about the optimized 2DCM development for image encryption and decryption. The resultants acquired with the proposed model are discussed briefly in Sect. 5. Finally, a strong conclusion is provided to the current research work in Sect. 6.

2 Literature Review

2.1 Related Works

In 2020, Yu et al. [18] have projected a novel IEA based on the PTSTFrFT and the H-CS. Here, the authors have split the original images into four sub-images for independent image encryption. They have encoded these sub-images with EU that was constructed from wave-based permutation. Then for ensuring the integrity of image information as well as for nonlinearity of phase truncation, they have linked the confusing phase information with amplitude information. The experimental evaluation of the projected model had revealed better key space and high sensitivity.

In 2019, Wang et al. [19] have introduced an innovative medical image encryption method on the basis of MGA and “coupled map lattices.” Initially, the authors have generated the secure cipher images with the aid of the coupled map lattice and have considered the count of the protected cipher images as the preliminary population. With this initial population, they have initialized the MGA. As a consequence, the computational time was lessened and entropy of the cipher images was improved.

In 2019, Xingyuan et al. [20] have projected a novel IEA based on the “ZigZag transform and LL compound chaotic system.” The authors have accompanied these two systems (“ZigZag transform and LL compound chaotic”) to split the original image into block sub-channel. Then, they have sorted the block sub-channel following the identification values using the sorting scrambling algorithm. This

model had destroyed the pixel values and has spread it with the adjacent-side XOR approach. This mode was said to have a better encryption effect, and it was resistant to common attacks.

2.2 Review

Table 1 [22] shows the features and challenges of the existing image encryption models. Among them, PTSTFrFT in [18] has an adequate amount of key space. But, the encrypted image correlation is weaker than the correlation of the original image. In [19], the MGA approach has low computational time. In addition, it does not have an essential key space. ZigZag transform and LL compound chaotic [20] can resist common attacks, but it possesses a low correlation. The security of encryption images is high in RT-enhanced chaotic tent map [21]. Despite this, it cannot resist chosen-plaintext attack. Further, DNA coding and chaotic operations in [22] have good capability to resist differential attacks. Apart from this advantage, here the cipher images are not completely different, and hence, the level of security needs further improvement. A simple and secure approach is constructed in [23] referred to ECC. Here, greater is the size of the key, it is complex to perform the “brute-force attack.” Thus, cannot be deployed in a large scale. Further, Lorenz chaotic system in [24] is robust to cropped attacks with high entropy, and hence, there is a chance for noise to get creep into the image. LDMLNCML in [25] lessens the correlation between the adjacent pixels and possesses good chaos high-sensitivity IEA. But, this technique is immune to high noise intensity.

3 Proposed Image Encryption Approach: A Short Overview

O2DCM is developed here for performing the task of encrypting the image data. The flow chart of the proposed image encryption model is illustrated in Fig. 1. The internal loop of the proposed model encompasses, “2-D logistic permutation, 2-D logistic diffusion, and 2-D logistic transposition.” Each of the stages acts alike the image cipher and altogether they form a “permutation-substitution network.”

The steps involved in image encryption and decryption process:

- Initially, an optimal 2D-LM is adopted for encrypting the input image I_{in} . Here, the diffusion and confusion properties are taken into consideration.
- A complicated chaotic map is utilized for generating the pseudorandom sequences.
- Since the chaotic mapping is extremely dependent on the “preliminary system constraints and the order of permutation,” it is exigent to describe or decide the most favorable preliminary parameters of the system. Fine-tuning of these preliminary constraints of the system is initiated as computationally incompetent.

Table 1 Features and challenges of existing image encryption techniques

Author [Citation]	Methodology	Features	Challenges
Yu et al. [18]	PTSTFrFT	<ul style="list-style-type: none"> ✓ Has a sufficient amount of keyspace and high key sensitivity ✓ Highly robust against widespread attacks ✓ Immune to the chosen-plaintext attack 	<ul style="list-style-type: none"> × The encrypted image correlation is weaker than the correlation of the original image
Wang et al. [19]	MGA	<ul style="list-style-type: none"> ✓ Secure against brute-force attack ✓ Low computational time 	<ul style="list-style-type: none"> × Do not have essential keyspace × The quality of decryption images are low
Xingyuan et al. [20]	ZigZag transform and LL compound chaotic	<ul style="list-style-type: none"> ✓ Can resist common attacks ✓ Can stand firm to occlusion attack 	<ul style="list-style-type: none"> × Low correlation
Zhu and sun [21]	RT-enhanced chaotic tent map	<ul style="list-style-type: none"> ✓ Security of encryption images is high 	<ul style="list-style-type: none"> × Cannot defend against the chosen-plaintext attack × Insensate to all the chaotic secret keys
Zhang and Wang [22]	DNA coding and chaotic operations	<ul style="list-style-type: none"> ✓ High sensitivity of plain-text ✓ Superior potential to stand firm for the degree of different attacks 	<ul style="list-style-type: none"> × Cipher images are not completely different
Guan et al. [23]	ECC	<ul style="list-style-type: none"> ✓ Simple and secure ✓ Greatest in terms of the encryption speed ✓ Resist the known-plaintext attack 	<ul style="list-style-type: none"> × Greater is the size of the key, it is complex to perform the brute-force attack
Thoms et al. [24]	Lorenz chaotic system	<ul style="list-style-type: none"> ✓ Robustness to cropped attacks 	<ul style="list-style-type: none"> × High entropy
Nematzadeh et al. [25]	LDMLNCML	<ul style="list-style-type: none"> ✓ Decrease the correlation between the adjacent pixels ✓ Possesses good chaos ✓ High-sensitivity IEA 	<ul style="list-style-type: none"> × High noise intensity

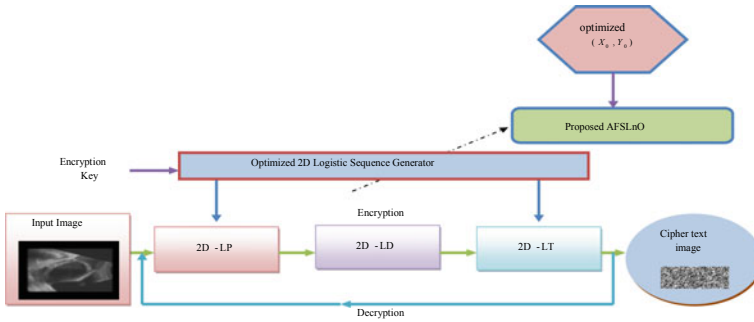


Fig. 1 Diagrammatic representation of the proposed image encryption and decrypt

- Thus, a new optimization algorithm referred as AF-SLnO is introduced for determining the most favorable CS parameters.
- The O2DCM model necessitates a security model to represent the “apriori ciphered image.” The key is represented as Key.
- Hence, in the chaotic key generation system, the entropy is maximized by fine-tuning the initial parameters (X_0, Y_0) with proposed AF-SLnO. Therefore, a two-dimensional logistic map becomes optimized and hence referred as O2DCM. The clipped image of O2DCM is denoted as I_c .
- As a consequence, optimal preliminary system constraints for the CS can be evaluated. Concerning these system parameters, image encryption can be achieved. The encrypted image is decrypted by reversing the process.

4 Optimized 2DCM Development for Image Encryption and Decryption

The 2DCM is more complicated than the 1DCM. The 2D-LM can be mathematically defined discretely as per Eq. (1). Here, R is the parameter of the system and by knowing about the (X_0, Y_0, R, j) , the trajectory at j th point can be determined as per Eq. (2).

$$2DCM = \begin{cases} X_{j+1} = R(3Y_j + 1)X_j(1 - X_j) \\ Y_{j+1} = R(3X_{j+1} + 1)Y_j(1 - Y_j) \end{cases} \quad (1)$$

$$\begin{cases} X_j = L_X^{2D}(X_0, Y_0, R, j) \\ Y_j = L_Y^{2D}(X_0, Y_0, R, j) \end{cases} \quad (2)$$

PP and CB: The 2D-LM as per Eq. (1) is a complicated “dynamical system.” Concerning R (system parameter), the evolution of the map takes place from one dynamics to other. On the basis of the variation in R (system parameter), the 2D-LM behavior is modeled as given in

As per the Table 2, it is clear that the system becomes chaotic, when $R \in (1.11, 1.19)$. The key “encryption and decryption” process is shown in Figs. 2 and 3, respectively.

The key decryption is nothing, but the reversal of the encryption process. Further, the “encryption and decryption process” is mathematically defined in Eqs. (3) and (4), respectively.

$$I_c = \text{Enc}(I_{in}, \text{Key}) \tag{3}$$

Table 2 Two-dimensional logistic map behavior with respect to system parameter

S.No	R (system parameter)	Behaviors of the map
1	$R \in (-1, 1)$	<ul style="list-style-type: none"> • Encompasses 1 attractive node and 2 saddle points • Makes X and Y axes to become unstable in the system
2	$R \in (1.11, 1.19)$	<ul style="list-style-type: none"> • In between the “invariant close curve,” there exist alterations with “oscillations, frequency locking, cyclic chaotic behaviors, contact bifurcations with basin boundaries, and single chaotic attractor”
2	$R = 1$	<ul style="list-style-type: none"> • The NeimarkHopf bifurcation is undergone by the attractive focus of the system
3	$R \in (1, 1.11)$	<ul style="list-style-type: none"> • The “attractive focus” of the system becomes repellent and tends to cause “oscillations” in the system
5	$R > 1.19$	<ul style="list-style-type: none"> • Unbalanced system

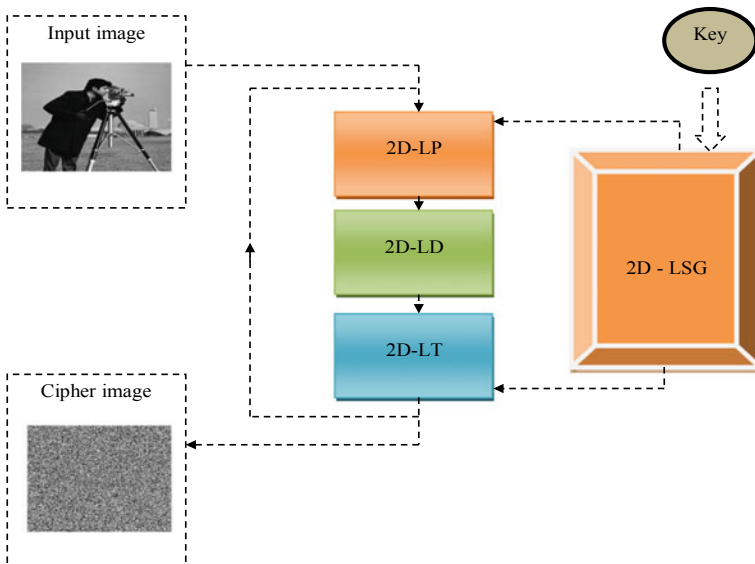


Fig. 2 Diagrammatic representation of the encryption process using 2D logistic mapping

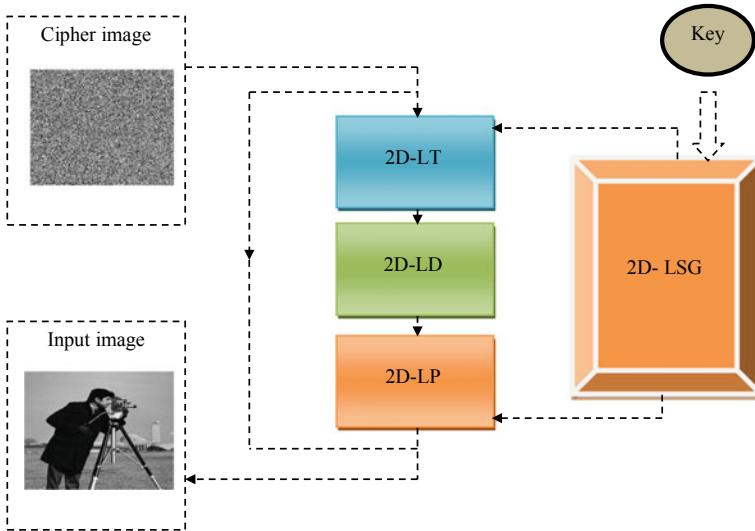


Fig. 3 Diagrammatic representation of the decryption process using 2D logistic mapping

$$I_{in} = Dec(I_c, Key) \tag{4}$$

where $I_{in} \rightarrow$ input image,
 $I_c \rightarrow$ cipher image and.
 Key \rightarrow Encryption key.

Key Schedule and 2D-LSG: The key Key is a “256-bit string” that encompasses five sections: “ $X_0, Y_0, R, T, B_1 \dots B_8$ ”. The key composition of Key is shown in Fig. 4.

Here, X_0, Y_0, R are the preliminary value and the constraints in 2DCM. Hence, they are fine-tuned by a new optimization algorithm referred as AF-SLnO. In addition, T and $B_1 \dots B_8$ are linear congruential generator parameters. Further, the coefficients of $B_1 \dots B_8$ comprises of 6-bit string $\{a_0, a_1, \dots, a_5\}$, which is translated into integers to obtain the requisite coefficients. For each of the rounds, the preliminary value is defined using Eq. (5).

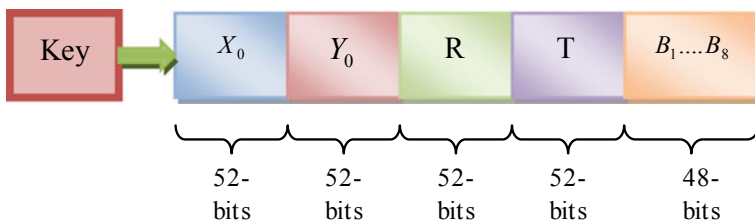


Fig. 4 Key composition of the encryption Key

$$\begin{cases} X_0^{\text{round\#}} = T + X_0 B_{(\text{round\# mod } 8)+1} \text{ mod } 1 \\ Y_0^{\text{round\#}} = T + Y_0 B_{(\text{round\# mod } 8)+1} \text{ mod } 1 \end{cases} \tag{5}$$

Thus, the most appropriate long chaotic sequence is generated using the long chaotic sequence $X_0^{\text{round\#}}, Y_0^{\text{round\#}}$ and R . The length of the long chaotic sequence is equivalent to the length of the input image (plain text) I_{in} .

The loop of the proposed model composes of three components: “Two-Dimensional logistic permutation, Two-Dimensional logic diffusion, and Two-Dimensional logistic transportation.” They are discussed in the subsequent section.

2D-LP: The dimension of input image I_{in} is $p \times q$. The count of pixels in I_{in} is pq . Let, (X_0, Y_0) be the preliminary value utilized in the round. Using Eq. (1) of 2D logistic map, the succession of “pair-wise X and Y ” can be generated. This sequence of X and Y be x_{seq} and y_{seq} , and it is mathematically defined in Eq. (6).

$$\begin{cases} x_{\text{seq}} = \{X_1, X_2, \dots, X_{pq}\} \\ y_{\text{seq}} = \{Y_1, Y_2, \dots, Y_{pq}\} \end{cases} \tag{6}$$

The elements of x_{seq} and y_{seq} having $p \times q$ elements are re-arranged in the form of a matrix and $p \times q$ matrices x and y are obtained, respectively. Further, the “bijective mapping” e_{π_x} and e_{π_y} are formed with R th of x . This is mathematically defined in Eqs. (7) and (8), respectively.

$$x_{R,j}^{\text{sorted}} = x_{R,e_x(j)} \tag{7}$$

$$X_{J,I_c}^{\text{sorted}} = X_{e_{y_j(j),I_c}} \tag{8}$$

Further, with the aid of Eqs. (9) and (10), the “row permutation matrix U^X and column permutation U^Y ” are obtained. Thus, it is easier to validate that the row permutation matrix is a permutation of the sequence $(1, 2, \dots, q)$. Alike this, column permutation is a permutation of the sequence $(1, 2, \dots, p)$.

$$U^X = [e_{\pi_x}^{R=1}, e_{\pi_x}^{R=2}, \dots, e_{\pi_x}^{R=p}]' \tag{9}$$

$$U^y = [e_{\pi_y}^{I_c=1}, e_{\pi_y}^{I_c=2}, \dots, e_{\pi_y}^{I_c=q}]' \tag{10}$$

The pseudo-code of the 2D Logistic Permutation is manifested in Algorithm 1.

Algorithm 1: Pseudo-code of the 2D Logistic Permutation Algorithm	
Input	U^X, U^Y and I_{in} (plaintext image)
Output	I_c
For $R = 1 : p$ do	
	For $I_c = 1 : q$ do
	$D_{R,I_c} = G_{U_{R,I_c}^X, C}$ // Permutation of pixel along the X axis
	End for
End for	
For $R = 1 : p$ do	
	For $I_c = 1 : q$ do
	$I_{C(R,I_c)} = D_{R,U_{R,I_c}^X, C}$ // Permutation of pixel along the Y axis
	End for
End for	

As a resultant of 2D logistic permutation, the pixel of the input plain text image I_{in} is shuffled well, and hence, the permuted image I_c^{perm} is unrecognizable.

2D-LD: the LD is deployed with the objective of achieving better diffusion properties for every $S \times S$ image block $I_{in(b)}$ of I_{in} . This is defined mathematically in Eq. (11). This takes place over the finite field GF (2^8). Mathematically L_d is defined in Eq. (13).

$$I_{c(b)} = (L_d \cdot I_{in(b)} \cdot L_d)_{2^8} \tag{11}$$

Here,

$L_d \rightarrow$ Maximal distance separation matrix that is acquired from the permutation matrices

$$G_b = (L_d^{-1} \cdot I_c^b \cdot L_d^{-1})_{2^8} \tag{12}$$

$$L_d = \begin{bmatrix} 4 & 2 & 1 & 3 \\ 1 & 3 & 4 & 2 \\ 2 & 4 & 3 & 1 \\ 3 & 1 & 2 & 4 \end{bmatrix}, (L_d^{-1})_{2^8} = \begin{bmatrix} 71 & 216 & 173 & 117 \\ 173 & 117 & 71 & 216 \\ 216 & 71 & 117 & 173 \\ 117 & 173 & 216 & 71 \end{bmatrix} \tag{13}$$

In the cipher rounds, the least number to have the dimension $p \times q$ changing pixels is represented in Eq. (14).

$$\#round_{min} = \log_{S \times S} p \times q = \log pq / 2 \log_2 S \tag{14}$$

Here, $S \rightarrow$ block size variable of I_{in} .

After processing of a certain rounds of sufficient clippers, even a smaller variation I_{in} leads to noteworthy transformations in ciphertext I_c . This goes ahead to initialize

the diffusion properties. Thus, I_{in} after the “two rounds of the diffusion process becomes incomprehensible.”

2D-LT: With respect to the “logistic sequence generated” acquired from the preceding stage, the pixels values changes by reference image F . The matrix z is the “sum of the matrices x and y ,” acquired from x_{seq} and y_{seq} in Eq. (15).

$$z = x + y \quad (15)$$

With respect to the reference image F , the image pixel values can be altered. This is accomplished as per Eq. (16). Here, the block b of the matrix z is a 4×4 block. Here, $f(b)$ denotes the block function, and it is expressed in Eq. (17).

$$F = f(b) \quad (16)$$

$$F = \begin{bmatrix} g_N(b_{1,1}) & g_H(b_{1,2}) & g_S(b_{1,3}) & g_Q(b_{1,4}) \\ g_H(b_{2,1}) & g_S(b_{2,2}) & g_Q(b_{2,3}) & g_N(b_{2,4}) \\ g_S(b_{3,1}) & g_Q(b_{3,2}) & g_N(b_{3,3}) & g_H(b_{3,4}) \\ g_Q(b_{4,1}) & g_N(b_{4,2}) & g_H(b_{4,3}) & g_S(b_{4,4}) \end{bmatrix} \quad (17)$$

The sub-functions $g_Q(\cdot), g_H(\cdot), g_N(\cdot)$ and $g_S(\cdot)$ are defined in Eqs. (18)–(21), respectively.

$$g_N(d) = T(d) \bmod W \quad (18)$$

$$g_H(d) = \lfloor T(d) \rfloor \bmod W \quad (19)$$

$$g_S(d) = T(d^2) \bmod W \quad (20)$$

$$g_Q(d) = T(2d) \bmod W \quad (21)$$

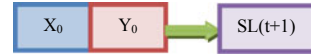
Here, $W \rightarrow$ count of allowable intensity scales of I_{in} format. As a resultant, the 2D-LT is achieved by shifting the individual pixels of I_{in} with the particular quantity of F .

In case of encryption with 2D logistic transposition, the ciphertext image is mathematically defined as per Eq. (22), while in the case of decryption with 2D-LT is shown in Eq. (23).

$$I_c = (I_{in} + 1) \bmod W \quad (22)$$

$$I_{in} = (I_c - 1) \bmod W \quad (23)$$

Fig. 5 Solution encoding



4.1 Objective Function and Solution Encoding

The preliminary parameters (X_0, Y_0) are optimized with new AF-SLNO model, to achieve the objective of entropy maximization of clipher image I_c . The solution encoding or input given to the algorithm is represented in Fig. 5. The objective function (Ob) of the current research work is mathematically defined in Eq. (24).

$$Ob = \max(\text{Entropy}(I_c)) \tag{24}$$

4.2 Proposed AF-SLNO

The traditional SLNO [26] was developed based on the raw inspiration acquired from the sea lions, the most intelligent animal. Since they are slower in convergence; there is a necessity to have self-improvement in it. The optimization algorithms have undergone various improvements in terms of many factors. One of them is by introducing adaptive operators or adaptive functions [27–30]. The new version of SLNO (AF-SLNO) is explained in the subsequent section:

- Step 1::** The overall population (Pop) of sea lions is initialized.
- Step 2::** Evaluate the fitness of the current solution $Fit(t)$ and compare it over the mean (or average) fitness of the solutions.
- Step 3:** if $Fit(t) < \text{mean}(Fit(t))$, then an Id is generated randomly within the range 1, 2 and 3 (i.e., $Id = \text{rand}$). Since the proposed work is based on the mean (or average fitness), this model is referred to AF-SLNO.
- Step 4::** If the current solution lies within the Id range 1($Id == 1$), the position of the search agents are updated using Eq. (25) of SLNO corresponding to prey detection and tracking phase.

$$\vec{Dis} = \left| 2\vec{F} \cdot \vec{E}(t) - \vec{SL}(t) \right| \tag{25}$$

- Here, $\vec{Dis} \rightarrow$ the distance between the objective sea lion search agent and prey.
- $\vec{F} \rightarrow$ random vector in the range 0 and 1.
- $\vec{E} \rightarrow$ positions vectors of the target prey.
- $\vec{SL} \rightarrow$ sea lion’s position vector.

Step 5: If the current solution lies within the Id range 2($Id == 2$), they get updated using the “Searching for prey (Exploration phase).” This is mathematically defined in Eq. (26).

$$\vec{SL}(t + 1) = \vec{SL}_{rnd}(t) - \vec{Dis}.. \vec{M} \tag{26}$$

Here,
 $\vec{SL}_{rnd} \rightarrow$ random sea lion.

Step 6: On the other hand, if the current solution takes the Id as 3(Id == 3), then the position of search agent is updated with the ‘‘Circle updating position’’ coming under the ‘‘Attacking phase (Exploitation phase).’’ This is mathematically defined in Eq. (27).

$$\vec{SL}(t + 1) = \left| \vec{E}(t) - \vec{SL}(t) \right| \cos(2\pi l) + \vec{F}(t) \tag{27}$$

Here,
 $\left| \vec{E}(t) - \vec{SL}(t) \right| \rightarrow$ distance between the most appropriate solution and sea lion (search agent).

$l \rightarrow$ Random number in the range $-1, 1$

Step 7: In case, if the mean fitness is lower than the current fitness, then the position of the current search agent is updated with a new update shown in Eq. (28).

$$\vec{SL}(t + 1) = L - bound + [U - bound - L - bound] * rand \tag{28}$$

Here,
 $L - bound \rightarrow$ Lower bound of the solution.
 $U - bound \rightarrow$ Upper bound of the solution.

The pseudo-code of the proposed AF-SLno model is exhibited in Algorithm 2.

Algorithm 2: Pseudo-code of the proposed AF-SLno model	
Initialization of overall population (<i>Pop</i>)	
Find the fitness (<i>Fit</i>) of the overall population	
if $Fit(t) < mean(Fit(t))$	
	$Id = rand(1,2,3)$
	If 1 $Id == 1$
	The positional update of the solution takes place using Eq. (24) corresponding to Prey detection and tracking phase
	Else If 2 $Id == 2$
	The positional update of the solution takes place using Eq. (25) corresponding to searching for prey (Exploration phase)
	Else If 3 $Id == 3$
	The positional update of the solution takes place using Eq. (26) corresponding to Circle updating position
	Else
	Update the solutions using Eq. (27)

The flow chart of the proposed model is shown in Fig. 6.
 Figure 6 Flow chart of the proposed AF-SLno.

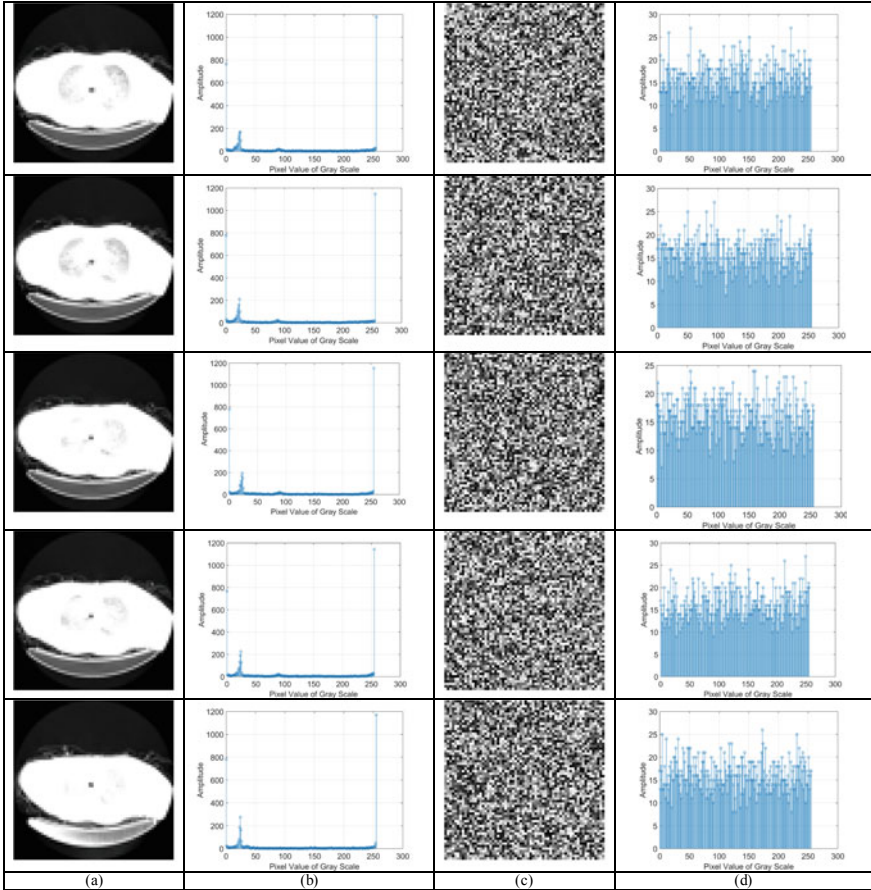
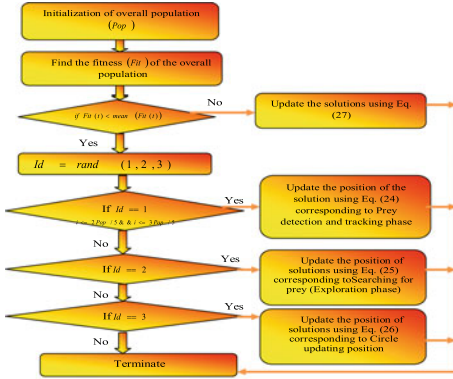


Fig. 6 Image Results under CT sample images. **a** Original CT images. **b** Histogram of original CT images. **c** Cipher CT images. **d** Histogram of cipher CT images

5 Results and Discussion

5.1 Simulation Procedure

The proposed image encryption model using the AF-SLNO was implemented in MATLAB, and the resultants acquired were noted. The O2DCM was investigated on “medical images such as CT, MRI and ultrasound images.” The datasets for CT, MRI are downloaded from “<https://www.osirix-viewer.com/>” and ultrasound is from “<https://www.ultrasoundcases.info/cases/abdomen-and-retroperitoneum/>” [Access Date: 2020–03-11]. The five ultrasound sample images with its cipher images are used. The security analysis for these images is demonstrated using standard security analysis such as entropy analysis, PSNR analysis, histogram analysis and adjacent pixel autocorrelation test. The presented work (AF-SLNO) is compared over the existing models like Standard, GWO [31], SLnO [32], PGWO [33], respectively.

5.2 Analysis on Information Entropy

The maximization of entropy being the objective is verified here for three image sets (CT, MRI and ultrasound images). The entropy of the presented work and existing work is shown in Table 3. On observing the entropy of the ultrasonic image, the presented work achieves the highest entropy in all five images. Further, for third image, the entropy of the presented work is 0.12, 0.007, 0.023 and 0.017% better than the existing works like standard, GWO, SLnO and PGWO, respectively. Then, in case of CT images, highest entropy (7.9683) is achieved by the presented work at fifth image and it is 0.14, 0.05, 0.007 and 0.05% better than the extant models like standard, GWO, SLnO and PGWO, respectively. Further, in the case of MRI images, the entropy of the presented work for fourth image is 7.9683, while the entropy of standard, GWO, SLnO and PGWO is 7.9571, 7.9653, 7.9668 and 7.9668, respectively. Thus, from the overall evaluation, it is obvious that the presented work shows the highest entropy value and hence reaches the objective.

6 Conclusion

The current research work has developed an O2DCM for encrypting the image data. The initial CS parameters were fine-tuned with a new optimization algorithm referred as AF-SLNO, which is an improved version of standard SLnO. The proposed AF-SLNO had maximized the “information entropy.” As a consequence, most favorable initial parameters for the CS were evaluated. The security improvement was demonstrated using standard security examination such as key sensitivity analysis,

Table 3 Analysis on information entropy for ultrasonic, CT and MRI images.

<i>Ultrasonic image</i>					
Approach	1	2	3	4	5
Standard	7.9586	7.9551	7.9575	7.9521	7.9551
GWO [31]	7.9662	7.9653	7.9645	7.9644	7.9654
PGWO [33]	7.9649	7.9668	7.9653	7.9668	7.9663
SlnO [32]	7.9656	7.9648	7.9658	7.966	7.9648
AF-SlnO	7.967	7.9665	7.9672	7.9643	7.9665
<i>CT image</i>					
Approach	1	2	3	4	5
Standard	7.9545	7.9618	7.9571	7.951	7.9571
GWO [31]	7.9673	7.9655	7.9678	7.9659	7.965
PGWO [33]	7.9673	7.9678	7.9679	7.9682	7.9677
SlnO [32]	7.9662	7.9655	7.9683	7.9653	7.9642
AF-SlnO	7.966	7.9654	7.9654	7.9683	7.9683
<i>MRI</i>					
Approach	1	2	3	4	5
Standard	7.9545	7.9618	7.9571	7.951	7.9571
GWO [31]	7.9665	7.966	7.9649	7.9643	7.9653
PGWO [33]	7.9681	7.9669	7.9649	7.9669	7.9668
SlnO [32]	7.9662	7.9655	7.9654	7.9653	7.9642
AF-SlnO	7.966	7.9654	7.9683	7.9683	7.9683

histogram analysis and adjacent pixel autocorrelation test. On observing the entropy of the ultrasonic image, the presented work achieves the highest entropy in all five images. Further, for third image, the entropy of the presented work is 0.12, 0.007, 0.023 and 0.017% better than the existing works like standard, GWO, SlnO and PGWO, respectively.

References

1. Dou S, Shen X, Zhou B, Wang L, Lin C (2019) Experimental research on optical image encryption system based on joint Fresnel transform correlator. *Opt Laser Technol* 112:56–64
2. Liu J, Bai T, Shen X, Dou S, Ying J (2018) Parallel encryption and hierarchical retrieval for multi-channel images using an optical joint Fresnel transform correlator. *Opt Commun* 426:497–505
3. Cai J, Shen X, Fan C, Zhou B (2019) Security-enhanced optical encryption based on JTC architecture with confused ciphertext. *Optik Commun*
4. Chen C, Sun K, He S (2020) An improved image encryption algorithm with finite computing precision. *Signal Proc* 168:107340

5. Huang ZJ, Cheng S, Gong LH, Zhou NR (2020) Nonlinear optical multi-image encryption scheme with two-dimensional linear canonical transform. *Opt Lasers Eng* 124:105821
6. Ghadirli HM, Nodehi A, Enayatifar R (2019) An overview of encryption algorithms in color images. *Signal Proc* 164:163–185
7. Leihong Z, Zhisheng Z, Yi K, Hualong Y, Dawei Z (2020) Research on double-layers optical information encryption based on ghost imaging. *Opt Commun* 455:124585
8. Babaei A, Motameni H, Enayatifar R (2019) A new permutation-diffusion-based image encryption technique using cellular automata and DNA sequence. *Optik Commun* 203:164000
9. Bentoutou Y, Bensikaddour EH, Taleb N, Bounoua N (2019) An improved image encryption algorithm for satellite applications. *Adv Space Res Commun*
10. Yang F, Mou J, Liu J, Ma C, Yan H (2020) Characteristic analysis of the fractional-order hyperchaotic complex system and its image encryption application. *Signal Proc* 169:107373
11. Ben Farah MA, Guesmi R, Kachouri A, Samet M (2020) A novel chaos based optical image encryption using fractional Fourier transform and DNA sequence operation. *Opt Laser Technol* 121:105777
12. Mahmud M, Rahman A, Lee M, Choi JY (2020) Evolutionary-based image encryption using RNA codons truth table. *Opt Laser Technol* 121:105818
13. Zhang D, Zhang F (2014) Chaotic encryption and decryption of JPEG image. *Optik—Int J Light Electron Opt* 125(2):717–720
14. Banik A, Shamsi Z, Laiphrakpam DS (2019) An encryption scheme for securing multiple medical images. *J Inf Secur Appl* 49:102398
15. Li CL, Li ZY, Feng W, Tong YN, Wei DQ (2019) Dynamical behavior and image encryption application of a memristor-based circuit system. *AEU—Int J Electron Commun* 110:152861
16. Liansheng S, Cong D, Xiao Z, Ailing T, Anand A (2019) Double-image encryption based on interference and logistic map under the framework of double random phase encoding. *Opt Lasers Eng* 122:113–122
17. Zhang X, Li W, Hu H, Dutta NK (2015) High-speed all-optical encryption and decryption based on two-photon absorption in semiconductor optical amplifiers. *IEEE/OSA J Opt Commun Netw* 7(4):276–285
18. Yu CC, Zhou NR, Gong LH, Nie Z (2020) Optical image encryption algorithm based on phase-truncated short-time fractional Fourier transform and hyper-chaotic system. *Opt Lasers Eng* 124:105816
19. Wang X, Zhao H, Feng L, Ye X, Zhang H (2019) High-sensitivity image encryption algorithm with random diffusion based on dynamic-coupled map lattices. *Opt Lasers Eng* 122:225–238
20. Xingyuan W, Junjian Z, Guanghui C (2019) An image encryption algorithm based on ZigZag transform and LL compound chaotic system. *Opt Laser Technol* 119
21. Zhu C, Sun K (2018) Cryptanalyzing and improving a novel color image encryption algorithm using RT-enhanced chaotic tent maps. *IEEE Access* 6:18759–18770
22. Zhang X, Wang X (2018) Digital image encryption algorithm based on elliptic curve public cryptosystem. *IEEE Access* 6:70025–70034
23. Guan M, Yang X, Hu W (2019) Chaotic image encryption algorithm using frequency-domain DNA encoding. *IET Image Proc* 13(9):1535–1539
24. Thoms GRW, Muresan R, Al-Dweik A (2019) Chaotic encryption algorithm with key controlled neural networks for intelligent transportation systems. *IEEE Access* 7:158697–158709
25. Nematzadeh H, Enayatifar R, Motameni H, 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
26. Rajakumar BR (2013) Impact of static and adaptive mutation techniques on genetic algorithm. *Int J Hybrid Intell Syst* 10(1):11–22. <https://doi.org/10.3233/HIS-120161>
27. Rajakumar BR (2013) Static and adaptive mutation techniques for genetic algorithm: a systematic comparative analysis. *Int J Comput Sci Eng* 8(2):180–193. <https://doi.org/10.1504/IJCSE.2013.053087>
28. Swamy SM, Rajakumar BR, Valarmathi IR (2013) Design of hybrid wind and photovoltaic power system using opposition-based genetic algorithm with cauchy mutation. In: *IET Chennai*

- fourth international conference on sustainable energy and intelligent systems (SEISCON 2013). Chennai, India. <https://doi.org/10.1049/ic.2013.0361>
29. George A, Rajakumar BR (2013) APOGA: an adaptive population pool size based genetic algorithm. AASRI Procedia—2013 AASRI conference on intelligent systems and control (ISC 2013), vol 4, pp 288–296. <https://doi.org/10.1016/j.aasri.2013.10.043>
 30. Rajakumar BR, George A (2012) A new adaptive mutation technique for genetic algorithm. In: proceedings of IEEE international conference on computational intelligence and computing research (ICCIC). Coimbatore, India, pp 1–7. <https://doi.org/10.1109/ICCIC.2012.6510293>
 31. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
 32. Masadeh R, Mahafzah BA, Sharieh A (2019) Sea lion optimization algorithm. *Int J Adv Comput Sci Appl (IJACSA)*, 10(5)
 33. Koppu S, Viswanatham VM (2018) Medical image security enhancement using two dimensional chaotic mapping optimized by self-adaptive grey wolf algorithm. *Evol Intell* 11:53–71

RDRLJ: Integrating Deep Learning Approach with Latent Semantic Analysis for Document Retrieval



Saicharan Gadamshetti, Gerard Deepak, A. Santhanavijayan,
and K. R. Venugopal

1 Introduction

Owing to the magnanimous size of the World Wide Web and its content that have exponentially increased in the recent past, there is a necessity for better and reliable techniques for discovering information from the web. There are some standard web content retrieval techniques such as the TF-IDF which often fails in terms of accuracy and scalability on large enough datasets. To accomplish this, there is a need for extraction of frequent interesting hidden patterns, compare, and rerank documents for which efficient tools and frameworks are required. Recent times suggest text clustering and analysis of relevant text are performance outcoming approaches in text analysis. Document clustering is used to accumulate same context documents into one cluster thereby grouping documents into many meaningful clusters. This makes user navigate online documents data productively. Similarly, if the data is labeled, then a similar effect of that of clustering can be achieved by applying classification schemes.

Use case includes searching for research papers in large databases like Scopus index.

Motivation: Traditional search engines compare similarity of user query with each document which is of high computational cost. Using only semantic analysis requires high vector spaces, so feature extraction using LSA removes irrelevant features within the initial stages. And, traditional retrieval algorithms have large semantic gap between user intention and retrieved documents.

S. Gadamshetti (✉) · G. Deepak · A. Santhanavijayan
Department of Computer Science and Engineering, National Institute of Technology
Tiruchirappalli, Tiruchirappalli, India

K. R. Venugopal
Bangalore University, Bengaluru, India

Contribution: An RDRLJ framework where each document is represented by set of topics after clustering. Semantic similarity check is performed using cosine similarity, and documents are reranked using Jaccard similarity, and obtained results suggest that relevant documents with reduced semantic gap are discovered.

Organization: Our paper is organized into following sections. Section 2 shows summary of the related research work. Section 3 deals with the proposed RDRLJ framework. Section 4 discusses implementation and evaluation metrics. Conclusion of the paper is provided in Sect. 5.

2 Related Work

In [1], Tang et al. proposed PTM framework where most occurring patterns are discovered removing meaningless information from documents. Because of noises in retrieved patterns and patterns are not detected from negative set text, results are very less. Pattern mining is mostly used approach in data mining with a spread of various algorithms, namely spade, FP tree, and prefix span. Post-processing of these extracted patterns includes compressing and combining patterns into various clusters. Primary open problem is detecting useful important patterns for data mining. In [2], Zaki et al. proposed replacing spade algorithm that locates patterns sequentially with high speeds. It splits parent subsystem into sub-classes and computes DFS and breadth first search (BFS) also. The approach yields redundant and unrelated patterns for which post-processing is required. In [3], Zhong et al. proposed novel method which organizes most occurring patterns into small representations to suggest only top k-patterns. In [4], Ramya et al. have proposed the DRDLC framework where documents are discovered using probabilistic model latent dirichlet allocations (LDA) and cosine similarity, but this module is restricted to fixed K (the number of topics is fixed) and not correlated topics (LDA distributions do not capture correlations). In [5], Yuefeng Li et al. have proposed relevant feature discovery framework using weighing features approach, but selecting relevant features may be a challenging task, and results are low. N-gram is informative and focuses on semantic/context than words which can be useful to get good ranking documents for users. Online data management is developed which uses NLP and context-free grammar approaches. In [6–18], several semantic applications and models have been discussed which are quite suitable to the proposed framework.

3 Proposed System Architecture

3.1 RDRLJ Framework

The proposed framework is divided into four distinct phases, namely the first phase involves pre-processing rectus corpus dataset and user query expansion using wordnet synonyms generation. In the second phase, document classification and clustering using LSTM and LSA are performed. Third phase deals with similarity check between user query and clustered documents by cosine similarity. In the fourth phase, reranking documents is performed using Jaccard similarity. One of the rarest novelties in the proposed scheme is the incorporation of a pair of similarity measures for document clustering at first, and thereafter incorporation of Jaccard similarity for reranking of documents (Fig. 1).

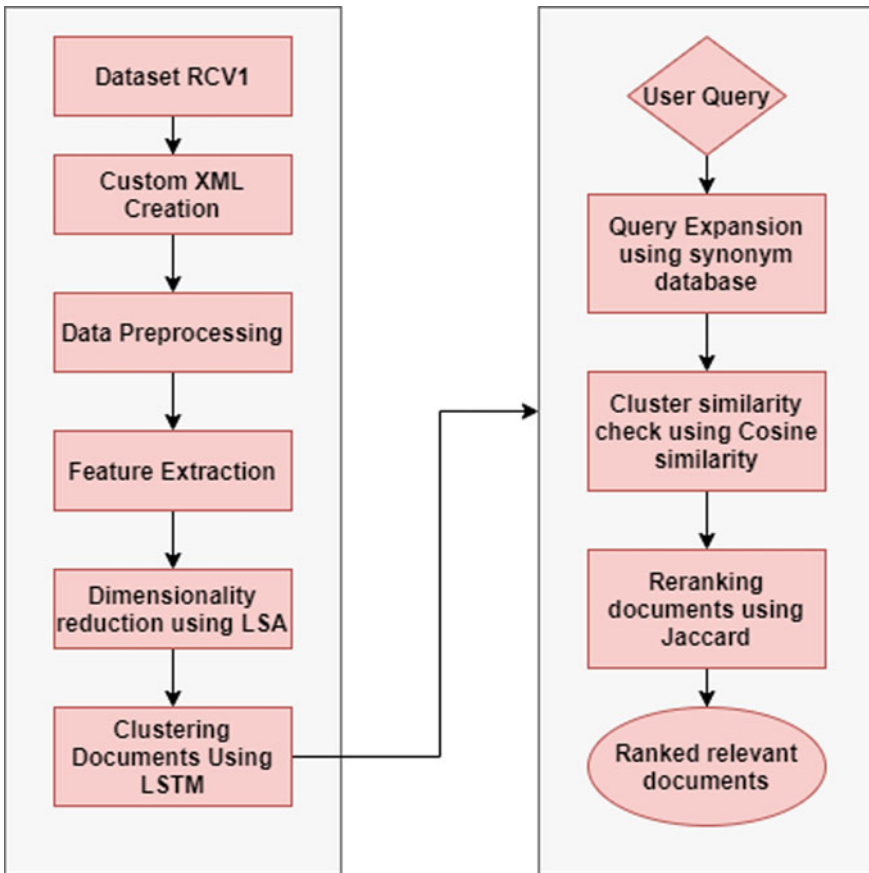


Fig. 1 General overview of the system

3.2 Data Processing Phase

Text pre-processing deals with data cleaning and converting it to set of tokens, and in later phases, those tokens are converted into numbers by vector space approaches. RCV1 dataset comprises of content, metadata, author info, headlines, and title in XML format. There is a need for the title and content of documents which are extracted and made as custom XML pages. Furthermore, the pre-processed dataset is passed through following modules, namely refining, tokenization, removal of stop words, stemming, and pruning. Refinement modules deal with removal of punctuation marks, URLs, special characters, translation to English language.

3.3 Vector Space Generation Phase

Vector space model [19] is used to represent documents. Each document is represented by word and document matrix where each row represents unique words and column represents each document. A_{ij} represents frequency of word i in document j . So, size of matrix is $m*n$ where m is number of unique words and n represents number of documents. The latent semantic analysis is for feature extraction and dimensionality reduction because web documents contain large dynamically growing data, representing each document in vector space requires many dimensions. Latent semantic analysis is known for representing document with significant relevant topics thereby decreasing dimensions. LSA comes under unsupervised statistical technique which maps semantic patterns among unique words. Noise and redundant data are compressed using singular value decomposition matrix operation. LSA assigns multi-labeled topics to each document. From machine learning to text mining and analysis, topic modeling is well-known approach to be used. It deals with discovering topics from corpus text and determines topics associated with every document and documents associated with every topic.

3.4 Classification of Documents Using LSTM

After performing LSA topic modeling, the feature vector is yielded which represents text from each document. There is a mandate need for clustering similar documents together which have the same contextual meaning. The approach uses long short-term memory (LSTM) [20] neural network, for classification of the documents.

3.5 Cosine Similarity to Retrieve Relevant Clusters

Similarity between two documents can be calculated using cosine similarity [21]. It uses dot product and magnitude of two vectors. Since cosine similarity deals with angle between the documents, when large documents are compared using Euclidian distance, we get high value for matching common article words, but they might not be similar concept-wise, so we used cosine similarity. Equation (1) is used to compute the cosine similarity in high dimensional space.

$$\text{Similarity}(X, Y) = \frac{X \cdot Y}{\|X\| \times \|Y\|} = \frac{\sum_{i=1}^n X_i \times Y_i}{\sqrt{\sum_{i=1}^n X_i^2 \times \sum_{i=1}^n Y_i^2}} \quad (1)$$

Output range will be from 0 to 90 degrees since term frequencies will not be negative. If angle between two documents is 0° , it means two vectors are almost similar. Semantic similarity between a cluster and user query is calculated using cosine similarity. This module is useful to retrieve documents from top relevant clusters. It compares two documents by measuring cosine of angle between them.

3.6 Reranking Using Jaccard Similarity

It is used to check similarity between user query and documents retrieved from relevant clusters. Reranking documents is important which can be preserved using Jaccard [22] similarity as it takes computationally less time. Equation (2) can be employed to compute the Jaccard similarity.

$$j_\mu(X, Y) = \frac{\mu(X \cap Y)}{\mu(X \cup Y)} \quad (2)$$

To suggest a greater number of recommendations, the user query must be expanded. This can be achieved by adding synonyms and related wordings/topics to the user query. For each user query, multiple queries are generated and larger set of related documents are retrieved. In this framework, we used lightweight wordnet ontology.

4 Implementation and Performance Evaluation

The proposed approach is implemented using Python 3, and Google Colab platform has been used with 12 Gb Nvidia Tesla k80 GPU. Neural networks are loaded from keras applications library. Performance of RDRLLJ framework is measured by precision, recall, f1-score, accuracy, and false discovery rate and compared with

Table 1 Proposed algorithm for RDRLJ framework

Input: User query
Begin
Offline phase
Step 1: Load rcv1 dataset and extract custom XML containing title and content of document
Step 2: Preprocess the data info using filtering, tokenization, stemming, and pruning
Step 3: Feature extraction using latent semantic analysis
Step 4: Cluster documents using LSTM
Online Phase
Step 5: Query expansion using synonym repository
Step 6: Similarity checking between user query and clusters to fetch top relevant clusters using cosine similarity
Step 6: Reranking retrieved documents using Jaccard similarity
End

Output: Semantically relevant documents discovery

standard other models like latent dirichlet allocation (LDA) [4] and relevance feature discovery (RFD) [5].

The proposed model using widely known rectus corpus volume-1 (RCV1) dataset which consists of 804,414 newswire documents classified into 103 categories. It represents data in XML format of which custom XMLs' are created carrying only title of document and content removing all metadata, XML tags, URL's. About 15,000 documents spanning all categories and applied text processing techniques and new dataset are created, saved offline for our experiment. Extracted unique terms from the dataset is stored in thesaurus which contains 66,314 terms by custom text parser. Table 1 depicts the algorithm for the RDRLJ framework.

Table 2 represents document retrieval results for each user query. Five queries of different contexts have taken and discovered corresponding documents. Average of all the queries is considered. Figure 2 shows the comparison results with existing models in information retrieval domain, namely LDA and RFD frameworks. Our model inculcated one more input, namely no. of recommendations, so as per the user preference top semantically related documents will be retrieved, and it represents average accuracy of all users queries. Figure 3 depicts the FDR comparison of the RDRLJ framework with other baseline approaches. Comparatively, our model

Table 2 Documents suggestion results for each user query

Query	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	FDR = 1-PPV
Q1	78.15	81.78	79.92	79.96	0.22
Q2	83.89	85.62	84.74	84.76	0.16
Q3	77.15	80.14	78.62	78.65	0.23
Q4	85.18	88.78	86.95	86.98	0.15
Q5	80.23	82.87	81.52	81.55	0.20
Average	80.92	83.84	82.35	82.38	0.19

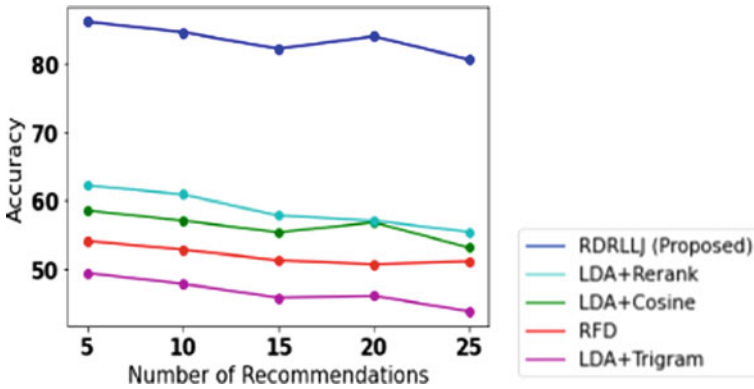


Fig. 2 Accuracy comparison for the proposed RDRLJ with other baseline strategies

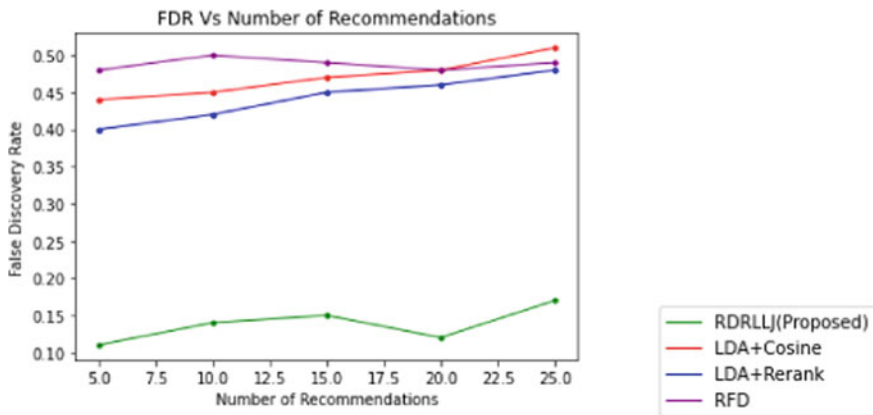


Fig. 3 FDR comparison of RDRLJ with other baseline strategies

outperforms LDA and RFD in terms of accuracy and response time. Owing to the incorporation of LSA, the response time is less due to dimensionality reduction in the initial phase by LSA. LSTM helps in keeping contextual information by iterating loops to make information flow from one level to the other. These loops help RNN perform better in clustering large volumes of data and improving accuracy thereby resulting in reducing semantic gap between user intention and documents suggestions. False discovery rate represents number of documents not related to user query, and this metric shows RFD has high value and RDRLJ has got the least valued function. Lesser FDR represents our framework discovered more prompt documents.

5 Conclusions

A RDRLJ framework that discovers documents of relevance by amalgamation of LSTM, LSA, and Jaccard similarity has been proposed. The approach amalgamates approaches like cosine similarity, LSTM, LSA, and Jaccard similarity. LSA helps in removing noise which is unimportant usage of some terms. LSA makes less computation as vectors dimensions is reduced in initial stages. LSTM preserves contextual information and clusters accordingly. Reranking plays a vital role since user tends to use top suggested documents more often, Jaccard serves the best for reranking, and it requires low computational power too. Periodical scan of Web sites will let updated site to inject into corresponding cluster dynamically. The RDRLJ framework yields an average accuracy of 82.38% with a very low FDR of 0.19, which makes it a best in-class methodology.

References

1. Wu ST, Li Y, Xu Y, Pham B, Chen P (2004) Automatic pattern taxonomy extraction for web mining. In: Proceedings of IEEE/WIC/ACM international conference on web intelligence, pp 242–248
2. Zaki MJ (2001) Spade: an efficient algorithm for mining frequent sequences. *Mach Learn* 42(12):31–60
3. Zhong N, Li Y, Wu S-T (2012) Effective pattern discovery for text mining. *IEEE Trans Knowl Data Eng* 24(1):30–44
4. Li Y, Algarni A, Albathan M, Shen Y, Bijaksana MA (2015) Relevance feature discovery for text mining. *IEEE Trans Knowl Data Eng* 15(7):1656–1669
5. Ramya RS, Ganesh Singh T, Sejal D, Venugopal KR, Iyengar SS, Patnaik LM (2018) DRDLC: discovering relevant documents using latent dirichlet allocation and cosine similarity. In: Proceedings of the 2018 v11 international conference on network, communication and computing. pp 87–91
6. Gerard D et al (2019) OntoQuest: an ontological strategy for automatic question generation for e-assessment using static and dynamic knowledge. In: 2019 Fifteenth international conference on information processing (ICINPRO). IEEE
7. Deepak G, Santhanavijayan A (2020) Onto best fit: a best-fit occurrence estimation strategy for RDF driven faceted semantic search. *Comput. Commun.*
8. Deepak G, Shwetha BN, Pushpa CN, Thriveni J, Venugopal KR (2018) A hybridised semantic trust-based framework for personalised web page recommendation. *Int J Comput Appl* 1–11
9. Deepak G, Kumar AA, Santhanavijayan A, Prakash N (2019) Design and evaluation of conceptual ontologies for electrochemistry as a domain. In: 2019 IEEE international WIE conference on electrical and computer engineering. pp 1–4
10. Santhanavijayan A, Naresh Kumar D, Deepak G, A novel hybridized strategy for machine translation of Indian languages. *Soft Comput Signal Proc* 363
11. Deepak G, Priyadarshini JS (2018) A hybrid semantic algorithm for web image retrieval incorporating ontology classification and user-driven query expansion. *Adv Big Data Cloud Comput* 41
12. Sairam H, et al. (2019) A novel approach for ontology focused inter-domain personalized search based on semantic set expansion. In: 2019 fifteenth international conference on information processing (ICINPRO). IEEE

13. Deepak G, Gulzar Z (2017) Ontoepds: enhanced and personalised differential semantic algorithm incorporating ontology driven query enrichment. *J Adv Res Dyn Control Syst* 9(Specia):567–582
14. Deepak G, Teja V, Santhanavijayan A (2020) A novel firefly driven scheme for resume parsing and matching based on entity linking paradigm. *J Discrete Math Sci Crypt* 23(1):157–165
15. Kumar N, Deepak G, Santhanavijayan A (2020) A novel semantic approach for intelligent response generation using emotion detection incorporating NPMI measure. *Procedia Comput Sci* 167:571–579
16. Deepak G, Priyadarshini S (2016) A hybrid framework for social tag recommendation using context driven social information. *Int J Soc Comput Cyber-Phys Syst* 1(4):312–325
17. Deepak G, Kumar N, Santhanavijayan A (2020) A Semantic approach for entity linking by diverse knowledge integration incorporating role-based chunking. *Procedia Comput Sci* 167:737–746
18. Kaushik IS, Deepak G, Santhanavijayan A (2020) QuantQueryEXP: A novel strategic approach for query expansion based on quantum computing principles. *J Discrete Math Sci Crypt* 23(2):573–584
19. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. *Commun ACM* 18(11):613–620
20. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
21. Sidorov G, Gelbukh A, Gómez-Adorno H, Pinto D (2014) Soft similarity and soft cosine measure: similarity of features in vector space model. *Computación y Sistemas* 18(3):491–504
22. Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société vaudoise des sciences naturelles* 37:547–579

Reverse Supply Chain Network for Plastic Waste Management



Rakshit Shetty, Neha Sharma, and Vishal A. Bhosale

1 Introduction

Earlier the manufacturing firms and the consumers were not answerable for the after-life of the utilized products. They were always ignorant when it came to the after process of those used products. The whereabouts of these used items were unknown, leading to pay little mind to the degree of their arrival, which leads to the process of burial and burning or other nature hazardous means in order to destroy these items [1]. In order to avoid that with plastics products we have used reverse supply phenomena in the existing forward supply chain for plastic waste. A closed-loop supply chain (CLSC) is formed when a normal forward supply chain is connected with a reverse supply concept and a new sustainable network is formed [2]. American Reverse Logistics Executive Council defined the term reverse logistics as “The process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in-process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal” [3]. In reverse supply the used product which is of no further use is collected by the recycling and other recovery centres for its further applications. A sum of these collected products are distributed once again post major significant fixes in dismantling outlets. Some reclaimed items are perceived to be conveyed to the producers to reproduce and sell to so-called the second clients as in [4]. Because of the expanding significance in recent years, the selection of reverse logistics has obtained

R. Shetty (✉) · N. Sharma · V. A. Bhosale

School of Mechanical and Civil Engineering, MIT-Academy of Engineering, Alandi, Pune, India

e-mail: rshetty@mitaoe.ac.in

N. Sharma

e-mail: ncsharma@mitaoe.ac.in

V. A. Bhosale

e-mail: vabhosale@mitaoe.ac.in

a lot of recognition from business and academia [5]. The effective management of reverse logistics operations increases profitability. The reusing stations will build their piece of the pie and become incorporated to entire remanufacturing industry by working together with remanufacturing business [6]. Many firms are nowadays working in order eliminate more of the waste by deriving a productive and effective supply chain network. It includes the factor of figuring out the usage and optimizing the utilization of available resources in the supply chain network as well as the to-and-fro between these dealers to optimize the physical flows which are connected to post process of reclaimed items [7]. Plastic recycling is not just a legal requirement but can also lead to increase in financial income for firms which invest in it. Generation of huge solid waste has been a top issue for many nations. It's about 130 billion kg of solid waste per year collected worldwide, and it's projected to increase in 2025 by 220 billion kg [8]. Government has rolled up their sleeves to overcome this situation and for the same have proposed various new rules and implications. Amongst which one is extended producers responsibility rule of plastic waste management act is a leading demand in change. This rule has brought not only responsibility for plastic producers and manufacturers but also has given a great platform for plastic waste collector, this resulted growth in employment too. Not only on environment prospective but also for this waste collection has brought up the change in idea of business and have generated greater opportunities to explore, these have strengthened circular economy and have had an impact on environmental issues. Transportation and cost are also means hurdle to deal with for designing a network model to come up with such situations. An extended supply chain network (SCN) is very much required in today's world in order for business firms to excel, which is conceivable only if the associations has a very much organized adaptable supply chain. Use of virgin plastics has been reducing and side by side alternate methods have been generated to find and replace possible packaging. The principle of 3R's came in to frame, i.e. reduce, recycle and remanufacture, which promotes reduction of plastic use and wastage of plastic too [9].

Thermoplastic polymers are to a great extent used for the creation of expendable plastic bundling, majority of which is not gathered for recycling. A noticeable special case is the single-use refreshment bottle made of polyethylene terephthalate (PET). PET containers are the most plentifully reused plastic product. R-PET is utilized generally in materials and in inflexible bundling applications and also mainly applicable in the assembling of comparative item frameworks. In today's world, plastic waste has sky rocketed. In order to take care of such plastics waste, a reverse supply chain model is merged in order to guide that plastic waste and increase its profitability by reuse recycle and remanufacture the waste. The model proposed in this paper shows a systematic flow of waste plastics from consumer back to the manufacturer as raw materials and for various other applications, hence by increasing the profitability of the company. This model is made keeping in mind most of the legislation in order to maintain a smooth and uniform flow. This helps in the environmental issues faced by today's generation and also saves raw material and decreases pollution by a very huge amount. The main aim of this model is to reduce plastic waste and use it in a convenient and environment friendly manner.

With the rise in production and the subsequent consumption also leads to a huge pileup of waste plastic, most often these are part of the product packaging or the single-use plastics, these are product of the supply chain activities. Recycling, of such plastics (one-time use virgin plastics), provides a very genuine and a promising solution to the growing concerns and market competition with the use of plastics and the effects of such wastes generated. The recycled plastics most often fetch a lower value than virgin ones, also due to contamination due to varied types of plastic collected, sorted from the other wastes and upon sorting being fed into the system. This research applied the reverse logistics along with the environmental engineering to chalk out plans for a better environmental sustainability as well as a sustainable model in itself for the business to create a better image as well as in a profit making way. This fundamental idea behind reverse supply chain can be utilized to reap the economic benefit from one of the legs of the 3R's of the plastic waste for an environmental perspective, whilst the economic benefits can be used to an overall business benefit and can provide a sustainable business model. The overall environmental and social impact would be worked as a most important motivation for implementing the reverse logistics strategy for sustainable development. In many countries, there are economic and legislative measures taken for pushing the reverse logistics cycle in order to preserve their ecological balance.

The variety in the plastic waste generated and the polymers included causes for the wastes to be degraded at various rates, and most often the contamination of different plastics, this causes for the plastic being recycled to have a varying performance, and thus, the end product is usually unsought for in the market. This causes for a concern because in an highly competitive market, development of product is a very huge and a necessary task in order for the survival, and due to heavy restrictions on the plastics manufacturing and the disposal of such solid waste around the world the recycling needs to be the key and also match with reference to the desired performance that should be parallel to the virgin plastic. Also, the process by which the plastic is being recycled affects the demand in the market with respect to the final specifications as in [10]. In recent developments across the industry (of recycling wastes), an alternative has been used to blend the waste plastic with virgin plastic, which acts as reinforcing to increase the specified mechanical properties of many alternates being recycled. Reduction in the waste at the primary source is always the main aim, this includes either alternate ways instead of traditional use of plastic products, and creating awareness in the society and providing products as alternative to traditional plastics is one of the main aims. Segregation of these waste at the source allows for less contamination of the recycled products and this is done through capacity building, these can also be part of the corporate and government social responsibilities by making technical and financial strategies, this also includes to interlink the stakeholders (the producer and the customer, and in most parts the government) [11]. The main two factors for achieving an efficient distribution networks are needs of the client and cost for those respective needs [12].

The purpose of this research is to investigate the research questions such as, how to collect waste plastic? And what will be the mathematical model?

The main objectives of this paper are;

- I. Propose a network model for collection of all plastic waste.
- II. Construct mathematical modelling of current problem.

After manufacturing, plastic packaging products are supplied to DC'S and then to retailers, and further it is provided to consumer. This is commonly known as straight supply, but when it comes to reverse supply of the same chain consumer plays the role of waste supplier which comes from various sources such as household, society, offices, school, collages, market, hospitals, cinema halls, etc. Which are further collected by municipalities' in which the current research uses two types of process which called as source and post separation. In source separation, useful plastic waste is collected by giving buy back offers to customers so as to collect good quality plastic in good condition, whereas remaining waste is collected by post separation from municipalities which includes all type of plastic.

Further according to condition and types and capacity to use of particular plastic they are separated at segregation centre and then good quality plastic is given to manufacturer for reuse, and for recycling. Non-recyclable and non-disposable plastic are given to construction companies, and rest which can be used for energy generation are supplied to incineration plants.

The rest of this paper is organized as follows. Section 2 reviews the literature on circular SC. Legislations are explained in Sect. 3. Methodology is presented in Sect. 4. Section 5 explains the problem formulation. Discussion is briefed in Sects. 6 and 7 concludes the paper.

2 Literature Review

Nowadays, green chain is a focusing platform for organizations, and closed-loop supply chain is applied almost in all the fields due to its effectiveness. Closed-loop supply chain has gained a lot of importance in sustainability matters. Hence, a systematic literature review, i.e. material collection, a descriptive analysis, category selection, and evaluation stage has been implemented from which the investigated models were based on game theory in [13]. The CLSC is not only used in product oriented business but also eco-friendly customer oriented or service oriented organization for all small and big companies in information and communication technologies (ICT) [14]. Alcayaga et al. [15] focused that CLSC is very dynamically suitable option for circular economy, gaining profits from products which are smart and having connectivity through internet. Das et al. [9] have established 3R principle which promotes reuse, reduce and recycle the waste produced and also discussed geographical and nation's economic status. The take away food packaging mostly made of materials like polypropylene; aluminium and extruded polyether are very bad. Plastic and its adverse effect are discussed which includes problems like global warming [16]. Iacovidou et al. [17] have specified the methods which focus on recovery of resources from waste within circular economy and also has maintained effective and transparent analysis. Geueke et al. [18] have examined and calculated about the effect

of recycle and reuse of plastic waste and discussed whether it really help to clean landfill and other environment problems of greenhouse emission, global economy, fossil fuel consumption, etc. Recycle and reuse are now a days is considered as important measure to manage packaging waste [19].

Plastic recycling is a legal necessity as well as it leads to environmental advantages. Rather than using crude oil for virgin plastics it would be rather beneficial to recycle the used plastics from the consumer and results in lower environmental impact comparatively [20]. However, plastic collection laws and protocols are different for many nations which could lead to changes in distribution networks and models. Although due to several positive points related in applying reverse chain supply sometimes firms are reluctant due to its obstacles related to it. Hence, a model surrounding the obstacles is made available in which all these obstacles were studied, and key points were made in order to lower the risks affiliated and critical decisions [21]. In order for a systematic flow of the reclaimed products to its end application and execute an optimal reverse chain supply, an integrated model is also important in order to evaluate the selection criteria of the partners such as the FDP (forward distribution partner) and other 3rd party reverse logistics providers [22]. Also to have proper co-ordination between the executing members of reverse logistics, two way tariff contracts are necessary for optimized operations between the parties that are the remanufactures and the collector [23]. Reverse logistics (RL) have become an important factor for firms to gain commercial profitability and a boon for environment friendly factors.

Although when the markets down or at uncertain condition it is always a plus point to have the fixed cost and transportation cost to be minimum, hence the model should take into consideration, the location of remanufacturing units and distribution centre should have the optimal minimum distance between the centres in order for a sustainable flow of products which could prove to be very beneficial to the organization in [23]. It is not easy for firms to execute an optimal sustainable reverse chain supply since many risks are involved with it. Thus, for the organizations to increase the profits and be competitive, it is necessary to analyse the risk, whilst operating it. Risk such as environmental, time and inventory management quantity and various other factors could have risks, whilst operating reverse logistics [24]. Therefore, it is very crucial to have a strategic network design in which it takes in both the sectors, the waste management and reverse supply chain and provide an integrated result where the waste management can benefit from the models designed in the area of reverse supply chain [25]. Green supply chain management is a nature-friendly supply chain which provides the goods to the customer by reducing the waste and because of rising environment issues and rising EPR, green supply chain has become a centre of attractions. Normally, transportation fleets and queuing system have not been considered much which is necessary in order to tackle the transportation problem for sustainable closed-loop supply chain [26]. The problems in which how the local government conducts solid waste management in the middle-east and north African areas [27]. These areas are still using the old waste disposal methods leaving large amount of carbon footprint since recycling activities and other services such as collecting services are insufficient. Awareness and benefits of green sustainable CLSC network

are must in order to decrease these waste generation and carbon foot print. Costs such as facility operational cost transportation cost and logistics maintenance inventory cost have to be recovered which can be done only by maintaining an optimal flow in order for firms to execute a fully functioning reverse chain supply [28].

Zhou and Wang [29] designed a reverse supply chain network model in which the reclaimed products can be further operated and brought back into the market through various means once it gets repaired or remanufactured in plants. Similarly, a closed-loop supply chain in order to recover the used products and the per cent gain from it as compared to recycling and other operations suggested in [30]. Whilst designing a reverse supply chain network once should consider the CO₂ emission, whilst transportation in order to gain efficiency and in economical and environment friendly way [20]. As these steps will need expenditure in order to execute the extended product responsibility there should be an optimized model in order to achieve cost minimization [31]. Not only limited to those aspects but also the customers should be given some buy back offers in order to make the customers obey those protocols in a convenient manner in order to keep the flow going, further it gave five tasks which improves the transportation services and gain efficient work flow and stability, i.e. transportation reduction intensity and emissions, lower area usage, choice of carrier, better working condition for employers [32]. A basic study and comparison of research papers in the field of reverse supply chain and closed-loop supply chain and various gaps are identified and analysing the multi objective problems, utilizing new approaches, and applying more green and sustainable for future upcoming approaches is mentioned in [33]. Generally in developing countries, about 30–70 per cent of the waste is discarded without any further inspection or waste management [34]. But if we see over the past decades due to environmental problems and for eliminating these problems, the study of reverse logistics have obtained a lot of attention due to its economic benefits and the rising environmental problems which has lead the government to take out strict policies and regulations and organizations have started to recognize that the public is moving towards the green revolution, or green, image [35]. In the perspective of Razmi et al. [36], redesigning a reverse supply chain network or model is very complex since it requires to carry out the changes in a planned manner by maintaining the operations. Guide et al. [37] optimized the CLSC network into phases which includes the first phase clarifies the remanufacturing techniques problems related to it coming to the second and third phase throws light on the manufacturing the reclaimed product for a sustainable and proper flow of the products. The fourth phase concentrates on completing the loop from start to end for the end to life of the product. And lastly where the prices and cost for reclaimed product will be decided in order for firms to collect the products for further application and have a sustainable business.

3 Legislation

This paper focuses on the fact that problems faced due to the lack of work on end use of various types of plastic stuck in environment and also to reduce its effect. The paper deals with the major problems caused by improper planning and poor law adoption. We designed a reverse supply chain network in back support to various legislations taken up in concern to plastic, its use and also to environment by referring the below legislations

- Natural green tribunal act 2010 [38]
- The hazardous waste management and handling rule 2016 [39]
- Plastic waste management rules 2016 [40]

In 2016, the plastic waste management rule was implemented so as to reduce and solve critical issues faced due to use of plastic. Drafting committee came up with new rules which include highly important factor—carry bags made up of virgin plastic or of recycled plastic shall not be less than 50 microns in thickness, recycling and reuse of plastic waste should be done according to Indian standards. The Indian road congress have promoted and given guidelines to use non-recyclable plastic waste in road construction. Use of plastic is not limited to urban area but also in rural areas. So the government has also added some rules and penalty for use, manufacture and sale of plastic for both areas [40]. Here, major highlight was extended producers responsibility rule. To bring responsibility in producers and generator of plastic and to reduce the use of virgin plastics and apply alternatives to them and also ask to promote recycling and reuse of plastic instead are discussed. They have also mention to design collect back systems and circular economy was another factor. In 2016, all the people waste generators, local body, gram panchayat, producers and importers are applicable. User fees are also added to use of plastic persecuted by ULB (Urban Local Body). The collect back system needs to be installed within 6 months in consultation with all local/ state urban development departments and must be implemented within two years thereafter. All the idea, plan and design of system and sheet must be submitted to SPCB/CPCB (State/Central Pollution Control Board), they also introduced plastic use fees 4000/month or 48,000/year. The different types of plastics are listed in Table 1.

4 Methodology

Important aspect for the environment is the treatment and usage of certain products like used and disposed plastics. The type of plastic to be identified and collected with respect to its category and its post treatment accordingly is very challenging. All of the developing nations need to take serious measures in order to overcome this task by conducting research iterations in the field of recyclability of these products. Trivedi et al. [42] drafted that in India almost 12 billion tonnes of plastic items are

Table 1 The seven types of plastic include [41]

Plastic type	Description
Polyethylene terephthalate (PETE or PET)	This plastic is used for making soda bottles, medicine jars, bean bags, rope tote bags, water bottles, carpet, milk jugs and motor oil container, etc. These plastic can be recycled only once
High-density polyethylene (HDPE)	It is used for shampoo and conditioner bottles, soap and detergent containers, plumbing pipes; bleach container and grocery bags etc. It can be recycled twice
Polyvinyl chloride (PVC)	The usage of these plastic is for making window frame, duct, gutter, shoe, sewage pipes and cling films, etc. PVC can be recycled three times
Low-density polyethylene (LDPE)	This is used to produce grocery bags, frozen food bags, flexible container lids, Tupperware, kitchen ware, etc. These plastic can be recycled and used for four times
Polypropylene (PP)	This type of plastic use to make products like stadium cups, bottle caps, takeout container, disposable cups and plates, plastic food boxes, etc. It can be recycled five times
Polystyrene or styrofoam (PS)	This includes plastic cutlery, packing foam, various compact discs made out of plastics, etc. These product can be recycled 6 times
Others	Rest of the other plastics (includes: acrylic, acrylonitrile butadiene, styrene, fibreglass and nylon, etc.) these other plastic includes, baby drinking container, big size water containers with multi gallon dimensions, exterior lighting fixture, goggles, etc. This type of plastic product has capacity to recycle up to 7 times

used by its people amongst which 70 per cent of its plastics are dumped or disposed of as waste without any further treatment leading to environmental problems. It's not just the government's task or responsibility but mind-set of the people that needs to be changed and should inculcate a certain set of responsibilities when it comes for waste generation [43]. As the amounts of manufacturing of such plastics are so high, an established network for its waste treatment, i.e. recycling and recovery of its plastic products should also be designed depending on its type. Hence, recycling of such plastic becomes a key point for environmental safety and economic benefits. Bing et al. [44] stated that the plastics collected can once again be reused in industrial operations with lesser environmental impacts by replacing the virgin plastics which used crude oil and has carbon footprint. Along with the help of our literature survey mentioned, we have derived a model in order to direct the used plastics back to the plastic supplier. In this model, we have connected a reverse supply chain network in an existing forward supply chain network in order to collect the plastics from

various departments using buyback offer in source separation block. Generally, the local municipal corporation is responsible for the solid waste management. When it comes to collecting the waste in India the local government has its own vehicles like garbage trucks and animal driven carts in rural areas and villages. Used plastic products are generally collected by the waste pickers or rag pickers. Such waste pickers also could be made to work for business models by the plastics collected by these sources and, itinerant waste buyers where plastics could be reclaimed depending on the type desired [43]. It is observed that lack of financial incentives for source separation results in poor results in this particular department. Hence, a buy back option has been made available in order to boost the source separation from houses and other departments. Post separation does not sort the reclaimed plastics into different types but differentiates those types of plastic products from the rest in order to be segregated into different categories in the later process. As it is very crucial to determine the waste plastic collected and its decision on its further application due to its type whether it could be high-quality hard plastics and films or low-contaminated plastics, these necessary evaluation and sampling should be done in these particular centres in order to carry out its further treatment [45]. These parts of sorting and segregation for the further applications of these plastic will play a key role in order to maintain the network and give a sustainable supply chain for the organizations to operate efficiently.

In order to segregate the plastics in their particular department, a segregation block has been introduced in the following study. These departments include basic 3R policy in which we have introduced the recovery system instead of reduce in order to increase the profitability from the wasted plastics. The 3Rs benefits the economy and environment by recycling the reclaimed plastic products will not only reduce the waste but also decrease the plastic material production and save the carbon footprint and the petrochemical feed stock related to it. The other type of reclaimed plastic waste which is contaminated can also be redirected and be applicable for other application such as incineration process for power development and road development which could lead to profitability [46]. The proposed theoretical model has been shown in Fig. 1.

5 Model Description

This paper attempts to design a reverse supply chain network for discarded plastic products which are once manufactured and are dumped and recyclable plastic products. The forward supply chain network is modified, and a reverse supply chain is added. In the forward supply chain network, there is an existing manufacturing plant from where products are transported to distribution centre. Now the product is transported to retailer established in the proximity of distribution centre. The customers buy products from retailers.

STEP 1: In our analysis, we consider a reverse supply chain where we set up a reclamation centre. Here, citizens living in the proximity of reclamation centre

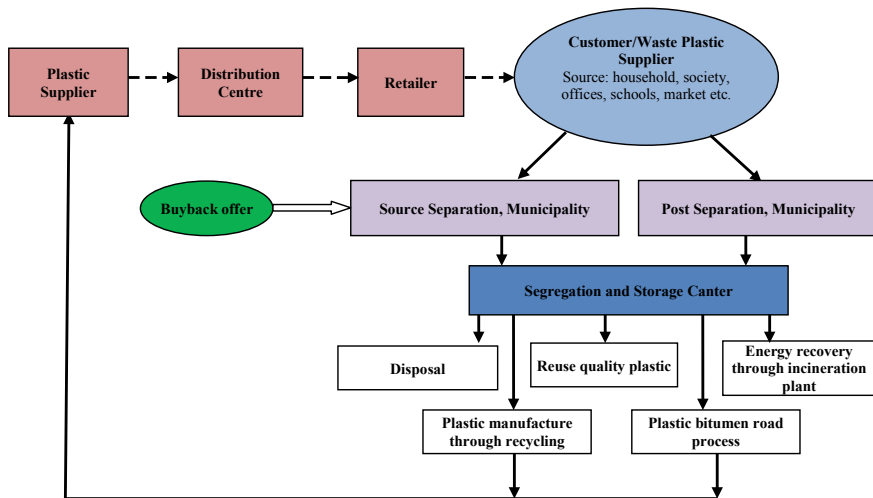


Fig. 1 Reverse supply chain network for plastic waste management

trade single-use plastic products and recyclable plastic products in exchange of a certain amount of money at the reclamation centre or their own concerns about the environment. The rate of single-use plastic and recycled plastic is fixed and determined by collector.

STEP 2: After the collection from reclamation centre, the plastic is transported to collection centre where it is cleaned and sorted into four different categories:

- I. PET (Polyethylene Terephthalate): Water bottles and soda bottles, salad trays, milk bottles, shampoo bottles.
- II. PP (Polypropylene): Margarine tubs, microwaveable meal trays, ice-cream trays, detergent bottles.
- III. Mix of hard plastic: PVC (polyvinyl chloride), PS (polystyrene), non-bottle PET and falsely sorted PE, PP and PET.
- IV. Film: Carrier bags, bin liners and packaging films.

STEP 3: After cleaning and sorting the plastic into different streams, it is transported to the manufacturers who wish to reuse the single-use plastics and recycle the PET plastics. The remaining lot is sent to governmental disposal grounds.

Following assumptions are made, whilst designing the model.

- I. The proposed model is developed for one collection centre and one reclamation centre.
- II. Collection effort cost information and information about used product supply market are well known to collector.
- III. Transportation intervals are constant between all the facilities.
- IV. The proposed model is developed for all kinds of plastic.
- V. Total capacity of collection centre is always greater than or equal to capacity at reclamation centre.

VI. Random parameter such as demand and return rates are neglected.

From the description it can be summarized that proposed model is developed for multi-product, single-stage distribution.

6 Problem Formulation

This section deals with formulations of reverse supply chain model integrated with a circular supply chain, for systematic and optimum flow of plastic waste. The proposed RSC model is shown in Fig. 1, which shows the flow of plastic back from the customer to the supplier by recycling, considering the fixed cost, shipping cost.

The notations and parameter used in designing the RSC model are listed as follows.

Notations

- P Set of plastic products indexed by p
- I Set of potential location of source separation municipality i .
- J Set of potential location of segregation centre indexed by j .
- K Set of potential location of post separation municipality indexed by k
- T Set of period indexed by t
- Y Set of age of product indexed by y
- Q Set of quality of returned products indexed by q

Parameters

- $f(i)$ Fixed cost of source separation in a potential location per unit capacity.
- $f(j)$ Fixed cost of segregation centre at a potential location per unit capacity.
- $f(k)$ Fixed cost of post separation at a potential location per unit capacity.
- $f(ijt)$ Unit shipping cost from location i to location j in a time period t .
- $f(kjt)$ Unit shipping cost from location k to location j in a time period t .
- $g(jst)$ Disposal rate of non-plastic waste at a landfill or a bio gas plant from location j to location s in a time period t .
- $f(jut)$ Unit shipping cost of a set of plastic from location j to location u in a time period t .
- $f(jvt)$ Unit shipping cost of a set of plastic from location j to location v in a time period t .
- $f(jwt)$ Unit shipping cost of a set of plastic from location j to location w in a time period t .

- $f(jqt)$ Unit shipping cost of a set of plastic from location j to location q in a time period t .
- α Fixed cost to maintain a segregation centre.
- β Fixed cost paid to maintain holding cost.
- $UP(qyt)$ Used plastic quantity q which are y period old at a time period t .
- $BB(qy)$ Buy back price of a used plastic of quantity q which is y period old.

Variables

- x_i, x_j, x_k Integer variable used for showing quantities of product p at a source separation municipality, segregation centre post separation municipality and
- x_s, x_u, x_v, x_w, x_z Integer variable used for showing quality of product p after segregation at various clients respectively.
- z_i, z_j, z_k Binary variables = 1 if the source separation centre, post separation centre and segregation are open, otherwise 0.

The model aims to minimize the transportation cost in order to gain optimum flow of the plastic waste back to the supplier. Hence the derived equation for the transportation is as follows

$$\begin{aligned}
 \text{MIN}(f_1) = & \sum_{i \in I} f_i \times z_i + \sum_{j \in J} f_j \times z_j \\
 & + \sum_{k \in K} f_k \times z_k + \sum_{p \in P} \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} f_{ijt} \times x_{pij} \\
 & + \sum_{p \in P} \sum_{t \in T} \sum_{k \in K} \sum_{j \in J} f_{kjt} \times x_{pkj} \\
 & + \sum_{s \in S} \sum_{p \in P} \sum_{t \in T} \sum_{j \in J} \sum_{d \in D} g_{jstd} \times x_{pjs} \\
 & + \sum_{u \in U} \sum_{p \in P} \sum_{j \in J} \sum_{t \in T} f_{jut} \times x_{pju} \\
 & + \sum_{v \in V} \sum_{p \in P} \sum_{t \in T} \sum_{j \in J} f_{jvt} \times x_{pju} \\
 & + \sum_{w \in W} \sum_{p \in P} \sum_{t \in T} \sum_{j \in J} f_{jw t} \times x_{pju} \\
 & + \sum_{b \in B} \sum_{p \in P} \sum_{t \in T} \sum_{j \in J} f_{jbt} \times x_{pjb}
 \end{aligned} \tag{1}$$

Fixed cost:

$$\text{Segregation and sorting cost is given by : } \alpha \times x_p \quad (2)$$

$$\text{Storage cost : } \beta \times x_p \quad (3)$$

Buyback price is given by:

$$\text{RETBB}_{qt} = \sum_{y \in Y} (\text{UP}_{qyt} \times \text{BB}_{qy}) \quad (4)$$

Total cost:

$$\sum_{t \in T} \cdot \sum_{q \in Q} \cdot \sum_{y \in Y} (\text{RETBB}_{qt} \times \text{BB}_{qy}) + \sum_{t \in T} \cdot \sum_{q \in Q} (\text{RETBB}_{qt} \times \text{ACB}) \quad (5)$$

Constraints:

The above equations are subjected to following constraints keeping in mind the legislations mentioned before.

$$(x_s, x_u, x_v, x_w, x_z) \geq 0 \quad (6)$$

$$x_i \geq 0 \quad (7)$$

$$x_k \geq 0 \quad (8)$$

$$\sum_{i \in I} x_i + \sum_{k \in K} x_k \leq \sum_{j \in J} x_j \quad (9)$$

$$\sum_{i \in I} x_i + \sum_{k \in K} x_k = \sum_{s \in S} x_s + \sum_{u \in U} x_u + \sum_{v \in V} x_v + \sum_{w \in W} x_w + \sum_{z \in Z} x_z \quad (10)$$

$$\sum_{j \in J} x_j - \sum_{d \in D} x_d = \sum_{s \in S} x_s + \sum_{u \in U} x_u + \sum_{v \in V} x_v + \sum_{w \in W} x_w + \sum_{z \in Z} x_z \quad (11)$$

After all the plastics are collected at the storage centre, Eq. 6 can be used to ensure the set of plastics which are collected after the sorting, into its various types are dispatched accordingly. Same with constraints of Eqs. 7 and 8 which ensures the set of plastics which can be collected at source separation and segregation centre. Equation 9 indicates that the sum amount of plastic which has been collected by the source separation centre and the type of plastics at post separation centre after sorting will always be less than equal to the amount of plastics finally stored at segregation centre. Equations 10 and 11 ensure that only the amount and type of plastics which will be collected at the post and source separation unit will be sorted and dispatched

accordingly to the distributors mentioned in the model network and the rest of the plastic waste which has no further operation or use to be disposed off at the disposal.

7 Conclusion

The main motto of this paper is to eliminate the plastic waste and improve circular economy in such a way that it satisfies the need of government, producers, customer and environmental conditions. The proposed network throws light on the systematic way of reusing and recycling plastic products. The plastic generated can be reclaimed again by suppliers in order to increase the profitability and reduce plastic reprocessing emissions, raw plastic usage and carbon emissions. These need an integrated design model to come up with all the flaws. People and the manufacturer need to change their strategies to reduce, recycle and reuse so as to reduce production of virgin plastic. These resulted in great activeness in all the plastic producers to come up front and start for recollection of their plastic products. This has led to employment and profits for waste collector. This reverse supply chain network is capable to reclaim, segregate and provide plastic to their respective producers. In account, it helps to increase the profit margin. In upcoming years may be it will grasp more value in field of business and self-employment. To reduce the worse effect of plastic waste and to satisfy the rule under plastic waste management act, we have designed an integrated network model which comprises of collection centre and are directed towards suitable end use of different types of plastic. The installed two separation centres help to differentiate the plastic easily without compromising its quality. By using this model, the major account of plastic which is non decomposable is used for road construction by the Indian national congress. Recyclable and reuse able plastic are returned to the plastic suppliers for further productions. Plastic which can be used for energy generation are supplied to plants and remaining plastics which cannot be used anywhere are disposed under the law of government. By this, we can reduce the use of virgin plastic and improve environmental health and increase the profit of this circular economy.

References

1. Alizadeh M, Makui A, Paydar MM (2020) Forward and reverse supply chain network design for consumer medical supplies considering biological risk. *Comput Ind Eng* 140:106229. <https://doi.org/10.1016/j.cie.2019.106229>
2. Rezapour S, Farahani RZ, Fahimnia B, Govindan K, Mansouri Y (2015) Competitive closed-loop supply chain network design with price-dependent demands. *J Clean Prod* 93:251–272. <https://doi.org/10.1016/j.jclepro.2014.12.095>
3. Tibben-Lembke R, Rogers D (2002) Differences between forward and reverse logistics in a retail environment. *Surg Endosc Other Interv Tech* 7:271–282
4. Pourjavad E, Mayorga RV (2018) An optimization model for network design of a closed-loop supply chain: a study for a glass manufacturing industry. *Int J Manag Sci Eng Manag.*:169–179. <https://doi.org/10.1080/17509653.2018.1512387>

5. Agrawal S, Singh RK, Murtaza Q (2015) A literature review and perspectives in reverse logistics. *Resour Conserv Recycl* 97:76–92. <https://doi.org/10.1016/j.resconrec.2015.02.009>
6. Tian G, Zhang H, Feng Y, Jia H, Zhang C, Jiang Z, Li P (2017) Operation patterns analysis of automotive components remanufacturing industry development in China. *J Clean Prod* 164:1363–1375. <https://doi.org/10.1016/j.jclepro.2017.07.028>
7. Rajendran S, Ravindran AR (2019) Inventory management of platelets along blood supply chain to minimize wastage and shortage. *Comput Ind Eng* 130:714–730. <https://doi.org/10.1016/j.cie.2019.03.010>
8. Mahajan J, Vakharia AJ (2016) Waste management: a reverse supply chain perspective. *J Decis Mak* 41:1–12
9. Das S, Lee S-H, Kumar P, Ki-Hyun K, Lee SS, Bhattacharya SS (2019) Solid waste management: scope and the challenge of sustainability. *J Clean Prod* 228:658–678
10. Al-Salem SM, Evangelisti S, Lettieri P (2014) Life cycle assessment of alternative technologies for municipal solid waste and plastic solid waste management in the Greater London area. *Chem Eng J* 244:391–402. <https://doi.org/10.1016/j.cej.2014.01.066>
11. Rajamanikam R, Poyyamoli G, Kumar S, Lekshmi R (2014) The role of non-governmental organizations in residential solid waste management: a case study of Puducherry, a coastal city of India. *Waste Manag Res* 32:867–881. <https://doi.org/10.1177/0734242x14544353>
12. Chopra S, Meindl P (2017) Supply chain management. Strategy. Planning & operation. Das summa summarum des management, pp 265–275
13. Shekarian E (2019) A review of factors affecting closed-loop supply chain models. *J Clean Prod* 253:119823. <https://doi.org/10.1016/j.jclepro.2019.119823>
14. Heyes G, Sharmina M, Mendoza JMF, Gallego-Schmid A, Azapagic A (2018) Developing and implementing circular economy business models in service-oriented technology companies. *J Clean Prod* 177:621–632. <https://doi.org/10.1016/j.jclepro.2017.12.168>
15. Alcayaga A, Hansen EG, Wiener M (2019) Towards a framework of smart-circular systems: an integrative literature review. *J Clean Prod* 221:622–634. <https://doi.org/10.1016/j.jclepro.2019.02.085>
16. Gallego-Schmid A, Mendoza JMF, Azapagic A (2019) Environmental impacts of takeaway food containers. *J Clean Prod* 211:417–427. <https://doi.org/10.1016/j.jclepro.2018.11.220>
17. Iacovidou E, Velis CA, Purnell P, Zwirner O, Brown A, Hahladakis J, Williams PT (2017) Metrics for optimising the multi-dimensional value of resources recovered from waste in a circular economy: a critical review. *J Clean Prod* 166:910–938. <https://doi.org/10.1016/j.jclepro.2017.07.100>
18. Ross S, Evans D (2003) The environmental effect of reusing and recycling a plastic-based packaging system. *J Clean Prod* 11:561–571. [https://doi.org/10.1016/s0959-6526\(02\)00089-6](https://doi.org/10.1016/s0959-6526(02)00089-6)
19. Geueke B, Groh K, Muncke J (2018) Food packaging in the circular economy: overview of chemical safety aspects for commonly used materials. *J Clean Prod* 193:491–505. <https://doi.org/10.1016/j.jclepro.2018.05.005>
20. Tseng L-K, Lim MK, Wong W-P, Chen Y-C, Zhan Y (2018) A framework for evaluating the performance of sustainable service supply chain management under uncertainty. *Int J Prod Econ* 195:359–372. <https://doi.org/10.1016/j.ijpe.2016.09.002>
21. Jindal A, Sangwan KS (2011) Development of an interpretive structural model of barriers to reverse logistics implementation in Indian industry. *Globalized Solutions for Sustainability in Manufacturing* 17:448–453. https://doi.org/10.1007/978-3-642-19692-8_77
22. Govindan K, Darbari JD, Agarwal V, Jha PC (2017) Fuzzy multi-objective approach for optimal selection of suppliers and transportation decisions in an eco-efficient closed loop supply chain network. *J Clean Prod* 165:1598–1619. <https://doi.org/10.1016/j.jclepro.2017.06.180>
23. Zhang Y, Alshraideh H, Diabat A (2018) A stochastic reverse logistics production routing model with environmental considerations. *Ann Oper Res* 271:1023–1044. <https://doi.org/10.1007/s10479-018-3045-2>
24. Senthil S, Murugananthan K, Ramesh A (2018) Analysis and prioritisation of risks in a reverse logistics network using hybrid multi-criteria decision making methods. *J Clean Prod* 179:1–752. <https://doi.org/10.1016/j.jclepro.2017.12.095>

25. Engeland JV, Beliën J, Boeck LD, Jaeger SD (2018) Literature review: strategic network optimization models in waste reverse supply chain. *Omega* 91:102012. <https://doi.org/10.1016/j.omega.2018.12.001>
26. Mohtashami Z, Aghsami A, Jolai F (2019) A green closed loop supply chain design using queuing system for reducing environmental impact and energy consumption. *J Clean Prod*:118452. <https://doi.org/10.1016/j.jclepro.2019.118452>
27. Negm AM, Shareef N (2020) *Waste Management in MENA regions*. Springer Water. <https://doi.org/10.1007/978-3-030-18350-9>
28. Sellitto MA (2018) Reverse logistics activities in three companies of the process industry. *J Clean Prod* 187:923–993. <https://doi.org/10.1016/j.jclepro.2018.03.262>
29. Zhou Y, Wang S (2008) Generic model of reverse logistics network design. *J Transp Syst Eng Inf Technol* 8:71–78. [https://doi.org/10.1016/s1570-6672\(08\)60025-2](https://doi.org/10.1016/s1570-6672(08)60025-2)
30. Krikke H, Bloemhof-Ruwaard J, Van Wassenhove LN (2003) Concurrent product and closed-loop supply chain design with an application to refrigerators. *Int J Prod Res* 41:3689–3719. <https://doi.org/10.1080/0020754031000120087>
31. Al-Salem M, Diabat A, Dalalah D, Alrefaei M (2016) A closed-loop supply chain management problem: reformulation and piecewise linearization. *J Manuf Syst* 40:1–8. <https://doi.org/10.1016/j.jmsy.2016.04.001>
32. Large RO, Kramer N, Hartmann RK (2013) Procurement of logistics services and sustainable development in Europe: fields of activity and empirical results. *J Purch Supply Manag* 19:122–133. <https://doi.org/10.1016/j.pursup.2013.05.002>
33. Govindan K, Soleimani H, Kannan D (2015) Reverse logistics and closed-loop supply chain: a comprehensive review to explore the future. *Eur J Oper Res* 240:603–626. <https://doi.org/10.1016/j.ejor.2014.07.012>
34. Ezeah C, Fazakerley JA, Roberts CL (2013) Emerging trends in informal sector recycling in developing and transition countries. *J Waste Manag* 33:2509–2519. <https://doi.org/10.1016/j.wasman.2013.06.020>
35. Ferrer G, Swaminathan JM (2006) Managing new and remanufactured products. *Manage Sci* 52:15–26. <https://doi.org/10.1287/mnsc.1050.0465>
36. Razmi J, Zahedi-Anaraki A, Zakerinia M (2013) A bi-objective stochastic optimization model for reliable warehouse network redesign. *Math Comput Model* 58:1804–1813. <https://doi.org/10.1016/j.mcm.2013.03.009>
37. Guide VDR, Van Wassenhove LN (2009) The evolution of closed-loop supply chain research. *Oper Res* 57:10–18. <https://doi.org/10.1287/opre.1080.0628>
38. Natural green tribunal act 2010. (https://indiacode.nic.in/handle/123456789/2025?view_type=browse&sam_handle=123456789/1362)
39. The hazardous waste management rules 2016. (<https://vikaspedia.in/energy/environment/waste-management/plastic-waste-management-rules-2016>)
40. Plastic waste management rules 2016. (<https://vikaspedia.in/energy/environment/waste-management/hazardous-waste/environment-ministry-notifies-hazardous-waste-management-rules-2016>)
41. Ghuge J, Surale S, Patil BM, Bhutekar SB (2019) Utilization of waste plastic in manufacturing of paver blocks. *Int Res J Eng Technol* 6:1967–1970
42. Trivedi M, Mathur M, Johri P, Singh A, Tiwari RK (2019) Waste management: a paradigm shift'. *Environ Concerns Sustain Dev*:337–363. https://doi.org/10.1007/978-981-13-6358-0_14
43. Kumar A, Samadder SR, Kumar N, and Singh C (2018) Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling. *Waste Manag* 79:781–790. <https://doi.org/10.1016/j.wasman.2018.08.045>
44. Bing X, Bloemhof-Ruwaard JM, van der Vorst JGAJ (2012) Sustainable reverse logistics network design for household plastic waste. *Flex Serv Manuf J* 26(1–2):119–142. <https://doi.org/10.1007/s10696-012-9149-0>

45. Faraca G, Astrup T (2019) Plastic waste from recycling centres: characterisation and evaluation of plastic recyclability. *J Waste Manag* 95:388–398. <https://doi.org/10.1016/j.wasman.2019.06.038>
46. Wichai-utcha N, Chavalparit O (2018) 3Rs Policy and plastic waste management in Thailand. *J Mater Cycles Waste Manage* 21:10–22. <https://doi.org/10.1007/s10163-018-0781-y>

Real-Time Implementation and Analysis of Different Adaptive Enhancement Algorithms Using Embedded Hardware Boards



Jharna Majumdar, Manish Verma, Prajwal Shah, Gagan Karthik, Srinath Ramachandhran, and Thribhuvan Gupta

1 Introduction

Implementation of image and video processing algorithms on embedded hardware boards has a wide variety of applications in many domains of research and practical application. Real-time implementation on hardware boards can be heavy for the processors of the CPU to process serially, causing it to be a very time-consuming task making it very undesirable and inefficient. This problem can be solved by using parallel computing techniques. In this paper, the methods used are multithreading using domain decomposition and task decomposition. The above-mentioned methods optimize the working of the algorithms on the hardware as they divide the load among the processors of the CPU, ensuring quicker execution, and make it a less time-consuming task. Further optimization can be made by following some basic thumb rules for software optimization. Adaptive enhancement is a process in which the image is processed, i.e., the histogram of the image is equalized according to the environmental factors such as brightness and contrast. Here, histogram equalization is the technique where the contrast of the image is enhanced by stretching the most frequent pixel value over the range of the histogram. Several algorithms have been proposed for adaptive enhancement of images, of which comparisons are made of the following algorithms: (BHEP) [3], (BPDHE) [6], and (MHE) [7] for gray-scale input video. Similarly for color video the following algorithms are used: (HDHE) [5],

J. Majumdar (✉)

Head Centre for Robotics Research, Nitte Meenakshi Institute of Technology, Bangalore 560064, India

e-mail: jharna.majumdar@nmit.ac.in

M. Verma · P. Shah · G. Karthik · S. Ramachandhran · T. Gupta

Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Bangalore 560064, India

(QDAPHE) [4], and (CEMM) [7]. The comparison is made on the basis of execution time, frames per second, and QM parameters, respectively.

The hardware boards used for the comparisons are UDOO Quad and NVIDIA Kepler™ Jetson TK1. UDOO Quad board is a single-board computer that can run Android or Linux OS and also features an Arduino-compatible microcontroller embedded onboard. It is a powerful prototyping board for software development and design. Easy to use, it allows to develop projects with minimum knowledge of hardware design. NVIDIA Kepler™ Jetson TK1 gives everything you need to unlock the power of the GPU for embedded systems applications. It is built around the revolutionary NVIDIA Tegra® K1 SoC and uses the same NVIDIA Kepler™ computing core designed into supercomputers around the world. This gives you a fully functional NVIDIA CUDA® platform for quickly developing and deploying compute-intensive systems for computer vision, robotics, medicine, and more. The next section of this paper explains description about the two embedded hardware boards. The third section explains about the different algorithms which were implemented in a real-time dark environment and the related optimization methods to reduce the time complexity of these algorithms. The final sections describe the experimental analysis and results, followed by acknowledgement and references.

2 Hardware Description

The Advantage of Implementing adaptive enhancement algorithms on image and video processing leads to its accessibility for Real-time applications such as Military and Defence Applications and Security applications. Such Adaptive Enhancement Algorithms might not be efficient to overcome the glitches that come in Real-time implementation. Here hardware implementation plays a vital role in a real scenario.

The two hardware boards that are used for analyzing the extent of adaptive enhancement algorithms usability in real-time scenario are as follows.

2.1 *UDOO Board*

UDOO board is a single-board computer with an integrated Arduino Due compatible microcontroller. UDOO Quad board can run on Android or Linux OS. It is a powerful prototyping board for software development and design. It has a quad-core processor with operating frequency 1 GHz and comprises a CPU which is specified as ATMEL SAM3X8E ARM Cortex M3 CPU. It has a RAM of DDR3 1 GB. Power supply is limited to 6–15 V, and in addition to this, external battery connector is available.

2.2 NVIDIA Jetson TK1

NVIDIA Jetson is embedded with Tegra K1 SoC (CPU + GPU + ISP on a single processor) that is available directly in the chip form. The embedded GPU gives an extra edge for embedded systems applications. This includes fully functional NVIDIA CUDA platform for quickly developing and deploying computer-intensive systems for computer vision, robotics, medicine, and more. It has a NVIDIA Kepler “GK20a” GPU with 192 SM3.2 CUDA cores and comprises NVIDIA “4-Plus-1” 2.32 GHz ARM quad-core Cortex-A15 CPU. It has a memory of DRAM:2 GB DDR3L 933 MHz with a in-built storage of 16 GB fast eMMC 4.51. It has been established by two types of USB ports, i.e., USB 3.0, a full-size Type-A female socket and USB 2.0, a micro-AB female socket. It has a separate HDMI port too for transferring video signals out.

The embedded boards used differ in their operating frequency and their CPU processing, and in this paper, we have compared only the CPU processing (Figs. 1 and 2).

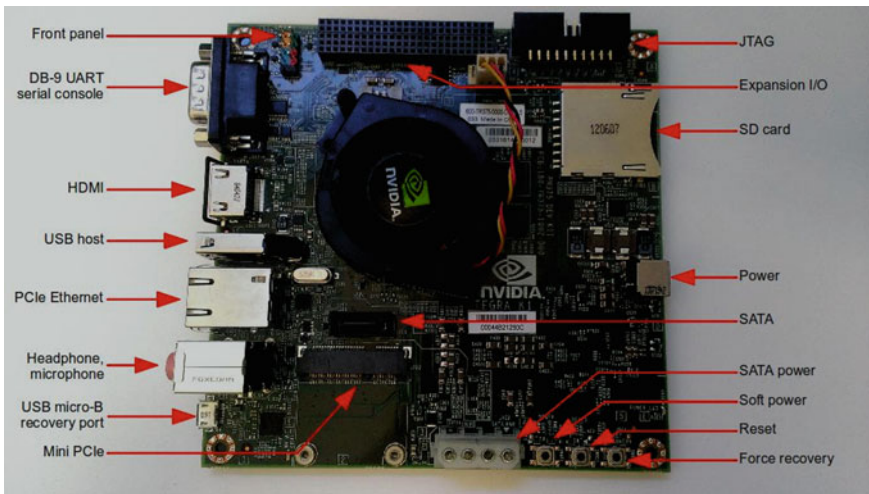


Fig. 1 Labeled parts of Nvidia Jetson TK1

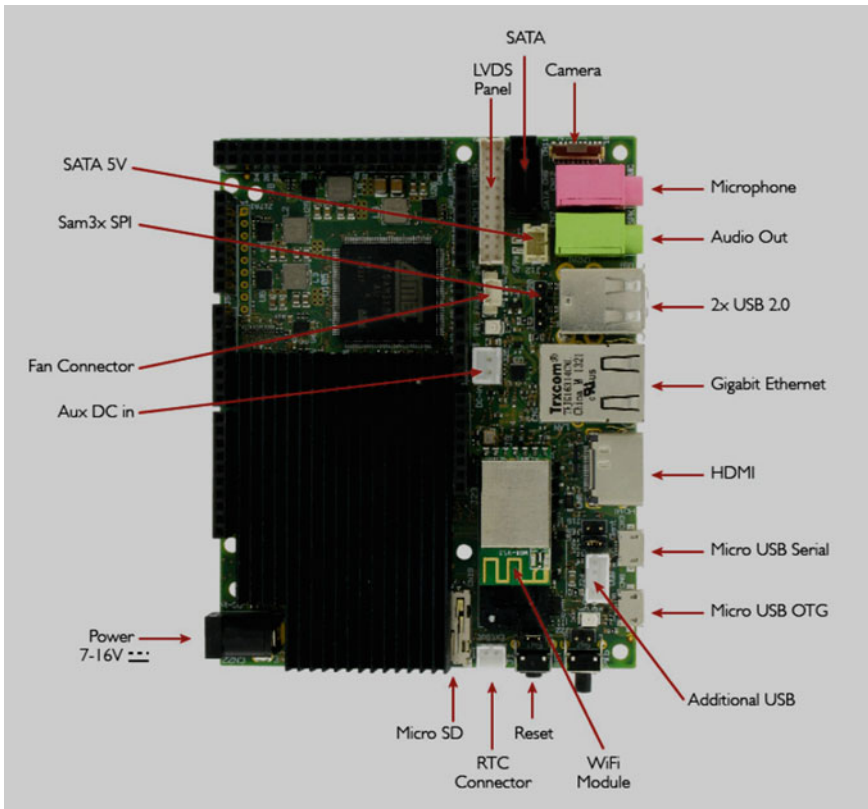


Fig. 2 Labeled parts of UDOO Quad

3 Algorithms

3.1 Algorithms Used for Gray Scale

3.1.1 Brightness Preserving Dynamic Histogram Equalization (BPDHE)

Read the image and apply the Gaussian filter. Detect the local maxima of the smoothed image. Partition the histogram based on the local maxima. Map each partition into updated dynamic range. Equalize each partition by finding pdf and cdf of the histogram. Then, the partitioned histograms are joined together to form the final histogram of the output image.

3.1.2 Multi Peak Histogram Equalization with Brightness Preserving (MHE)

The input image or video frame is read, and the histogram is obtained along with its probability density function. Nine consecutive probabilities of the brightness levels are averaged for the new probability of the central brightness level. The histogram is then smoothened, and then, the breakpoints are used to detect the peaks of the histogram. The peaks are then equalized independently and joined together to form the histogram of the output image.

3.1.3 Bi-Histogram Equalization using Plateau Limits (BHEP)

The input image is read, and the histogram of the image is obtained and is divided into two equal parts. Two threshold values are calculated for each half, and the histogram values greater than the threshold values are clipped, thus enhancing the histogram. The clipped histogram halves are then subjected to different transform functions. The sub-histograms are then joined together to form the final output image.

3.2 Algorithms used for Color

3.2.1 Color Image Enhancement Based on Hue Differential Histogram Equalization (HDHE)

In this algorithm, hue information is used for image enhancement, preservation of hue is crucial for color image, hence it is impossible to conduct processing of hue, but changes in hue component are quiet significant when considered a shadow region, therefore, this information of hue change can be used by calculating differential hue information, and this can help in significant enhancement of the color image.

3.2.2 Quadrant Dynamic with Automatic Plateau Limits Histogram Equalization for Image Enhancement (QDAPHE)

The input image is read, and the histogram of the image is obtained and is divided into four equal parts. Four threshold values are calculated for each quarter, and the histogram values greater than the threshold values are clipped, thus enhancing the histogram. The clipped histogram halves are then subjected to different transform functions. The sub-histograms are then joined together to form the final output image.

3.2.3 Contrast Enhancement Based on Mathematical Morphology (CEMM)

Mathematical morphology is a tool used for extraction of image components that are useful for image processing. In this algorithm, erosion and dilation are used which help in smoothening of the edges and filling gaps in an image.

3.3 Optimization

In a lay man's term, optimization means action of making the best or most effective use of a situation or resource. A program is optimized so that it becomes of a smaller size, consumes less memory, executes more rapidly or with fewer operations. Higher optimization levels perform more global transformations on the program and apply more expensive analysis algorithms in order to generate faster and more compact code. The expense of compilation time and the resulting improvement in execution time both depend on the particular application and the hardware environment. There are two different ways of optimization, namely software optimization and hardware optimization.

Software optimization is a process of efficiently editing some part of the codes, such that the overall complexity of the code decreases which results in making the system work more efficient with lesser number of resources. There are various methodologies that we followed while adapting software optimization, which are as follows.

- (1) Minimizing the use of conditional statement if-elseif-else and jump/branch statements.
- (2) Calculating the log values at the beginning of the program and formed a lookup table.
- (3) Prefer passing the structures by value rather than reference.
- (4) Prefer use of shift operators “>>” or “<<” for multiplication and division.

Hardware optimization is a method to employ the resources available in the embedded hardware board efficiently for the given program. For the same, we performed multithreading, we followed two approaches to implement multithreading, i.e., task decomposition and domain decomposition. To implement the algorithms efficiently using task decomposition, we distributed each of the task within an algorithm to different threads of the video frame so that overall execution time of the algorithm reduces. Whereas using domain decomposition for these algorithms, we divided each video frame into three different parts and executed all the operations separately and combining them at the end by averaging the neighboring elements.

4 Experiment and Results

The main objective of the work presented in this paper is to analyze the performance of different adaptive enhancement algorithms in real time on embedded hardware boards. With the optimization using task and domain decomposition, the overall performance of these algorithms is increased significantly. The results of the timing analysis and frame analysis with comparison to each algorithm on a real-time video and an image are presented here. The results have been showcased in two parts: **Part(A)** displays results and analysis for gray-scale input video, whereas **Part(B)** showcases the analysis and output for color video and images.

4.1 Quality Metric Analysis

Quality metric parameters play a vital role to compare and analyze the overall attributes of an image. It describes the overall betterment of the image with reference to the visual attributes, i.e., sharpness, tone rendition, and color. To compare the described algorithms, we calculated the following QM parameters.

1. Global contrast
2. Quality index
3. Absolute mean brightness error
4. Entropy error rate
5. Relative entropy
6. Structural similarity index [10].

Table 1 showcases the comparison of QM analysis for Part A. Part (A) output and analysis (Fig. 3).

Comparison of all the algorithms with quality metric parameters (Table 1).

Table 1 Denotes best value of QM parameter

Methods	Quality metric parameters		
	BPDHE	BHEP	MHE
ERR	0.020985	0.0472299	0.0009849
GC	4030.93	22.217	5201.95
QI	4.01285	5.67641	3.69542
AMBE	66.9086	10.3623	63.3572
RE	-0.82099	-0.256893	-0.10059
SSIM	0.8323	0.7636	0.92

4.2 Domain Versus Task Decomposition

For optimization of the adaptive enhancement algorithms, two methods were used, namely domain and task decomposition, and the observed values of both the methodologies are compared in Fig. 1 (UDOO Quad) and Fig. 2 (Nvidia Jetson TK1) (Figs. 4 and 5).

Comparison of performance analysis on UDOO Quad and Jetson TK1 with a video resolution of 480×480 pixels and total frames of 108,305.

1. BHEP (Table 2).

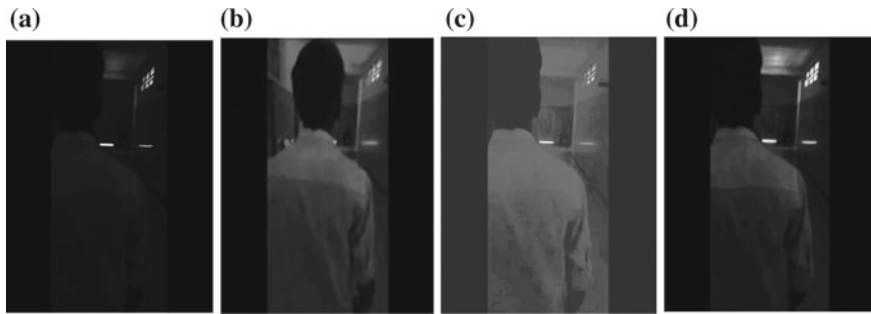


Fig. 3 a Input video shot, b BHEP output, c MHE output, d BPDHE output

Fig. 4 Graphical representation of comparison of timing analysis for task and domain decomposition for UDOO Quad

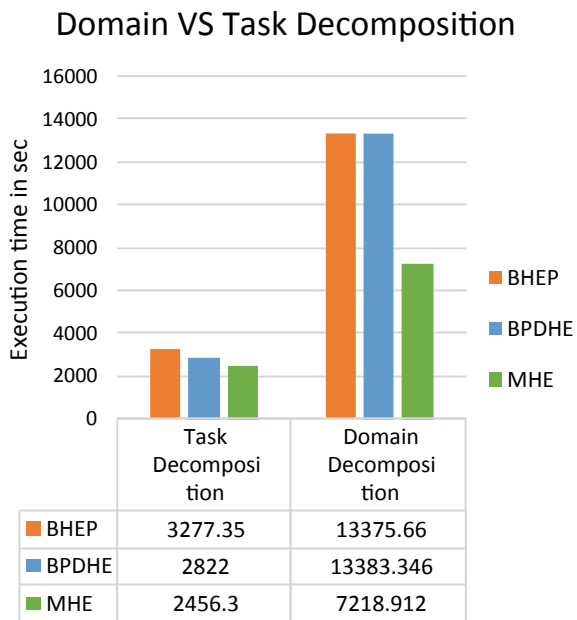


Fig. 5 Graphical representation of comparison of timing analysis for task and domain decomposition for Nvidia Jetson TK1

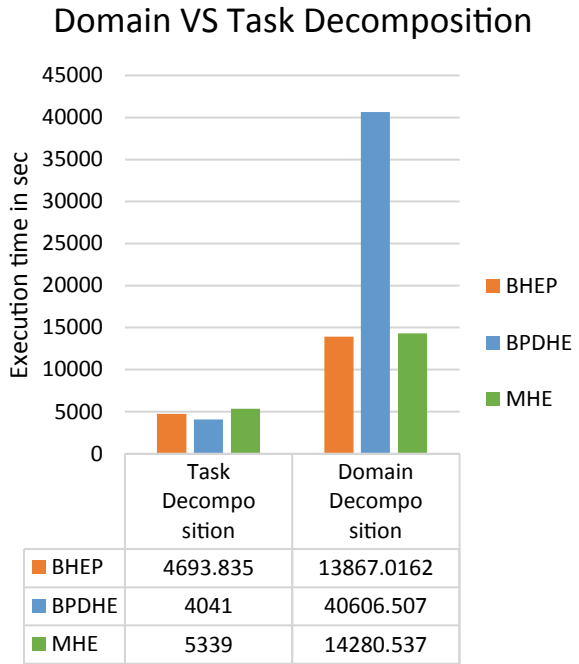


Table 2 Comparison of performance analysis on UDOO Quad and Jetson TK1 for BHEP

No. of threads	UDOO Quad		Jetson TK1	
	Time taken (sec)	FPS	Time taken (sec)	FPS
2	9387.675	12	3277.35	33
3	9766.679	11	3278.00	32
4	10,468.875	10	3366.00	32

2. BPDHE (Table 3).

3. MHE (Table 4).

Table 3 Comparison of performance analysis on UDOO Quad and Jetson TK1 (BPDHE)

No. of threads	UDOO quad		Jetson TK1	
	Time taken (sec)	FPS	Time taken (sec)	FPS
2	4041	26	2822	38
3	4202	25	2765	39
4	4215	25	2570	42

Table 4 Comparison of performance analysis on UDOO Quad and Jetson TK1 for MHE

No. of threads	UDOO Quad		Jetson TK1	
	Time taken (sec)	FPS	Time taken (sec)	FPS
2	10,702	10	4912.6	22
3	10,678	10	4925.3	22
4	10,798	10	4930.5	21

Table 5 Comparison of performance analysis on UDOO Quad and Jetson TK1 with two threads per processor

Methods	UDOO Quad	Jetson TK1
	Execution time in sec	
BHEP	27,511.636	6325.00
BPDHE	2101.023	1373.30
MHE	22,516.600	5187.80

Tables showcase the observed values of timing analysis and frame per second (FPS) analysis of the algorithms on UDOO Quad and Nvidia Jetson TK1, with the graphical representations of the (FPS) of respective tables.

Comparison of more number of threads to each processor.

In this analysis, we processed more number of threads to each processor statistically 2, 3, 4 threads per processor so as to observe the behavior of each algorithm for its performance with more than one threads per processor.

- A. Two threads per processor (total eight threads) (Table 5).
- B. Three threads per processor (total 12 threads) (Table 6).
- C. Four threads per processor (total 16 threads) (Table 7).

Table 6 Comparison of performance analysis on UDOO Quad and Jetson TK1 with three threads per processor

Methods	UDOO Quad	Jetson TK1
	Execution time in sec	
BHEP	28,141.28	6184.000
BPDHE	3412.622	1361.97
MHE	22,911.92	5187.50

Table 7 Comparison of performance analysis on UDOO Quad and Jetson TK1 with four threads per processor

Methods	UDOO quad	Jetson TK1
	Execution time in sec	
BHEP	29,054.00	6130.00
BPDHE	4080.33	1392.00
MHE	23,737.20	5155.00

Table 8 Comparing the optimal values for the execution time in seconds for UDOO Quad and Jetson TK1. Part (B) output and analysis (Table 8) (Figs. 6, 7, and 8)

No. of threads	UDOO quad	Jetson TK1
	Execution Time in sec	
1	15,920.83	9530.84
2	1527.60	1340.00
3	1559.54	1317.00
4	1568.49	1315.00
8	5111.99	1373.30
12	3412.62	1361.97
16	4080.33	1392.00

Fig. 6 Graphical representation of Table 8

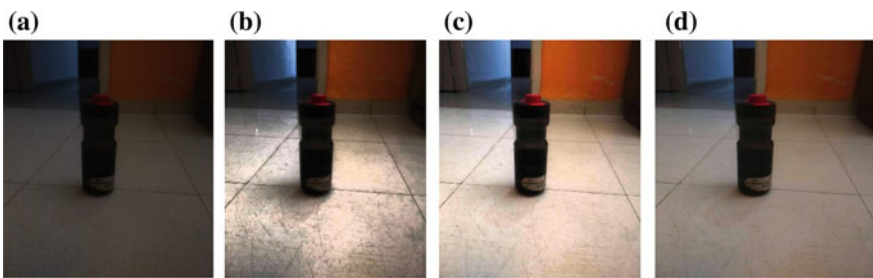
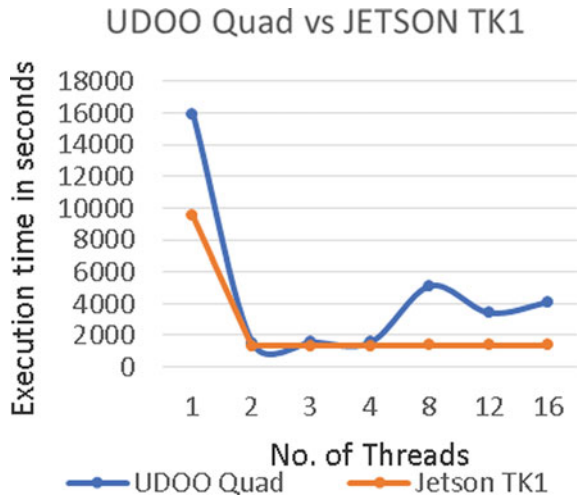


Fig. 7 a Input image, b HDHE output, c QDAPHE output, d CEMM output

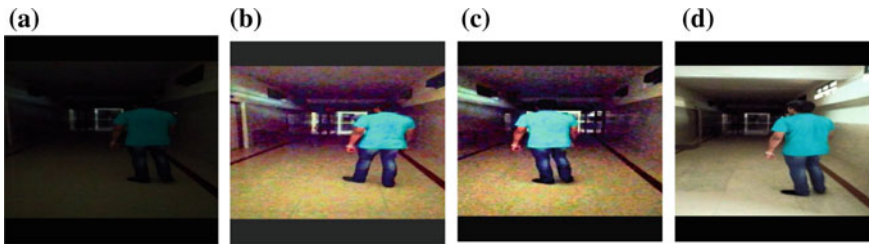


Fig. 8 a Input video shot, b HDHE output, c QDAPHE output, d CEMM output

Table 9 Comparison of all the algorithm with quality metric parameters (Table 9)

Methods	Quality metric parameters		
	HDHE	QDAPHE	CEMM
ERR	0.000205327	- 0.0305316	-0.0236203
GC	4540.48	4249.5	4545.2
QI	25.134	16.3228	19.0208
AMBE	73.3871	109.87	73.2983
RE	-0.179169	0.0134539	-0.175172
SSIM	0.79069	0.8383	0.8242

Table 10 Comparison of performance analysis on computer with a video resolution of 480×480 pixels and total frames of 52,825 (Table 10)

Methods	Without optimization	With optimization
	Execution time in sec	
HDHE	3380.8	3364.95
QDAPHE	3380.8	3333.25
CEMM	3301.56	3290.99

5 Conclusion

In this paper, we have compared and analyzed real-time implementation of above-stated adaptive enhancement algorithms on embedded hardware boards—UDOO Quad and Nvidia Jetson TK1. With the adaptation of various optimization techniques, we have observed a drastic and effective outputs.

The comparison of these algorithms in real-time with regards to Quality metric parameters we observed Multi Peak Histogram Equalization (MHE) to be the optimum algorithm for gray-scale video input and Quadrant Dynamic with Automatic Plateau Histogram Equalization (QDAPHE) for color video input. With this analysis, anyone can implement this algorithm to enhance and navigate through a dark environment and track the target object.

Acknowledgements Our sincere thanks to Vision Group of Science and Technology (VGST), Govt. of Karnataka, to acknowledge our research and provide us the financial support to carry out

our research at NMIT. We express our sincere thanks to our colleagues at Centre for Robotics and Research, NMIT, for providing support.

Finally, our sincere gratitude goes to Professor N.R. Shetty, Director NMIT and Dr. H.C. Nagaraj, Principal, NMIT, for providing the infrastructural support and a whole hearted encouragement to carry out research at NMIT.

References

1. Wangt H, Chen Y, Fang T, Tyan J, Ahuja N (2006) Gradient adaptive image restoration and enhancement
2. Ooi CH, Kong NSPK, Ibrahim H (2009) Bi-histogram equalization with plateau limits for digital image enhancement
3. Jagatheeswari P, Suresh Kumar S, Mary Linda M (2014) Quadrant dynamic with automatic plateau limit histogram equalization for image enhancement
4. Ibrahim H, Kong NSPK (2007) Brightness preserving dynamic histogram equalization for image contrast enhancement
5. Wongsritong K, Kittayaruasiriwat K, Cheevasuit F, Dhejan K, Somboonkaew (1998) A Contrast enhancement using multipeak histogram equalization with brightness preserving
6. Purushothaman J, Kamiyama M, Taguchi A (2016) Color image enhancement based on Hue differential histogram equalization. In: 2016 international symposium on intelligent signal processing and communication systems (ISPACS), Phuket, 2016, pp. 1–5. <https://doi.org/10.1109/ISPACS.2016.7824720>
7. Antony M, Sathiaselan, JGR (2018) Contrast enhancement of grayscale and color images using adaptive techniques
8. Vinyals O, Friedland G, Mirghafori N (2007) Revisiting a basic function on current CPU's : a fast logarithmic implementation with adjustable accuracy, TR-07-002
9. C code optimization examples for the stare core SC3850 core, Freescale semiconductor application note, Document number:AN3674 Rev 0,04/2010
10. Wang Z, Bovik AC, Shiekh HR, Simoncelli. EP (2004) Image quality assessment: from error visibility to structural similarity
11. Jhama M, Adarsh C, Harshpreet S, Rahul VC (2019) Real-time performance analysis of retinex algorithm on embedded boards for robotics application. In: Proceedings of the third international conference on advanced informatics for computing research (ICAICR '19). Association for computing machinery, New York, NY, USA, Article 15, pp. 1–8. <https://doi.org/10.1145/3339311.3339326>
12. <https://www.udoo.org/docs/Introduction/Introduction.html>
13. <https://developer.nvidia.com/embedded/buy/jetson-devkit>

Proxy Re-Encryption Using Vector Decomposition



A. U. Kaviya and I. Praveen

1 Introduction

Public key cryptography (PKC) obtained enormous growth in the recent decades. The significance of PKC can be identified by its applications in various cryptographic primitives such as identity-based encryption (IBE), oblivious transfer (OT) and group signatures. Proxy re-encryption is another such primitive. Most of the proxy re-encryption schemes are based on public key cryptography. Proxy re-encryption is a mechanism that re-encrypts the ciphertext without decrypting the ciphertext. That is, proxy encrypts the ciphertext again, without the knowledge of plaintext. Identity-based proxy re-encryption uses unique identity such as email id and finger print with some credentials.

In this paper, we propose a proxy re-encryption scheme based on the hardness of vector decomposition problem (VDP). VDP was proposed by Yoshida et al. [1]. The concept of VDP was further developed by OKamoto and Takashima [2]. In this paper, they proposed a homomorphic encryption and various types of signatures. Based on this, Nidhin et al. [3] proposed an access control scheme for encrypted data. We use this idea for access control in our proposed scheme.

In the proposed scheme, the proxy obtains ciphertexts C_1, C_2, \dots, C_n corresponding to plaintexts M_1, M_2, \dots, M_n . She encrypts them and upload key in server. Now, when the sender Alice wants to share these encrypted messages with the receiver

A. U. Kaviya

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

I. Praveen (✉)

Department of Mathematics, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

e-mail: i_praveen@cb.amrita.edu

Bob, Alice allows the server to re-encrypt her encrypted messages without decrypting them.

The world is looking for schemes with strong security because of so many forgeries. Nowadays, technology like cloud computing is also open to public since it is more vulnerable in accessing some of the databases by using sql injection, etc.... Presently, so many proxy re-encryption schemes are available in the literature for using in the fields of business, health care, etc. Still it has some risk in the amount of trust required. So, great measures have to be taken with strong cryptographic calculations to provide proper security. Here, we propose a scheme based on vector decomposition problem which can serve as a solution to such issues. The hardness of the proposed system ensures the best authorisation and authentication.

The remaining part of the paper is organised as follows: In Sect. 2, preliminaries and basic definitions are briefly explained. Proposed scheme and correctness of the scheme is presented in Sect. 3. In Sect. 4, we conclude the paper.

2 Preliminaries

2.1 Bilinear Pairing

Bilinear pairing is a function from a pair of groups to another group. Let P_1, P_2, P_T be from any groups of prime order. Let p_j be a generator of P_j , for $j = 1, 2$. An efficiently computable bilinear map $e : P_1 \times P_2 \rightarrow P_T$, is called a bilinear pairing if it satisfies the the following properties:

- Bilinearity: For all $p_j \in P_j$, for $j = 1, 2$ and $a, b \in \mathbb{F}$, $e(ap_1, bp_2) = e(p, p)^{ab}$.
- Non-degenerate: There exists $p_j \in P_j$, $j = 1, 2$ such that $e(p_1, p_2) \neq 1$.

2.2 Vector Decomposition Problem

VDP was introduced by Yoshida [1, 4] and generalised to higher dimensions by Okamoto and Takashima [2]. There are various constructions that are available in the literature using VDP [5–7].

Definition 2.1 *Generalised Computational Vector Decomposition Problem (gCVDP):* Suppose \mathbb{V} be an n -dimensional vector space with bases $\{K_1, K_2 \dots K_n\}$. With respect to the given basis, for a given $Q \in \mathbb{V}$, VDP is to find $R \in \mathbb{V}$ such that $R \in \langle K_1, K_2, \dots K_m \rangle$ and $Q - R \in \langle K_{m+1}, K_{m+2} \dots K_n \rangle$, $m < n$.

V - l_1 dimensional F_T vector space with security parameter K and $l_1 > l_2$.

Given $b_0, \dots b_{l_2-1} \in V$, $x_0, \dots x_{l_1-1} \in F_r$.

$$V = \sum_{i=0}^{l_1-1} x_i b_i.$$

Computational Vector Decomposition Problem $CVD P_{l_1, l_2}$ is to find $w = \sum_{i=0}^{l_1-1} x_i b_i$ which is in a subspace of V generated by vectors b_0, \dots, b_{l_2-1} .

Okamoto and Takashima generalised VDP into higher dimensions using a novel concept called distortion eigenvector space.

Definition 2.2 *Distortion Eigenvector space.*

V : l -dimensional vector space F_r .

A : a_0, \dots, a_{l_1-1} be a basis of F_r vector space V and f a polynomial time computable automorphism V .

Each a_i is an eigenvector of f . That is $f(a_i) = \lambda_i a_i$. Also, each a_i has a different eigenvalue.

There exist polynomial time computable endomorphism $\phi_{i,j}(a_j) = a_i$.

$A = a_0, \dots, a_{l_1-1}$ is called. (distortion eigenvector basis.)

V : l -dimensional vector space over F_r is called a distortion eigenvector space if

1. V has a distortion eigenvector basis. 2. There exists a skew symmetric non-degenerate bilinear pairing.

$e: v \times v \rightarrow \mu_r$ i.e. μ_r is multiplicative cyclic group of order r .

There exist a polynomial time computable automorphism ρ on V such $e(v, \rho(v)) \neq 1$ for any v in a quadratic hypersurface of $V \cong (F_r)^l$.

Lemma 1 If $B = \{B_1, B_2, \dots, B_n\}$ be the distortion eigenvector basis of \mathbb{V} with B_i having eigenvalue λ_i of T , the below polynomial of T is a projection operator:

$$Pr_j(v) = \prod_{i \neq j} \frac{(T - \lambda_i)(v)}{(\lambda_j - \lambda_i)}$$

So, $Pr_j(B_k) = 0$ for $k \neq j$ and $Pr_j(B_j) = B_j$.

Consider the non-singular matrix $X = (x_{ij})$ such that $P_i = \sum_{j=1}^n x_{ij} B_j$. Since X is non-singular, $P = (P_1, P_2, \dots, P_n)$ forms a basis for \mathbb{V} . Let $v' = \sum_{j=1}^n y_j P_j$ be a vector in \mathbb{V} . Our aim is to decompose v' as the sum of two vectors v'_1 and v'_2 where v'_1 lies in the span of $(P_1, P_2, \dots, P_m), m < n$ and v'_2 in the span of $(P_{m+1}, P_{m+2}, \dots, P_n)$. Let $X^{-1} = (h_{ij})$, then lemma 3 of Okamoto and Takashima [2] proves that the function

$$VDeco(v', \langle P_j \rangle, X, \langle P_1, P_2, \dots, P_n \rangle) = \sum_{i=1}^n \sum_{k=1}^n h_{ij} x_{j k} t_{ki} (Pr_i(v))$$

can be used to accomplish the above aim.

2.3 Related Work

The paper [8] deals with proximity match of the user’s profile using attribute-based method. The user’s interest was represented in vectors, if the two vectors of the both

user match then they can communicate. The adversary model considered honest-but-curious framework. Here, the two cloud servers will never collude with each other. The adversary cannot alter the vector since it is following dot product which has to prove unity. The honest-but-curious adversary is nothing but the participant is legal following the protocol properly and also attempts to gather information from the received messages. It is using Elgamal encryption which contains additive homomorphic properties. The Elgamal encryption works based on bilinear mapping.

The paper presented by Nidhin et al. [3] suggested a method for access control based on RBAC. The administrator keeps the list of all the roles and the role assignment of users. The master authority is a cloud storage system which can be subdivided into public and private clouds. Private cloud is used for storage, and public cloud is used for communication to any real-time system. Access control using a role based system assigns roles to users according to the permissions to access a particular file. So, depending on the privilege, roles are assigned. We use this mechanism in our scheme for controlling the access of the users.

In 2007, Green and Ateniese [9] published an identity-based proxy re-encryption. The assumptions used in the security proofs are hard to compute because of decisional Diffie Hellman assumptions.

A proxy re-encryption scheme using blockchain technology is proposed in [10] for sharing data between IoT devices. Before starting the system, every user has to register in the blockchain. The registration detail is stored in smart contract. The main elements of the scheme are sensor, user, cloud and blockchain. The encrypted data is uploaded in the cloud. Later, a proxy re-encrypts the encrypted message which is then sent to the receiver who has made the query.

Mostly all proxy re-encryption schemes use bilinear pairing [11]. But there exists some schemes which do not require bilinear pairings.

The paper [12] is fully based on directionality, that is, distinguishes the difference between unidirectional and bidirectional schemes. Unidirectional proxy re-encryption is made to resist chosen ciphertext attacks and collision attacks. The paper [12] also briefly explains all types of proxy re-encryption schemes. Type-based proxy re-encryption and threshold-based proxy re-encryption are bidirectional schemes. Identity-based, key private, attribute-based, conditional and time-based proxy re-encryptions are unidirectional schemes.

In 2012, Madaan and Agrawal [13] suggested for implementing proxy re-encryption scheme by using Linux platform. Infrastructure consists of four nodes (main server nodes, storage nodes, various data nodes and administrative nodes). Each node has three scripting files written in PHP language. They also mention the disadvantage of using Hadoop technology for proxy re-encryption scheme. So, the Linux-based platform reduces the computation part which made easy to get the output.

3 Proposed System

In the proposed re-encryption scheme, the re-encryption is done by the cloud storage, which is the master authority of the network. This cloud storage is further subdivided as public cloud and private cloud in a similar manner in [3, 14]. The public cloud used to interact with users and performs the re-encryption. All the significant data will be stored in the private cloud and communicate only with the public cloud. The sender, Alice, sends encrypted plaintext to the public cloud. Public cloud will decrypt and re-encrypt and store private cloud. We use the scheme [3] for the access control of users. Before getting the private key for re-encryption, authentication and verification process is done. When the receiver, Bob, make a query for getting the access of some information kept in the cloud storage, the public cloud communicates with the private cloud through an imaginary channel. This is for verifying whether the user is having permission to access the requested data. If Bob is having the access to the data, it sends the re-encrypted data to Bob. Bob decrypts the data. We use a similar encryption process used by Okamoto and Takashima [2].

3.1 The Proposed Proxy Re-Encryption Scheme

- **Set-up**

V : three-dimensional distortion eigenvector space.

$U_1 = Q_1, Q_2, Q_3$: distortion eigenvector basis.

$V_1 = P_1, P_2, P_3$ basis generated form U_1 using the non-singular matrix X of order 3.

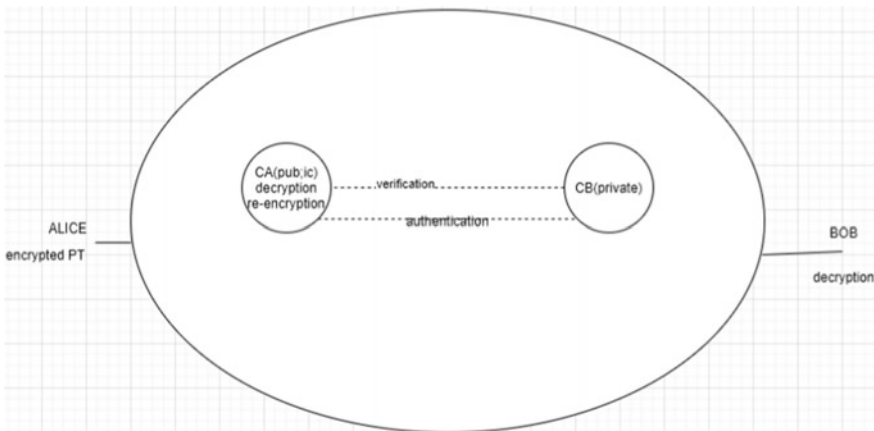


Fig. 1 System architecture

- Key generation
 - Each party runs the set-up algorithm and generates the following:
 - The public keys of Alice are (A_1, A_2, A_3) , and secret key is the matrix X_A .
 - The public keys of Bob are (B_1, B_2, B_3) , and secret key is the matrix X_B .
 - The public keys of server are (P_1, P_2, P_3) , and secret key is the matrix X_S .
- Re-encryption Key generation
 - This algorithm is run by server. The server uses the algorithms **Keygen** and **ManRole** given.
 - Nidhin et al. [3] generate roles and permissions to each user.
 - Let r_A be the secret information of sender (say, Alice) and r_B be the secret information of receiver (say, Bob).
 - The server sends this information to each user initially, through a secure channel. (This is a one-time process and not required in each communication.)
- Encryption
 - This is run by the sender. The sender (Alice) embeds the secret m in $\langle B_1 \rangle$, say tB_1 as the first coordinate, computes $C = tr_A B_1 + t_2 r_A B_2 + t_3 r_A B_3$, where (t, t_2, t_3) are random, computes $C^{01} = t_1 r_A A_1 + t_2 r_A A_2 + t_3 r_A A_3$, $C^{02} = t_1 r_A P_1 + t_2 r_A P_2 + t_3 r_A P_3$, $C^{03} = t_1 r_A P_1 + t_2 r_A P_2 + t_3 r_A P_3$ and $T^0 = t_1 r_A B_1 + t_2 r_A B_2 + t_3 r_A B_3$, with a random t_1 and also computes $C^0 = t_1 r_A P_1 + t_2 r_A P_2 + t_3 r_A P_3$.
 - Sends $(C, C^0, C^{01}, C^{02}, C^{03}, T^0)$ to the cloud server.
- Decryption
 - This algorithm is run by the server. Server computes $VDeco(C^0, \langle P_j \rangle, X_s, \langle P_1, P_2, P_3 \rangle) = t_j r_A P_j$, for $j = 1, 2, 3$.
 - Computes $C^{0j} - \sum_{i \neq j} t_i r_A P_i = C_1^{0j}$.
 - Server verifies

$$e(P_j, (r_A^{-1})C_1^{0j}) = (e(r_A^{-1})(t_j r_A P_j), A_j) \quad \text{for } j = 1, 2, 3. \quad (1)$$
 - The server also verifies

$$e(P_j, T^0 - (t_2 r_A P_2 + t_3 r_A P_3)) = e(t_j r_A P_j, B_1) \quad \text{for some } j = 2, 3. \quad (2)$$
 - If these tests are satisfied, then server performs the decryption $(r_A)^{-1}C$.
- Re-encryption
 - This algorithm is run by server. Server computes $C^* = r_B (r_A^{-1}C)$ and $T^1 = T_1 r_B B_1 + T_2 r_B B_2 + T_3 r_B B_3$.
 - Computes $C^{11} = T_1 r_B A_1 + T_2 r_B A_2 + T_3 r_B A_3$, $C^{12} = T_1 r_B B_1 + T_2 r_B A_2 + T_3 r_B B_3$, $C^{13} = T_1 r_B B_1 + T_2 r_B B_2 + T_3 r_B A_3$, T_1, T_2 and T_3 are random.
 - Also, computes $C^1 = T_1 r_B P_1 + T_2 r_B B_2 + T_3 r_B B_3$.

- Sends $(C^*, C^1, C^{11}, C^{12}, C^{13}, T^1)$ to Bob.
- Re-decryption
 - This algorithm is run by receiver (Bob).
 - Compute $(r_B^{-1})C^*$ and $VDeco(C^*, \langle B_1 \rangle, X_B, \langle B_1, B_2, B_3 \rangle) = tB_1$.
 - The first coordinate is the required secret to be shared.
 - Receiver computes $VDeco(T^1, \langle B_j \rangle, X_s, \langle B_1, B_2, B_3 \rangle) = T_j r_A B_j$, for $j = 1, 2, 3$.
 - Computes $C^{1j} - \sum_{i \neq j} T_i r_B B_i = C_1^{1j}$.
 - Receiver verifies

$$e(B_j, (r_B^{-1})C_1^{1j}) = (e(r_B^{-1})(T_j r_B B_j), A_j) \quad \text{for } j = 1, 2, 3 \quad (3)$$

- The receiver also verifies

$$e(B_j, C^1 - (T_2 r_B B_2 + T_3 r_B B_3)) = e(T_j r_B B_j, P_1) \quad (4)$$

for some $j = 1, 2, 3$.

4 Conclusion and Future Work

The present world of digitization, communications through Internet, involves high amount of risk. To reduce the risk/cost of data transfer and to enhance the speed of communication, it is suggested to store the data in a cloud storage and provide access only to appropriate users. The trust on this storage and their resistance capacity towards attacks is significant in such scenarios. Hence, it is advisable to store this data in the encrypted form. Since a stronger encryption is required, involvement of third party is required in many situations. But this makes the amount of trust invested on the third party significant, and it is preferable that the data is unknown to this trusted third party. As a solution, we propose a proxy re-encryption scheme. The security of our scheme is based on computational in-feasibility of vector decomposition problem. The scheme uses low bandwidth as well as communication cost. Though the computational capacity of the receiver seems to be high, it is of polynomial time and can be compromised while comparing with the security. In our scheme, the amount of trust required in third party is low compared to other proxy re-encryption schemes.

References

1. Yoshida M (2003) Inseparable multiplex transmission using the pairing on elliptic curves and its application to watermarking. In: Proc. Fifth conference on algebraic geometry, number

- theory, coding theory and cryptography. University of Tokyo
2. Okamoto T, Takashima K (2008) Homomorphic encryption and signatures from vector decomposition. In: International conference on pairing-based cryptography. Springer, pp 57–74
 3. Nidhin D, Praveen I, Praveen K (2016) Role-based access control for encrypted data using vector decomposition. In: Proceedings of the international conference on soft computing systems. Springer, pp 123–131
 4. Yoshida M (2003) Vector decomposition problem and the trapdoor inseparable multiplex transmission scheme based the problem. In: The 2003 symposium on cryptography and information security SCIS'2003
 5. Namboodiri KS, Praveen I (2019) An efficient batch verification scheme for secure vehicular communication using bilinear pairings. In: International conference on soft computing and signal processing. Springer, pp 711–720
 6. Praveen I, Rajeev K, Sethumadhavan M (2016) An authenticated key agreement scheme using vector decomposition. *Def Sci J* 66(6):594
 7. Kumar M, Praveen I (2015) A fully simulatable oblivious transfer scheme using vector decomposition. In: Intelligent computing, communication and devices. Springer, pp 131–137
 8. Gao C-Z, Cheng Q, Li X, Xia S-B (2019) Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network. *Clust Comput* 22(1):1655–1663
 9. Green M, Ateniese G (2007) Identity-based proxy re-encryption. In: International conference on applied cryptography and network security. Springer, pp 288–306
 10. Manzoor A, Liyanage M, Braeke A, Kanhere SS, Ylianttila M (2019) Blockchain based proxy reencryption scheme for secure iot data sharing. In: 2019 IEEE international conference on blockchain and cryptocurrency (ICBC). IEEE, pp 99–103
 11. Shao J, Cao Z (2009) Cca-secure proxy re-encryption without pairings. In: International workshop on public key cryptography. Springer, pp 357–376
 12. Khurshid A, Khan FG, Khan AN (2016) A comparison of proxy re-encryption schemes-a survey. *Int J Comput Sci Inf Secur* 14(6):392
 13. Madaan S, Agrawal R (2012) Implementation of identity based distributed cloud storage encryption scheme using php and c languages on linux platform. In: 2012 2nd IEEE international conference on parallel, distributed and grid computing. IEEE, pp 268–271
 14. Sandhu RS, Coyne EJ, Feinstein HL, Youman CE (1996) Role-based access control models. *Computer* 29(2):38–47

Correction to: A New Kaiser-Bessel Constant Modulus Technique for Smart Antenna Beamforming



K. S. Shashidhara, Veerendra Dakulagi, Jasmineeth Kaur, Kim Ho Yeap, Mandeep Singh, and Ratneshwar Kumar Ratnesh

Correction to:
Chapter “A New Kaiser-Bessel Constant Modulus Technique for Smart Antenna Beamforming” in:
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_56

The Chapter 56 was inadvertently published with the following belated correction. In the affiliation of co-author “Ratneshwar Kumar Ratnesh”, the department has been changed from “Department of Applied Physics” to “Department of Electronics and Communication Engineering”. The Correction chapter has been updated with this change.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-16-1342-5_56

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
N. R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, Lecture Notes in Electrical Engineering 790,
https://doi.org/10.1007/978-981-16-1342-5_83

C1