Lecture Notes in Networks and Systems 356

Yu-Chen Hu
Shailesh Tiwari
Munesh C. Trivedi
K. K. Mishra *Editors*

Ambient Communications and Computer Systems

Proceedings of RACCCS 2021

Springer

Lecture Notes in Networks and Systems

Volume 356

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

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

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

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

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

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

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

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

Yu-Chen Hu · Shailesh Tiwari · Munesh C. Trivedi · K. K. Mishra
Editors

Ambient Communications and Computer Systems

Proceedings of RACCCS 2021

Springer

Editors
Yu-Chen Hu
Department of Computer Science
and Information Management
Providence University
Taichung, Taiwan

Munesh C. Trivedi
National Institute of Technology Agartala
Tripura, India

Shailesh Tiwari
Krishna Engineering College
Ghaziabad, India

K. K. Mishra
Computer Science and Engineering
Department
MNNIT
Allahabad, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-16-7951-3 ISBN 978-981-16-7952-0 (eBook)
https://doi.org/10.1007/978-981-16-7952-0

Preface

The RACCCS 2021 is a major multidisciplinary conference organized to provide a forum for researchers, educators, engineers, and government officials involved in the general areas of communication, computational sciences, and technology to disseminate their latest research results and exchange views on the future research directions of these fields, to exchange computer science and integrate its practice and application of the academic ideas, to improve the academic depth of computer science and its application, and to provide an international communication platform for educational technology and scientific research for the universities and engineering field experts and professionals.

Nowadays, globalization of academic and applied research is growing with speedy pace. Computer, communication, and computational sciences are the heating areas with lot of thrust. Keeping this ideology in preference, the fourth version of the International Conference on Recent Advancement in Computer, Communication, and Computational Sciences (RACCCS 2021) has been organized at **Aryabhatta College of Engineering and Research Center, Ajmer, India, during August 20–21, 2021**.

Ajmer is situated in the heart of India, just over 130 km southwest of Jaipur, a burgeoning town on the shore of the Ana Sagar Lake, flanked by barren hills. Ajmer has historical strategic importance and was sacked by Mohammed Gauri on one of his periodic forays from Afghanistan. Later, it became a favorite residence of the mighty Mughals. The city was handed over to the British in 1818, becoming one of the few places in Rajasthan controlled directly by the British rather than being part of a princely state. The British chose Ajmer as the site for Mayo College, a prestigious school opened in 1875 exclusively for the Indian Princes, but today open to all those who can afford the fees. Ajmer is a perfect place that can be symbolized for demonstration of Indian culture, ethics, and display of perfect blend of wide plethora of diverse religion, community, culture, linguistics, etc., all coexisting and flourishing in peace and harmony. This city is known for the famous Dargah Sharif, Pushkar Lake, Brahma Temple, and many more evidences of history.

This is the fifth time Aryabhatta College of Engineering and Research Center, Ajmer, India, is organizing international conference based on the theme of computer,

communication, and computational sciences, with a foreseen objective of enhancing the research activities at a large scale. Technical program committee and advisory board of RACCCS include eminent academicians, researchers, and practitioners from abroad as well as from all over the nation.

RACCCS 2021 received around 230+ submissions from 488 authors of five different countries such as Vietnam, Malaysia, China, Bangladesh, and Oman. Each submission has been gone through the plagiarism check. On the basis of plagiarism report, each submission was rigorously reviewed by atleast two reviewers with an average of 2.90 per reviewer. Even some submissions have more than two reviews. On the basis of these reviews, 58 high-quality papers were selected for publication in this proceedings volume, with an acceptance rate of 25.2%.

We are thankful to the speakers, delegates, and the authors for their participation and their interest in RACCCS as a platform to share their ideas and innovations. We are also thankful to Prof. Dr. Janusz Kacprzyk, Series Editor, LNNS, Springer, and Mr. Aninda Bose, Senior Editor, Springer Nature, for providing continuous guidance and support. Also, we extend our heartfelt gratitude to the reviewers and technical program committee members for showing their concern and efforts in the review process. We are indeed thankful to everyone directly or indirectly associated with the conference organizing team leading it toward the success.

Although utmost care has been taken in compilation and editing, however, a few errors may still occur. We request the participants to bear with such errors and lapses (if any). We wish you all the best.

Ajmer, India

Editors
Yu-Chen Hu
Shailesh Tiwari
Munesh C. Trivedi
K. K. Mishra

Contents

Editors and Contributors

About the Editors

Prof. Yu-Chen Hu received his Ph.D. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan in 1999. Currently, Dr. Hu is a professor in the Department of Computer Science and Information Management, Providence University, Sha-Lu, Taiwan. He is a senior member of IEEE. He is also a member of Computer Vision, Graphics, and Image Processing (CVGIP), Chinese Cryptology and Information Security Association (CCISA), Computer Science and Information Management (CSIM), and Phi Tau Phi Society of the Republic of China. He joints the editorial boards of several other journals. His research interests include digital forensics, information hiding, image and signal processing, data compression, information security, and data engineering.

Prof. Shailesh Tiwari Ph.D., currently works as a director, Krishna Engineering College, and a professor in Computer Science and Engineering Department, Krishna Engineering College, Ghaziabad, India. He is an alumnus of Motilal Nehru National Institute of Technology Allahabad, India. His primary areas of research are software testing, implementation of optimization algorithms, and machine learning techniques in various engineering problems. He has published more than 80 publications in International Journals and in Proceedings of International Conferences of repute. He has edited special issues of several Scopus, SCI, and E-SCI-indexed journals. He has also edited several books published by Springer. He has published four Indian patents as IPRs. He has organized several international conferences under the banner of IEEE, ACM, and Springer. He is a senior member of IEEE, member of IEEE Computer Society, fellow of Institution of Engineers (FIE).

Dr. Munesh C. Trivedi currently working as an associate professor in the Department of CSE, NIT Agartala, Tripura, India. He has published 20 text books and 80 research publications in different International Journals and Proceedings of International Conferences of repute. He has received Young Scientist and numerous awards from different national as well international forum. He has organized several international conferences technically sponsored by IEEE, ACM, and Springer. He is on the review panel of IEEE Computer Society, International Journal of Network Security, Pattern Recognition Letter, and Computer and Education (Elsevier's Journal). He is an executive committee member of IEEE India Council and IEEE Asia Pacific Region 10.

Dr. K. K. Mishra is currently working as an assistant professor, Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, India. His primary area of research includes evolutionary algorithms, optimization techniques and design, and analysis of algorithms. He has also published more than 50 publications in International Journals and in Proceedings of Internal Conferences of repute. He is serving as a program committee member of several conferences and also edited Scopus and SCI-indexed journals. He is also a member of reviewer board of Applied Intelligence Journal, Springer.

Contributors

Yuva Teja Aagollu Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India

R. Abishek Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Abin T. Abraham Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

Sanskriti Agarwal Department of Computer Science & Engineering, ABES Engineering College, Uttar Pradesh, Ghaziabad, India

Vasu Agarwal Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

Abhineet Anand Institute of Engineering and Technology, Chitkara University, Punjab, India

Samanthapudi Anil Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

H. B. Anita Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

K. S. Archana Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

S. Arnold Jerome Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

M. Arulprakash Department of CSE, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India

R. H. Aswathy Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

G. Balamurugan Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry University, Kalapet, Pondicherry, India

Samir K. Bandyopadhyay Lincoln University College, Kota Bharu, Kelantan, Malaysia

Nipun Bansal Department of Computer Science Engineering, Delhi Technological University (Formerly Known As DCE), New Delhi, India

T. Baskar Research Scholar, Department of Computer Science and Applications, SCSVMV University, Kanchipuram, India

Swayami Bera Computer Science Engineering, SRMIST, Chennai, Kattankulathur, India

A. Rohan Bernard Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

SV. Shri Bharathi Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

R. Bharathy Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

Anand A. Bhaskar Sir Padampat Singhania University, Bhatewar, Rajasthan, India

Amiya Bhaumik Lincoln University College, Kota Bharu, Kelantan, Malaysia

Suresh Reddy Bhumireddy Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Payal Bose Lincoln University College, Kota Bharu, Kelantan, Malaysia

Upasana Chakraborty CSE Department, SRM Institute of Science and Technology, Chennai, India

Shipra Choudhary Department of Computer Science Engineering, Delhi Technological University (Formerly Known As DCE), New Delhi, India

Sangram Sanjayrao Dandge School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India

G. Dhanush Eswar Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Ashwani Kumar Dhingra Department of Mechanical Engineering, University Institute of Engineering & Technology Maharshi Dayanand University, Rohtak, Haryana, India

Sunita Dhingra Department of Computer Science & Engineering, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, Haryana, India

Nguyen Thuy Dung Thai Nguyen University, University of Information and Communications Technology, Thai Nguyen, Vietnam

P. Mohamed Fayaz Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

Atul Garg Institute of Engineering and Technology, Chitkara University, Punjab, India

Shruti Garg CSE Department, SRM Institute of Science and Technology, Chennai, India

Vinay Gautam School of Computing, DIT, Dehradun, India

M. Gayathri SRM Institute of Science and Technology, Chennai, India

Angelina Geetha Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Biraj Ghosh Computer Science Engineering, SRMIST, Chennai, Kattankulathur, India

Pinaki Ghosh Institute of Engineering and Technology, Chitkara University, Punjab, India

Pradeep K. Goyal Delhi Technological University, Delhi, India

Vishal Goyal GLA University, Mathura, India

Satinder Bal Gupta Department of Computer Science, Indira Gandhi University, Rewari, Haryana, India

Sylwia Gwoździewicz Jacob of Paradies University, Gorzow Wielkopolski, Poland

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

Pon Harshavardhanan School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India

Dinh Tran Ngoc Hien Ho Chi Minh University of Technology, Thành ph´ô H`ô Chí Minh, Vietnam

Le Thi Thanh Huong Dai Nam University, Hanoi, Vietnam

Dinh Tran Ngoc Huy Banking University HCMC, Ho Chi Minh City Vietnam—International University of Japan, Niigata, Japan

Anjali Jain Jaypee Institute Of Information Technology, Noida, India;
Kiet Group of Institutions, Ghaziabad, India

J. Jayabharathy Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry University, Kalapet, Pondicherry, India

R. Jebakumar Department of CSE, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, India

Libin M. Joseph Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

Ranjana Joshi GLA University, Mathura, India

S. Kalaivani Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

M. Kannan Department of Computer Science and Applications, SCSVMV University, Kanchipuram, India

G. G. Krishna Kanth Clians Labs Pvt. Ltd., Hyderabad, India

Rahul Karmakar Department of Computer Science, The University of Burdwan, Burdwan, India

Vasantharaj Karunakaran Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Kelambakkam, India

K. Aldrin Karunharan International Maritime College Oman, Karunya University, Coimbatore, India

P. Keerthivasan Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

Nishad Kharkar Department of Computer, Saraswati College of Engineerirng, Kharghar, India

K. Mohana Krishnan Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

G. Ranjith Kumar Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Nagresh Kumar Deparment of Computer Science, Banasthali Vidyapith, Vanasthali, Rajasthan, India

Sanjeev Kumar KIET Group of Institutions, Delhi-NCR, Ghaziabad, India

Vikram Kumar Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Vinod Kumar Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

K. Sai Vijaya Lakshmi Computer Science and Engineering, GITAM (deemed to be University), Hyderabad, India

M. Latha Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

Trung-Hieu Le Dai Nam University, Hanoi, Vietnam

Umesh Kumar Lilhore Institute of Engineering and Technology, Chitkara University, Punjab, India

K. Little Flower Computer Science and Engineering, GITAM (deemed to be University), Hyderabad, India

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

Prashant Mani SRM Institute of Science and Technology, Modingar, UP, India

T. Manikandan Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

Jakkamsetti Bharath Manoj School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India

X. Anitha Mary International Maritime College Oman, Karunya University, Coimbatore, India

A. Menaga School of Management Studies, Vels Institute of Science & Technology & Advanced Studies (VISTAS), Chennai, India

Anurag Mishra Department of Computer Science & Engineering, ABES Engineering College, Uttar Pradesh, Ghaziabad, India;
Dr. APJ Abdul Kalam University, Lucknow, India;
GLA University Mathura, Mathura, India

K. K. Mishra Department of Computer Science, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, UP, India

Mohit Misra SRM Institute of Science and Technology, Modingar, UP, India

G. Mohana Priya Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

Sarada Nidamanuri Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Nishant Nijhawan Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

Anand Bhushan Pandey Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

Anurag Pandey Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Himanshu Pandey Faculty of Engineering & Technology, University of Lucknow, Lucknow, U.P., India

Manjusha Pandey School of Computer Engineering, KIIT-DU, Bhubaneswar, India

Suhasani Parvartikar Department of Computer, Saraswati College of Engineerirng, Kharghar, India

Utkarsh Patiyal Department of Computer Science Engineering, Delhi Technological University (Formerly Known As DCE), New Delhi, India

Sandeep Poddar Lincoln University College, Kota Bharu, Kelantan, Malaysia

R. S. Ponmagal Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India

K. Pradeep Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

P. Pranideep Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India

Patel Priyanka Department of Computer, Saraswati College of Engineerirng, Kharghar, India

Shaik Rafi Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India

C. Rajesh Babu CSE Department, SRM Institute of Science and Technology, Chennai, India

D. Rajesh Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India

S. Ramamoorthy Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India

R. Ramya Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India

P. Ranjana School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India

Siddharth Rautaray School of Computer Engineering, KIIT-DU, Bhubaneswar, India

Naga Chakradhar Rao Rayasam Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India

Janga Pavan Krishna Reddy School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India

M. Devi Keerthi Reddy Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

P. Renupriya Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

S. Revathy Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Anil Kumar Rodda Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

M. Ruby Evangelin School of Management Studies, Vels Institute of Science Technology and Advanced Studies, Chennai, India

Samadrita Saha Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India

K. Sakthivel Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Jasminder Sandhu Institute of Engineering and Technology, Chitkara University, Punjab, India

S. Santhi Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, TamilNadu, India

R. Saravanan Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

D. Sasikiran Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Kalicheti Sathvika School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India

S. Sathya Priya Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Aditya Sharma Department of Computer Science & Engineering, ABES Engineering College, Uttar Pradesh, Ghaziabad, India

Ashish Sharma GLA University Mathura, Mathura, India

Rohan Sharma School of Computer Engineering, KIIT-DU, Bhubaneswar, India

Sanjay Kumar Sharma Deparment of Computer Science, Banasthali Vidyapith, Vanasthali, Rajasthan, India

Savita Sharma Government Polytechnic College, Ajmer, Rajasthan, India

S. V. Shri Bharathi Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India

Sarita Simaiya Institute of Engineering and Technology, Chitkara University, Punjab, India

Aditya Singh Department of Computer, Saraswati College of Engineerirng, Kharghar, India

Akanksha Singh Sir Padampat Singhania University, Bhatewar, Rajasthan, India

Arun Kumar Singh Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

Binit Kumar Singh School of Computer Engineering, KIIT-DU, Bhubaneswar, India

Deepak Kumar Singh KEC, Ghaziabad, India

Jagendra Singh Bennett University, Greater Noida, India

Rupali Singh Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

Alka Singhal Jaypee Institute Of Information Technology, Noida, India

K. Sivakumar Department of Mathematics, Saveetha School of Engineering, SIMATS, Thandalam, Chennai, Tamil Nadu, India

Samarth Srivastava SRM Institute of Science and Technology, Chennai, India

B. S. Sudhan Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

K. Sujith Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Sukritin Department of Computer Science & Engineering, ABES Engineering College, Uttar Pradesh, Ghaziabad, India

Smrithy R. Sunil Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

P. Suresh Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

Pavan Teja Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

R. Yamini Teja Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Sonal Kumari Thakur Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India

Nguyen Thi Phuong Thanh Thai Nguyen University, University of Information and Communications Technology, Thai Nguyen, Vietnam

S. Thavasi Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, TamilNadu, India

G. Thilak Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

E. J. Thomson Fredrik Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India

Shailesh Tiwari ABES Engineering College, Ghaziabad, India;
Krishna Engineering College, Ghaziabad, UP, India

Ashish Tripathi Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

Munesh C. Trivedi NIT, Agartala, India

Naresh Kumar Trivedi Institute of Engineering and Technology, Chitkara University, Punjab, India

Nguyen Dinh Trung National Economics University (NEU), Hanoi, Vietnam

Vaibhav Tyagi Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

N. Umakanth Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, TamilNadu, India

Sai Laasya Vabilisetty Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India

Hoang Van Thuc Thai Nguyen University of Information and Communication Technology, Thái Nguyên, Vietnam

D. Vanusha Department of CSE, SRMIST, Chennai, Kattankulathur, India

Oshi Varma SRM Institute of Science and Technology, Chennai, India

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

D. Venkatesan Department of Mathematics, Saveetha School of Engineering, SIMATS, Thandalam, Chennai, Tamil Nadu, India

Sahil Verma Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India

V. Vetrivel Department of Mathematics, Saveetha School of Engineering, SIMATS, Thandalam, Chennai, Tamil Nadu, India

Aruna Yadav Bennett University, Greater Noida, India

A. Mohammed Shahid Zafer Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai, India

A Review of Physical Unclonable Functions (PUFs) and Its Applications in IoT Environment

Aruna Yadav, Sanjeev Kumar, and Jagendra Singh

Abstract This paper describes various studies about physical unclonable functions, and it inspires use of physical unclonable functions over conventional security mechanisms and compares them in many aspects. It categorizes physical unclonable functions as strong physical unclonable functions and weak physical unclonable functions. For any communication in a network, authentication scheme for nodes, server, router, and network gateway is presented and procedure of communication is explained and presented through architecture. This paper explained problems faced by smart devices due to attacks on security. Finally, this paper reviews various emerging concepts of physical unclonable functions and its advancement.

Keywords Physical unclonable functions · Cryptographic hardware · Hash function · Authentication · FPGA (field programmable gate arrays) · Randomness

1 Introduction

Physical unclonable functions provide improved security in terms of low-cost authentication and in key creation applications. Physical unclonable functions are used for secret key storage without secured EEPROMS. Physical unclonable functions can do this because they do not store secret in digital memory rather; they find out some undisclosed features from physical characteristics of integrated circuit like manufacturing differences of gate delay as a physical characteristic. There are many reasons physical unclonable functions are better than standard secure digital storage. Physical

A. Yadav (✉) · J. Singh
Bennett University, Greater Noida, India
e-mail: e19soe808@bennett.edu.in

J. Singh
e-mail: jagendra.singh@bennett.edu.in

S. Kumar
KIET Group of Institutions, Delhi-NCR, Ghaziabad, India
e-mail: sanjeev.yadav@kiet.edu

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_1

unclonable functions use cryptographic hardware which are nominal in price. Physical unclonable function consumes less power and easy to fabricate. Generally, invasive attacks are possible only if attacker able to change physical characteristic, that is main reason to secure physical unclonable functions continuous active anti-tamper mechanism is not needed. Physical unclonable functions derive secrets manufacturing variability of Integrated circuits. No two integrated circuits can be same even there must be some slight difference two integrated circuits even if manufacturing and masking process is same. There are many applications that integrate physical unclonable functions into integrated circuits. Physical unclonable functions can be categorized as: strong physical unclonable functions and weak physical unclonable functions. For authentication purpose, strong physical unclonable functions are used. In storage purpose, weak physical unclonable function works efficiently. Focus of physical unclonable function is to decide c in $f(c)$, where x is parameter in which function is dependent and that parameter lies on manufacturing variability. Basic difference between strong and weak physical unclonable functions lies on domain of $f(c)$. For strong physical unclonable functions, domain is large, while for weak physical unclonable functions, domain is small.

Consider c as challenge for function; then response given by function is r, $r = f(c)$; physical unclonable functions ensures unpredictability: An adversary cannot predict its response until it recognizes intrinsic properties; robustness: Response of physical unclonable functions is steady over time; uncolorability: An adversary cannot mimic the behavior of weak physical unclonable functions on another device. Since weak physical unclonable functions have less numbers of challenge response pairs, these pairs stored secretly. Mostly weak physical unclonable functions are used for secret key derivation. Physical unclonable functions provide secured storage that is why physical unclonable function response is never exposed. So, response can be used as secret key to encrypt or decrypt data. While in strong physical unclonable functions authenticated directly without using cryptographic hardware, responses are stable to environment, attackers cannot expect response. Weak physical unclonable functions also able to provide authentication with the use of HMAC. Models of weak physical unclonable functions a.

Both weak and strong physical unclonable functions rely on analog physical properties of the fabricated circuit. Example of strong physical unclonable function architectures: Optical physical unclonable function: It was first implementation of a strong physical unclonable function.

Another example of strong physical unclonable function is Arbiter physical unclonable function: There are some limitations due to macroscopic approach of optical physical unclonable function. Silicon implementations were used for strong physical unclonable function, manufacturing variability works as challenge, and that is input for unclonable randomness. In case of Arbiter physical unclonable function, manufacturing variability in the gate delay of each multiplexer produces an edge at the latch and that latch behaves like Arbiter [1]. There are large numbers of concepts for physical unclonable function that are emerging. It is also shown that rate of new physical unclonable function is increasing. It is also observed that concept behind it is randomness. There are several physical unclonable function concepts

for electronic application, so are many options of physical unclonable function, by comparing merits and demerits best suitable option, can be chosen. Choice depends on application.

Examples of weak physical unclonable function are static RAM and ring oscillator physical unclonable function.

In low-cost authentication physical unclonable functions: A strong physical unclonable function can be replacement for secured memory and cryptographic hardware on an embedded system. Strong physical unclonable functions do not demand for anti-tamper circuit structure, any cryptographic hardware, or any additional thing. Strong physical unclonable functions demand less power, area. Conceptually a planned, steady broad evaluation is mandatory to regulate the optimal physical unclonable functions concept and explicit modification for any given purpose should outstanding for all.

2 Literature Survey

Recently, number of concepts related to physical unclonable functions increased, and concepts utilize explicit source of randomness, concerning distinctive material and skills in field of physical unclonable functions and using new materials and technologies in the field of physical unclonable function [2].

Hammouri et al. [3] presented tamper proof lightweight challenge response authentication scheme that varies according to noisy-level physical unclonable functions. Compared to earlier projected structure, this scheme does not require cryptographic hash function which is more expensive for ultra-low power applications. The security against passive attacks is claimed considering that no polynomial time algorithm presents for enquiring threshold of almost half cases.

Reza et al. [4] described application scenario where all previous physical unclonable function-based authentication schemes failed and then used converse physical unclonable function-based authentication scheme.

Wang et al. [5] proposed strong physical unclonable functions that are secured against machine learning attacks on conventional and quantum computers. Lattice physical unclonable functions built through weak physical unclonable functions. Lattice physical unclonable functions are lightweight, digital uses concept of pseudo-random number generator. Delvaux [6] explained physical unclonable function which is a circuit whose input–output conduct is dependent on manufacturing random variations and investigated significance of 21 physical unclonable function-based protocols. Work out proficient predictive model of Arbiter physical unclonable functions.

Calhoun [7] enlightened physical unclonable function-based e-cash protocol and analyzed arithmetical features of the produced authentication bit strings and specified that physical unclonable function cash is robust, scalable and can be implemented in commercially hardware.

Mukhopadhyay et al. [8] illustrated Internet of Things (IoT) which is network of large number of exclusive distinguishable intercommunicating smart devices. Lightweight authentication protocol for these smart devices should be protected against physical attacks. Explored design tasks, operating principles and defense approaches.

Kong and Koushanfar [9] illustrated strong physical unclonable function which is a circuit structure that abstracts large number of unique chip signs and described information for process variation (process variation produces inherent randomness in silicon structure and disturbs threshold voltage consequential power consumption and delay around silicon chip), circuit aging technique (circuit aging is process in which efficiency of circuit is degraded by its usage) and delay model.

Yin and Qu [10] explained silicon-based physical unclonable function that operates the differences through silicon manufacturing process to abstract information that is exclusive for each chip. Many research shows that physical unclonable function can be used in security of applications.

Bin chen [11] described problem of secret key generation over noisy and biased physical unclonable function to resolve this problem. Polar code-based syndrome structure is applied to break the concept of polarization to cooperation of secret key's randomness and decodability so it will be able to minimize the effect of bias on secret key theft.

Guo et al. [12] explained that physical unclonable function provides challenge–response sets for authentication of devices, but many conventional strong physical unclonable function designs represented by the Arbiter physical unclonable function are hard to implement on FPGA and then suggested new lightweight strong physical unclonable function that can enthusiastically reconstruct while maintaining high entropy and large CRP space.

Rostami [13] illustrated and analyzed two lightweight and secured protocols based on substring matching of physical unclonable function response strings to accomplish authentication and session key exchange; simultaneously, this protocol ensures robustness against noise in the physical unclonable function response string.

Chatterjee et al. [14] developed a secure physical unclonable function-based authentication mechanism and identity without certificate identity protocol. Asymmetric behavior of protocol resolved problems are due to challenge–response pair-based physical unclonable function authentication scheme; it is advantageous over security susceptibility of low hardware.

Kumar et al. [15] explored that ensuring data security in smart devices from external attacks is a tedious job. Generally, traditional cryptographic techniques preserved keys in nonvolatile memory that is susceptible to physical attacks. Physical unclonable functions have capability to save smart devices from physical attacks as keys are not stored in nonvolatile memory and tough to regenerate. In new generation of electronic devices, usage of semiconductor material, for example, carbon nanotubes or 2D materials are preferred reason, is better electrical, optical, mechanical, and thermal properties.

Gu et al. [16] proposed TCR physical unclonable function design dependent on tristate inverter matrix. This design has feature of ultra-lightweight and reconfigurable as compared to RO physical unclonable function design. MOS simulation and FPGA implementation showed and verified the better, consistent, and unique behavior of TCR physical unclonable functions.

Venkata et al. [17] illustrated issues regarding Internet of Medical Things. Adding extra security features to existing electronic devices must not enhance power consumption and diminish battery life. Proposed design of physical unclonable function was advantageous for authentication techniques that is power optimized hybrid oscillator Arbiter physical unclonable function. This physical unclonable function design was verified through 32 nm FinFET and Dpoingless junctionless FETs.

Das et al. [18] discussed new physically secure lightweight anonymous user authentication protocol for Internet of Things applying physical unclonable function; this work is carried out through the tough analysis of ROR model and verified formal security through AVISPA tool and informal security. Real-world implementation of this scheme is analyzed using NS3 simulation tool.

Byun [19] presented novel authentication technique for two parties containing two authentication factors (1. Own physical unclonable function embedded device and 2. long-term secret.) willing to authenticate each other through mutually decided session key over a distributed network. A novel concept DEVICE ORACLE is discovered and presented distinct physical unclonable function embedded AKE ensuring verified security.

Barbareschi et al. [20] extended existing protocol PHEMAP which ensures mutual hardware authentication based on physical unclonable function for one-to-many scenario to CE-dependent Internet of Things scenario, mainly emphasized on complete mutual authentication scheme dependent on PHEMAP whose significance to achieve requirements and restrictions of three-tier Internet of Things (Table 1).

3 Classification of Physical Unclonable Functions

3.1 Classification Dependent on Used Physical Employment

Physical unclonable functions used till now can be roughly categorized into four major groups based on the technology role in physical engagement: optical physical unclonable function, silicon physical unclonable function, coating physical unclonable function, and acoustic physical unclonable function (PUF).

Optical physical unclonable function (OPUF):
It first proposed physical unclonable function, though initially projected as the physical characterization of a particular one-way function used for cryptography. Principal constituent of the optical physical unclonable function shows clear token using arbitrarily fixed smattering elements. After radiation through the laser, a multifaceted image using sunny and shadowy adverts ascends, known as “speckle pattern”. A

Table 1 Physical unclonable functions' evolution

Advancement in physical unclonable functions
Integrated Circuit identification using device mismatch in year 2000
Physical One-way function in year 2001
Physical Random Function in year 2002
Arbiter physical unclonable function and Feed Forward Arbiter in 2004
Coating physical unclonable function in 2006
Latch physical unclonable function, XOR-Arbiter physical unclonable function, Ring Oscillator physical unclonable function and Static RAM physical unclonable function. in 2007
Butterfly physical unclonable function, D Flip-Flop physical unclonable function, Tristate Buffer physical unclonable function, Lightweight Secure physical unclonable functions in 2008
Power Grid physical unclonable function in 2009
Glitch physical unclonable function, SHIC (Super High Information Content) physical unclonable functions in 2010
Mecca physical unclonable function, Bistable Ring physical unclonable function, Ultra-low Power Current-Based physical unclonable function, Current Starved Inverter Chain physical unclonable function and Logically Reconfigurable physical unclonable function in 2011
Converse Physical function in 2012
Light weight authentication protocol 2014
Robustness against noise in the physical unclonable function 2015
TCR physical unclonable function in 2017
optimized hybrid oscillator arbiter physical unclonable function in 2018
Distinct physical unclonable function embedded AKE in 2019

Gabor filter fits in this scenario as a worthy feature extractor specially for these types of configurations; then filter production is the rejoinder of such type of optical physical unclonable function, though the physical factors of the laser (location, wavelength, and orientation) set up the challenge. Composite environment of the communication of laser light through sprinkling particles, the reactions are extremely arbitrary and exclusive. The great requirement of the response arranged the precise microscopic physical particulars of the optical token reasons two similarly formed tokens to display different reactions to the identical challenge, then escapes a specific token and they can be duplicated through great accuracy.

Silicon physical unclonable function:
It exploits strong CMOS manufacturing discrepancies that is the outcome provided inescapable inadequacies in recent integrated circuit manufacturing methods. Manufacturing discrepancy of factors like dopant absorptions and line breadths are obvious like differences in terms of timing performance among cases of the identical integrated circuits. Such time differences dignified for an appropriate circuit system, if required then encoded. Preferably, a silicon physical unclonable function should not involve a nonconformity from the normal CMOS treating steps and implementable with the help of EDA design flows. It is noticed that such type of physical unclonable

functions is inclined toward temperature deviations that finally reward schemes to work the system appropriately.

Coating physical unclonable function:
It is a physical unclonable function in which a defensive coating for integrated circuits is the basis of unpredictability. The dense coating substance is fixed in di-electric elements, which have arbitrary specifications regarding location, shape, and dimension. Underneath the coating layer, an arrangement of sensors for metal wires is applied to scope the capacitance for coating. Calculated values are arbitrary just cause of the arbitrariness is there in the coating, generates the responses to their challenges, all are specific to the voltage of a definite amplitude and frequency, practical to a section of the sensor array. As there are implicit randomness, di-electric constituents and these values are tough to generate two values as all sensors generate similar values. Although there are some disadvantages to use coating physical unclonable function is that it allows restricted number of challenge–response pairs. While the advantage of coating physical unclonable function is less price of their fabrication, an advantage of coating physical unclonable function is that adaptable to identifying some physical changes.

Acoustic physical unclonable function:
This physical unclonable function is constructed based on properties belongs to acoustical wave propagation. Any oscillating voltage of electrical signal is converted into mechanical vibration through a transducer. These vibrations transmitted as a sound wave in a specific solid medium and disseminates on the arbitrarily distributed inconsistencies of the solid medium. The reflections of these sound waves are dignified by another transducer which changes these vibrations again in the form of electric signal. It shows that the reflections of each token are exclusive.

3.2 Classification Based on Response Collection Mechanism

Intrinsic physical unclonable function-
Intrinsic physical unclonable function is embedded inside the hardware system, it preserves without any modification done to produce randomness, and obviously any opponent cannot read it exactly due to this feature that a reader will not disturb its output value. The measurement settings are inbuilt in the physical unclonable function and is integrated on chip. Such type of physical unclonable functions (like Silicon physical unclonable functions) are comparatively easy to build and control because these PUFs usually do not required any preprocessing and post-processing, although some error correction for noise is required, but these PUFs are prone to security susceptibilities.

Controlled-physical unclonable function-
Controlled-physical unclonable function relates to an algorithm in a very special way and that algorithm controls these types of PUFs. Any changes in the algorithm

are reflected in the controlled-physical unclonable function. This algorithm discourages the reader to directly challenge the physical unclonable function and limits the information about responses that is seen by the outside world. A controlled physical unclonable function makes enumerating physical unclonable function responses harder and stops man-in-the-middle attacks. In addition, some physical unclonable functions require external or extra logic to explicitly introduce randomness (such as the optical and coating physical unclonable function) can also be considered under this category.

4 Authentication Scheme

We discuss a scheme that defines working of authentication protocol. A network corresponding several nodes (named as node1, node2, node3, noden) acts as sources and receiver of information. Network consists of some server nodes (named as SN1, SN2,…SNi) put upon to manage the communication of each node in its range and forward the data to a router (named as router1, router2…, routerj). The router therefore forwarded the information to the cloud through and through the network gateway. The data nodes have limited computational resources, while the server nodes and advanced nodes are comparatively resourceful. It is pretended that the server nodes and the data nodes are dispersed in the way that each data node is always inside the read range of one server node. Every data node, server node and router need to be authenticated so mount a physical unclonable function instance on them. Written account that in the given architecture, there is no definite key storage available for the nodes. We depicted our protocol for data nodes and server nodes. The protocol grows through the following steps:

Firstly, enrollment phase is performed, formerly each data node is assigned to a particular server node. Through this, a CRP database created for the physical unclonable functions contained in each node in the server node corresponding to the data node. In next phase if two nodes want to communicate under same server, their server node authenticates those nodes, then generation of their public/private key pairs occurred, and then share secure key pair. In last phase, secure communication executed, the two nodes send and receive the message securely over the network using the keys, without any intervention of the server. If a node1 wants to communicate to node2, then server1 needs to authenticate both of these nodes before key generation and communication between these two nodes. Similarly, if server1 wants to communicate with server2, then router1 needs to authenticate both these servers before key generation and communication. In the same way if a router1 wants to communicate with router2, then network gateway needs to authenticate these two routers, before key generation and communication between these two routers. If a node wants to move from one server to another server, then all things will come through by server2 now onward as mentioned in Fig. 1.

In enrollment phase, server sends random challenge to the mobile data node; then node employs challenge c to its physical unclonable function and produces response

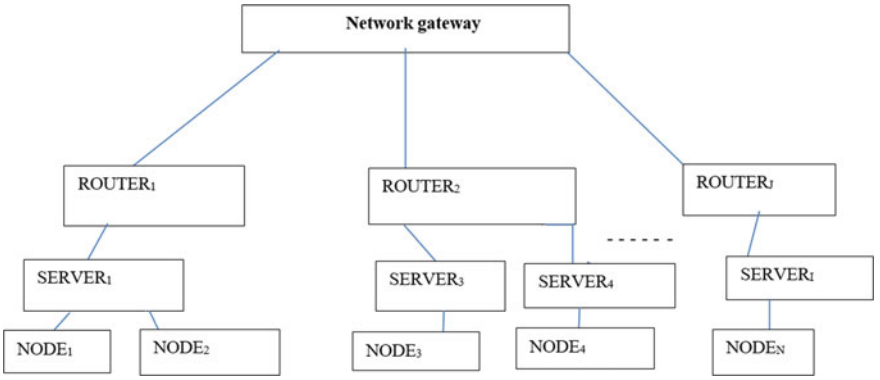


Fig. 1 Communication architecture of different types of nodes

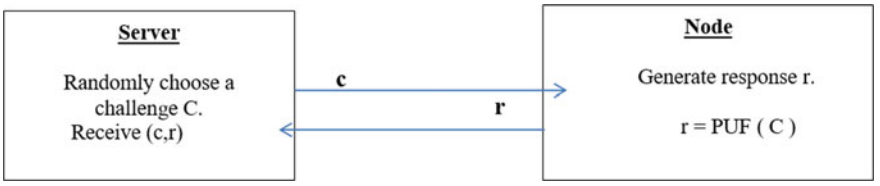


Fig. 2 Enrollment phase

r and send back to its server, then server update its database by (c, r) pair as shown in Fig. 2. These types of many challenge response pairs are stored in database according to the memory of server. This database is secured.

Second phase Authentication Phase: with the help of challenges $c1$, $c2$ for node1 and $c3$, $c4$ for node 2 and corresponding responses $r1$, $r2$ for node1 and $r3$, $r4$ for node2 applied hash function on them calculated $\Delta 1$ and $\Delta 2$ for any particular timestamp.

$$Q1 = \text{Hash function } (\text{puf1}(c1) \,||\text{puf1}(c2) \,||ts)$$
$$Q2 = \text{Hash function } (c1 \oplus c2)$$
$$\text{Private key} = t.Q1$$
$$\text{Public key} = t.Q2$$

Similarly, we will calculate for node2.

Last phase communication phase: In this phase, node1 sends message to node2. This communication is secured against CPA, CCA, repudiation attack [21].

5 Security of Physical Unclonable Function

New technologies do not bring only new facilities but also bring some weakness and unpredicted problems. We will discuss diverse physical unclonable function particularized attack.

These are some of the attacks:

Fault attacks on physical unclonable functions and fuzzy extractors:

Fault attacks initiate a suspected performance in that device when using over intense situation or compel it to function in a manner to enclose fault into it, provoking to work at immoderate temperature that will significantly change physical unclonable function random noise conduct on the far side of error correction potentiality of the circuit [18, 20, 22].

Side channel attacks on physical unclonable functions: Side channel attacks are hardware-dependent attack those wants to analyze hardware and interpret cryptographic key. As data processed in hardware are interrelated to each other, after analysis of hardware behavior attackers can go through analysis of data also. Attackers discover characteristics of section for example power consumption while analyzing secret information to be extracted [23].

3. Modeling attacks on delay-based physical unclonable functions: Goal of these types of attack is to figure out mathematics approximation for definite set of challenge response pair.

So that afterward approximation can help in forecasting of responses for unknown challenges with comparatively higher accuracy. Delay-supported physical unclonable function for example arbiter physical unclonable function and its variance is sensitized to these types of modeling attacks imputable to linear delay circuits [8].

6 Secured Physical Unclonable Functions Dependency

Physical unclonable functions have two important security parameters: unpredictability and unclonable property. Unclonable property is very much required characteristic that is not attained using traditional cryptographic methods. It can be divided into two forms again: mathematically unclonable and physically unclonable. A physical unclonable function is physically unclonable if the manufacturer cannot create a physical duplicate of the physical unclonable function with alike challenge–response pair. It is termed as existential unclonable. Generally, all silicon-based physical unclonable functions hold this property. Physical unclonable function is mathematically unclonable.

Also it is not probable to generate a mathematical approximation which exhibits the original Physical unclonable function conduct nearby previous one. There is no known silicon physical unclonable function that is mathematically unclonable. This mathematical unpredictability states that opponent cannot forecast same challenge response pair from already existing set of challenge–response pairs. Typically, *fn* (the

random function) is dignified through these security experiments which consist of an important knowledge as well as a challenge phase. In the learning phase, the opponent examines the behavior of the random function *fn* for particular input challenges × 1, × 2, …, *xn* (it can be specified by opponent). Finally, in the challenge stage, the opponent provides values in the set of (*x*, *fn* (*x*)) for some *x* × 1, × 2, …, *xn* surely. Normally, unpredictability is evaluated through of entropy or can specified as average min-entropy of physical unclonable function distribution [24].

7 Current Trend for Authentication Protocol

7.1 Ultra-Lightweight Authentication Protocol

Rather than to design an authentication protocol by using ultra-lightweight weight operations, it can be difficult when employs it from scrape. It is better to design ultra-lightweight authentication protocol using traditional challenge–response pair authentication protocol using cryptographic primitive for ultra-lightweight construction [13]. Ultra-lightweight primitives were not applicable to DES (data encryption standard it is a symmetric key algorithm), AES (advanced encryption standard is also known as Rijndael algorithm), and IDEA (International Data Encryption Algorithm is also known as improved Proposed Encryption Standard), though ultra-lightweight primitives are applicable to low-cost primitives designed for this purpose [25, 26].

7.2 Human-Dependent Authentication Protocol

This human-dependent authentication protocol needs lightweight evaluation dependent on learning parity with noise. Although human-dependent authentication protocol has many specific features, this protocol requires additional protective environment, when any third-party adversary involved in between message communication and tries to modify the message [17]. Then there were lot of works done to avoid man-in-middle attack in human-dependent authentication protocol with different versions like random-human-dependent authentication protocol, trusted-human-dependent authentication protocol, bilinear-human-dependent authentication protocol, and tree-human-dependent authentication protocol [10, 22].

8 Conclusion

We studied various research papers related to physical unclonable functions and their different models; illustrated their variations based on applications, devices, materials,

and attacks; and covered comparative analysis of physical unclonable functions with traditional techniques in many aspects.

We undergo through metrics of physical unclonable functions for any application and discussed authentication scheme of data nodes if they want to communicate each other considering all the conditions whether data nodes are under the same server, router, network gateway, or in different constraints. Current study on physical unclonable functions primarily concentrated on apprehension of the capabilities and limitations of physical unclonable function by detailed study of possible attacks on physical unclonable functions. This paper covers applications of physical unclonable functions in IoT environment as well as current trends of authentication protocol in scope of ultra-lightweight authentication protocol and human-dependent authentication protocol.

References

1. Herder C, Yu MD, Koushanfar F, Devadas S (2014) Physical unclonable functions and applications: A tutorial. Proc IEEE 102(8):1126–1141
2. McGrath T, Bagci IE, Wang ZM, Roedig U, Young RJ (2019) A PUF taxonomy. Appl Phys Rev 6(1)
3. Hammouri G, Öztürk E, Sunar B (2008) A tamper-proof and lightweight authentication scheme. Pervasive Mob Comput 4(6):807–818
4. Kocabaş Ü, Peter A, Katzenbeisser S, Sadeghi AR (2012) Converse PUF-based authentication. Lect Notes Comput Sci (including Subser. Lect Notes Artif Intell Lect Notes Bioinf) 7344:142–158, LNCS
5. Wang Y, Xi X, Orshansky M (2019) Lattice PUF: a strong physical unclonable function provably secure against machine learning attacks 2019
6. Delvaux J (2019) Machine-learning attacks on PolyPUFs, OB-PUFs, RPUFs, LHS-PUFs, and PUF-FSMs. IEEE Trans Inf Forensics Secur 14(8):2043–2058
7. Calhoun J, Minwalla C, Helmich C, Saqib F, Che W, Plusquellic J (2019) Physical Unclonable function (PUF)-based e-Cash transaction protocol (PUF-Cash). Cryptography 3(3):18
8. Mukhopadhyay D, Chakraborty RS, Nguyen PH, Sahoo DP (2015) Tutorial T7: physically unclonable function: a promising security primitive for Internet of Things, pp 14–15
9. Kong J, Koushanfar F (2014) Processor-based strong physical unclonable functions with aging-based response tuning. IEEE Trans Emerg Top Comput 2(1):16–29
10. Yin CE, Qu G (2014) Obtaining statistically random information from silicon physical unclonable functions. IEEE Trans Emerg Top Comput 2(2):96–106
11. Chen B, Willems FMJ (2019) Secret Key Generation over biased physical unclonable functions with polar codes. IEEE Internet Things J 6(1):435–445
12. Hou S, Guo Y, Li S (2019) A lightweight LFSR-based strong physical unclonable function design on FPGA. IEEE Access 7:64778–64787
13. Rostami M, Majzoobi M, Koushanfar F, Wallach DS, Devadas S (2014) Robust and reverse-engineering resilient PUF authentication and key-exchange by substring matching. IEEE Trans Emerg Top Comput 2(1):37–49
14. Chatterjee U et al (2019) Building PUF based authentication and key exchange protocol for IoT without explicit CRPs in verifier database. IEEE Trans Dependable Secur Comput 16(3):424–437
15. Kumar N, Chen J, Kar M, Sitaraman SK, Mukhopadhyay S, Kumar S (2019) Multigated carbon nanotube field effect transistors-based physically unclonable functions as security keys. IEEE Internet Things J 6(1):325–334

16. Cui Y, Gu C, Wang C, O'Neill M, Liu W (2018) Ultra-lightweight and reconfigurable tristate inverter based physical unclonable function design. IEEE Access 6:28478–28487
17. Yanambaka VP, Mohanty SP, Kougianos E, Puthal D (2019) PMsec: physical unclonable function-based robust and lightweight authentication in the internet of medical things. IEEE Trans Consum Electron 65(3):388–397
18. Banerjee S, Odelu V, Das AK, Chattopadhyay S, Rodrigues JJPC, Park Y (2019) Physically secure lightweight anonymous user authentication protocol for internet of things using physically unclonable functions. IEEE Access 7:85627–85644
19. Byun JW (2019) End-to-end authenticated key exchange based on different physical unclonable functions. IEEE Access 7:102951–102965
20. Barbareschi M, De Benedictis A, La Montagna E, Mazzeo A, Mazzocca N (2019) A PUF-based mutual authentication scheme for cloud-edges IoT systems. Futur Gener Comput Syst 101:246–261
21. Chatterjee U, Chakraborty RS, Mukhopadhyay D (2017) A PUF-based secure communication protocol for IoT. ACM Trans Embed Comput Syst 16(3)
22. Yilmaz Y, Gunn SR, Halak B (2018) Lightweight PUF-based authentication protocol for IoT devices. In: 2018 IEEE 3rd international verification and security workshop, IVSW 2018, pp 38–43
23. Merli D, Stumpf F, Sigl G (2013) Robust authentication using physically unclonable functions. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinf) 3(1):19–41
24. Oztürk E, Hammouri G, Sunar B (2008) Towards robust low cost authentication for pervasive devices. In: 2008 Sixth annual IEEE international conference on pervasive computing and communications. PerCom 2008, pp 170–178
25. Schoenmakers B (1998) Security aspects of the ecashTM payment system. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform) 1528:338–352
26. Van Herrewege A et al. (2012) Reverse fuzzy extractors: enabling lightweight mutual authentication for PUF-enabled RFIDs. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform) 7397:374–389, LNCS

Role of Transfer Learning in Glioma Grading

Ranjana Joshi, Munesh C. Trivedi, Vishal Goyal, and Deepak Kumar Singh

Abstract Treatment planning of Glioma case is highly dependent on precise grading of Glioma. As per World Health Organization guidelines, tumor grading should be decided with the help of histopathology reports, i.e., biopsy report. Biopsy report may contain sampling errors which may result in incorrect grading. The biopsy report is also subject to inter- and intra-observer variability. To overcome the limitations of biopsy procedure involved in Glioma, grading process motivated the clinicians and scientists to look for machine learning, deep leaning and transfer learning approaches in Glioma grading process. As these advanced concepts were applied on MR images to carry out Glioma grading task, it not only make the procedure fast but also helps in precise grading. These approaches in contrast to biopsy can be performed repeatedly. This paper presents an overview of potential of transfer learning approaches in Glioma grading process.

Keywords Brain tumor · Biopsy · Magnetic resonance imaging · Transfer learning · Spectroscopy

1 Introduction

Brain is considered to be one of the complex organs of the body. The blood brain barrier protects the brain cells which on the other hand makes the brain chemistry invisible. Brain tumor occurs when there is formation of abnormal cells. These abnormal cells may push healthy areas of brain affecting normal functioning of the brain [1]. Till now, there are 120 types of known brain tumor described in literature. There are two types of brain tumor: primary and secondary brain tumor [2]. The type of tumor is called primary brain tumor when tumor starts within brain.

R. Joshi · V. Goyal
GLA University, Mathura, India

M. C. Trivedi (✉)
NIT, Agartala, India

D. K. Singh
KEC, Ghaziabad, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_2

In simple words, source of tumor cells lies within the brain itself. When source of brain tumor cells is from other parts of the body other than brain, such abnormal formation of cell mass results in secondary brain tumor. Secondary brain tumor is also called metastasis tumor. The severity of brain tumors can be judged in two ways: benign and malignant. Benign brain tumors are considered to be non-cancerous, and malignant brain tumors are considered to be cancerous. Benign tumors are generally less harmful in comparison to malignant tumor. Benign tumor has less growth and has less chances to infiltrate while malignant tumor grow quickly and may spread to other parts of the brain [3]. Studies are conducting nowadays to account for cancer causes. Definite reasons of cancer causes are unknown till today [4–16].

General practice of tumor treatment involves: detection, treatment followed by rehabilitization [3, 17–19]. For detecting brain tumor, clinician may ask for neurological test. Neurological test includes checking balance, vision, etc. Clinician may ask for imaging test also. Imaging includes magnetic resonance imaging (MR), computed tomography (CT) and positron emission tomography (PET). MR is preferred choice because of its advantage over other imaging modalities (Fig. 1).

MR provides better soft tissue contrast, and it uses non-ionizing radiations. MR can produce more detailed images of brain [17]. Advanced MR sequences such as functional MRI, perfusion MRI, diffusion MRI and spectroscopy may help clinician to understand the tumor characteristics in much better way. Team of clinician and scientist after tumor assessment with the help of MRI may conduct surgery in order to remove whole tumor if possible [12–14]. Resection of tumor depends on its location in the brain. If tumor lies in the interior part of the brain and it is difficult to carry out the surgery, then clinician may plan for the stereotactic biopsies to collect the tumor cells. The cells are collected from resected tissues and send for further analysis in pathology laboratory. In the pathology laboratory, these cells are analyzed for tumor

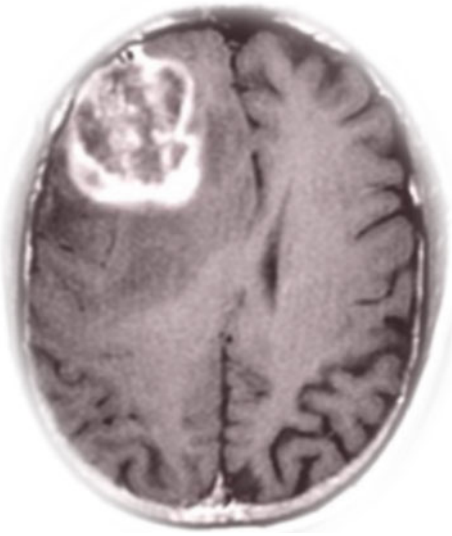


Fig. 1 Representative image of Glioblastoma Multiforme (high grade glioma) [20]

grade and its type. Once tumor type and grade are decided, the clinician may plan the further treatment which includes chemotherapy, radiotherapy or combination of these [12–16]. Majority of adult population around the globe suffers from the Gliomas [1–5, 9, 17–25]. World Health Organization classifies Gliomas into four grades, i.e., Grade I, Grade II, Grade III and Grade IV. Grade I and Grade II Gliomas are considered to be low grade Gliomas (LGG), and Grade III and Grade IV are considered to be high grade Gliomas (HGG). HGGs are considered to be malignant in nature. Grade I and Grade II Gliomas have considerable prognoses [2] in comparison to Grade IV. Different Grades have different treatment strategy. Grade I tumors are mostly cured by surgery [17]. Different grades have different survival rates [17]. It is important to diagnose the Gliomas grade accurately to plan the treatment accurately. Biopsy reports are considered to be gold standard in Gliomas grading. Increasing number of Gliomas cases and sampling errors involved in preparing biopsies report motivated the clinician and scientist to look for alternatives available to diagnose the Gliomas grade. Clinicians and scientists are also looking for machine learning and deep learning concepts in order to grade the Gliomas accurately using less time (Fig. 2).

Transfer learning is an emerging concept in which a pretrained model is fine tuned to carry out the grading task. Advantage of using transfer learning over machine learning is that it does not require handcrafted features. Also, pretrained model learning can be utilized to carry out the new task. In simple words, pretrained model does not require training from scratch. In medical field, data is scare, and use of transfer learning is worth. In transfer learning, internal layers of pretrained network are used to extract the features, and final layer is for carrying out the desired task. The final layer in pretrained model is commonly known as adaptation layer.

The rest of the paper is organized as follows: Section 2 discusses the most commonly used transfer learning model in Glioma grading. Section 3 describes the

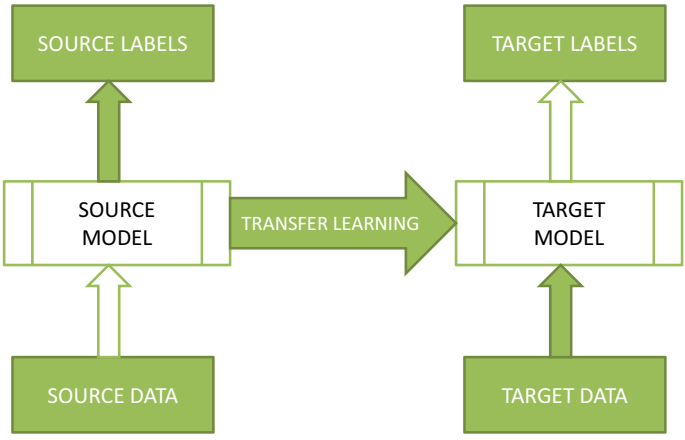


Fig. 2 Transfer learning approach [20]

available literature which has used the transfer learning concept in Glioma grading process. Section 4 concludes the paper.

2 Popular Transfer Leraning Models

AlexNet has about 60 million parameters. The AlexNet consists of five layers for performing convolution, three layers for performing max-pooling operation, two layers for performing normalization and one is output layer (softmax layer). The input size to AlexNet network is 224 × 224 × 3. GoogLeNet is another popular transfer learning approach used in Glioma grading process. GoogLeNet contains 27 layers. The GoogLeNet architecture uses 1 × 1 convolutions. Using 1 × 1 convolutions decreases the number of parameters. The input size to GoogLeNet network is 224 × 224. It is also known as inception V_1. visual geometry group (VGG) 16 architecture contains 16 layers. These 16 layers contain convolution and pooling layers. The last layer, i.e., output layer, is fully connected layer. The VGG network is much deeper network in comparison to AlexNet. VGG 16 has total 138 million parameters. VGG was conceptualized by its creators in order to reduce the parameters. These are several variants of VGG network such as VGG 16 and VGG 19. ResNet solves the problem of vanishing gradient which allows the network to go deeper. The most commonly used variants are ResNet 50 and ResNet 101.

3 Related Work

This section discusses about the literature which uses transfer learning concept in Glioma grading process. Table 1 represents the summary.

In study [2], AlexNet were fine tuned over Grade II, Grade III and Grade IV to classify Gliomas into Grade II, Grade III and Grade IV. The study includes 30 cases in Grade II patients, 43 cases in Grade III and 57 cases in Grade IV. Data augmentation was also done with the help of auto augmentation approach. The reported accuracy was equal to 97.9% in classifying Gliomas into Grade II, Grade III and Grade IV. In study [22], both AlexNet and GoogLeNet were fine tuned over 113 Gliomas cases. They also have performed fivefold cross-validation test. With the help of GoogLeNet, the author achieved the accuracy equal to 97.9%. The author also compared the performance of Glioma grading using the both the pretrained network, i.e., AlexNet and GoogLeNet. In study [22], author reported that GoogLeNet performs better than AlexNet. In study [1], authors used transfer learning concept to classify brain tumors into benign and malignant. They tuned ResNet 101, ResNet 50, AlexNet, GoogLeNet and SqueezeNet. As per their reported results, AlexNet performs better in comparison to other pretrained model. In study [1], author reported that GoogLeNet takes more training time as compared to all other pretrained networks.

Table 1 Summary of studies based on transfer learning model

References	Considered cases in each class	Pretrained model	Evaluation metrics (%)
[2]	*Grade II = 30* *Grade III = 43* *Grade IV = 57*	*AlexNet*	*Accuracy = 97.9*
[22]	Grade II = 25 Grade III = 27 Grade IV = 61	GoogleNet/AlexNet	Accuracy = 97.9
[1]	*Benign images = 224* *Malignant images = 696*	*AlexNet*	*Accuracy = 99.04*
[17]	Grade II = 50 Grade III = 58	Vgg 18	Accuracy = 89
[3]	*Glioma images = 1426* *Meningioma images = 708* *Pituitary images = 930*	*AlexNet, GoogLeNet, VGG 16, VGG 19, ResNet18, ResNet 101, ResNet-Inception-V2, SENet*	*Accuracy (AlexNet) = 98.71*
[18]	Malignant images = 155 Benign images = 98	VGG 16, ResNet 50, Inception-V3	Accuracy (VGG 16) = 96 Accuracy (ResNet) = 89 Accuracy (Inception) = 75
[19]	*Meningioma = 1426* *Glioma = 708* *Pituitary tumors = 930*	*Inception-V3, DensNet 201*	*Accuracy (Inception-V3) = 99.34* *Accuracy (DensNet201) = 99.51*
[23]	BraTS 2019	VGG-19	Accuracy (Training Accuracy) = 99.82 Accuracy (Validation Accuracy) = 96.32 Accuracy (Testing Accuracy) = 99.30
[24]	LGG = 76 HGG = 259	AlexNet	Area Under Receiver Operating Characteristic = 82.89
[25]	Grade II = 249 Grade III = 264 Grade IV = 607	ResNet 18	Accuracy = 73
[4]	GBM = 109 Metastases = 57	DNN	Accuracy = 95.5

In study [17], author used the concept of VGG18 pretrained neural network to classify the Gliomas into Grade II and Grade III. The study was conducted over 110 patients. The study includes 50 cases of Grade II patients and 58 cases of Grade III patients. The reported accuracy of developed classifier with the help of pretrained VGG 18 was equal to 89%. Author used cross-validation approach to validate their model. In study [3], authors used nine pretrained networks, i.e., AlexNet, GoogLeNet,

VGG 16, VGG 19, ResNet18, ResNet 101, ResNet-Inception-V2 and SENet. The last three layers of pretrained model were modified from 1000 outputs to three output labels, i.e., Gliomas, meningioma and pituitary. Authors reported that AlexNet and VGG 18 have achieved the same accuracy equal to the 98.55. Although these both pretrained classifier achieved the same accuracy, VGG 18 takes more time in training as compared to AlexNet. In study [18], authors developed a classifier which classifies the input into two classes, i.e., benign and malignant. The classifier was trained using 253 images. To increase the size of data set, data augmentation approaches such as flipping and rotation were used. The reported accuracy using these three fine tuned models, i.e., VGG-16, ResNet-50 and Inception-V3 models was 96%, 89% and 75%, respectively. In study [19], the author developed classifier which classifies the cases into three classes, i.e., Gliomas, meningioma and pituitary tumors. Two pretrained classifiers, Inception-v3 and DensNet 201, were fine tuned to carry out the classification task. The testing accuracy recorded with the help of Inception-v3 and DensNet was equal to 99.34% and 99.51%, respectively.

In study [23], authors used VGG 19 pretrained model to classify the cases into three categories, i.e. normal, LGGs and HGGs. BraTS 2019 data set w used for developing classifier. In study [24], authors used AlexNet to carry out the Glioma grading task, i.e., LGG versus HGG. The author kept the parameter settings same but the number of epoch has been incremented and was equal to 80. In study [25], authors used the concept of residual networks to carry out the distinguishing HGG from rest of the grades. The authors also developed the classifier which distinguishes the Grade III Glioma from Grade II Glioma. The use of residual network block solves the problem of vanishing gradient. ResNet 18 architecture was used to develop the classifier. ResNet 18 as its name implies is 18-layer deep architecture. In study [25], author mentioned that the developed classifier performs poor in case of distinguishing Grade III Glioma from Grade II Glioma.

4 Conclusion and Future Research Direction

Increasing cases of brain tumors in last five years are motivating factor for clinician and scientist to look for machine learning and deep learning approaches in Glioma grading process. The use of machine learning and deep learning approaches on MR images in Glioma grading not only makes the procedure fast but also overcomes the limitations of biopsy procedure. As biopsy is invasive in nature and may contain sampling errors, it is evident from the literature that around 30% cases of Glioma were diagnosed incorrect. Incorrect grading may affect the survival of the patients. Precise tumor grading helps clinician to plan the treatment accordingly. Several studies reported in literature show the potential of machine learning approaches in Glioma grading. Support vector machine and random forest were the most used approaches in Glioma grading. The development of machine learning model to carry

out Glioma grading task depends on the handcrafted features which can be considered as limitation of proposed machine learning approaches. The deep learning approaches were also proposed to carry out Glioma classification task but availability of limited dataset leaves the scope of validation of proposed concepts. The transfer learning concept can be thought as an alternative approach which overcomes the limitation of both approaches, i.e., machine learning and deep learning approaches. This paper highlights the role of transfer learning in classification of Gliomas, a kind of brain tumor.

References

1. Mehrotra R, Ansari MA, Agrawal R, Anand RS (2020) A transfer learning approach for AI-based classification of brain tumors. Mach Learn Appl 2:100003.https://doi.org/10.1016/j.mlwa.2020.100003
2. Lo C-M, Chen Y-C, Weng R-C, Hsieh K-C (2019) Intelligent glioma grading based on deep transfer learning of MRI radiomic features. Appl Sci 9(22):4926. https://doi.org/10.3390/app9224926
3. Chelghoum R, Ikhlef A, Hameurlaine A, Jacquir S (2020) Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. In: Maglogiannis I, Iliadis L, Pimenidis E (eds) Artificial intelligence applications and innovations. AIAI 2020. IFIP advances in information and communication technology, vol 583. Springer, Cham. https://doi.org/10.1007/978-3-030-49161-1_17
4. Trivizakis E, Papadakis GZ, Souglakos I, Papanikolaou N, Koumakis L, Spandidos DA, Tsatsakis A, Karantanas AH Marias K (2020) Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). Int J Oncol 57:43–53. https://doi.org/10.3892/ijo.2020.5063
5. Bae S, An C, Ahn SS et al (2020) Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. Sci Rep 10:12110. https://doi.org/10.1038/s41598-020-68980-6
6. Khalsa SSS, Hollon TC, Adapa A, Urias E, Srinivasan S, Jairath N, Szczepanski J, Ouillette P, Camelo-Piragua S, Orringer DA (2020) Automated histologic diagnosis of CNS tumors with machine learning. CNS Oncol 9(2):CNS56. https://doi.org/10.2217/cns-2020-0003
7. Lo´pez-Garcı´a G, Jerez JM, Franco L, Veredas FJ (2020) Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data. PLoS ONE 15(3):e0230536. https://doi.org/10.1371/journal.pone.0230536
8. Amiri S, Rekik I, Mahjoub MA (2016) Deep random forest-based learning transfer to SVM for brain tumor segmentation. In: 2016 2nd international conference on advanced technologies for signal and image processing (ATSIP). Monastir, pp 297–302. https://doi.org/10.1109/ATSIP.2016.7523095
9. Goodenberger ML, Jenkins RB (2012) Genetics of adult glioma. Cancer Genet 205(12):613–621
10. Banerjee S, Mitra S, Masulli F, Rovetta S (2019) Brain tumor detection and classification from multi-sequence MRI: study using ConvNets. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Springer International Publishing, Cham, pp 170–79
11. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: Proceedings of the 2009 IEEE conference on computer vision and pattern recognition. Miami, FL, pp 248–255, 20–25 June 2009

12. Taylor L, Nitschke G (2018) Improving deep learning using generic data augmentation. In: Proceedings of the 2018 IEEE symposium series on computational intelligence (SSCI). Bangalore, India. pp 1542–1547, 18–21 Nov 2018
13. Taylor LP (2010) Diagnosis, treatment, and prognosis of glioma: five new things. Neurology 75(18 Suppl 1):S28–S32
14. Shorten C (2018) Generative models for data augmentation—towards data science. https://towardsdatascience.com/generative-adversarialnetworks-for-data-augmentation-experiment-design-2873d586eb59. Accessed 21 Dec 2020
15. Liu S, Shah Z, Sav A et al (2020) Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. Sci Rep 10(1):7733
16. Berstad TJD, Riegler MA, Espeland H et al (2018) Tradeoffs using binary and multiclass neural network classification for medical multidisease detection. In: Proceedings of the 2018 IEEE international symposium on multimedia (ISM 2018). Taichung, Taiwan, pp 1–8, 10–12 Dec 2018
17. Naser MA, Deen MJ (2020) Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. Comput Biol Med 103758. https://doi.org/10.1016/j.compbiomed.2020.103758
18. Ali Khan H, Jue W, Mushtaq M, Umer Mushtaq M (2020) Brain tumor classification in MRI image using convolutional neural network. Math Biosci Eng 17(5):6203–6216. https://doi.org/10.3934/mbe.2020328
19. Noreen N, Palaniappan S, Qayyum A, Ahmad I, Imran M, Shoaib M (2020) A deep learning model based on concatenation approach for the diagnosis of brain tumor. IEEE Access 8:55135–55144. https://doi.org/10.1109/access.2020.2978629
20. Glioblastoma Multiforme—Symptoms, diagnosis and treatment options. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme. Accessed: 10 Jan 2021
21. Brain tumor—Symptoms and causes—Mayo clinic. https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084. Accessed: 1 Jan 2021
22. Yang Y, Yan LF, Zhang X et al (2018) Glioma grading on conventional MR images: a deep learning study with transfer learning. Front Neurosci 12:804. Published 15 Nov 2018.https://doi.org/10.3389/fnins.2018.00804
23. Ahuja S, Panigrahi BK, Gandhi T (2020) Transfer learning based brain tumor detection and segmentation using superpixel technique. In: 2020 international conference on contemporary computing and applications (IC3A). Lucknow, India, pp 244–249. https://doi.org/10.1109/IC3A48958.2020.233306
24. A transfer learning based active learning framework for brain tumor classification|Papers With Code. https://paperswithcode.com/paper/a-transfer-learning-based-active-learning. Accessed: 29 Nov 2020
25. Truong AH, Sharmanska V, Limbäck-Stanic C, Grech-Sollars M (2020) Optimization of deep learning methods for visualization of tumor heterogeneity and brain tumor grading through digital pathology. Neuro-Oncol Adv 2(1) Jan-Dec 110. https://doi.org/10.1093/noajnl/vdaa110

Current Approaches for Brain Tumor Segmentation

Ranjana Joshi, Munesh C. Trivedi, Vishal Goyal, and Deepak Kumar Singh

Abstract Abnormal formation of cell mass within the brain results in brain tumor. This abnormal formation may affect the normal functioning of the brain, hence affecting the survival rate. Treatment planning and follow-up depend on how accurately the grade and type of tumor are detected. Gold standard for detecting tumor grade and type is biopsy report. Biopsy is invasive in nature, and biopsy report may contain sampling errors. Defining tumor boundary over MRI is required to define clinical target volume (CTV). Manually locating tumor boundary is difficult and may contain subjectivity errors. These limitations motivated researcher to look for computer-based methods for diagnosing brain tumor type, its grades and defining CTV. The presented paper highlights the current trends in tumor segmentation approaches in order to define gross tumor volume (GTV) or CTV. It was established in literature with the help of table discussed in related work section that convolution-based approaches were more efficient in tumor segmentation and, hence, defining CTV or GTV.

Keywords Brain tumor · Clinical target volume · Tumor segmentation · Biopsy · Convolution neural network

1 Introduction

Brain tumor is considered to be life-threating disease. Abnormal formation of cells within the brain may result in formation of brain tumor. This abnormal formation may push the normal tissues inside the brain, hence affecting the normal functioning of the brain. Brain tumors can form in two ways.

(a) *Primary brain tumor:* when tumor cells start within the brain itself.

R. Joshi · V. Goyal
GLA University, Mathura, India

M. C. Trivedi (✉)
NIT, Agartala, India

D. K. Singh
KEC, Ghaziabad, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_3

(b) *Secondary brain tumor:* when tumor cells that form cell mass belong to other organs of the body. This type of tumor is also called metastasis.

Some tumors cells are considered to be cancerous and are known as malignant cells [1–20]. If tumor contains malignant cells, it will result in bad prognosis and less survival rate. Some tumor cells are non-cancerous and fall in category of benign tumor. Benign tumor has better survival rate in comparison to malignant tumor [1–5]. Clinical treatment management depends on the size, types and grades of brain tumor diagnosed. Magnetic resonance imaging is preferred choice for clinicians over computer tomography and positron emission technology. Size of brain tumors helps in defining clinical tumor volume which is needed for radiation therapy. Manually defining tumor size is difficult due to heterogeneous shape. Manual practice may contain subjectivity errors. Biopsy is considered to be gold standard for diagnosing brain tumor [1–10]. It was well established in literature that biopsy reports may contain sampling errors [1–21]. These sampling errors may result in incorrect grading, hence affecting survival rate of a patient (Fig. 1).

Limitations involved in tumor segmentation and grading process motivated clinical researchers to look for an alternative approaches for segmentation and diagnosing tumor type using imaging technique.

The objective of this presented paper is to give an overview of brain tumor segmentation approaches and challenges involved. The rest of paper is organized as follows. Section 2 discusses the related work done in domain of brain tumor segmentation. Section 3 concludes the paper and highlights the challenges involved in brain tumor segmentation.

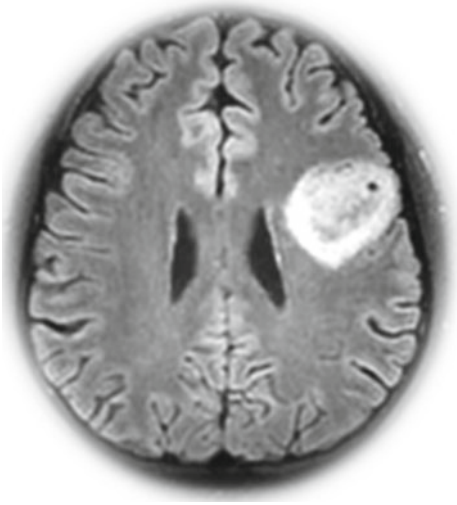


Fig. 1 MRI FLAIR image showing brain tumor (hyperintense region)

2 Related Work

This section discusses the current methods developed in the area of brain tumor segmentation. In study [1], authors proposed tumor segmentation based on the method of thresholding and watershed algorithm. Thresholding approach was used to segment the tumor, and watershed algorithm was used to segment the brain. In study [2], author used the late fusion convolution neural network (CNN) to segment the brain tumor. In late fusion CNN approach, independent CNN was used for every MRI sequence. In study [3], authors proposed the concept of tumor segmentation based on the automatic tumor thresholding and automatic tumor localization concept. In study [4], authors proposed the tumor components segmentation approach which was based on the 3D UNet deep learning model. For carrying out their study, BraTS 2019 dataset was used. In study [5], authors proposed the tumor segmentation approach which was based on the watershed algorithm and CNN. Watershed approach was used to carry out the tumor segmentation task, and CNN approach was used to carry out the tumor classification task. In the study [6, 11], authors used the CNN approach to carry out the tumor segmentation task. In study [7], authors presented a survey based on deep learning approaches. In study [8], authors used watershed algorithm for segmenting the tumor. In study [9, 10], for carrying out segmentation task, K-means algorithm was used followed by tumor classification.

In study [12], author used histogram-based approach for segmenting tumor. In study [13], authors used CNN concept to carry out the segmentation task. For improving the segmentation simultaneously reducing parameters, 1×1 layer was added to base network. Fuzzy C-means approach was used to carry out segmentation task in study [14]. In study [15, 17], authors presented a survey based on the concept of CNN. They concluded that CNN-based approach yields better results in comparison with other proposed approaches. In study [16], authors proposed the segmentation concept based on the concept of histogram and region growing approaches. In study [18], authors used the concept of UNet using HardELiSH activation for carrying out segmentation task. Authors used other activation functions such as ReLU, ELiSH and Swish for carrying out segmentation task, but HardELiSH results were found to be better. In study [19], authors used the concept of hybridized multilevel thresholding (Table 1).

3 Conclusion and Challenges

The objective of this carried out work was to present a summary of the latest methods developed in the area of brain tumor segmentation. This paper was divided into three major sections: introduction, related work followed by conclusion and challenges. Introduction part gives a glimpse of brain tumor and mentioned the limitations of existing practices. It tries to establish the motivation of clinicians involved for carrying out the segmentation task. The related work section discusses the recent

Table 1 Summary of implemented machine learning classifiers reported in literature along with their reported accuracy

References	Sample considered	Methods	Evaluation
[1]	Tumor images = 14	Thresholding, watershed method	Error = 10%
[2]	LGG = 75 HGG = 210	Late fusion CNN	Dice = 87.21%
[3]	Tumor images = 320	Automatic tumor thresholding, automatic tumor localization	Sensitivity = 91% Specificity = 99%
[4]	BraTS 2019 dataset	3D UNet	Dice (whole tumor) = 89.1%
[5]	BraTS database	Watershed, CNN	Dice = 93.5%
[6]	BraTS 2017 database	CNN	Dice = 89%
[8]	Tumor images = 36	Semantic segmentation (watershed approach)	NA
[9]	Internet available tumor images	K-means	NA
[10]	HGG = 20 images	K-means	Dice = 81%
[11]	NA	CNN	Dice = 77%
[12]	NA	Histogram-based approach	Dice = 80.56%
[13]	BraTS 2015 dataset	CNN	Dice = 90%
[14]	Cancer imaging archive	K-means, fuzzy C-means	Relative tumor area (FCM) = 93%
[15]	NA	Histogram-based, region growing technique	NA
[18]	BraTS 2015	UNet/HardELiSH activation	Dice = 87.4%
[19]	NA	Hybridized multilevel thresholding	NA

* *NA*—not available (not mentioned in their presented paper)

alternative approaches that were developed in order to carry out segmentation task. Machine learning and deep learning approaches-based methods show promising results and were base of many proposed methods in literature. Transfer learning-based approaches were also getting developed and can be well found in literature. Although lots of machine learning, deep learning, transfer learning and other methods were developed but due to heterogeneity involved in tumor cells and lack of availability for standard protocols in imaging acquisitions across vendors around the globe makes the segmentation task a difficult problem and proposed segmentation approaches difficult to compare respectively. This further opens the door for more research in area of brain tumor segmentation.

References

1. Wulandari A, Sigit R, Bachtiar MM (2018) Brain tumor segmentation to calculate percentage tumor using MRI. In: 2018 international electronics symposium on knowledge creation and intelligent computing (IES-KCIC), Bali, Indonesia, pp 292–296. https://doi.org/10.1109/KCIC.2018.8628591
2. Rahimpour M, Goffin K, Koole M (2019) Convolutional neural networks for brain tumor segmentation using different sets of MRI sequences. In: 2019 IEEE nuclear science symposium and medical imaging conference (NSS/MIC), Manchester, UK, pp 1–3. https://doi.org/10.1109/NSS/MIC42101.2019.9059769
3. Abdulraqeb ARA, Al-haidri WA, Sushkova LT (2018) A novel segmentation algorithm for MRI brain tumor images. In: 2018 ural symposium on biomedical engineering, radioelectronics and information technology (USBEREIT). Yekaterinburg, Russia, pp 1–4. https://doi.org/10.1109/USBEREIT.2018.8384535
4. Wu P, Chang Q (2020) Brain tumor segmentation on multimodal 3D-MRI using deep learning method. In: 2020 13th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI). Chengdu, China, pp 635–639. https://doi.org/10.1109/CISP-BMEI51763.2020.9263614
5. Jemimma TA, Vetharaj YJ (2018) Watershed algorithm based DAPP features for brain tumor segmentation and classification. In: 2018 international conference on smart systems and inventive technology (ICSSIT). Tirunelveli, India, pp 155–158. https://doi.org/10.1109/ICSSIT.2018.8748436
6. Derikvand F, Khotanlou H (2019) Patch and pixel based brain tumor segmentation in MRI images using convolutional neural networks. In: 2019 5th Iranian conference on signal processing and intelligent systems (ICSPIS). Shahrood, Iran, pp 1–5. https://doi.org/10.1109/ICSPIS48872.2019.9066097
7. Somasundaram S, Gobinath R (2019) Current trends on deep learning models for brain tumor segmentation and detection—a review. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). Faridabad, India, pp 217–221. https://doi.org/10.1109/COMITCon.2019.8862209
8. Hussain A, Khunteta A (2020) Semantic segmentation of brain tumor from MRI images and SVM classification using GLCM features. In: 2020 second international conference on inventive research in computing applications (ICIRCA). Coimbatore, India, pp 38–43. https://doi.org/10.1109/ICIRCA48905.2020.9183385
9. Lavanyadevi R, Machakowsalya M, Nivethitha J, Kumar AN (2017) Brain tumor classification and segmentation in MRI images using PNN. In: 2017 IEEE international conference on electrical, instrumentation and communication engineering (ICEICE). Karur, India, pp 1–6. https://doi.org/10.1109/ICEICE.2017.8191888
10. Imtiaz T, Rifat S, Fattah SA (2019) Automated brain tumor segmentation from MRI data based on local region analysis. In: 2019 IEEE international conference on biomedical engineering, computer and information technology for health (BECITHCON). Dhaka, Bangladesh, pp 107–110. https://doi.org/10.1109/BECITHCON48839.2019.9063199
11. Kumar S, Negi A, Singh JN, Verma H (2018) A deep learning for brain tumor MRI images semantic segmentation using FCN. In: 2018 4th international conference on computing communication and automation (ICCCA). Greater Noida, India, pp 1–4. https://doi.org/10.1109/CCAA.2018.8777675
12. Akter MK, Khan SM, Azad S, Fattah SA (2017)Automated brain tumor segmentation from mri data based on exploration of histogram characteristics of the cancerous hemisphere. In: 2017 IEEE region 10 humanitarian technology conference (R10-HTC). Dhaka, 2017, pp 815–818. https://doi.org/10.1109/R10-HTC.2017.8289080
13. Yang T, Song J (2018) An automatic brain tumor image segmentation method based on the U-net. In: 2018 IEEE 4th international conference on computer and communications (ICCC). Chengdu, China, pp 1600–1604. https://doi.org/10.1109/CompComm.2018.8780595

14. Srinivas B, Rao GS (2018) Unsupervised learning algorithms for MRI brain tumor segmentation. In: 2018 conference on signal processing and communication engineering systems (SPACES). Vijayawada, India, pp 181–184. https://doi.org/10.1109/SPACES.2018.8316341
15. Venu K, Natesan P, Sasipriyaa N, Poorani S (2018) Review on brain tumor segmentation methods using convolution neural network for MRI images. In: 2018 international conference on intelligent computing and communication for smart world (I2C2SW). Erode, India, pp 291–295. https://doi.org/10.1109/I2C2SW45816.2018.8997387
16. Jasmine RA, Rani PAJ (2016) A two phase segmentation algorithm for MRI brain tumor extraction. In: 2016 international conference on control, instrumentation, communication and computational technologies (ICCICCT). Kumaracoil, pp 437–440. https://doi.org/10.1109/ICCICCT.2016.7987989
17. Parihar AS (2017) A study on brain tumor segmentation using convolution neural network. In: 2017 international conference on inventive computing and informatics (ICICI). Coimbatore, India, pp 198–201. https://doi.org/10.1109/ICICI.2017.8365336
18. Salih MM, Salih ME, Ahmed MAA (2019) Enhancement of U-net performance in MRI brain tumour segmentation using HardELiSH activation function. In: 2019 international conference on computer, control, electrical, and electronics engineering (ICCCEEE). Khartoum, Sudan, pp 1–5. https://doi.org/10.1109/ICCCEEE46830.2019.9071235
19. Dawngliana M, Deb D, Handique M, Roy S (2015)Automatic brain tumor segmentation in MRI: Hybridized multilevel thresholding and level set. In: 2015 international symposium on advanced computing and communication (ISACC). Silchar, India, pp 219–223. https://doi.org/10.1109/ISACC.2015.7377345
20. Boughattas N, Berar M, Hamrouni K, Ruan S (2014) Brain tumor segmentation from multiple MRI sequences using multiple kernel learning. In: 2014 IEEE international conference on image processing (ICIP). Paris, France, pp 1887–1891. https://doi.org/10.1109/ICIP.2014.7025378
21. Hirahara D (2019) Preliminary assessment for the development of CADe system for brain tumor in MRI images utilizing transfer learning in Xception model. In: 2019 IEEE 8th global conference on consumer electronics (GCCE). Osaka, Japan, pp 922–924. https://doi.org/10.1109/GCCE46687.2019.9015529

Intelligent Traffic and Transport Management Systems: Categorization, Problems and Future Objectives

Mohit Misra, Prashant Mani, and Shailesh Tiwari

Abstract In recent years, large number of vehicles are running on the roads, but available traffic infrastructure is not able to beat it. Therefore, nowadays, traffic congestion is major concern, and it gives rise to happen more accidents. This problem creates adverse effects on society, including economic development, traffic accidents, rise in greenhouse productions, time loss and health issues. In this situation, it requires a modern and robust traffic management system to reduce traffic congestion and its adverse effects. There is a need of traffic management systems which consist of a set of modern applications which consist of application and management tools to expand the efficiency overall traffic. Also, such system should provide the safety to the transportation systems. This system works by gathering information from various types of sources, and then it analyzes the collected such information and identifies the problems that could degrade the traffic efficiency and then runs services to manage them. By considering all these aspects, this paper presents a classification system to analyze challenges and develop a future outlooks to generate a traffic management system.

Keywords Intelligent traffic and transport management system (ITTMS) · Road surface recognition · Condition monitoring · Machine learning algorithms

1 Introduction

In a developing infrastructure, roads are a piece of core unit in it, and it supports people and logistics movements, formation in the foundation of social and economic development and interconnecting cities, ports, projects and airports. There are various issues arising in this procedure, i.e., maintenance, development of roads, congestion and traffic and increasing population. For this, we propose a new system of roads

M. Misra (✉) · P. Mani
SRM Institute of Science and Technology, NCR Campus, Delhi Meerut Road, Modingar, UP, India

S. Tiwari
Krishna Engineering College, Ghaziabad, UP, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_4

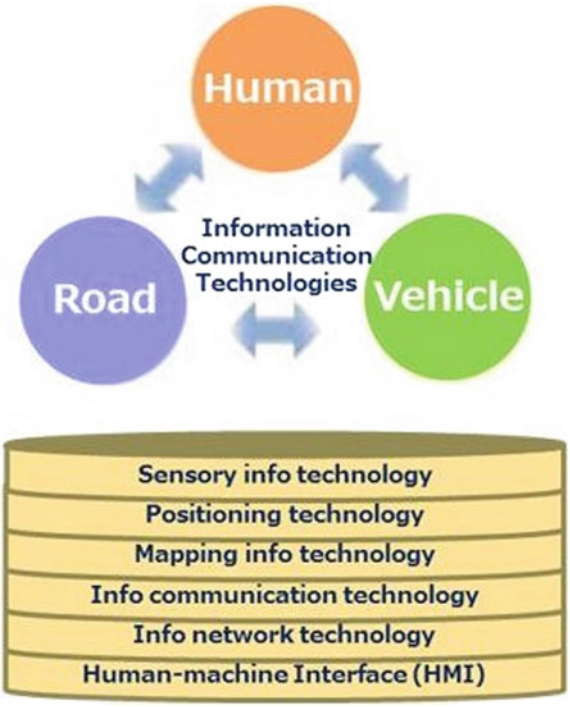


Fig. 1 Basic components used in ITS

which is a combination of information, communication and control techniques to integrate drivers, vehicles and roads properly. This system is called as intelligent transport system (ITS) which only aims to traffic-related issues along with many other ones (Fig. 1).

For which, we have set its applications focus on three issues of problems: Society, the road administrator and drivers. At societal level, it helps in decreasing accidents, environmental effects and reducing traffic. Similarly, at the level of road administrator, management of roads and traffic can be handled by identifying the hazardous locations and frequently congested areas using probe information collected by vehicles. And at last in case of drivers, we can get advantage in economic front such as reduced tolls, route detection and payments of tolls and parking. ITS provides just not only a convenient road traffic application along with the provision of the usage of variety of information and communication technologies that results to the generation of new employment and business opportunities along with economic development.

1.1 Problems Faced in Traffic Management System

Basic Problems Faced in Traditional Traffic management System:

Urbanization: Increase in population living in metropolitan and nearby cities for having better earning opportunities.

Increasing transportation demands: Along with increasing population, demands for better and cheap transportation also increase, due to which there is an increase of motor vehicles on roads.

Increasing pollution: Along with increasing vehicles, emission of pollutant gases like CO_2 and other gases occur as the usage of fuels increases in sufficient amount.

Increasing traffic issues: Road safety and congestion.

1.2 Categorization of Intelligent Traffic and Transport Management System

Categorization of ITTMS is done on basis of infrastructure-based and infrastructure-less approaches. All the research work done in this field so far is basically categorized in infrastructure-based and infrastructure-less approaches. Parameters focused for dividing various approaches are design, goals, identification strategy, re-routing strategy and network challenges (Fig. 2).

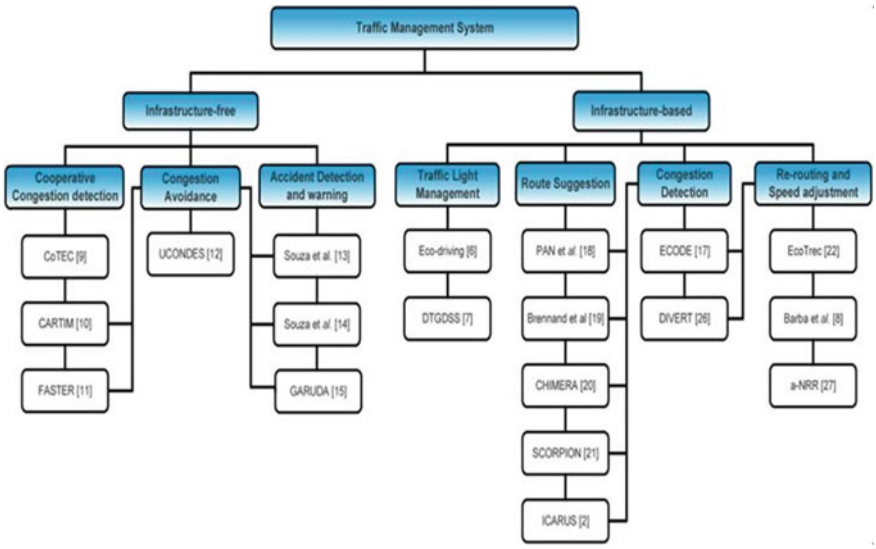


Fig. 2 ITTMS classification

2 Comparison of Various Infrastructure-Based and Infrastructure-less ITTMS Approaches on Design and Goal Strategy

See Tables 1 and 2.

3 Proposed Methodology

The challenge is to build a low-cost information platform for monitoring traffic flows and to bring that information from different parts of a city together so that it can be analyzed to enable better traffic management routines for mega-cities in low income countries. Present traffic flow monitoring techniques used in developed countries such as SCOOT cannot cope with the volume of traffic, which would go quickly wear out the SCOOT loops. In addition, newer systems such as ETSI-G5 technology for connected vehicles are very expensive. These systems work by having On Board Units (OBUs) and Roadside Units (RSUs), which are placed along the road infrastructure. Currently, an OBU and an RSU each cost over $1000 USD. So the solution being proposed is to use Bluetooth Low Power (BLE) devices. These devices communicate using Bluetooth technology and have a common coverage range of 70–100 m.

Since mobile phones also have Bluetooth interfaces, a lot of apps have been written to take data from the BLE device and use the mobile phone to communicate with a central server or navigational entity. Though what we are proposing does not remove such capability, the system we want to build is not be dependent on users having mobile phones. Instead, the BLE device will beacon every few seconds, and those beacons with their UUIDs will be picked up be WiFi/BLE gateways. The gateway devices cost £54 each, and this project will buy 50 of such gateways. These gateways will then use right wireless routers wireless routers, which will be connected to the University of Ghana's communications network. This is an Ethernet network running at 1 Gbps. The data will be routed to a central server where it will be stored and processed and used to display traffic flows in real time.

The WiFi/BLE gateways will be placed on two very busy roads that border the University of Ghana. The first is called National Highway—4 or N4, which runs from north to south along the east side of the campus. In fact, it forms a direct boundary with the campus for 3.5 km. The second road is called the Haatso-Atomic road and runs from east to west on the north side of the campus and forms a natural campus boundary for 1.5 km. So in this proposal, we would like to monitor traffic from both roads making this a total of 5 km of road infrastructure that will be monitored. The data gathered will be analyzed to show the traffic flow along these roads. The results will be displayed using a virtual network computing (VNC) server; hence, the results will be available to all commuters in real time. Information will also be

Table 1 .

Solutions of various ITTMS approaches		Design		Goals						
		V2V	V2I	Congestion De-	Accident detection	Traffic classification	Traffic	Data dissemination	Congestion avoid-	Speed adjustment
Infrastructure-less approaches	CoTEC [1]	Yes	No	Yes	No	Yes	No	No	No	No
	CARTIM [2]	Yes	No	Yes	No	Yes	Yes	No	Yes	No
	FASTER [3]	Yes	No	Yes	No	Yes	No	No	Yes	No
	UCONDES [4]	Yes	No	Yes	No	No	No	Yes	Yes	No
	De Souza et al. [5]	Yes	No	Yes	Yes	No	No	Yes	Yes	No
	De Souza and Villas [6]	Yes	No	Yes	Yes	No	No	Yes	Yes	No
	GARUDA [7]	Yes	No	No	Yes	No	Yes	Yes	Yes	No
	Eco-driving [8]	No	Yes	No	No	No	Yes	No	Yes	Yes
	DTGDSS [9]	No	Yes	No	No	No	Yes	No	Yes	No
Infrastructure-based	DSP [10]	No	Yes	Yes	No	No	Yes	No	No	No
	Brenand [1]	No	Yes	Yes	No	No	Yes	Yes	No	No
	CHIMERA [11]	No	Yes	Yes	No	No	Yes	Yes	Yes	No
	SCORPION [12]	No	Yes	Yes	No	No	Yes	No	Yes	No
	ICARUS [13]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
	ECODE [14]	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No
	DIVERT [15]	Yes	Yes	Yes	No	No	No	Yes	Yes	No
	EcoTrec [16]	Yes	Yes	No	Yes	No	No	No	Yes	No
	Barba et al. [17]	Yes	Yes	No	Yes	No	Yes	No	No	Yes

(continued)

Table 1 (continued)

Solutions of various ITTMS approaches		Design		Goals						
		V2V	V2I	Congestion De-	Accident detection	Traffic classification	Traffic	Data dissemination	Congestion avoid-	Speed adjustment
	a-NRR [18]	No	Yes	No	No	No	No	No	No	No

Table 2 .

Solutions of various		Identification strategy				Network challenges			Re-routing strategy		
ITTMS approaches		Fuzzy log	Neural network	Classifier	EA	Broadcast storm	Network particle	IEEE 802.11 P	Shortest path	Traffic balance	Rerouting
Infrastructure-less approaches	CoTEC [1]	Yes	No	No	No	No	No	No	No	No	No
	CARTIM [2]	Yes	No	No	No	Yes	No	No	Yes	Yes	No
	FASTER [3]	No	No	No	Yes	Yes	No	Yes	Yes	Yes	No
	UCONDES [4]	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No
	De Souza et al. [5]	No	No	No	No	No	No	No	Yes	No	No
	De Souza and Villas [6]	No	No	No	No	No	No	No	No	Yes	No
	GASUDA [7]	No	No	No	No	No	No	No	Yes	No	No
	Eco-driving [8]	No	No	No	Yes	Yes	Yes	No	Yes	Yes	No
	DTGDSS [9]	No	No	No	Yes	Yes	No	No	Yes	Yes	No
Infrastructure-based approaches	DSP [10]	No	No	No	Yes	No	No	Yes	No	No	No
	Brennand et al. [19]	No	No	No	Yes	No	No	No	Yes	No	No
	CHIMERA [11]	No	No	Yes	No	Yes	Yes	No	Yes	Yes	No
	SCORPION [12]	No	No	Yes	No	No	No	No	No	No	Yes
	ICARUS [13]	No	No	Yes	No	Yes	No	No	Yes	Yes	No
	ECODE [14]	No	No	No	No	Yes	No	No	Yes	Yes	No
	DIVERT [15]	No	No	No	No	No	No	Yes	No	No	No

(continued)

Table 2 (continued)

Solutions of various		Identification strategy				Network challenges			Re-routing strategy		
ITTMS approaches		Fuzzy log	Neural network	Classifier	EA	Broadcast storm	Network particle	IEEE 802.11 P	Shortest path	Traffic balance	Rerouting
	EcoTiec [16]	No	No	No	No	No	Yes	No	No	No	No
	Barba et al. [17]	No	No	No	No	No	No	No	Yes	Yes	No
	a-NRR [18]	No	No	YES	No	No	No	No	No	No	No

communicated using mobile phones as well as road signage such as overhead road gantries.

Proposed methodology has been organized into:

(a) **Research Design**: The study is both exploratory and descriptive in nature. Exploratory research design has identified the related variables of the study and defined the research problem. The conceptual framework incorporating the relevant variables is also proposed through exploratory research. Thereafter, descriptive research design has been used to empirically test the proposed conceptual framework of the study and subjecting the data to statistical analysis.

(b) **Objectives of the Study**:

- Proper monitoring of traffic flow on roads using a low-cost solution will help in providing information and raising warning messages on timely basis to all the road users.
- Identifying the pits, cracks on the roads, crocodile-shape roads, distraction caused due to any unwanted objects and unwanted breakers in order to prevent damage in the vehicles and road accidents. Main objective is to minimize the casualties/fatalities that happen on the road.
- Objective is to provide the safety on Indian Roads by associating the risk associated with roads and to specify the ranking of roads on the basis of quality of roads.
- By analyzing and retrieving the data from the vehicles regarding obstruction on the road in turn will be provided to the government in order to remove that obstruction. Objective is to provide congestion free and cost-effective driving experience to all road users.

(c) **Sampling Design**:

The sample for the study the city of Noida has been selected. We plan to collect the data regarding the roads condition and traffic conditions of the road near the Amity campus. To study will be performed on the road conditions using the sample size of 500 vehicles with BLE devices for communication and sensors like accelerometers embedded with them.

(d) **Data Collection Instruments**:

The basic data collection instruments are the sensor that is embedded on the vehicle.

1. Accelerometer
2. Gyro meter
3. Cameras
4. GPS
5. Smart phones with WiFi connectivity
6. Bluetooth
7. APP on smart phone for emergency messages.

(e) **Data Collection Method**:

The study is based on survey method of collecting initial information during the research process. Primary data is collected by researcher himself using survey

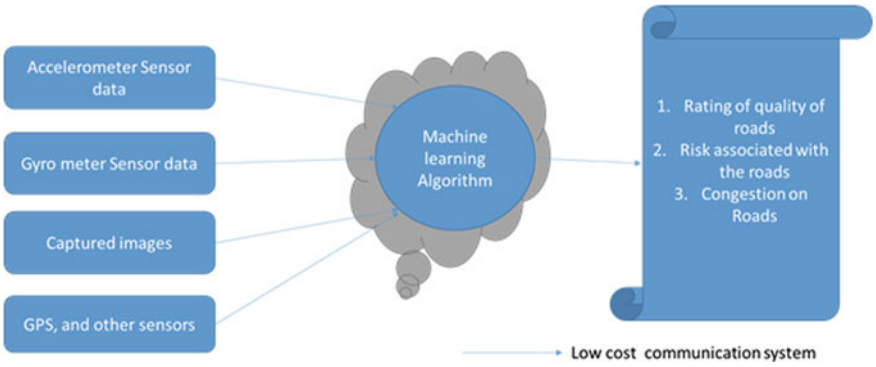


Fig. 3 Proposed model

method. It is a reliable way to collect the data because researcher knows where it came from and how it was collected and analyzed.

(f) **Statistical Techniques**:

The various data cleaning and analysis operations will be performed based on the data collected form the sensors and perform the analysis using statically operations used in the machine learning algorithm. The summarized tentative plan for the same has been described in the figure below.

(g) **Variables of the Study**:

Demographic variables such as gender, age, education, occupation, driving experience, driving training and type of vehicles are independent variables of the study, and factors such as risk perception and perception of driving tasks have been treated as exogenous variable and road safety attitude comprising five factors of traffic rules, personal responsibility, road safety and distraction on roads are treated as endogenous variable in the study (Fig. 3).

4 Challenges in ITTMS

In this section, various challenges faced in ITTMS are:

- Homogenous and heterogeneous data collection and integration of data.
- Managing large amount of data received from sensor is a tedious task.
- Monitoring and surveillance of traffic condition.
- Functionality similar to Google Maps for providing alternate route guidance.
- SCOOT technology can only be used where the traffic flow is very less such as Yamuna Express Way or Agra-Luck now Expressway.
- In addition to SCOOT, new technology was developed that is ETSI-G5 which is very expensive as it uses On Board Units and Road Side Units which are placed along with the road infrastructure.

- Neither SCOOT nor ETSI-G5 talks about the safety of roads. Only monitoring or surveillance of vehicles is done using these technologies. This seems to be providing a costly incomplete solution in terms of all types of traffic flow.
- Quality of the roads and accidents due to animals is not being considered in the literature as the factor for road accidents or damages in the vehicles.
- Study of traffic congestion at traffic lights was not discussed in the literature.

5 Conclusion and Future Scope

This paper basically focuses on the various challenges faced in the traffic management system. An improvement in ITTMS is expected from the researchers and industrialists by identifying the various challenges and research gaps. Various researchers have already proposed different research schemes in order to strengthen the traffic flow and congestion of traffic. Basic analysis of all the researchers who have done their research in the field of TMS is done, and some gaps are identified as future work for other researchers. Proper monitoring of traffic flow on roads using a low-cost solution will help in providing information and raising warning messages on timely basis to all the road users. Identifying the pits, cracks on the roads, crocodile shaped-roads, distraction caused due to any unwanted objects and unwanted breakers in order to prevent damage in the vehicles and road accidents. Main objective is to minimize the casualties/fatalities that happen on the road. Objective is to provide the safety on Indian Roads by associating the risk associated with roads and to specify the ranking of roads on the basis of quality of roads. By analyzing and retrieving the data from the vehicles regarding obstruction on the road in turn will be provided to the government in order to remove that obstruction. Objective is to provide congestion free and cost-effective driving experience to all road users.

References

1. Bauza R, Gozálvez J (2013) Traffic congestion detection in large-scales scenarios using V2V communications. J Netw Comput Appl 36(5):1295–1307
2. Araujo GB, Queiroz MM, de LP Duarte-'ueiredo FF et al (2014) CARTIM: a proposal toward identification and minimization of vehicular traffic congestion for VANET. In: 2014 IEEE symposium on computers and communications (ISCC). Funchal, pp 1–6, 23–26 June 2014. http://dx.doi.org/10.1109/ISCC.2014.6912491
3. De Souza AM, Villas L (2016) A fully-distributed traffic management system to improve the overall traffic efficiency. In: 2016 ACM international conference on modeling, analysis and simulation of wireless and mobile systems, Malta. ACM, New York, 13–17 Nov 2016
4. Meneguette RI, Ueyama J, Filho GPR et al (2015) Enhancing intelligence in inter-vehicle communications to detectand reduce congestion in urban centers. In: 20th IEEE symposium on computers and communication (ISCC), Larnaca. New York, IEEE, pp 662–667, 6–9 July 2015

5. De Souza AM, Boukerche A, Maia G et al (2014) Decreasing greenhouse emissions through an intelligent traffic information system based on inter-vehicle communication. In: Proceedings of the 12th ACM international symposiumon mobility management and wireless access (MobiWac'14). ACM, New York, pp 91–98. http://doi.acm.org/10.1145/2642668.2642677
6. De Souza AM, Villas LA (2015) A new solution based on inter-vehicle communication to reduce traffic jam in highway environment. IEEE Lat Am Trans 13(3):721–726
7. De Souza AM, Yokoyama RS, da Fonseca NLS et al (2015) Garuda: a new geographical accident aware solution to reduce urban congestion. In: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing (CIT/IUCC/DASC/PICOM), Liverpool. IEEE, New York, pp 596–602, 26–28 Oct 2015
8. De Souza AM, Yokoyama RS, Maia G et al (2015) Minimizing traffic jams in urban centers using vehicular ad hoc net-works. In 2015 7th international conference on new technologies, mobility and security (NTMS), Paris. IEEE, New York, pp 1–5, 27–29 July 2015
9. Lee W-H, Lai YC, Chen PY (2012) Decision-tree based green driving suggestion system for carbon emission reduction. In: 2012 12th international conference on ITS telecommunications (ITST), Taipei, Taiwan, IEEE, New York, pp 486–491, 5–8 Nov 2012
10. Pan J, Khan M, Popa SI et al (2012) Proactive vehicle re-routing strategies for congestion avoidance. In: 2012 IEEE 8th international conference on distributed computing in sensor systems (DCOSS), Hangzhou, China. IEEE, New York, pp 265–272, 16–18 May 2012
11. De Souza AM, Yokoyama RS, Maia G et al (2016) Real-time path planning to prevent traffic jamthroughanintelligent transportationsystem. In: 2016 IEEE symposium on computers and communication(ISCC), Messina. IEEE, NewYork, pp 726–731, 27–30 June 2016
12. De Souza AM, Yokoyama RS, Botega LC et al (2015) Scorpion: a solution using cooperative rerouting to prevent congestion and improve traffic condition. In: 2015 IEEE ınternational conference on computer and information technology, Liverpool. IEEE, New York, pp 497–503, 26–28 Oct 2015
13. Doolan R, Muntean GM (2016) EcoTrec—a novel VANET-based approach to reducing vehicle emissions. IEEE T Intell Transp (99):1–13
14. Younes M, Boukerche A (2013) Efficient traffic congestion detection protocol for next generation vanets. In: 2013 IEEE international conference on communications (ICC), Budapest. IEEE, New York, pp 3764–3768, 9–13 June 2013
15. Pan J, Popa IS Borcea C (2016) Divert:a distributed vehicular traffic re-routing system for congestion avoidance. IEEE T Mobile Comput (99)
16. Chen N, Xu X (2013) Information-fusion method for urban traffic flow based on evidence theory combining with fuzzy rough set. J Theor Appl Inform Tech 49(2):560–566
17. Barba C, Mateos M, Soto P et al (2012) Smart city for vanets using warning messages, traffic statistics and intelligent traffic lights. In: 2012 IEEE intelligent vehicles symposium (IV), Alcala´deHenares. IEEE, New York, 3–7 June 2012, pp 902–907
18. Wang S, Djahel S, McManis J (2015) An adaptive and vanets-based next road re-routing system forunexpected urbantrafficcongestionavoidance. In: 2015 IEEE vehicular networking conference(VNC), Columbus, OH. IEEE, New York, pp 196–203, 16–18 Dec 2015
19. Brennand CA, de Souza AM, Maia G et al (2015) An intelligent transportation system for detection and control of congested roads in urban centers. In: 20th IEEE symposium on computers and communication (ISCC), Larnaca. IEEE, NewYork, pp 476–481, 6–9 July 2015

An IoT-based Surveillance Method

Vaibhav Tyagi, Rupali Singh, Ashish Tripathi, Vasu Agarwal, and K. K. Mishra

Abstract Crimes have long been a nerve-racking ordeal to pursue a remedy, not just for the government but also for the whole community. It is undeniably true that there is a dramatic increase in crime rate among many nations. This rise stems from many factors. There appears to be some evidence that administrative lock-ups do not minimize crime rates. The atmosphere plays a significant role in shaping a criminal. There are so many slums where criminals live today. It is recorded that a child whose parent is guilty is more likely to be involved in social tragedies as a result of living among gamblers, drug dealers, or even murderers, to name only a few. It would seem, besides the family, that schools would take responsibility for sharpening an individual. Numerous juvenile criminals are supposed to be a bully when they are students. The inattention from teachers and parents, particularly during the student's adolescence, makes the struggling students more vulnerable to hardened criminals. To summarize, the increase in crime rate could be attributed to a number of factors including the low extreme penalties and vice enclosures. To reduce the number of crime rates in India and provide evidence to the police, we came up with a solution for which we created an Internet of Things (IoT)-based surveillance app. As the use of mobile applications is increasing at a rapid rate in India, we have designed this application so as to reduce the crime rate and provide safety measures and awareness among the society of people who are thinking of committing a crime.

Keywords Crime · Security · App · Safety · Surveillance

V. Tyagi (✉) · R. Singh · A. Tripathi · V. Agarwal
Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

K. K. Mishra
Department of Computer Science, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, UP 211002, India
e-mail: kkm@mnnit.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_5

1 Introduction

Today, with the rapid population growth and progress of urban centers and villages, the crime chart is also on the rise as shown in Fig. 1. This tremendous increase in crime is a matter of great concern and alarm to all of us. The frequent and repetitive thefts, robberies, murders, rapes, substance trafficking, smuggling, and so on have rendered ordinary people unrest. They feel very insecure and unprotected in the presence of anti- social and negative elements. The scammers have been trained or even have contacts and ties around the country and across the world [1]. The need for the hour is to sensitize the nation's citizens to the problem and seek their help in reducing the increasing crime [2]. The worrying pattern of female crime persists from the pattern seen in the past. A decrement in rape cases has been observed in 2017 as compared to 2016. The number of female crimes which were reported including various crimes such as rape, suicide, murder, acid attack, etc. was 3.5 lakhs [3]. The line chart below in Fig. 1 depicts the increment of crime against female in India in the recent years. A lot of work has been undertaken to lower the crime rate, but nothing has been successful to date. Crime analyzes of the historical past are expected to expose the nuances of the crime dataset. This mechanism would help the law enforcement agencies in arresting criminals and coordinating the approaches for crime prevention [5]. The capability to foresee potential crime on the basis of past, time, and position will serve either from strategic or from operational backgrounds as a powerful source of intelligence for them. Nonetheless, it is a difficult challenge to forecast potential crime correctly for a higher outcome due to the growing rates of crime in the modern days. Therefore, some methods to reduce the number of crime rates are important.

Mobile media has been a central tool in inspiring individuals to tackle a variety of societal problems in the past decade. Computers have been an extravagant and

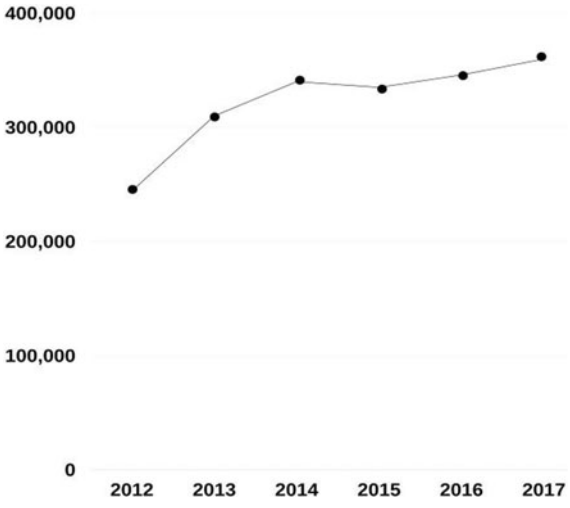


Fig. 1 Crime against female in India [4]

unreachable for a majority of citizens in many low-income nations, but cell phone technologies and smartphones in particular allow users to access and connect to statistics and resources like finance, business market analytics [6]. Other emergency monitoring applications rely on helping crime survivors, such as encouraging a female to alert the police and others anytime she thinks she is in a position where assistance is required. This study has been done to reduce the crime rate by building an IoT-based security and surveillance android application. Considering the current situation, some features such as a safest path are given for the women to travel safely from one place to another place. We have discussed it in the future sections. The paper is organized into six sections. Section 2 comprises the background research about increasing crime rates in India. Categories of crimes are discussed in Sect. 3. Different crime-prediction methods are discussed in Sect. 4. While the Sect. 5 contains the proposed work. Finally, the paper is concluded in Sect. 6.

2 Background

The National Crime Records Bureau (NCRB) released data for 2017 on a much delayed case. Although it was without some critical categories of data, the data included were clearly laid out more than in the 2016 crime report. About three dozen new categories and subcategories of crimes have been added under different headings by the NCRB. At least four areas where major data diversification can be seen are crimes against women and children, violence against Dalits (lowly caste), corruption cases, and the time taken by police and courts to resolve cases. The NCRB implemented categories of cybercrimes against women and children for the first time [7]. According to the report, a total of 9,89,070 cases of human-body felony were registered, which were responsible for 32.2% of total IPC crimes in 2017, of which approximately 4 lakhs 95 thousand cases of hurt contributed for maximum cases, i.e., 50.0% followed by negligence cases causing death (1,42,793), molestation on women (86,000) representing 14.2% and 8.6% [8]. In terms of percentage, during 2017, kidnapping and abduction (41%) along with cases under Children's Protection from Sexual Crimes Act of 2012 (25.3%) were registered as the major crimes under 'Crime Against Children' which also included child abuse [9]. In 2018, India appeared marginally safer compared to 2017, with the crime rate (incident per lakh population) for IPC crimes decreasing to 236.7 from 237.7 in 2017, the lowest crime rate ever in the country. The crime rate has been gradually rising since 2005, with the exception of 2016 and 2009 when it decreased [10]. According to the 2018 edition of the National Crime Records Bureau's (NCRB'Crime)'s in India' statistics, crime in India increased by 1.3% in 2018 compared to 2017, with more than 50 lakh identifiable offenses detected. According to the National Crime Records Bureau (NCRB), which is part of the Ministry of Home Affairs, a total of 5072 lakh crimes were recorded in 2018, with 3130 lakhs falling under the Indian Penal Code (IPC) and 1940 lakhs falling under Special Local Laws (SLL). The overall number of incidents recorded in 2017 grew by 1.3%, while the crime rate per lakh population decreased from

388.5 in 2017 to 383.4 in 2018, according to the study [11]. In 2018, the cumulative of kidnapping and abduction cases was found out to be 104 lakhs across the nation, reflecting an increase of 10% over 95,892 similar cases in 2017 [12]. As observed, the prominent reason for largest number of murders is disputes (9622) which is followed by personal quarrel and revenge (3874) and benefit (2994). An increase in accidental deaths has also been seen from 3,96,584 in 2017 to 4,11,824 in 2018—a spike of 31.1% from 30.3% in 2017. In its disclaimer section, the NCRB study 'Crime in India-2018' said that clarifications were pending from few states, namely Sikkim, Kolkata, Arunachal Pradesh, West Bengal, and Meghalaya on the results. The data from these states and the town can, therefore, be considered as provisional [13].

3 Categories of Crimes

3.1 Crime by Location

Location has a huge effect on the Indian crime. In 2012, Uttar Pradesh registered Indian states with the highest identifiable crime rate of 455.8, while Nagaland posted the lowest level (47.7) [14, 15]. National Crime Statistics Bureau had estimated the level as the number of incidents per 100,000 populations.

3.2 Crime Against Women

In India, violence against women refers to physical or sexual abuse of Indian women, which is generally perpetrated by a man. Domestic harassment, sexual assault, and murder are all prevalent types of violence against women in India. To be considered abuse against women, the offense must be done exclusively because the victim is female [16].

3.3 Organized Crime

In India, organized crime refers to elements of organized crime which originate in India and are also active in various parts around the globe. The mob also participates in illegal ventures originating in India as well as abroad [17, 18]. The Indian mob also applies to certain wealthy families with illegal aspects [19].

3.4 Cybercrime

Cybercrime, often known as computer-oriented crime, is a type of crime that involves the use of a computer and a network. The computer might have been either the perpetrator or the victim of a crime. Cybercrime may jeopardize an individual's or a country's safety and financial well-being. If sensitive information is intercepted or released, legally or otherwise, there are many questions regarding cybercrime [20]. Cybercrime growth rates had dropped in 2015 and 2016 before a dramatic spike was reported in 2017. Cybercrime increased by 69% in 2014 compared to 2013. In 2015, there was a rise of 20.5% in cybercrime in India with 11,592 cybercrimes as compared to 2014. In 2016, cybercrime rate continued to decline by 6.3% as 12,317 crimes were reported [21].

4 Related Work

Below are few applications that have been developed for the surveillance-related issue earlier.

4.1 Safetipin

Safetipin is one of the best choices when it comes to women's security apps. The app is developed with the idea of personal protection in mind. It includes all the necessary features such as GPS monitoring, directions to safe places, emergency contact numbers. The application also pins the protected areas together with their safety ratings to go when some problem occurs. This also helps users to detect dangerous areas and support others. Apart from English, Safetipin is available in Hindi, Bahasa, and Spanish [22, 23]. The application is free for all and uses crowd-sourcing for data collection. We have the Safety Audit at the core of the app which is a method for assessing a specific area based on the criteria of physical and social infrastructure [24].

4.2 Eyewatch SOS for Women

The safety app Eyewatch SOS captures the surrounding's video as well as audio of the user and delivers it to the registered contacts along with a warning message. The app has some appreciable features such as position accuracy, confirmation features for protection, and ability to operate without GPRS. The user can notify their near

ones by pressing the I am Safe button on reaching the destination securely. This is on Google Play Store and iTunes

4.3 bSafe

The bSafe is a fully featured-in-one safety application for women that provides functionality varieties. The bSafe alarm feature of the app delivers accurate video and audio of the surrounding along with the accurate location to the pre-selected contacts. The 'timer-alarm' feature enables to set an auto-alarm to keep your guardian updated about your whereabouts. To avoid any awkward situation, you can fake a call using the fake calling feature of the app. The user can be digitally monitored using the GPS tracking until reaching the location safely through the 'follow me 'feature. The app is available on the iTunes and Google Play Store. The business is working with warning centers, mobile carriers, and smartphone manufacturers in the USA, Asia, Europe, and Latin America, as well as non- profit corporations and organizations in the USA with an emphasis on surveillance and protection [25].

4.4 VithU App

For emergency situations, the VithU app is a step launched by Channel [V] [26]. Nobody has the time to call an emergency number when they are at risk. Using this app, a possible sufferer can avoid the time-consuming dialing process and can send an instant message by pressing the power button twice. Every two minutes, the contacts that have been chosen are notified with a warning message. The warning message contains the updated location which is auto-updated every time the warning is sent.

4.5 Nirbhaya: Be Fearless App

For emergency situations, another women safety app named the Nirbhaya app can be used to deliver a warning through a call or message with a single touch to your pre-chosen contacts. It delivers your exact position when triggered. With every 300 m you pass, the location is updated. Various different features such as 'shake to alert' and geo-fence are also included in the app. The power button can be used to send a warning signal instantly. It may also be used to convey conditions such as accident, aside from attempted molestation by means of predefined text messages. The user will be asked to request personal information about residential address and emergency contact numbers [27].

4.6 My Police Department

The uniqueness of the MyPD app which is available on iOS and Android for free is in its simplicity. An app that links you to over 200 U.S. wide police stations. This app is filled with forty different apps which fight against crime. These apps have features such as access to 'wanted' databases, missing child records, a record of molesters [28]. MyPD is currently in its fourth version. It is sleek and responsive [29].

5 Proposed Method

In a developing nation like India, the rising levels of crime play an important role in hindering the nation's development as a country [30]. Progress in the conduct of crime also has a greater effect on international investment as well as the country's reputation in the global community. Therefore, there is a strong need to track and evaluate, as well as control, the crimes that are happening across the country. Keeping in mind the need to reduce crime in India, we developed an IoT-based app that can play an important role in helping the government take better action to preserve law and order in the nation and also use its valuable resources effectively. To reduce the number of crimes and to provide evidence to police for investigation, this app is created which has features such as SOS button, safest path, instant calling buttons, awareness section, missing or wanted section, smart offline tools. The SOS button is a single tap button which can be accessed in an emergency situation. The safest path will provide you with the safest possible route for you in case of any emergency situation to go for one place to another. The instant calling buttons can be used to connect to the nearest essential services during emergency such as police, ambulance. Another feature of the app is missing or wanted section which displays all the missing or wanted people in your area. All these sections have been discussed in detail below.

5.1 SOS Button/Emergency Button

Fig 2 shows the SOS button. The features of SOS button are as follows.

- SOS button can be accessed by smart watch, power button, voice assistant, IoT-based jewelry you are wearing, the rupture of the cloth by using IoT technology.
- SOS button is clicked to send an alert message of your live location to the nearest police vans and police station.
- When the button is pressed, live audio and video recording also starts which is saved in your phone as well as onto the cloud.
- This feature will not allow to delete the evidence (audio and video recording) which definitely leaves the potential criminal with no option other than not committing the crime.

Fig. 2 Screenshot of SOS button of the proposed app

- When the SOS button is clicked, the app is taking full control over the phone, and the phone cannot be accessed anymore by anyone including the person who clicked it though it can be accessed if you stop the live video.

5.2 Safest Path

'Map provides you with the shortest path, we provide you with the safest path.' For anyone in need at any time, our app is capable of providing the safest path to move around freely. Based on several parameters like crime rate of the area, light, openness, visibility, people, security, public transport, feeling, which determine which path is more specifically safer. Fig 3 shows the safest route provided by the app.

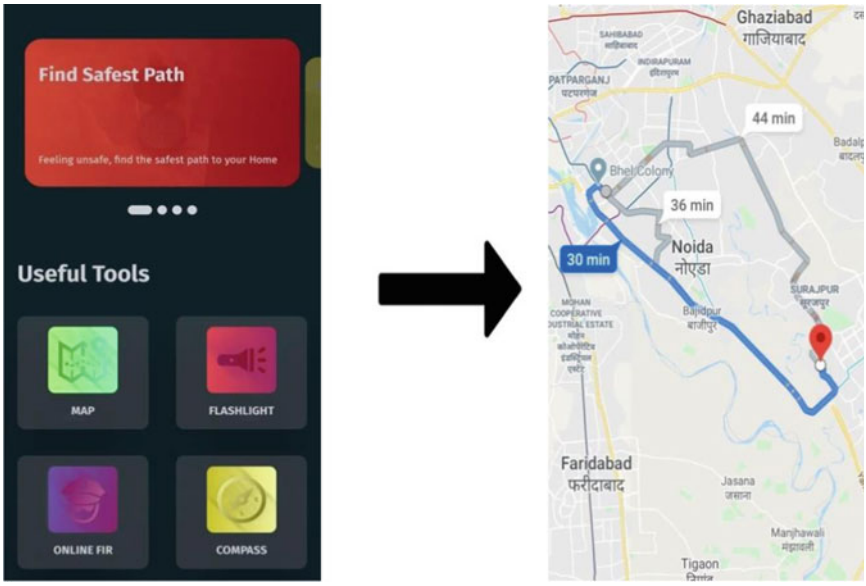


Fig. 3 Screenshot of the safest route provided by our app

5.3 Instant Calling Buttons

Instant call buttons are provided to approach for police, fire brigade, ambulance, bomb squads as soon as possible. When an instant calling button is clicked, it sends information to the nearest office as shown in Fig 4.

5.4 Missing or Wanted Section

MongoDB will be holding the data for all the missing or wanted people throughout, and then region-wise data will be displayed to the people. Such that if someone lives in Delhi, that person will be shown the missing or wanted people of that region only. Thus, missing or wanted sections will be changed according to region to region. Using this, common people will be involved to reduce the crime rate in India and to help to solve the crimes faster. Geo-server will be used for generating, sharing, and managing data. Fig 5 depicts the workflow of missing or wanted section.

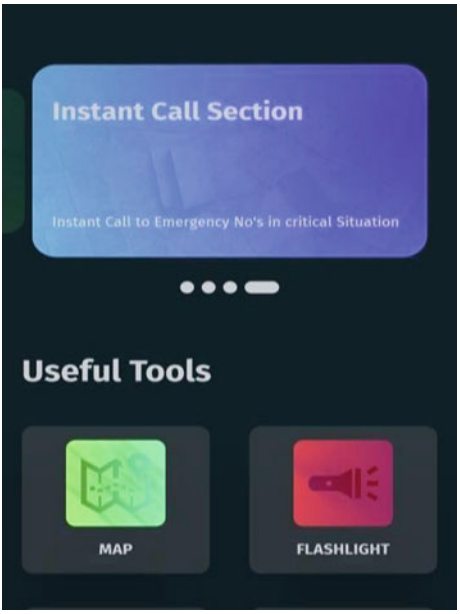


Fig. 4 Screenshot of instant call section of the app

Fig. 5 Workflow of missing or wanted section

5.5 Offline Smart Tools

Some useful offline smart tools available in the proposed app are that can be emergency situations as shown in Fig. 6.

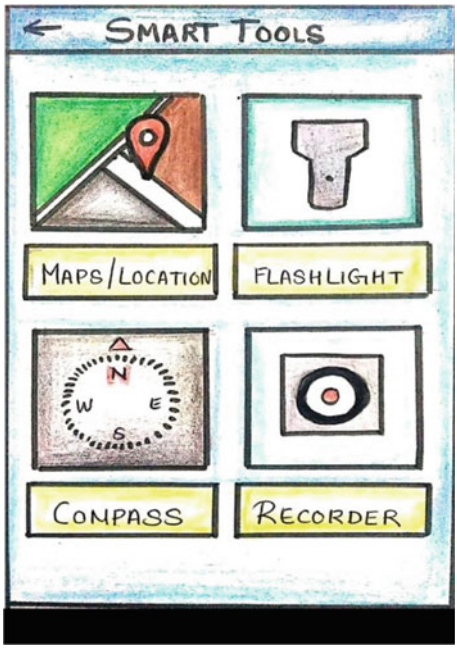


Fig. 6 List of useful tools available in our app

5.6 Benefits of the Proposed App

None of the applications which are discussed in the related work section are able to help in solving crime as effectively as our application. Benefits of the proposed application are as follows:

- One stop solution for the public to inform the police in real time.
- Can work even without Internet effectively in many use cases—geolocation has the ability to track your location without needing an Internet connection
- This is a very cost-effective application—we are using Apache Gateway, MongoDB, Android platform, languages used JAVA, for SMS gateway: OZEKI NG (.NET framework) all of which make the app cost efficient.
- It is integrated with some unique features such as safest path, missing/wanted section, SOS button.

6 Conclusion

Crime is a socio-economic problem which affects quality of life and economic development based on the nature of culture and environment. Previous crime forecast analysis has shown that variables such as schooling, poverty, jobs, and environment influence crime rate. Different studies have discussed crime prevention concerns,

but none have proven particularly successful. So, in this research paper, we have proposed an IoT-based surveillance android application. The proposed system will help the police solve crimes faster, and police will be able to use the common people as an asset to solve crimes. This is a one stop solution to inform police about the crime in real time. The whole system is low cost and multi-functional. Sometimes, because of the SOS button, the potential criminal might not even commit crime which results in reduction of the crime rate. Police will also get evidence to solve the case at a faster rate using our application.

References

1. Increasing Crime Rate in India, https://myessaymaster.com/essays/19851-increasing-crime-rate-in-india-essay1/
2. Brainly, https://brainly.in/question/3424876
3. YKA, https://www.youthkiawaaz.com/2019/10/ncrb-data-reveals-that-crimes-against-women-in-india-are-on-the-rise/
4. News Click, https://www.newsclick.in/achhe-din-crimes-against-women-dalits-and-adivasis-surge
5. Shamsuddin NHM, Ali NA, Alwee R (2017) An overview on crime prediction methods. In:2017 6th ICT international student project conference (ICT-ISPC). Skudai, pp 1–5, https://doi.org/10.1109/ICT-ISPC.2017.8075335
6. Viswanath B (2015) SafetiPin: an innovative mobile app to collect data on women's safety in Indian cities. Gend Dev 23. https://doi.org/10.1080/13552074.2015.1013669
7. Crime Statistics, http://mospi.nic.in/sites/default/files/Statistical_year_book_india_chapters/ch37.pdf
8. Brainly, https://brainly.in/question/8429992
9. Wikipedia, https://en.wikipedia.org/wiki/Crime_in_India
10. Times of India, https://timesofindia.indiatimes.com/india/crime-rate-drops-but-not-for-women-kids/articleshow/73187216.cms
11. Macrotrends, https://www.macrotrends.net/countries/IND/india/crime-rate-statistics
12. India Today, https://www.indiatoday.in/news-analysis/story/ncrb-2018-data-have-indians-become-less-criminal-more-civilised1635601-2020-01-10
13. Livemint, https://www.livemint.com/news/india/crime-in-india-increased-by-1-3-in-2018-ncrb-11578562577195.html
14. Wayback Machine, https://web.archive.org/web/20140620023952/, http://ncrb.nic.in/CD-CII2012/Statistics2012.pdf
15. Times of India, https://timesofindia.indiatimes.com/india/Kerala-is-countrys-most-crime-prone-state-NCRB-statisticsshow/articleshow/14364473.cms
16. Wikipedia, https://en.wikipedia.org/wiki/Violence_against_women_in_India
17. Can Narendra Modi bring Dawood Ibrahim back to India from Pakistan?. dna. Retrieved 26 Dec 2014. https://www.dnaindia.com/mumbai/report-can-narendra-modi-bring-dawood-ibrahim-back-to-india-from-pakistan1982500
18. Archived copy. Archived from the original on 2007-10-22. Retrieved 11 Oct 2006. https://web.archive.org/web/20071022174950/, http://flonnet.com/fl2004/stories/20030228003910800.htm
19. Know about Haji Mastan, the first underworld don of Mumbai. India TV News. Retrieved 26 December 2014. IndiaTV, https://www.indiatvnews.com/news/india/know-about-haji-mastan-the-first-underworld-don-of-mumbai-30938.html
20. Wikipedia, https://en.wikipedia.org/wiki/Cybercrime

21. The Hindu, https://www.thehindu.com/data/cyber-crime-cases-in-india-jumped-77-in-2017-compared-to-2016/article29889061.ece#:~:text=India%20recin%202017
22. Hongkiat, https://www.hongkiat.com/blog/android-personal-safety-women-apps/
23. Business World, http://www.businessworld.in/article/10-Safety-Apps-For-Women/12-06-2018-151793/
24. Safetipin, https://safetipin.com/
25. Getbsafe, https://getbsafe.com/
26. NothingGeek, https://nothinggeek.com/download-vithu-app-feelsafe/#:~:text=VithU%20app%20is%20an%20initiative,SOS%20message%20to%20their%20guardian&text=Apart%20from%20this%2C%20the%20app,the%20Crime%20Scene%20in%20India
27. Gadgets 360, https://gadgets.ndtv.com/apps/news/nirbhaya-app-launched-to-help-women-in-distress-314786
28. Concordia University, https://online.concordia.edu/criminal-justice-news/12-awesome-apps-to-help-fight-crime/
29. My Police Department App, http://mypdapp.com/
30. Ansari S, Verma A, Dadkhah KM (2015) Crime rates in India: a trend analysis. Int Crim Justice Rev 25(4):318–336. https://doi.org/10.1177/1057567715596047

A Comparative Study of PSO, PSO Variants, and Random Scheduling in Solving Workflow Scheduling Problem in Cloud Computing Environment

Ashish Tripathi, K. K. Mishra, Anand Bhushan Pandey, Arun Kumar Singh, and Vaibhav Tyagi

Abstract To obtain an optimal solution for an optimization problem, the most important and, hence, the crucial step is to make the right choice of the optimization algorithm from a diverse range of algorithms available. Cloud computing is a methodology that dynamically as well as simultaneously provides services and allocates resources to remotely residing users through Internet-based tools using the pay-for-use model. The unlimited storage, ease of use, backup and amp, recovery, and security are some of the features of a cloud computing environment which make its high demand even higher. This demand is the reason that calls for better cloud optimization which minimizes the processing cost using scheduling policies and algorithms. The objective here is to minimize the total cloudlet processing cost using an optimal scheduling algorithm on virtual machines. In this paper, the optimization heuristics named particle swarm optimization, its variants, and random scheduling are compared. The comparison of the results shows that different variants of the PSO heuristic perform better in comparison with random scheduling. Among different variants of PSO used for task scheduling purposes, PSO using constriction factor and PSO variant using both linear decreasing inertia weight (LDIW) and amp; constriction factor is found to perform better than others.

Keywords M particle swarm optimization · CloudSim · Task scheduling · PSO heuristic · PSO variants · Random scheduling · Clouding computing

A. Tripathi · A. B. Pandey · A. K. Singh · V. Tyagi (✉)
Department of Information Technology, G. L. Bajaj Institute of Technology and Management, Greater Noida, India

K. K. Mishra
Department of Computer Science, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, UP 211002, India
e-mail: kkm@mnnit.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_6

1 Introduction

Cloud computing is the delivery of computing services, which are accessed from the Internet, using web-based tools and applications. It enables the remote access of services, and it uses a pay-for-use model for utilizing their service. It is a computing methodology, which involves dynamic, instantaneous allocation of resources whenever they are needed. Some of the benefits of cloud computing over traditional hosting are that it provides unlimited storage, backup, and recovery, security, scalability, speed, energy-efficiency, and ease of access. Due to the aforementioned advantages, its demand is on a rapid rise. This increase in demands of cloud computing calls for better cloud optimization. Thus, it is highly important to study better scheduling policies in the cloud. Several evolutionary algorithms have been employed for this purpose, and we implemented different variations of PSO heuristic and observed their performance using CloudSim. The problem at hand is to minimize the total cost of processing cloudlets by optimal scheduling on given virtual machines. Our hypothesis is that we can obtain better cost using particle swarm optimization to schedule our tasks as compared to the random scheduling algorithm.

Particle swarm optimization is an optimization heuristic that is inspired by the study of a flock of birds. The birds search for their food in a flock where the direction of every bird depends on its best position and the best position of all birds in the swarm. It is a widely popular heuristic used in solving various optimization problems. Relay protection coordination in distribution networks and economic dispatch of generators in the grid are defined as two power system-related optimization problems that can be solved using PSO. PSO is used in the railway domain mainly for network layout planning, scheduling of trains, and active controls. PSO variants such as genetic PSO, chaotic PSO, and quantum-behaved PSO are also used in the railway domain. Electronic chip design is another area where PSO can play a vital role. Cloudlet scheduling has a number of challenges associated with it. Some of these challenges may include ensuring efficient resource utilization, heterogeneous physical nodes in cloud data centers, and dynamic and fluctuating workloads.

2 Literature Survey

Pandey et al. [1]: They advocated the use of particle swarm optimization to schedule different applications to cloud resources taking into consideration both data transfer cost as well as data computation cost. They tried to study the effects of varying the computation and communication cost. They also compared the cost calculated using PSO and best resource selection algorithm.

Senthil et al. [2]: They discussed a version of particle swarm optimization heuristic which combines the global best and the local best model and named it GLBest-PSO. The GLBest-PSO uses a global–local best inertia weight (GLBest-IW) with

a global–local best acceleration coefficient (GLBest-Acc). The velocity equation of the GLBest-PSO is also simplified.

Kennedy [3]: They put forward the concept of using PSO and other evolutionary algorithms to solve the different nonlinear optimization problems. They also described the use of PSO in neural network training for feature selection and described the benchmark testing and other proposed applications.

Angeline [4]: He tried to differentiate between the performance of evolutionary and particle swarm techniques and their working methods. Four different nonlinear functions were studied to highlight the differences between the two heuristics.

Shi and Eberhart [5]: First, they studied the impact of inertia weight and maximum velocity on the performance of the particle swarm optimization, and then provided guidelines on basis of which these two parameters should be selected.

Xie et al. [6]: In distributed environment, workflow scheduling is an NP-hard problem; task scheduling cannot be performed well having multiple optimization goals using existing approaches. PSO has the advantages of having a simpler algorithm, fewer parameters, and higher convergence speed, so it is applied to workflow scheduling. A novel directional and non-local-convergent particle swarm optimization (DNCPSO) is proposed which uses nonlinear inertia weight along with mutation and selection operations which reduces the makespan and cost.

Wang et al. [7]: Because of a large number of resources available for the users', cloud computing brings new challenges in scheduling and optimization, and cost and makespan are the issues to focus upon in workflow scheduling. A low cost and fast makespan method are needed for users. An immune particle swarm optimization algorithm (IMPSO) is proposed, which improves the quality and enhances the speed of the optimization. The proposed IMPSO solves the problem of slow convergence of PSO, which falls in local optimization.

Nagar et al. [8]: The authors propose a new PEFT genetic algorithm approach which reduces the execution time on the workflow scheduling model 'predict earliest finish time.' In this approach, GA focuses on to optimize chromosome's objective to obtain the best suitable mutated children. The GA optimizes the execution time when a feasible solution is obtained. Experimental results prove that this approach finds a better solution in lesser time.

Hosseinzadeh et al. [9]: The paper focuses on metaheuristic multi-objective optimization context and presents an overview of the multi-objective scheduling strategies designed to be used in different cloud computing environments. A comparison of the multi-objective scheduling schemes is also provided.

3 Proposed Work

3.1 Particle Swarm Optimization with Cloud

3.1.1 Initial Parameters Involved

Number of cloudlets, number of virtual machines, number of hosts, task length of each cloudlet, MIPS rating of VM, output file size of each cloudlet, execution cost of each VM, number of iterations, number of data centers, and number of particles [10].

3.1.2 Assumptions

For our simulation purposes, we assumed that there are only one data center and one user. All the hosts are present in the same data center. Furthermore, we assumed that there are two hosts involved. Cloudlet ids and VM ids are indexed from zero. No cloudlet migration takes place. A cloudlet when scheduled on a VM cannot switch to another VM even if it is unoccupied.

3.1.3 Particle

The particle is a 1-D array of length equal to the number of tasks. Therefore, the number of dimensions for the particle is the number of tasks. The value in each array index is the corresponding VM id. Here, the index of the array represents the cloudlet id, and the array value represents the VM id. The particle is also referred to as mapping of cloudlets to VM [11].

3.1.4 Fitness Function for Independent Tasks

Fitness function uses execution cost provided, task length, MIPS rating of VMs to calculate the cost based on given particle.

1. Execution time of all cloudlets is calculated using task length of all tasks and MIPS of all VMs using Eq. (1).

$$\begin{aligned} &\forall_i | i \in \text{no. of cloudlets} \\ &\forall_j | j \in \text{no. of VM}_s \\ &\text{Exec. Time}_{i,j} = \frac{\text{Task length}_i}{\text{MIPS}_i} \end{aligned} \tag{1}$$

2. Execution cost of at VMs is calculated using Eq. (2).

$$\forall_i | i \in \text{no. of cloudlets}$$
$$\text{VM cost}_{\text{Map}(i)} = \text{Exec. time}_{i,\text{Map}(i)} * \text{Exec. cost}_{\text{Map}(i)} \quad (2)$$

3. The total cost of all cloudlet execution is calculated using cost at all VM by using Eq. (3).

$$\forall_i | j \in \text{no. of } M_s$$
$$\text{Cost} = \sum_{j=0}^{\text{No. of VM}_s} \text{VM cost}_j \quad (3)$$

A directed acyclic graph (DAG) is used to represent the task dependency [12]. DAG is represented by an adjacency matrix where $M\ [i, j] = 1$ represents that task i sends output file to task j and $M\ [i, j] = 0$ represents that it does not.

3.1.5 Fitness Function for Dependent Tasks

Fitness function uses execution cost provided, task length, MIPS rating of VMs, output file size of each task, data transfer speed, communication cost, DAG to calculate the cost based on the given particle.

1. Execution cost of given cloudlets on VMs is calculated using task lengths and MIPS of all VMs given by Eq. (1).
2. The directed acyclic graph (DAG) is used to represent the relationship between the cloudlets. A task is said to depend on other task(s) when it needs the output file of preceding task for its execution. It is represented by Eq. (4).

$$\text{DAG}_{i,k} = \begin{cases} 1 \text{ if task } k \text{ depends on task } i \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

3. Task output file matrix denoted by Eq. (5) represents the amount of data cloudlet i has to send to cloudlet k for cloudlet k to start executing.

$$\forall i, k | i, k \in \text{no. of cloudlets and } k > i$$
$$\text{Task output file}_{i,k} = \text{Output file size}_i * \text{DAG}_{i,k} \quad (5)$$

4. Communication time between two dependent tasks is calculated using task output file matrix and data transfer speed represented by Eq. (6).

$$\forall i, k | i, k \in \text{no. of cloudlets and } k > i$$
$$\text{Comm. time}_{i,k} = \frac{\text{Task output file}_{i,k}}{\text{Data transfer speed}_{i,k}} \quad (6)$$

5. Waiting time of tasks is also a contributing factor for total cost. The waiting time of every cloudlet is calculated as in Eq. (7).

$$\begin{aligned} &\forall i, k | i, k \in \text{no. of cloudlets and } k > i \\ &\text{Wait time}_k = \max(W\text{ait time}_k,\ \text{Wait time}_i \\ &\quad + \text{Exec. time}_{i,\text{map}(i)} + \text{Comm. time}_{i,k}) \end{aligned} \tag{7}$$

6. The total cost is the sum of execution cost at all VMs, waiting cost of all cloudlets, and communication cost involved due to transfer of data between cloudlets. Total cost is calculated as represented by Eq. (10).

$$\begin{aligned} &\forall_i | i \in \text{no. of cloudlets} \\ &\text{VM cost}_{\text{Map}(i)} = \text{Exec. time}_{i,\text{Map}(i)} * \text{Exec. cost}_{\text{Map}(i)} \\ &\forall i, k | i, k \in \text{no. of cloudlets and } k > i \end{aligned} \tag{8}$$

$$\begin{aligned} &\text{VM cost}_{\text{Map}(i)} = \text{VM cost}_{\text{Map}(i)} \\ &\quad + \text{Comm. time}_{i,k} * \text{Comm. cost}_{i,k} \\ &\forall j | j \in \text{no. of VM}_s \\ &\text{Cost} = \sum_{j=0}^{\text{No. of VM}_s} \text{VM cost}_j \end{aligned} \tag{9}$$

$$\begin{aligned} &\forall i | i \in \text{no. of cloudlets} \\ &\text{Cost} = \text{Cost} + \sum_{i=0}^{\text{No. of cloudlets}} \text{Wait time}_i * \text{Wait cost}_{\text{Map}(i)} \end{aligned} \tag{10}$$

3.2 Approach Used for Simulating Task Scheduling of Independent Tasks

The initial parameters are taken, and CloudSim is set up for simulation. The parameters are passed to define the fitness function that calculates cost of particle based on the execution time of cloudlet and execution cost of VM. A swarm of particles is created with random positions and velocities. The best fitness (cost) is calculated, and the particle is used for scheduling cloudlets to VMs by passing mapping to data center broker. A random mapping is created, and the corresponding cost is calculated. Both the costs are compared.

3.3 Approach Used for Simulating Task Scheduling of Dependent Tasks

The initial parameters and task dependency graph are taken, and CloudSim is set up for simulation. The parameters are passed to define the fitness function, which calculates the cost of particle based on the execution time of cloudlets, communication time between cloudlets, waiting time of cloudlets, and corresponding costs [13]. A swarm of particles is created with random positions and velocities. The best fitness (cost) is calculated, and the particle is used for scheduling cloudlets to VMs by passing mapping to data center broker. A random mapping is created, and the corresponding cost is calculated. Both the costs are compared.

3.4 Realization of Task Scheduling Using Oracle VirtualBox

The following approach can be used for the execution of tasks in different VMs according to the mappings found by different variants of PSO heuristic and random scheduling. The overview of the setup for task scheduling in VirtualBox and execution workflow is shown in Figs. 1 and 2, respectively.

- First, set up the required number of VM instances using Oracle VirtualBox.
- The specifications of the VMs must be same as provided while simulation in CloudSim for accurate results.
- For the purpose of transferring files from the host system to the VM instances, a shared folder was set up between the VMs and host system.
- The required numbers of cloudlets is kept in the shared folder. The details of cloudlets must be kept same as mentioned while simulation in CloudSim.
- The host system runs PSO heuristic and random scheduling algorithm and generates the corresponding mapping of cloudlets to VMs.

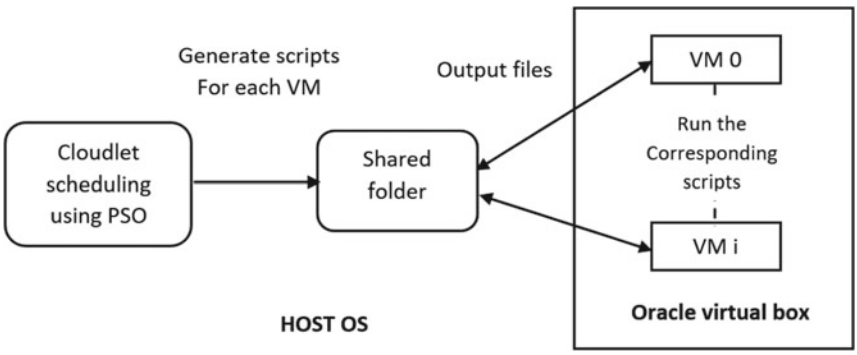


Fig. 1 Overview of the setup

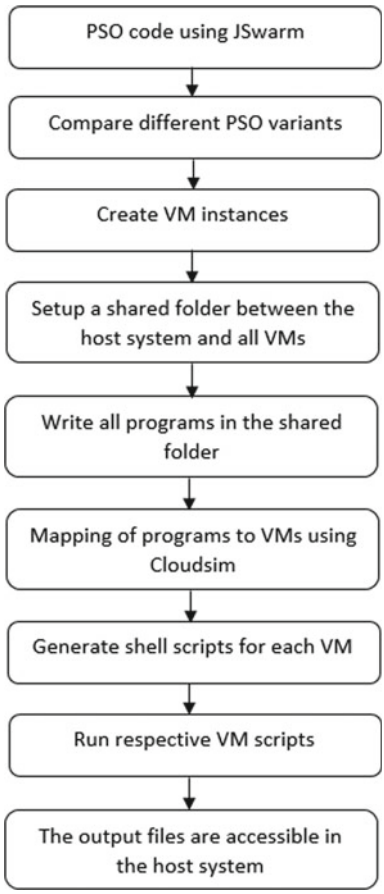


Fig. 2 Task execution workflow

- Shell scripts corresponding to every VM are generated keeping in mind the dependency of tasks and to calculate the time of execution of all programs on VM.
- The shell scripts are used to run the programs on VMs, and corresponding output files are generated which the host system can access from the shared folder.

4 Results Analysis and Observations

4.1 Graphical User Interface

The results are obtained from analysis of performance of different PSO heuristics. The 5.1 experimental setup: language used—Java, operating system—Windows 10,

RAM—8 GB, CloudSim version—3.0.3, JSwarm version—2.08, Oracle VirtualBox version—5.1.28, and Ubuntu OS version—14.04.3.

- Parameters involved in comparing PSO: (number of dimensions—2, 5, 10), (number of particles—30), (number of iterations—1500), and (number of runs per function—7).
- Parameters involved in comparing multiswarm PSO: (number of dimensions—2, 5, 10), (number of particles—5), (number of swarms—6), (number of iterations—4000), and (number of runs per function—5).
- Parameters involved in cloudlet scheduling: The results were calculated for following details (Figs. 3 and 4).

(Number of cloudlets—10), (number of VMs—8), (number of particles—30), (number of iterations—10,000), (number of runs per PSO variant—30), (task length of cloudlets (in mi)—3000, 2000, 1000, 5000, 4000, 3500, 2500, 1500, 6000, 1300), (output file size of cloudlets (in mi)—300, 400, 100, 500, 350, 700, 400, 800, 1000, 550), (millions of instructions per second (MIPS) rating of virtual machines (in mi/s)—80, 50, 10, 10, 5, 15, 20, 60), (execution cost of VMs per second (in dollars per second)—6, 10, 2, 0.5, 0.5,4.5, 2, 7), (communication cost per second between any two cloudlets (in dollars per second)—3), (data transfer speed between any two VMs (in mi/s)—80), (waiting cost of VMs per second (in dollars per second)—6, 10, 2, 0.5, 0.5, 4.5, 2, 7), and (task dependency—explained in Fig. 5). Here, mi/s stands for millions of instructions per second (Figs. 6 and 7).

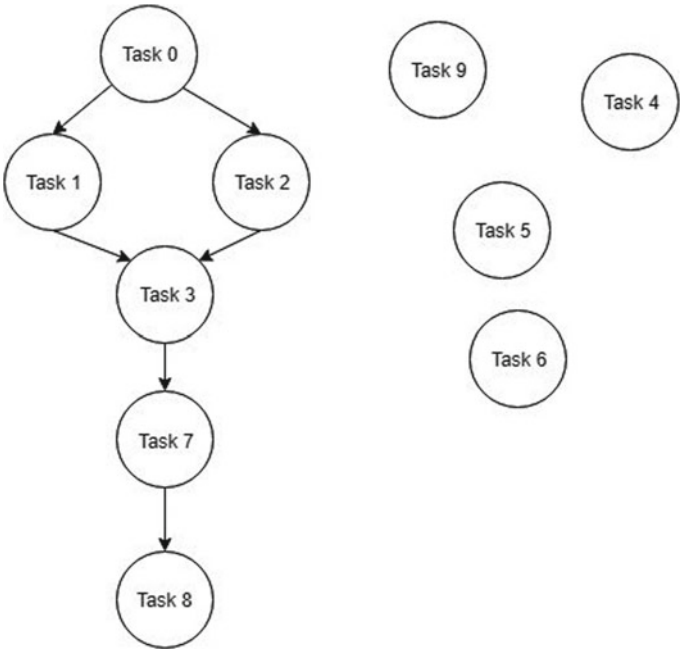


Fig. 3 Relationship of tasks considered for calculating results of dependent tasks

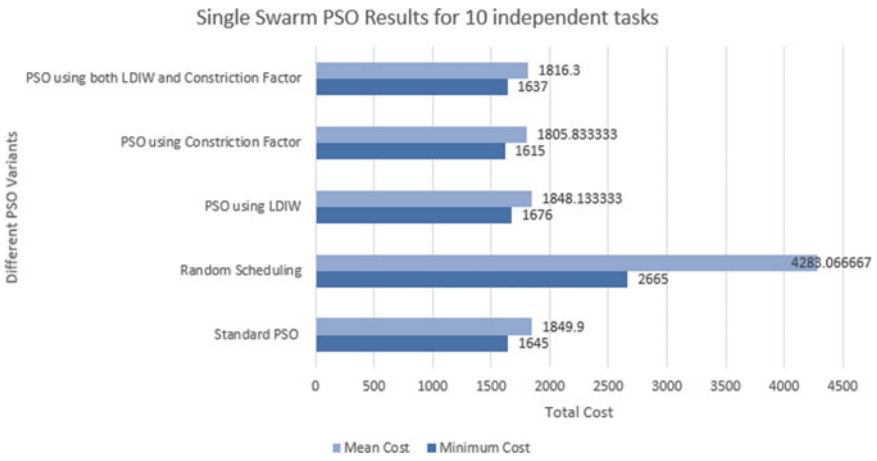


Fig. 4 Results of 30 runs of cloud simulations for different PSO variants having 10,000 iterations each and random scheduling for 10 independent tasks on 8 VMs

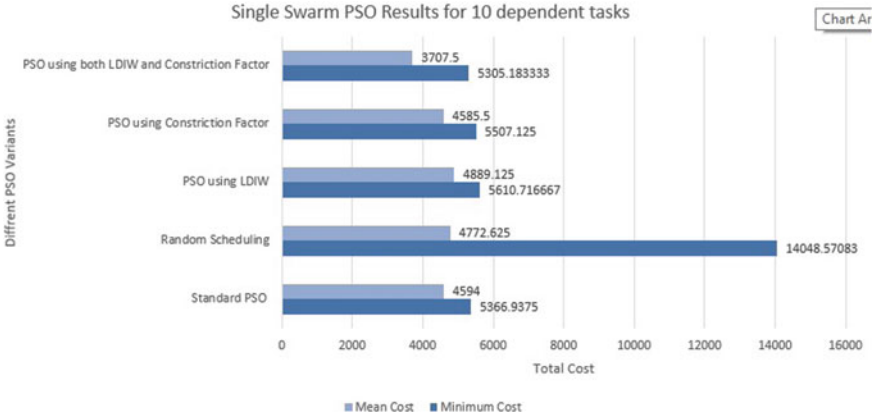


Fig. 5 Results of 30 runs of cloud simulations for different PSO variants having 10,000 iterations each and random scheduling for 10 dependent tasks on 8 VMs

It was observed that the cost of cloudlet execution by using the optimal mapping of cloudlets to VM found by PSO is usually less than mapping given through random scheduling. Using the above observations, we formulate the result that different variants of PSO heuristic performed better in comparison to random scheduling. Among different variants of PSO used for task scheduling purposes, PSO using constriction factor and PSO variant using both linear descending inertia weight (LDIW) and constriction factor are found to perform better than others. The mapping of tasks to VMs found by the PSO heuristic is then used to assign tasks to Ubuntu VM instances in Oracle VirtualBox for both dependent and independent tasks.

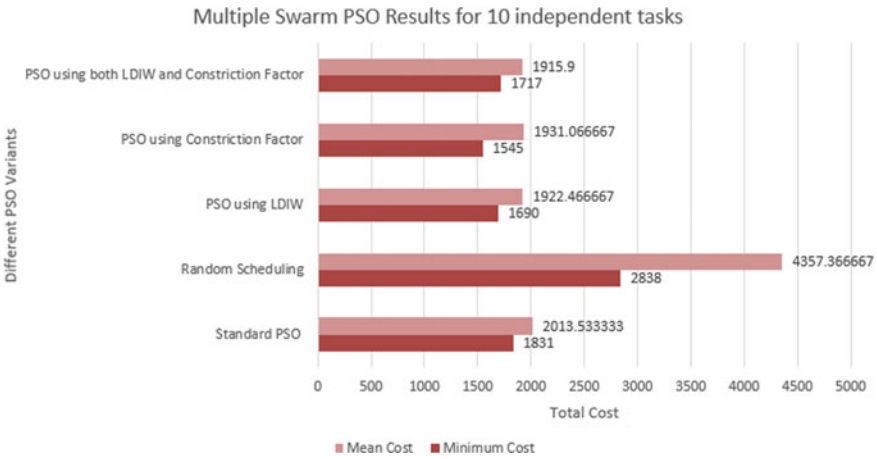


Fig. 6 Results of 30 runs of cloud simulations for different PSO variants having 10,000 iterations each and random scheduling for 10 independent tasks on 8 VMs

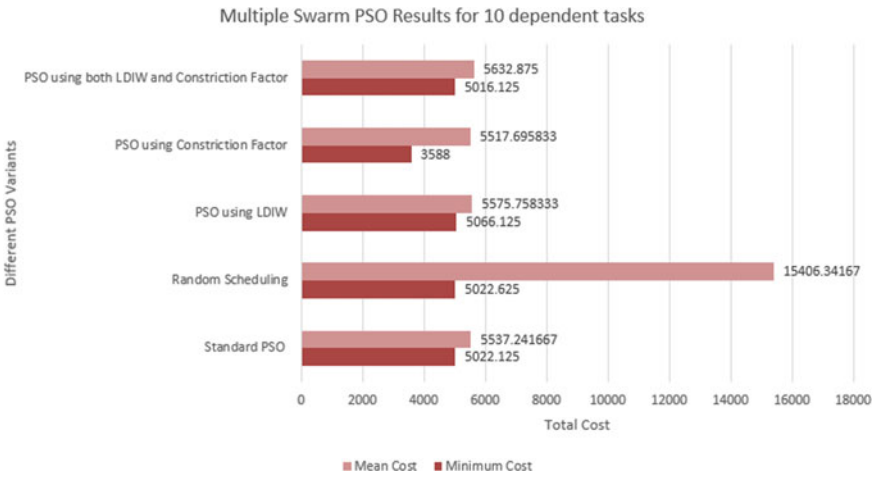


Fig. 7 Results of 30 runs of cloud simulations for different PSO variants having 10,000 iterations each and random scheduling for 10 dependent tasks on 8 VMs

5 Conclusion and Future Work

The different variants of PSO heuristics were tested, and three different approaches to tackle the cost minimization problem of cloudlet scheduling were studied: First using a PSO variant involving a single swarm, PSO variant involving multiple swarms, and one approach using random scheduling were simulated. Recently, some more heuristics have begun to be used in cloudlet scheduling. Some of them include ant colony optimization (ACO), cuckoo search. We can try to implement cloudlet scheduling against these heuristics and compare their results against PSO. Due to computational constraints, we could only run and verify our PSO up to ten dimensions. In the future, we could try to run our PSO for 30 and 50 dimensions and run it on physical machines instead of virtual simulation. A certain number of machines can be taken, our tested code can be implemented on them, and we can compare our simulated results with the real-time results.

References

1. Pandey S, Wu L, Guru SM, Buyya RA (2010) Particle swarm optimization-based heuristic for scheduling workflow applications in cloud com- puting environments. In: 2010 24th IEEE international conference on advanced information networking and applications (AINA). IEEE, pp 400–407
2. Arumugam MS, Rao M, Chandramohan A (2008) A new and improved version of particle swarm optimization algorithm with global–localbest parameters. Knowl Inf Syst 16(3):331–357
3. Kennedy JER (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, vol 4, IEEE Press, pp 1942–1948
4. Angeline PJ (1998) Evolutionary optimization versus particle swarm optimization: philosophy and performance differences. In: International Conference on Evolutionary Programming, Springer, Berlin, pp 601–610
5. Shi Y, Eberhart RC (1998) Parameter selection in particle swarm optimization. In: International conference on evolutionary programming. Springer, Berlin, pp 591–600
6. Xie Y, Zhu Y, Wang Y, Cheng Y, Xu R, Sani AS, Yuan D, Yang Y (2019) A novel directional and non-local-convergent particle swarm optimization based workflow scheduling in cloud– edge environment. Futur Gener Comput Syst 97:361–378
7. Wang P, Lei Y, Agbedanu PR, Zhang Z (2020) Makespan-driven workflow scheduling in clouds using immune-based PSO algorithm. IEEE Access 8:29281–29290
8. Nagar R, Gupta DK, Singh RM (2018) Time effective workflow scheduling using genetic algorithm in cloud computing. Int J Inf Technol Comput Sci 10(1):68–75
9. Hosseinzadeh M, Ghafour MY, Hama HK, Vo B, Khoshnevis A (2020) Multi-objective task and workflow scheduling approaches in cloud computing: a comprehensive review. J Grid Comput 1–30
10. The cloudsim framework: Modelling and simulating the cloud environment. https://opensourceforu.com/2014/03/cloudsim-framework-modelling-simulating-cloud-environment/
11. Resource allocation policy in cloudsim environment image in dynamic virtual machine allocation policy in cloud computing complying with service level agreement using cloudsim. https://iopscience.iop.org/article/10.1088/1757-899X/263/4/042016/pdf

12. Hu P, Rong L, Liang-lin C, Li-xian L (2011) Multiple swarms multi-objective particle swarm optimization based on decomposition. Procedia Eng 15:3371–3375
13. Lu Y, Liang M, Ye Z, Cao L (2015) Improved particle swarm optimization algorithm and its application in text feature selection. Appl Soft Comput 35:629–636

Early Detection of Road Abnormalities to Ensure Road Safety Using Mobile Sensors

Mohit Misra, Prashant Mani, and Shailesh Tiwari

Abstract In today's scenario of continuous development and reconstruction of the localities and surrounding buildings, factories, companies, flyovers, etc. and increase in the vehicular traffic have resulted in building speed breakers. Although these were made in order to control the speed of the vehicles for the safety of the people and mainly to reduce road accident casualties. But on the contrary the incorrect construction or unnecessary construction of breakers in front of their properties is leading to an increase of accidents. One another demerit to the road safety is the poor building of roads and its potholes. One way is that various organizations should do their work properly so as to improve road safety. But, rather than waiting for organizations to do the maintenance work we can work for our safety through various tools, which works on software codes. Since, slowly and slowly the lifestyle is becoming technology based, the major focus here is the designing of a system for detecting both bad road conditions as well as speed breakers with an already bonding. The approach which has been used in the paper is a real-time approach and has been developed by android service which actually runs in the background and also it relies on the application of Google maps which is present in the smartphone. The service will basically provide an early warning if a speed breaker or a bumpy road is about to be present in the way. Also it provides an alternative route as well as a better route apart from these already alerts to the user.

Keywords SVM · Android application · Accelerometer · Speed breakers · Bad roads

M. Misra (✉) · P. Mani
SRM Institute of Science and Technology, NCR Campus, Delhi Meerut Road, Modingar, UP, India

S. Tiwari
Krishna Engineering College, Ghaziabad, UP, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_7

1 Introduction

Due to the growth of population as well as city expansion there is an increase in the weight of traffic as well as the volume due to which there is an deterioration of different transportation infrastructures. Also the different causes for such a destruction our various effects of bad construction material use, deficiency present in the design of the road, climate change. In combination all these issues basically result in the appearance of different anomalies in the roads like bumps, potholes, cracks which can be found in different expressways highways and streets present worldwide. The effects caused due to these anomalies can elevate from being costly damages to the vehicles to various fatal accidents.

Not only these anomalies but illegal construction of the safety breakers is also becoming a threat leading to deterioration of the road safety. Speed breakers are generally constructed in order to provide safety for the pedestrians in different zones in order to control the limit of speed for the vehicles which helps in avoiding of accidents (Table 1).

There are a lot number of unauthorized speed breakers that have been laid unnecessarily which are known to not follow the actual standard size which has been proposed by the national highway authorities. The speed breakers are very common in the countries like India due to the fact that different signboards like speed limit, stop, yield, etc. will not be able to work because of a lack of the traffic to enforcement resources.

This generally happens in the country because of various reasons such as higher speed of vehicle, negligence of the driver, lesser visibility at night. There are different incidents that have been reported where vehicles such as cars, scooters, motorcycles are much more vulnerable because some speed breakers which are unnoticed may cause them to lose balance and can lead to severe accidents as well as damages.

In the year 2016, India had 300,000 road deaths which is actually double the amount of the estimation by the government which was one 151,000. In accordance to 2018, global road safety report provided by world health organization which basically highlighted the lack of the amount of data on fertility road accidents. According to

Table 1 Comparative analysis of last 5 years road accidents due to road conditions

Year	Total number of road accidents	% change	Total numbers of person killed	% change	Total number of person injured	% change
2014	489,400	2.21	139,671	4.21	493,000	4.41
2015	502,321	2.46	146,133	4.63	500,278	1.38
2016	480,652	−4.14	150,785	3.18	495,294	−1.13
2017	464,910	−3.28	147,913	−1.9	475,274	−4.78
2018	467,044	0.46	151,417	2.37	469,297	−0.33
2019	510,000	2.41	164,418	2.51	488,212	−2.21
2020	515,000	3.21	168,413	3.12	476,298	−2.25

the data of the government about 1.5,00,00 people have lost their lives in 2017 due to road accidents which basically means that 17 people die due to road accident every hour.

The number of fatalities is actually one third of the amount of accidents which basically results in about 53 road accidents in every hour. The worst state affected is Uttar Pradesh with 20,124 people who lost their lives in such a road accidents, which has been followed by Tamil Nadu where 16,157 people have lost their lives. Also Rajasthan, Karnataka and Maharashtra are a part of the top five states. Delhi is considered to be one of the safest places for driving considering the various numbers of fatalities that take place.

Different solutions have been considered in order to automatically detect as well as report any kind of anomalies in the road to different government agencies so that the maintenance tasks can be accelerated. For example, by making the use of different computer vision methods which are based on texture differences as well as shape segmentation in order to identify the potholes. Approaches similar to this like contour information, edges, adopting shape, etc. have been considered (Figs. 1 and 2).

In today's time majority of people possess android mobile phones or smartphones which are well equipped with various inbuil

t apps such as navigation, Google maps, sensation systems like Gyro sensor, accelerometer and magnetometer are enabled with GPS and always connected to Internet. The technology of smartphones has been adopted in order to tackle such a problem provided its geo referencing, sensing capabilities.

The accelerometers in smartphones can detect the movement of the device in such a way that if the vehicle comes across any irregular road surface like a bomb or a pothole, the accelerometer will record the occurrence of this event. The major problem then is the identification of the series of continuous readings of the accelerometer when any kind of anomaly occurs. It does becomes important to emphasize that even though the problem for the identification of road anomalies using smart phone is well defined but still there are some issues that have prevented the company from getting much more experience and a complete perspective. It also enables and motivates the development of an android application which will use the data that has been generated by these sensors present in the smartphones from the user to alert the user about the bumps and bad road situation.

2 Literature Work

Although a lot of methodologies as well as strategies have been considered in order to tackle the problem by using accelerometer sensors. This work basically aims a comparison and evaluation of various high impact works that have motivated major developments in this field. Many of the existing systems use hardware additional to the device in order to detect different speed breakers which can be costly. The existing approaches have been discussed in this section.

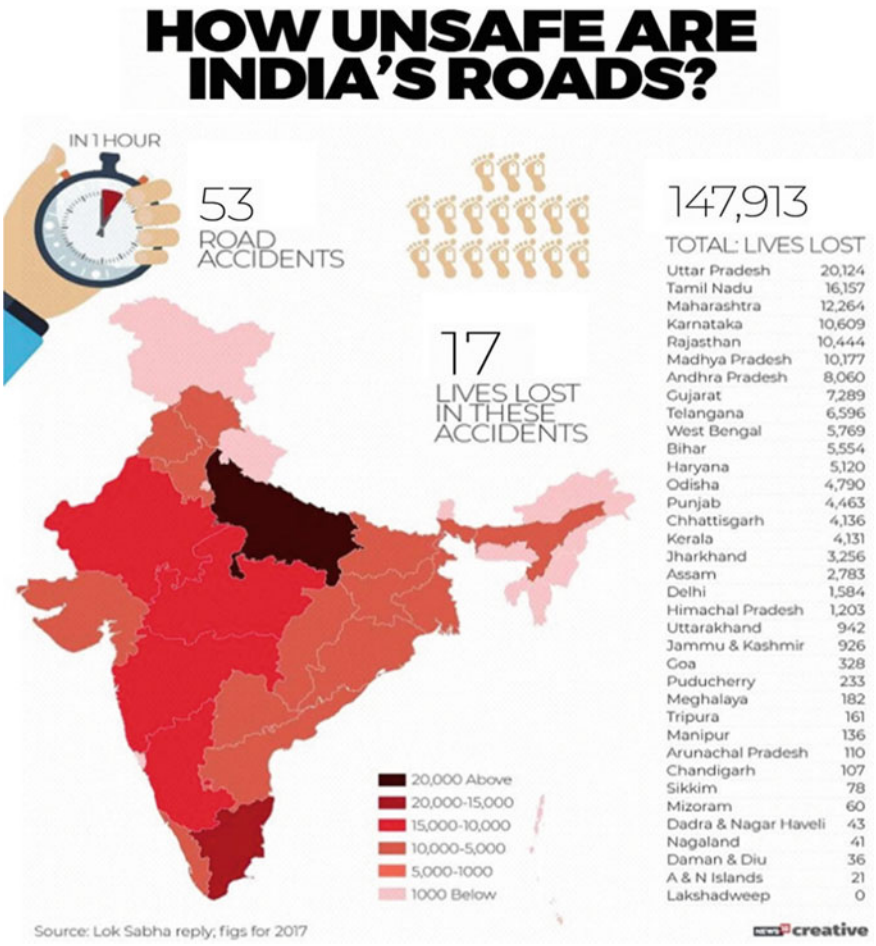


Fig. 1 State-wise road accident analysis

Accelerometers as well as magnetometer are the main sensors which are being utilized by the researchers and developers to overcome or to propose a solution to tackle this problem in order to reduce the casualty rate of per 17 h. Since, the magnetometers are expensive and not all current available smartphone or android phones do not have this. So, initially various works and developments were done and tested with accelerometers. Different types of anomalies can exist when working with vertical acceleration [1–3].

Thus, there is a realization that it is important what type of anomaly it is. Because there are many anomalies such as cracks, metal bumps, potholes, uneven roads, the main task is thus to assign it an accurate label, i.e., it results in a multiclass identification problem. To make this identification process less hectic, the problem of identification is being dissolved into a binary situation [4, 5]. That is there is

only requirement of identifying whether there exists any anomaly or not without considering its type on the road.

In *Pothole Patrol's* [6] approach, the authors used tri-axial accelerometers and GPS which are installed at fixed positions in seven taxis that travel a distance of 9730 kms approximately which collects data while driving.

The main purpose of this was the processing along with the threshold-based filters for detection of different road anomalies as well as the differentiation of potholes from other anomalies. The conclusion basically presents that less than 0.2% of non-identified potholes have been reported and most of their detection is data that is uncontrolled corresponding to the actual potholes.

Another approach, in which the authors and researchers considered using the smart phones for sensing in order to provide the monitoring of traffic conditions in the city. Rather than identifying various other anomalies, this approach addresses the detection of bumps and potholes. The proposal provided is actually based on the application of two different procedures consecutively. Z-PEAK and Z-SUS [7]. The authors observed and concluded this approach with 20–30% false negative rate.

Another system proposed in [8], in this system accelerometer and magnetometer is the sensors that have been used. The accelerometer has been used for the collection of data for traffic monitoring as well as the detection of different road bombs. Whereas the magnetometer has been used for the reorientation of access off the vehicle as well as the device since the placement of the device inside the vehicle could be at a different arbitrary orientation. The first stage ensures that the geometric axes as well as the axis of the phone should be aligned.

In the next stage, the axes of the new device are aligned with the axes of the vehicle. In order to carry out this approach the GPS has been used to determine as well as locate the direction in which the motion of the vehicle would be there in order to find the angle of the motion for the vehicle using the magnetic north in order to transform and rotate the axis of the vehicle according to the axis of the device.

Conclusion of this approach is, in order to detect bombs, the system provided a false negative rate of 10% and for the detection operating system actually produces a first negative rate of 21.6% and a positive rate of 2.7%. The main limitations of such a system or that there is a requirement for magnetometer in order to reorient but the magnetometer is actually not present in all the phones and it also can cause an increase in the consumption of battery.

Eriksson et al. [6] used a signal processing as well as machine learning approach to detect as well as classify several specific road anomalies. The paper describes the issue of having other unwanted anomalies manifest in the accelerometer signal, for instance abrupt turning or stopping.

Different characteristics, such as speed and filters, such as a high-pass filter, were used to reject such events. The training of the learning algorithm is based on the peak X axis and Z axis acceleration values and the instantaneous velocity of the car. To help remove false positives and increase the overall accuracy of correct anomaly detection, the paper describes a process where an anomaly is reported only when several other detectors have also detected an anomaly in the same spot.

Overall, the paper concludes, that by using training data, which had been carefully examined, the described pothole detection system achieved a false positive rate of 0.2% in controlled experiment.

Medniset al. [9] in a research paper gave a system that can be used to detect potholes. The system consists of smart phones with OS that have accelerometer in it for detecting the acceleration. It is made up of sensors and simple algorithms. One of the limitation on this work is that it uses sensors and accelerometer and data is collected using hardware only.

Remote sensors and analog devices are used to get the accelerometer data in this study [7]. The peak value influenced the results of the *Z*-Thresh approach and *G*-Zero approach which generated more false positives. Frequency and timing are the contributing factors in the *Z*-DIFF AND STDEV-*Z* approach. According to [10] the contrasting and design of devices is required to be studied upon.

Many amount of resources were allocated for this system for having efficiency as a mobile device. Also it had requirement of very huge amount of computation power. So because of this reason the sensing methods involving mobile phones are considered more efficient for detecting potholes. But the issue was that the previous approaches had obtained large number of false positives while detecting the potholes [11–13]. So Mednis's methodology [9] was proposed to improve the accuracy of the process of pothole detection. There was an increase in accuracy by this process.

- *Z*-Thresh: It is based on *Z* axis value of accelerometer and gyroscope. Basic idea is to identify the change in gravity value during the movement of vehicle on different type of roads [10]. The value of *Z* axis drops below -9.8 m/s^2 when vehicle enters the pot hole. The value of *Z* axis increases from -9.8 m/s^2 when vehicle passes through the speed breaker. In case of staggered and bumpy road this value will fluctuate for quickly.
- *Z*-DIFF: This technique uses the two consecutive reading of accelerometer to estimate the difference and compute the velocity of change in the values [10]. The sharp increase and decrease will signifies the pothole or speed breaker. However, if the frequency of such events will signifies it has staggered or bumpy road. The major limitation of this approach is evaluate the difference in time between two points. Time difference affects the accuracy of this approach.
- *Z*-STDEV: lt considers the maximum standard deviation of the *Z* axis accelerometer data as the threshold. The major issue with this technique to define the value *K* that is the number of records used to identify the pothole or speed breaker. The value of *K* influences the standard deviation.
- *G*-Zero: This approach considered all the 3 axis data of accelerometer and finds the lower bound and upper bound to detect the pot hole and speed breaker for accuracy improvement. The upper bound for the pot hole will be the minimum value of all the three accelerometer values and lower bound will be the maximum value of all the three accelerometer values. The upper bound for the speed breaker will be the maximum value of all the three accelerometer values and lower bound will be the minimum value of all the three accelerometer values.

3 Proposed Methodology

Naive baye's algorithm is a supervised machine learning algorithm which evaluates the probability classification by evaluating the number values and frequencies of data set. It is a very simple probability classifier algorithm. It takes into consideration that all the features are independent of each other.

Many researches have been carried forward to evaluate the quality of roads. Many recognition systems have been proposed to detect and analyze the roads by using sensors on mobile phones [14] said that it requires additional hardware for this purpose. In the research paper, the proposed system can be used to detect potholes by the readings of the accelerometers. The total process is divided into two phases. training and prediction (Fig. 2).

In this system, the collection of data and GPS plotting are done using a accelerometer sensor and a gyroscope. Three machine learning algorithms have been used for this purpose which are SVM, decision tree and naïve bayes algorithm.

The system is able to produce an accuracy of 98.6% which is very good. The results are so good because of the use of both accelerometer and gyroscope. The use of smartphone is very efficient as it can be used as an alternate to special sensors. It is also very user friendly and is highly scalable. It gives the users of the application ample amount of knowledge about the ways of transportation.

After more research in this field this project can be used to improve the condition of roads in developing countries. This system can work with complete efficiency and can also maintain a whole record of the condition of roads (Fig. 3).

Real-Time Pothole Detection Method: The proposed real-time pothole detection method based on mobile sensing includes three steps [10]:

1. Accelerometer data normalization,
2. Pothole detection approaches and
3. Pothole location determination.

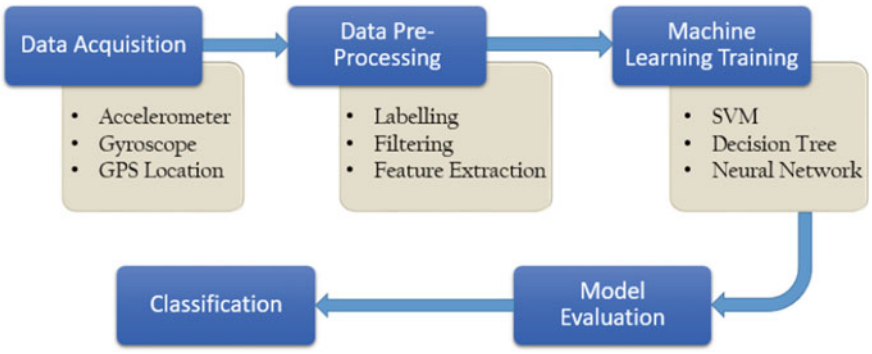


Fig. 2 Block diagram of the proposed model

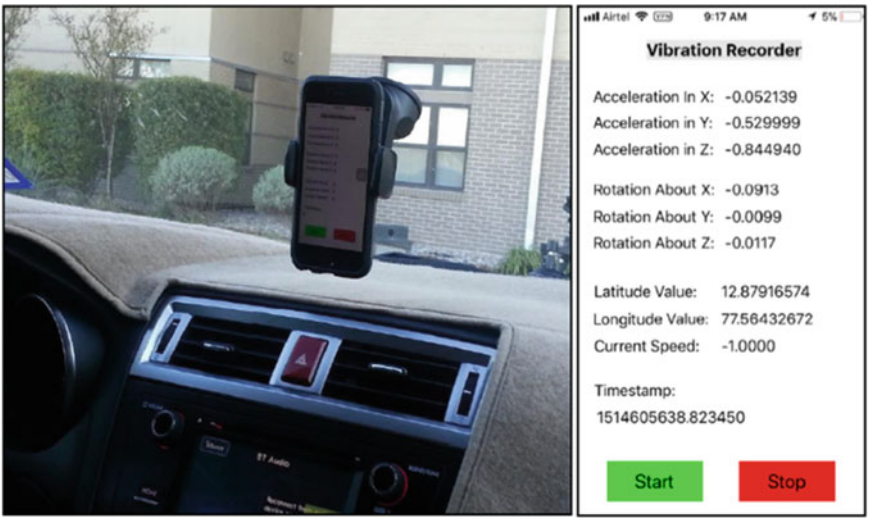


Fig. 3 Deployment of mobile sensors in the vehicle

Problem Definition:

Some shortcomings are in previous pothole detection methods as follows:

1. Mobile devices should be equipped with the specific angle;
2. High false positives may be generated with considering only one threshold for pothole detection;
3. The precise pothole location has not been investigation.
4. There should be 10 readings in 1 microsecond in order to get the exact location of the potholes, speed breakers or any other obstruction on the road that caused the deviation.

Therefore, the above study provides the real-time road anomaly detection approach which depends on mobile sensors to grasp and normalize the accelerometer stats from mobile devices.

4 Conclusion and Future Work

The smartphone-based approach is very convenient as it eliminate the requirement to deploy special sensors in vehicles. It is the edge of high expandability as smartphone users grow with time. The road -sense application is a try to provide its users with better information about the course of their transportation. With further work in this field, it is possible for this project to play ardent part in improving road state in growing countries. To this end, this method can be used to create a personal road type caution system that maintains a record of road state. In coming time the aim should

be set to improve the road type detection using detecting other road abnormality and trying other ML classifiers.

Intelligent Transportation System ITMS proposes a real-time pothole detection method using various mathematical approaches and various mobile sensing devices. This method uses the accelerometer and gyroscope sensor reading obtained from the mobile devices. Additionally, a real-time pothole detection approach is united with the *Z*-Thresh and *G*-Zero method for decreasing the false positives of potholes detection. To obtain the exact location of the potholes an spatial Interpolation method is proposed. Mathematical results and experiments shows that the proposed approach can accurately detect the exact location of the potholes without the false positive and accuracy will be very much closer to the 100%. Hence, the proposed real-time pothole detection approach will definitely reduce the road accidents as well as damage in the vehicles and make life of the end users safe.

References

1. Yu X, Salari E (2011) Pavement pothole detection and severity measurement using laser imaging. In: Proceedings of the IEEE international conference on electro/information technology (EIT '11). Mankato, Minn, USA, pp 1–5, May 2011
2. de Zoysa K, Keppitiyagama C, Seneviratne GP, Shihan WWAT (2007) A public transport system-based sensor network for road surface condition monitoring. In: Proceedings of the workshop on networked systems for developing Regions, Kyoto, Japan, Aug 2007
3. Mohan P, Padmanabhan VN, Ramjee R (2011) Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: Proceeding of the 6th ACM conferences embedded networks sensor systems. Raleigh, NC, USA, pp 323–336
4. Zviedris R, Elsts A, Strazdins G, Mednis A, Selavo L (2010) LynxNet: wild animal monitoring using sensor networks. In: Proceedings of the 4th international conference on real-world wireless sensor networks. Colombo, Sri Lanka
5. Singh P et al (2013) Using mobile phone sensors to detect driving behavior. In: Proceedings of the 3rd ACM symposium on computing for development. ACM, p 53
6. Eriksson J, Girod L, Hull B, Newton R, Madden S, Balakrishnan H (2008) The pothole patrol: using a mobile sensor network for road surface monitoring. In: Proceeding of the 6th international conference mobile system applications services. New York, NY, USA, pp 29–39
7. Kiran M et al (2017) An early detection-warning system to identify speed breakers and bumpy roads using sensors in smartphones. Int J Electr Comput Eng (IJECE) 7(3):1377–1384
8. Lin J, Liu Y (2010) Potholes detection based on SVM in the pavement distress image. In: Proceedings of the 9th international symposium on distributed computing and applications to business, engineering and science (DCABES '10). Hong Kong, pp 544–547, Aug 2010
9. Strazdins MG, Zviedris R, Kanonirs G, Selavo L (2011) Real time pothole detection using android smartphones with accelerometers. In: Proceeding of the international conferences distribution computing sensor systems. Workshops, Barcelona, Spain, pp 1–6
10. Wang H-W, Chen C-H, Cheng D-Y, Lin C-H, Lo C-C (2014) A real-time pothole detection approach for intelligent transportation system. Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Received 14 Aug 2014; 4 Dec 2014; Accepted 9 Dec 2014
11. Hautakangas H, Nieminen J (2011) Data mining for pothole detection. In: Progradu seminar. University of Jyväskylä
12. Ukrit MF et al (2016) Location based services with location centric profiles. Int J Electr Comput Eng (IJECE) 6(6):3001–3005

13. Ranjan M et al (2016) A cloud based secure voting system using homomorphic encryption for android platform. Int J Electr Comput Eng (IJECE) 6(6):2994–3000
14. Allouch A, Koubâa A, Abbes T, Ammar A (2017) RoadSense: smartphone application to estimate road conditions using accelerometer and gyroscope. IEEE Sens J 17(13)

Smart Farming: An IOT Based Automation

Sunita Dhingra, Ashwani Kumar Dhingra, and Satinder Bal Gupta

Abstract Revolution in farming is highly demanded to produce sufficient quantity and quality of food to feed continuously increasing population with limited natural resources like arable land, water etc. It seems just impossible with traditional methods of farming due to continuously shrinking arable land, harsh weather conditions, water scarcity, and less rural labor. Highly developed technologies like agricultural robots, sensors, communication methods, big data, cloud computing, and Internet of things (IOT) need to be included in farming to utilize these limited resources efficiently by automation. A number of challenges such as untrained manpower and standards infancy encountered with inclusion of technologies. So, a detailed study of these technologies and identification of farming stages which can be improved, if made technology driven, is required. The paper highlights the application of technology in farming and aims to identify the farming stages where use of technology has been proven efficient.

Keywords Internet of Things (IOT) · Automation · Sensors · Technology · Smart Farming

1 Introduction

Growth of developing countries depends on agriculture as more than 70% of population belong to rural areas and be dependent on agriculture for their livelihood [12]. So, the development of agriculture industry is the need of today to improve the economy

S. Dhingra (✉)
Department of Computer Science & Engineering, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, Haryana, India
e-mail: sunitadhingramdu@rediff.com

A. K. Dhingra
Department of Mechanical Engineering, University Institute of Engineering & Technology Maharshi Dayanand University, Rohtak, Haryana, India

S. B. Gupta
Department of Computer Science, Indira Gandhi University, Rewari, Haryana, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_8

steadily. Most of the farmers use traditional methods of farming including a lot of human intervention from sowing to harvesting. In traditional farming, farmers spent most of the time in monitoring the crop states as they have to stopover agriculture sites regularly all through the crop life to have a better plan about the crop setting [15]. This work of monitoring can be easily done by today's sensing and communication technologies without having the farmers to be physically in the field. Likewise, automation, resource optimization, higher yields, and climate effects are the factors which demand for use of technology in agriculture. Recent technology, Internet of things (IOT), is developing as a boon to farming because of its basic communication infrastructure and a variety of services like home and distant data acquisition, cloud-based intelligent information scrutiny and resolution, user interfacing, and agriculture operation mechanization [15]. So, the concept of using latest technologies like IOT, big data, and cloud computing in farming to make it more proficient by dropping human involvement through automation is smart farming. In process of automation, data is collected using sensors, processed using controllers, and automated using actuators [16].

Internet of things (IOT) also known as Internet of everything (IOE) is transforming the way we live. It allows things-to-things, human-to-thing, and human-to-human interaction via Internet [17]. After the smarter cities, more efficient industries, more connected cars, inclusion of IOT in agriculture are the need of today to reduce waste and increase productivity by automation. Modernization is the need of agriculture, and the stages at which theses advance technologies can be included are soil sampling and mapping, irrigation, fertilizers, crop disease and pest management, yield monitoring, forecasting, and harvesting to enhance the overall efficiency. A number of tools and technologies available for the modernization of farming are wireless sensors, harvesting robots, IOT-based tractors, communication, smartphones, cloud computing [15]. The aim of this paper is to identify the farming stages where use of technology is proven to be a boon to enhance productivity and efficiency by researchers.

2 Smart Farming Process

The motivating strength of smart farming is IOT which makes the farming process data-driven by integrating the smart machines and sensors on farms. These devices need to gather and process data in a continual sequence that enables the farmers to rapidly respond to rising issues and changes in ambient situation. This cycle involves four phases which are:

- Observation: Observational facts from crops, farm animals, soil, and atmosphere is recorded by sensors.
- Diagnostics: The recorded data is fed to precise software with predefined decision policy and models that discover the conditions of examined objective and any need or deficiencies.

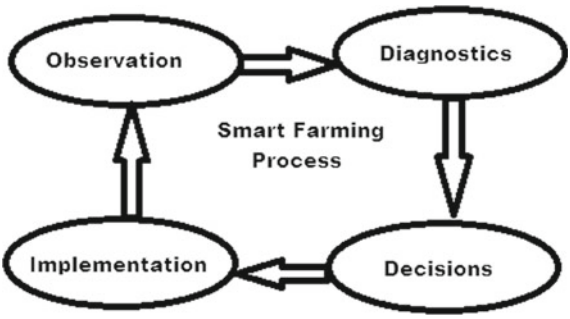


Fig. 1 Process of smart farming

- Decisions: Once issues are publicized, the software decides whether site specific management is essential and if so, which.
- Implementation: The management is to be performed by way of accurate action of machines (Fig. 1).

IOT is coming with revolutionary change in farming methods creating new opportunities. But every new opportunity comes with challenges, so does the smart farming as the technology is going to be integrated with traditional farming practices. The major challenges in implementing technology for smart farming are untrained manpower, high investments, fear of modern machinery, coverage and connectivity, and standards infancy [15].

3 Prior Work

Improvement in farming is the need of today and involvement of technology is the solution from various researchers. A large number of researchers are working on this burning field to improve the livelihood of backbone of a country, i.e., farmers. Some of the work identifying the area and benefits of involvement of technology in farming is presented below:

Prathibha et al. [1] proposed work on smart monitoring system based on IOT in smart agriculture. They found that monitoring environmental factors plays key role to get better the yield of efficient crops. They designed and synthesized smart agriculture system by using CC3200 single chip interfaced with camera for monitoring temperature and moisture in agricultural field. The captured pictures were sent through MMS to farmers mobile. They concluded that the developed scheme was more efficient for farmers.

Dwarkani et al. [2] proposed a new approach for farming by involving smart sensing and irrigation system for measurement of moisture and pH of soil and nutrient content to spray the essential quantity of manure, water, and green compost. They concluded that their smart sensing system provided precise results to manage the water and nutrient spray by smart irrigation system.

Patil and Kale [3] developed a concurrent monitoring system to measure properties such as moisture, pH, and temperature and based on those properties implemented a decision support review system for pest and disease alarm, crop disease detection etc. Their smart farming system also included the remote control of field from wherever at any time using mobile as well as Web application. They found that farmers received updated information to cope with weather and soil conditions and epidemics of pests using proposed model. They also concluded that real time as well as historical environmental conditions are required to accomplish resourceful supervision and management of resources.

Sreeram et al. [4] proposed a semi-automated farming system with aim to put less efforts for large fields by implementing a technology induced smart agro environment and ways to look after the crops during inevitable natural calamity. They focused on three main issues of farming, i.e., shearing furrows, manual watering, and security against animal grazing and provided way out for these problems by monitoring and controlling activities through mobiles via GSM or DTMF technology. They concluded that the proposed solution will get better the quality of living of farmers beside the facts of difficulty in implementation at initial stage and require extensive agricultural techniques.

Mat et al. [5] presented the basic requirements of IOT for smart farming. They built a smart farming system to keep an eye on the crop ground using sensors and automatic irrigation system. They tested their smart farming system for mushroom cultivation and IOT-based irrigation system. They analytically compared the traditional mushroom house and IOT-based mushroom house and concluded that IOT-based mushroom house produced good quality mushroom as compared to traditional mushroom house within similar age of period. They also concluded that there is saving of water fertilizers in automatic irrigation system as compared to scheduled irrigation mode. They come up with an idea to use smart farming for organic farming and family farming.

Dagar et al. [6] proposed IOT-based poly house which is a completely enclosed arrangement to protect the field from harmful sunrays, storms, and outside factors like insects resulting into less need of insecticides. They focused on two main domains of electricity and water for improvement in farming. They built a simple structural design of IOT sensors like soil pH and moisture sensor, humidity sensor, air temperature sensor etc. to gather information and forward it over wireless network to servers which can take actions accordingly. They concluded that an efficient IOT architecture can result in good quality and amount of production by saving assets like water and electricity.

Yoon et al. [7] proposed a smart wired/wireless system which resolve the space and power constraint crisis of wired system by means of low-power Bluetooth and low-power wide area networks (LPWAN) communication modules. They implemented their monitoring and control functions via IOT-dedicated protocol MQ telemetry transport. They also improved the inter-module message exchange by applying easy techniques to bit loss which may happen in wireless networks. They concluded that their projected scheme can be a foundation for big data and artificial intelligence farming in terms of environmental and growth data collection.

Heble et al. [8] proposed a power- and cost-efficient IOT system for smart farming by using in house built IITH (Indian Institute of Technology Hyderabad) mote as sink and sensor node. They designed an inexpensive soil humidity sensor to assess soil parameters. They evaluated their network with other proposed similar networks for cultivation monitoring in terms of power and cost and found that their projected network consumes a smaller amount of power and has long-lasting life span at a lower price.

Zamora-Izquierdo et al. [9] developed a flexible multi-tier scheme to envelop the severe PA (Precision Agriculture) needs by means of mechanization, IOT, edge, and cloud computing through virtualization. The three tiers of system include a CPS local layer associated with greenhouse, a new edge computing plane for monitoring and managing PA tasks and a cloud sector with computing and data investigative resources for crop care decisions. They implemented their system in real greenhouse in southeast of Spain for two cycles of tomato crops. The first cycle validated the architecture and second cycles proved the improvement over regular open crop. They concluded that the proposed system was good in terms of water saving and nutrient value.

Muangprathub et al. [10] designed a system with three key parts of hardware, Web application, and mobile application to get better yields and to diminish cost. The hardware part included the control box and sensors to obtain data on crops. The second part, Web application, used the association rule of data mining to manage the particulars of crop and field data. Finally, the third part, mobile application was designed to manage the crop watering in automatic as well as manual mode. They installed the proposed system in Suratthani province, Thailand. They concluded that the proposed system was helpful in reducing cost and increasing productivity by appropriately maintaining the moisture content in soil.

Abouzahir et al. [11] analyzed the popular segmentation techniques independent of plant species and vigorous to open-air enlightenment in terms of robustness and execution time. The key objective of this comparison was to develop lightweight intelligent system for real-time uses in agriculture. They concluded that CIVE and NDI are good for identifying plant regions in complicated atmosphere with less calculation time, but for weed identification, a classification algorithm is also required which will increase the computation time. They found that FCM and BPNN perform better in identifying crop plants and weed if specific information like consistent color is provided.

Das et al. [12] proposed three methodologies of agriculture using IOT to make it a smart agriculture system. These are soil moisture and pests supervision, intelligent seeds corporation, and well-organized food corporation of India. They assured that the proposed system will control financial loss by reducing manual labor, surplus of seeds in go down and declining of scarce seeds, dampness and rodents effective loss in FCI.

Koksal and Tekinerdogan [13] defined a novel IOT-based FMIS (Farm Management Information System) which is a key element of smart farming to support the automation of each operation. The proposed architectural design was able to derive the customized application FMIS architecture. They tested their system through two

case studies, one for IOT-enabled wheat production and other for IOT-enabled tomato FMIS. They concluded that proposed approach was effective as well as practical in comparison to the approaches from literature.

Cadavid et al. [14] come up with an addition of trendy open-source platform Thingsboard (used to gather and handle sensors provided information) with an idea to concentrate on sensors, decision support system, and distantly controlled independent devices so that the requirements of Colombian farming business be met. The proposed structural design of this addition was validated on real-life scenario and identified an exceptionally unsafe potato disease in real time to activate the simulated controller of automated drones in response.

Ayaz et al. [15] considered all aspects from traditional farming to smart farming by including prospective of wireless sensors and IOT, challenges expected, IOT devices and communication techniques all the way through the crop phases from sowing to harvesting to recognize present and future trends of IOT in agriculture. They also discussed a variety of architectures and platforms based on IOT for farming applications. They concluded that IOT-based smart farming is the need for efficient utilization of every inch of field to make the most of crop production.

Madushanki et al. [16] analyzed IOT applications in farming and agriculture industry to provide overview of technologies, sub-verticals, and sensor data compilation. They reviewed 60 papers from 2016 to 2018 and concluded that water management is utmost considered sub-vertical, most significant sensor data collection dimension is environmental temperature and Wi-Fi is the highest demanding technology.

Lakhwani et al. [17] studied and analyzed the applications of IOT in farming and forestry by thorough review. They found that use of IOT in cultivation can help the farmers to have a remote control of their farms without heavy farm labor. Also framers can do intelligent analysis and better management of agricultural soil, crop, and environmental information.

Rezk et al. [18] proposed crop productivity and drought predicting technique (WPART) by merging a wrapper feature selection approach and PART classification. They tested the proposed approach on five datasets and analyzed that the approach is more precise in predicting drought with the yield of crops like Soybean, Bajra, Jowar, and Sugarcane. They also concluded that the proposed technique is better in terms of sensitivity, precision, and F-Score metrics than existing methods.

Al-Ali et al. [19] designed an IOT-based solar power system for smart irrigation. The proposed system utilizes a controller with built in Wi-Fi connectivity to read the humidity, soil moisture, and temperature sensors and trigger proper signals to activate irrigation pumps. In order to prevent the burning in pump motor, underground water level is also monitored by the controller. They tested the prototype system by isolating from power supply and making the system fully rely on solar power.

4 Discussion

From literature, it is derived that there are several stages of farming which can be improved, if made technology driven. Some common stages addressed are water management, irrigation management, soil management, crop management, livestock management, energy management, pest and diseases management, waste management, poultry management, cattle detection management, warehouse management, and nutrient management. Table 1 represents the stages of farming addressed by various researchers to improve quality and quantity of production.

Table 1 IOT-enabled farming stages and their benefits

S. No.	Author	Sub-area addressed	Data measures	Benefits of proposed work
1	Prathibha et al. [1]	Weather monitoring, soil management	Environmental temperature, humidity	Can detect the temperature and humidity for yield improvement
2	Dwarkani et al. [2]	Water management, nutrition management	Soil moisture, soil pH, soil nutrients	Better water and nutrition management
3	Patil and Kale [3]	Pest control, crop disease identification	Temperature, soil moisture, soil pH	Remote monitoring and updated information for pests epidemics, weather and soil conditions
4	Sreeram et al. [4]	Automatic irrigation, cattle detection management, natural disasters management	Water level, plant health	Lessen human efforts, remote monitoring
5	Mat et al. [5]	Crop management, automatic irrigation	Water level, plant health	Remote monitoring, better quality crop
6	Dagar et al. [6]	Water management, electricity management	Temperature, humidity, soil moisture, soil pH	Improved production, better resource management
7	Yoon et al. [7]	Power and space management	Environmental and growth factors	Remote monitoring and control, better data collection and analysis

(continued)

Table 1 (continued)

S. No.	Author	Sub-area addressed	Data measures	Benefits of proposed work
8	Heble et al. [8]	Power and cost management	Soil parameters	Improved Technology in terms of cost by developing low-cost soil sensor
9	Zamora-Izquierdo et al. [9]	Water and nutrition management, growth measurement	Environmental and growth factors	Better quality crop, better resource (water, nutrients) management
10	Muangprathub et al. [10]	Crop quality, Cost management	Soil moisture	Increased productivity, Reduced cost
11	Abouzahir et al. [11]	Intelligent farming	Weed and plant regions	Better plant region and weed identification
12	Das et al. [12]	Soil and pest management, waste management	Soil moisture, seed and pests requirement	Reduced cost, Reduced manual labor, reduced wastage of seeds
13	Koksal and Tekinerdogan [13]	Crop management, soil and pests management	Soil moisture, seed and pests requirement, environmental and growth factors	Automation of each operation like data acquisition, monitoring, planning etc., Better quality
14	Cadavid et al. [14]	Disease management, smart farming	Environmental factors	Automation, better decision-making with in-time disease detection
15	Rezk et al. [18]	Crop management and drought prediction	Environmental and growth factors	Improved productivity, better decision making
16	Al-Ali et al. [19]	Irrigation management	Soil moisture, temperature, humidity	Better resource (Water) management

5 Conclusion

Internet of things (IOT) is developing as a boon in farming along with its various applications due to strong communication infrastructure, data acquisition techniques, quick information examination and resolution building, and farming operation mechanization. So, basic process of smart farming along with challenges is presented. This paper considered practical work from various researchers and represented sub-areas

addressed, data measures used, and benefits gained by them to identify the farming stages which can be improved by technology and found that technology-driven every stage like water management, irrigation management, weather monitoring, pest and disease management, soil management etc. is efficient than traditional farming. Once the farming stages are identified, one can learn the basics of technologies involved and discover novel traditions in farming using automation.

References

1. Prathibha SR, Hongal A, Jyothi MP (2017) IOT based monitoring system in smart agriculture. In: International conference on recent advances in electronics and communication technology (ICRAECT) Bangalore. https://doi.org/10.1109/ICRAECR.2017.52
2. Dwarkani CM, Ram GR, Jagannathan S, Priyatharshini R (2015) Smart farming system using sensors for agricultural task automation. In: IEEE technological innovation in ICT for agriculture and rural development (TIAR),Chennai. https://doi.org/10.1109/TIAR.2015.7358530
3. Patil KA, Kale NR (2016) A model for smart agriculture using IoT. In: International conference on global trends in signal processing, information computing and communication (ICGTSPICC), Jalgaon. https://doi.org/10.1109/ICGTSPICC.2016.7955360
4. Sreeram K, Kumar RS, Bhagavath SV, Muthumeenakshi K, Radha S (2017) Smart farming—a prototype for field monitoring and automation in agriculture. In: International conference on wireless communications, signal processing and networking (WiSPNET) Chennai. https://doi.org/10.1109/WiSPNET.2017.8300148
5. Mat I, Mohd Kassim MR, Harun AN, Yusoff IM (2018) Smart agriculture using internet of things. In: IEEE conference on open systems (ICOS), Langkawi Island, Malaysia. https://doi.org/10.1109/ICOS.2018.8632817
6. Dagar R, Som S, Khatri SK (2018) Smart farming-IOT in agriculture. In: International conference on inventive research in computing applications (ICIRCA), Coimbatore. https://doi.org/10.1109/ICIRA.2018.8597264
7. Yoon C, Huh M, Kang S, Park J, Lee C (2018) Implement smart farm with IoT technology. In: 20th International conference on advanced communication technology (ICACT), Chuncheon-si Gangwon-do, Korea (South). https://doi.org/10.23919/ICACT.2018.8323908
8. Heble S, Kumar A, Prasad KVVD, Samirana S, Rajalakshmi P, Desai UB (2018) A low power IoT network for smart agriculture. In: IEEE 4th world forum on internet of things (WF-IoT), Singapore. https://doi.org/10.1109/WF-IoT.2018.8355152
9. Zamora-Izquierdo MA, Santa J, Martínez JA, Martínez V, Skarmeta AF (2019) Smart farming IoT platform based on edge and cloud computing. Biosys Eng. https://doi.org/10.1016/j.biosystemseng.2018.10.014
10. Muangprathub J, Boonnam N, Kajornkasirat S, Lekbangpong N, Wanichsombat A, Nillaor P (2019) IoT and agriculture data analysis for smart farm. Comput Electron Agric. https://doi.org/10.1016/j.compag.2018.12.011
11. Abouzahir S, Sadik M, Sabir E (2017) IoT-empowered smart agriculture: a real-time light-weight embedded segmentation system. Lect Notes Comput Sci. https://doi.org/10.1007/978-3-319-68179-5_28
12. Das RK, Panda M, Dash SS (2018) Smart agriculture system in India using internet of things. Adv Intell Syst Comput. https://doi.org/10.1007/978-981-13-0514-6_25
13. Koksal O, Tekinerdogan B (2018) Architecture design approach for IoT-based farm management information systems. Prec Agric. https://doi.org/10.1007/s11119-018-09624-8
14. Cadavid H, Garzón W, Pérez A, López G, Mendivelso C, Ramírez C (2018) Towards a smart farming platform: from IoT-based crop sensing to data analytics. Adv Comput. https://doi.org/10.1007/978-3-319-98998-3_19

15. Ayaz M, Ammad-Uddin M, Sharif Z, Mansour A, Aggoune EM (2019) Internet-of-Things (IoT)-based smart agriculture: toward making the fields talk. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2932609
16. Madushanki AAR, Halgamuge MN, Wirasagoda WAHS, Syed A (2019) Adoption of the Internet of Things (IoT) in agriculture and smart farming towards urban greening: a review. Int J Adv Comput Sci Appl 10:11–28
17. Lakhwani K, Gianey H, Agarwal N, Gupta S (2018) Development of IoT for smart agriculture a review. Emerg Trends Expert Appl Secur. https://doi.org/10.1007/978-981-13-2285-3_50
18. Rezk NG, Hemdan EE-D, Attia A-F, El-Sayed A, El-Rashidy MA (2020) An efficient IoT based smart farming system using machine learning algorithms. Multimedia Tools Appl. https://doi.org/10.1007/s11042-020-09740-6
19. Al-Ali AR, Al Nabulsi A, Mukhopadhyay S, Awal MS, Fernandes S, Ailabouni K (2020) IoT-solar energy powered smart farm irrigation system. J Electron Sci Technol. https://doi.org/10.1016/j.jnlest.2020.100017

Length of Stay Prediction of Patients Suffering from Different Kind of Disease to Manage Resource and Manpower of Hospitals

Rohan Sharma, Binit Kumar Singh, Siddharth Rautaray, and Manjusha Pandey

Abstract The sole purpose of this research is to explore the usage of deep neural network techniques and implement a model for predicting the length of stay of patients getting admitted to medical healthcare. In our research, we have compared the severity of various diseases rather than focusing on a single type of disease, which makes this research different from already existing studies. The research throws limelight on previous works on length of stay (LOS) prediction and different approaches of multiple authors on the same. The discussed model is broadly based on three sub criteria which acts as a starting parameter to the prediction model, which first classifies LOS of different patients suffering from different disorders; label them into three categories based on LOS namely <20 (mild), 20–50 (severe non-fatal), >50 (extreme fatal), based on the LOS category of the patient the management authority and healthcare provider can easily calculate the usage of various different utilities, sporadically available equipment and manpower and allocate them accordingly.

Keywords Length of stay · Predictive models · Classification · Hospital resource management

1 Introduction

Previous research using ML/DL has been limited to predicting LOS for specific diseases, and some authors have attempted to determine the same for certain patients or conditions. However, the research done in [1] focuses on patients regardless of

R. Sharma (✉) · B. K. Singh · S. Rautaray · M. Pandey
School of Computer Engineering, KIIT-DU, Bhubaneswar, India
e-mail: 1805152@kiit.ac.in

B. K. Singh
e-mail: 1806208@kiit.ac.in

S. Rautaray
e-mail: siddharthfcs@kiit.ac.in

M. Pandey
e-mail: manjushafcs@kiit.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_9

what diseases they are suffering from. In contrast, we propose a more complex yet efficient way of handling patients with different diseases and conditions and severity of their illnesses.

In the below discussion, we contour different deep learning algorithms for the effective prediction of LOS of a patient suffering from a specific malady. Taking into consideration the result of certain factors like the accessibility of additional rooms, type of admission, the severity of health problems, age of the patient, etc., and using completely different conventional algorithms, we have tried to predict the length of stay (LOS) of a patient. This prediction can successively facilitate the potency of healthcare management in hospitals. Because of the high number of patients suffering ailments to chronic diseases, it is difficult for the public to access hospitals, where resources are limited and the number of individuals seeking treatment is high. In this regard, it is essential to ensure that the patients have access to the necessary resources within their prescribed time frame. There are many of us facing challenges to find a proper bed or perhaps be admitted to the hospital due to an insufficient allocation of resources; therefore, to beat this downside, this model not only specifies the length that the patients are admitted to the hospital but also conjointly helps people to induce their finances economically.

A new Harvard study [2] demonstrated that the new corona virus could create overcrowding in many parts of the United States of America if nothing expands hospital capacity. In 40 percent of markets around the country, hospitals would not be able to create enough space for all the patients who became sick with COVID-19, although they might empty their beds of alternative patients. That statistic assumes that forty percent of adults become infected with the virus over twelve months, a situation represented as "moderate" by the team behind the calculations. These numbers do not seem to be precise predictions. In many ways, they reflect a worst-case situation since they do not consider the efforts hospitals will create to quickly increase during an associate degree emergency. Around the country, hospitals have begun canceling elective operations and speeding home patients with less critical ailments. Those efforts may increase the number of free beds available for corona virus patients. In a very half-dozen interviews, hospital executives calculated that they might increase their capability between twenty percent and seventy percent.

Better management of patients, i.e., keeping a low length of stay, will not only aid the patient but also come in handy for improving several sections that square measure unnoticed within the hospitals. Those majorly are better management of assets or resources the hospitals are using which can be equally distributed to all or any the destitute patients if they get to understand the length of keep of the patients, another is a sensible and planned patient flow returning and being discharged from the hospitals, and another is nurse allocations, who are the second most vital a person once the doctor to the patient while under treatment in the hospital [3, 4].

This analysis will play a significant role in the massive number of individuals within the society, creating the economic use of technology within the correct time and place. This product guarantees a correct solution to management in hospitals and lets the patient's family know when they can get their friend discharged. During these difficult times, the requirement for resource management within the health

sector is a must. Moreover, this analysis will facilitate the hospitals to manage their resources like the required kinds of instrumentation square measure allotted for a selected person and ultimately serving the patient to place the finance within the health optimally as required. This model achieves our goal of serving the hospitals and aid because it takes a few features as input like the department of admission, their age, and type of admission, and, therefore, the severity of health problem and predicts the specified output we aim for, i.e., the length of stay of the patient and classifying them to entirely different labels of time they are taking according to the input variables.

2 Literature Survey

The idea of predicting the length of stay of a patient or discussing it since the 1960s has been discussed [5]. It is an all-around advantage for the hospitals or the various healthcare centers to know how long a patient will be under medical support and using multiple resources of the hospitals or healthcare centers. We are aiming for this research work to predict the length of stay of any patient depending on the type or the severity of the disease irrespective of the disease the patient is suffering from, which was never attempted by any of the authors.

In recent years, several studies [1] have been conducted on similar topics. They have helped us understand the problem statement better in a more detailed manner and develop better solutions. Some of the previous works we came across are as follows. In a research done by Kelly [6] on the length of stay and readmissions of patients, which was impacted by developments in colorectal cancer surgery. This research mainly investigated predictors of LOS and emergency readmission after the initial surgery episode on a single group of patients who have cancer rather than on multiple cases of multiple diseases registered in a hospital. In the LOS predictions, they determined that age, higher level of comorbidities, and marital status increase a patient's LOS.

Gustafson [7], in his explanation of predicting LOS, put forward five different types of methodology to do the same and eventually compared each of them. Several criteria of effectiveness and cost of the medical care system were used to measure and evaluate the performance of the methodologies he proposed. As a result, it was found out that the public is demanding more of the improved services of the medical care system but is unwilling to pay its cost. An accurate prediction of a patient's LOS can be reached by examining factors like faculty, equipment, human resources, and the type of bed allocated to the patient. Sessler [8] discussed how patients have low blood pressure; the low bispectral index is hospitalized for a more extended period. Logistic regression found that a "triple-low value of MAP, BIS, and MAC were strongly correlated with an extended LOS. BIS refers to a measure of the depth of anesthesia, and MAC refers to the concentration of anesthesia vapors used and MAP means arterial pressure. The analysis done by Wang [9] aims to identify the patients' extended stay after acute exacerbation of COPD based on variables on

admission. They also established a predictive model for LOS for earlier identification of patients who have to stay hospitalized for a more extended period. The hospitals correctly manage the medical equipment and human resources. In order to evaluate the predictors of LOS, they collected demographic and clinical data on 599 patients discharged in a given period and used logistic regression analyses. Results from this was that admission between Thursday and Saturday, heart attacks, diabetes, and high arterial PCO2. Furthermore, a low serum albumin level was associated with a prolonged LOS.

The research done by Rezai [10] was primarily focused on the length of stay of patients with cardiac arrests or problems and predicting the length of stay of those kinds of patients which gives the hospital management an upper hand to give better services and higher satisfaction and assists them with their management of resources properly and treat as many patients at a given period. They used three classification algorithms: decision tree, support vector machine, and artificial neural networks. The result of his analysis showed that marital status was again affecting the patients' length of stay; single patients had LOS lower than five days, whereas married ones had LOS greater than ten days. More study was done by Appelros [11], which showed that patients suffering from lung disorders and hemorrhage with drug consumption extends the patient's length of stay. Even comorbidities such as being a current smoker and having social security type insurance in coronary artery patients led to longer LOS than other subjects. Kramer [12] discussed the excessive usage of the resources by the patients due to their prolonged length of stay in ICUs (intensive care unit), early identification of which can lead to better and proper utilization of resources and reduce the ICU stay of a patient. They performed their intensive analysis on a record of 343,555 admissions to 31 ICUs in 31 U.S hospitals. They developed a multivariate regression model for predicting the LOS of the patients. They also claimed that their model could be used to benchmark ICU performance and resource management. Some of these important factors include whether mechanical ventilation was required, PaO2:FiO2 ratio, assorted physiological components, and sedation.

3 Hospital Stay Prediction Model

To obtain the length of stay (LOS) of patients admitted in different departments in a hospital and based on the age and severity of the illness of patients, through the use of data analysis methods, and with the help of machine learning and neural networks, the proposed work intended to create a back-end model which will help to predict the length of stay by learning the patterns from historic as well as present experience and implementing the same to improve and enhance the accuracy of the predictions.

In the first instance, we implemented the supervised part of the machine learning in which we used simple classification techniques, and on the other hand, we used bucketization techniques to predict the output we got by implementing classification into different labels of the length of stay. So, with the help of those predictions our

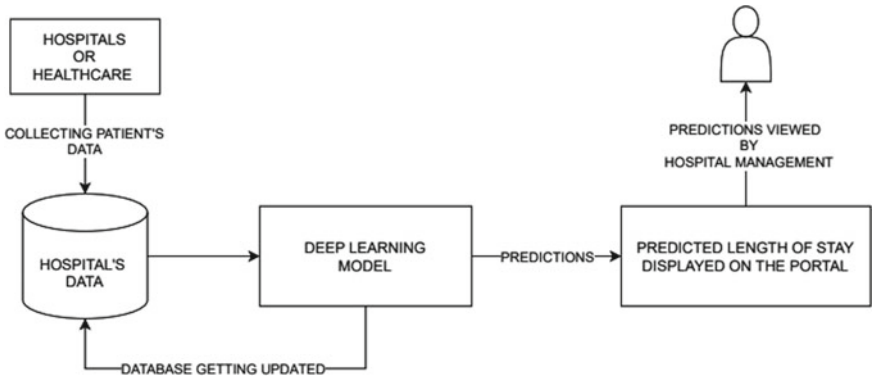


Fig. 1 Los prediction workflow

model, as well as the user using the model, gets a better understanding of how to group together and handle similar kinds of patients getting admitted to the hospital for better utilization of resources and manage other resources associated with them, which can all be seen by the hospital's management on its portal, so as not to get overwhelmed when so many patients present in a critical situation. The proposed working model of the complete setup is shown in Fig. 1 which helps us to get an understanding of its simplicity yet efficiency and effectiveness.

3.1 Step 1: Data Analysis and Model Implementation

To predict the length of stay of a patient, which is a discrete variable, we used classification models, classification is a supervised machine learning technique used to learn from different categorical data and the patterns out of it and predict the category to which the new entry belongs to. In our model, the dataset which we scraped out has different kinds of categorical columns which has various categories such as the department of the patient being admitted their age, severity of the illness, which is one of the most critical columns when we talk about patients as the whole procedure of treatment mainly depends on the severity and the age of the patient severity of the disease the patient is suffering from. That is the sole reason we have opted to work on those features only and not take other features into consideration that do not affect the resources and treatment procedure of a patient.

As far as the dataset is concerned, it is a huge collection of homogeneous data obtained from various different sources grouped together. On further processing and inspecting the data, primarily three main features as discussed above were drawn into account for model creation. Finally, different classification algorithms were used for model training and produced varying results as shown in Table 1. Upon tuning the hyper parameters of different models, i.e., logistic regression, SVM, K-NN, decision

Table 1 Accuracies achieved from different algorithms

Classification ML algorithm used	Accuracy achieved (%)
Logistic regression	59.03
Support vector machine	60.19
K-NN	61.05
Decision Tree	61.78
Random Forest Classifier	62.34

tree, and random forest. Out of which, random forest generated the best overall accuracy of 62%.

Also, we have implemented an artificial neural network with a different number of layers which contains many efficient neurons with dropout layers which will not allow our model to overfit over the train set and will be able to predict values on the final test set or the data entered by the user in real day to day scenarios without getting biased on a single feature or criteria. As predicted values, which needs to be almost correct, are very essential for the given scenario as it is a model which is going to be used in hospitals and healthcare centers giving an overall accuracy of 66% which will improve over time as more and more data gets accumulated. Hence, the results and support discussions are based on an artificial neural network which was found to be better than other algorithms we tried to implement, such as support vector machine and random forest.

3.2 Step 2: Mitigating Fatal and Non-fatal Stage

After the predictions are obtained from the classification model (ANN), we then labeled the patients based on their probable number of days of stay at the hospital (LOS), which suffices the objective of the study to identify different groups of patients who require varied amounts of time for bed possession and required resources for the patient's recovery. Moreover, resource allotment is faster for both the patient and the hospital management. Furthermore, labeling plays a vital role to display the future group of patients for which hospitals will need to manage their resources and therefore, take the best decision for the patient's faster treatment. In this research, according to predicted output obtained from the model, we followed the same path which was discussed by the author in the study [13]. According to which, patients under medical observation with LOS of less than 20 days were labeled as mild, more than 20 days, less than 50 days were labeled as severe non-fatal, and patients with LOS of more than 50 days were labeled as fatal.

3.3 Step 3: Experimental Setup

Following all the predictions made by the model, the next step is to deploy the model using Python Flask as the backend framework and the streamlite API to deploy the model on the hospital's management portal. Further, we propose using SQLyog to store the trained state of the model and the train and test data results. This will allow for a fast and accurate data retrieval from the server in real time.

4 Result Analysis

As final model artificial neural network (ANN) was implemented by us which consists of one input layer and six hidden layers with 1024 to 128 neurons on subsequent layers, ReLU activation function is used for back propagation among successive layers, and a dropout layer after each layer which prevents the over-fitting of our model. The p-value is calibrated such that activated and deactivated neurons put together predicts desired results for each consecutive layers resulting in highest possible accuracy which can be obtained with the acquired dataset. Now, the output layer uses Softmax as an activation function, and then the model is compiled with categorical_crossentropy loss function. The optimizer used here is an Adam, which is considered to be one of the best, and we used accuracy metric to determine the model's accuracy.

From Fig. 2, we conclude that the train accuracy remained almost static during the epoch cycle of 20 epochs maintaining an accuracy of 66.78%. The test accuracy increased gradually during the first two epochs from 63.55 to 65.64% and remained

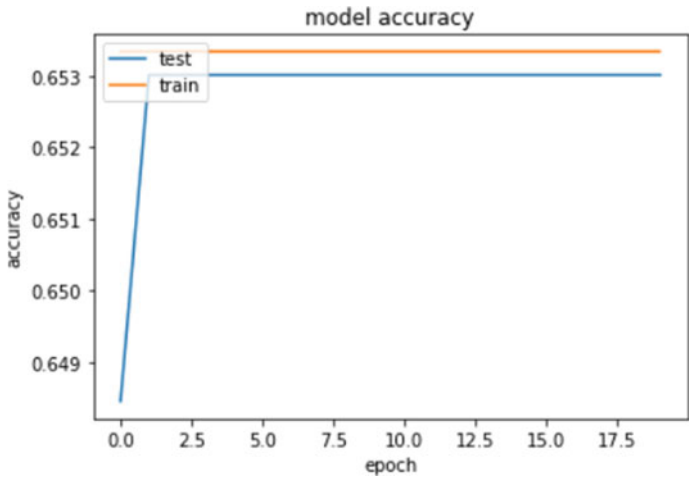


Fig. 2 Accuracy versus epochs

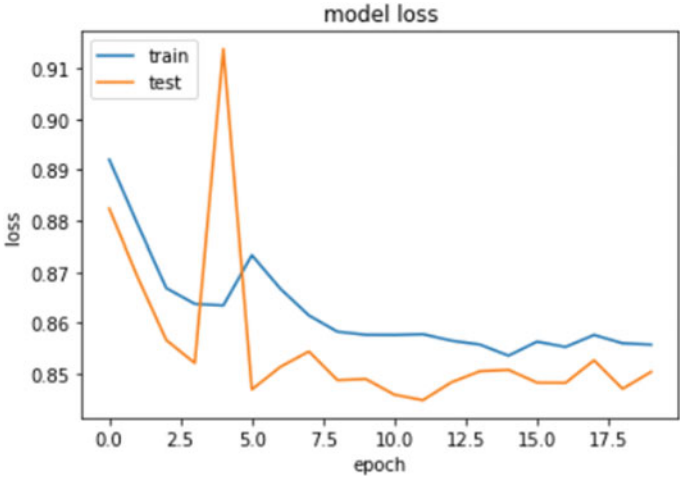


Fig. 3 Loss versus epochs

fairly consistent for the remaining epochs, i.e., 3 to 20, in the epoch cycle. The loss curve remained fairly low and kept on gradually decreasing for both test and train set during the entire epoch cycle except for the epoch period of 2 to 5 for train and 4 to 6 for test set where a sudden spike in the loss curve is observed in Fig. 3, due to anomaly in the data.

Although, the accuracy of the model can be increased through proper data collection and segregation. The intent of the result discussion is aimed at providing justification for the feasibility of the proposed idea.

5 Conclusion

In this research, by taking into consideration the most vital factors that are of utmost importance for the hospitals to provide immediate and viable treatment for the patients, which ultimately led to the prediction of total length of stay (LOS) that a particular patient would spend in the hospital. Also, the proposed work is a classification model, which is obtained by comparing different classification algorithms out of which the random forest model fitted reasonably well to the data. However, as the data is time dependent, we chose ANN as the final model. The result obtained from the model was found to be having an accuracy of 66% with a loss of 0.85. After obtaining the classification result, we labeled the category in which a patient will fall according to the predictions obtained from the model and, eventually, predict the LOS class of the patient.

Though, the scope and ability of the model is much more beyond the scope of this research. On the other hand, if the data provided to the discussed model is

directly drawn from the hospitals or medical care centers, without alteration, can be used more effectively. As the data we extracted was mere segregation from different online sources and not directly from any of the hospitals, the accuracy of the discussed model could not attain its maximum efficiency. The results and analysis section draws light on the potentiality of AI/ML and deep learning techniques in not only predicting the LOS of patient's but also in a more generalized medical healthcare ecosystem, which puts forward the countless dimensions of AI in the medical field and healthcare sector and establish its utility and accountability on the same.

References

1. Gentimis T, Ala'J A, Durante A, Cook K, Steele R (2017) Predicting hospital length of stay using neural networks on mimic III data. In: 2017 IEEE 15th international conference on dependable, autonomic and secure computing
2. https://www.nytimes.com/interactive/2020/03/17/upshot/hospital-bed-shortages-coronavirus.html
3. Robinson GH, Davis LE, Leifer RP (1966) Prediction of hospital length of stay. Health Serv Res Winter(1):287
4. Robinson GH, Davis LE, Johnson GC (1966) The physician as an estimator of hospital stay. Hum Factors 8(3):201–208
5. Azari A, Janeja VP, Mohseni A (2012) Predicting hospital length of stay (PHLOS): Multi-tiered data mining approach. In: 2012 IEEE 12th international conference on data mining workshops
6. Kelly M, Sharp L, Dwane F, Kelleher T, Comber H (2012) Factors predicting hospital length-of-stay and readmission after colorectal resection: a population-based study of elective and emergency admissions. BMC Health Serv Res 12(1):77
7. Gustafson DH (2002) Length of stay prediction and explanation. Health Serv Res 37(3):631–645
8. Sessler DI, Sigl JC, Kelley SD, Chamoun NG, Manberg PJ, Saager L, Kurz A, Greenwald S (2012) Hospital stay and mortality are increased in patients having a "triple low" of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. J Am Soc Anesthesiologists, pp 1195–1203. Web of Science
9. Wang Y, Stavem K, Dahl FA, Humerfelt S, Haugen T (2014) Factors associated with a prolonged length of stay after acute exacerbation of chronic obstructive pulmonary disease (aecopd). Int J COPD, pp 99–105. Dovepress
10. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F (2013) Use of data mining techniques to determine and predict the length of stay of cardiac patients. Healthc Inf Res, pp 121–129. The Korean Society of Medical Informatics
11. Appelros P (2007) Prediction of the length of stay for stroke patients. Acta Neurol Scand 116(1):15–19. [PubMed] [Google Scholar]
12. Kramer A, Zimmerman J (2010) A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. BMC Med Inf Decis Making, pp 10–27. BioMed Central
13. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P (2008) Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Med Care 46(3):232–239

Next Word Prediction Using Hindi Language

Sanskriti Agarwal, Sukritin, Aditya Sharma, and Anurag Mishra

Abstract Natural language generation is a process that concerns on generating human understandable language. This study provides method to guess next word from previous sequence of words that are in Hindi language. This process reduces the keystrokes of a user by predicting next word. In this two machine learning technology, BERT (Bidirectional Encoder Representations from Transformers) Model and ML (Masked Language) model is used to predict next word from previous words. Next word prediction is a technology that takes input and simplify typing process as of suggest user the next word according to the understanding of previous words. It makes typing less time consuming and error free. In current times, everyone is writing data digitally and to make typing efficient in Hindi language we require such systems. There are various systems related to English Language. Similar systems are required for Hindi language as Hindi is a more frequent language in India.

Keywords Machine learning · Next word prediction · BERT model · Prediction · Natural language processing · Masked language model · Fine tuning

S. Agarwal (✉) · Sukritin · A. Sharma · A. Mishra
Department of Computer Science & Engineering, ABES Engineering College, Uttar Pradesh, Ghaziabad, India
e-mail: sanskriti.17bcs1@abes.ac.in

Sukritin
e-mail: shukritin.17bcs1128@abes.ac.in

A. Sharma
e-mail: aditya.17bcs1010@abes.ac.in

A. Mishra
e-mail: anurag.mishra@abes.ac.in

A. Mishra
Dr. APJ Abdul Kalam University, Lucknow, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_10

1 Introduction

Considering the current digital scenario, everyone works on digital writing platforms, where they need to write the different kind of documents. Using the technologies, we can make this work easy. Next word prediction is a process that predicts next word according to the understanding of current sequence of words. In this project, we are creating the system that does this process, i.e., next word prediction in Hindi language. There are various machine learning technologies which we can use to complete this process. Next word prediction comes under the machine learning technology that is Natural language processing. Natural language processing means to do the modeling of language that is convert the data into the natural language. There are various applications of NLP such as automation of text, language translator, chatbots, voice assistants, etc. Automatic text generation is same as next word generation as we are generating new text from the input.

Hindi is a most frequent language that is spoken in India. Therefore, creating system that predicts words in Hindi will be very helpful for people living in India. As compared to English language processing done on Hindi language is tough because there are various symbols and mantras. So, to make processing of language easy and less complex using words to get less confusing records.

This paper proposes the method to create the system that tells user the predicted next words according to the current words from which user can choose word accordingly.

There are various advantages of this system such as:

1. Make typing less time consuming.
2. Make typing error-free.
3. Helps the user to learn language in case of new learner.

The below sections tell more about the system.

2 Purpose

People are very imaginative yet sometime they lack new ideas or cannot come up with new ideas by themselves. As we read a sentence, we perceive each word depending on our knowledge and understanding of past words. As human beings, we have the ability to think, and our consideration proceeds. But this is not possible with the case of Conventional Neural Networks (CNN) as they do not have the ability to collect and remember the previous inputs.

The Recurrent Neural Networks (RNN) find its use in the text message, address, e-mail, and the like. Although we have the ability to convey our thoughts continuously but this is not possible for Traditional Neural Networks to convey. They fail in the course of such thoughts.

The principle purpose of our project is to provide this kind of ability to digital system using machine learning technologies. As everything is shifting toward digital world, this type of systems make writing efficient on digital platform.

This system provides efficiency not only at individual level but can also provide efficiency at organization level. This system will make typing easy and error-free by predicting next sequence of words with the help of understanding of previous words.

3 Motivation

In this era, where many things have gone digital but the written words still occupy importance and authenticity. We regularly deal with the task that involves writing on a digital platform. This may include writing long text messages, addresses, e-mails, or other documents. Generally, every person cannot write these texts with ease in a single go and in the right way. So by making this entire process more efficient and accurate, it can be used more efficiently and which will take less time.

The next word prediction tool lessens the probability of having mistakes, while typing a paragraph or text. It has proven to be faster and beneficial in typing proficiently. The prediction has proven itself to be a helping hand for the needy who cannot think of a new word instantly or the person who is having slow typing speed.

Talking about our country where people are more fluent in speaking the Hindi language, it develops a need for Hindi Prediction Tools. It is difficult to write a whole sentence in Hindi without making errors and mistakes. These tools can decrease the odds of having errors, while typing.

4 Related Previous Work

Sharma et al. [1] suggest the use of LSTM and Bi-LSTM with an accuracy of 59.46% and 81.07%, respectively, for the task of next word prediction in Hindi language.

They proposed that the sentences is splitted into 6:1 ratio. It predicts the upcoming word based on the basis of previous sequence of 6 words. The initial 6 words are fed to the neural network, and the 7th word is predicted as the output of the model.

The recent approaches for next word prediction are based on the probability distribution of the Language Model [2]. It uses REEDS (Relevance and Enhances Entropy based Dempster Shafer Approach) to reduce the computational burden and to reduce the long learning time of the neural network.

The proposed method performs better than the traditional neural network. It is computationally efficient and reduces training time and computation burden.

Statistical predictors [3] are used for text prediction, and these are mainly n-gram language models. N-gram generates predicted words sequence based on the recently used words and frequency of words.

It increases the keystroke savings by 43.05% when compared to other existing word prediction system like 16.82% (Google). The text entry rate (wpm) also increased to 12.56 as compared to 4.84 (Google).

N-Gram approach doesn't consider grammar and structure of the sentence [4]. It only takes sequences of words into consideration.

Syntactic N-Gram uses different approach from traditional N-Gram approach. It consider syntactic connection between words, thus, each word is linked to its original neighbors, ignoring the randomness that is produce by the structure of sequence of text.

The language model can be used for various tasks like NER, Question—Answering, etc. The Question —Answering (QA) system [5] is used to answer the query of the user in Hindi language. It uses the classification algorithms KNN and nearest neighbor algorithm. It uses Smith Waterman similarity function which outperforms other similarity functions. The KNN algorithm and Nearest Neighbor (NN) algorithm gives 97% and 93.3% accuracy, respectively.

According to [6], Recurrent Neural Networks are used to generate output sequences based on the previous input sequences.

It also shows the non-effectiveness of gradient based learning algorithm. These algorithms are not effective for learning long-term dependencies and also suffers vanishing gradient problem which makes input to output mapping difficult.

The limitation of vanishing gradient in RNN is solved using the LSTM model. LSTM (Long Short-Term Memory) is used to predict the next word for Assamese phonetic transcription [7]. The maximum accuracy achieved is 72.10%. The neural network structure used—3 layer—256 neuron in each layer. The learning rate used is 0.001 and number of epochs are 77,000.

According to [8], BERT can be pre-trained to learn pre-training loss. The pre-training phase uses non-annotated data (unsupervised learning) to learn the weights of the model. BERT base model has 12 hidden layers, 768 hidden nodes, 12 attention heads and total of 110 M parameters.

Kushal et al. [9] proposed transformer architecture that have NLP capabilities as classifications and answering.

Anoop et al. [10] introduced IITs English-Hindi Parallel Corpus that exhibits, compilation of parallel corpora previously available in the public domain as well as new parallel corpora.

This pre-trained model can be fine-tuned to perform various NLP tasks like classification, question answering, named entity recognition, etc. (Table 1).

5 Proposed System

Our system have three main components first component is collection of data the second component is modeling of data and third component id integrating of data.

In first component the data is collected which is used for training and testing according to which final output to user is given, i.e., predicted words.

Table 1 Previous works algorithms

S. No.	Year	Title	Algorithms used
1	2020	Indic-Transformers: An Analysis of Transformer Language Models for Indian Languages	BERT, Distill BERT, RoBERT and XLM- RoBERTa
2	2019	Next Word Prediction in Hindi Using Deep Learning Techniques	LSTM and Bi- LSTM
3	2019	REEDS-Relevance and Enhanced Entropy based Dempster Shafer Approach for Next Word Prediction using Language Model	REEDS optimization for bi-gram for statistical model
4	2019	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	Pre-training BERT and Fine tuning BERT
5	2018	A run based approach for next word prediction in Assamese phonetic transcription,"	RNN (Recurrent Neural Network) model
6	2016	A unique word prediction system for text entry in Hindi	Syntactic N-Gram Approach
7	2016	Performance evaluation of different similarity functions and classification methods using web based Hindi language question answering system	KNN Algorithm and NN algorithm
8	2014	Word prediction system for text entry in Hindi	N-gram approach
9	1994	Learning long-term dependencies with gradient descent is difficult	RNN (Recurrent Neural Network) model

In second component different machine learning models are used to train our collected data. In our system, BERT Model and Masked language Model are used.

In third component basically we have integrate our system with GUI which helps user to enter the words and our system predict the next words that user can type.

The second component is the main component as it provides the main functioning of system. Figure 1 shows the system architecture of our system.

5.1 Dataset

Data set that has been used was developed by IIT Bombay at the Indian Language Technology Center. The data set was collected from surfeit of already available sources and compilations.

This data set represents English—Hindi compilation which can be used for translations. In this model Hindi text is used. There are 15,61,841 number of sentences in data set. The number of sentences we have used in this project is 7000. Each

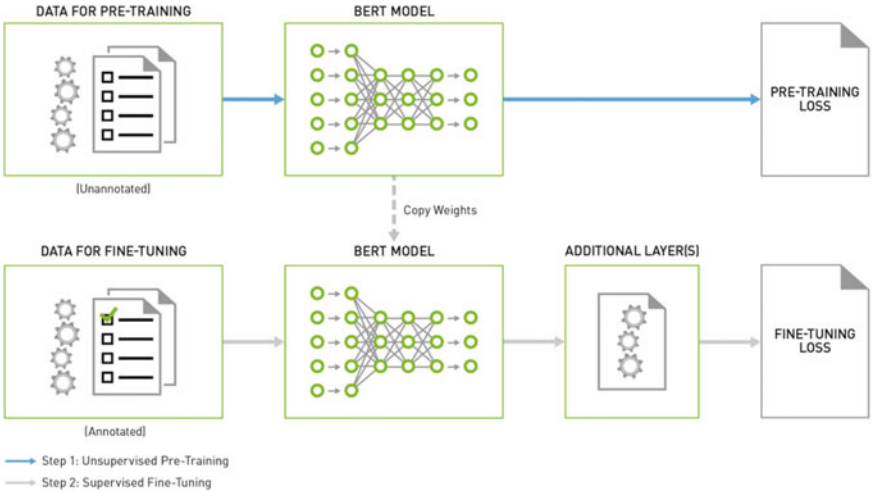


Fig. 1 System Architecture

sentence consist of minimum of seven words. In project for further computation sentences that have length equal to six words are used. The words used in the project are in Devanagri Hindi language. The data set consists of many short sentences, i.e., of length less than seven words.

5.2 BERT Model

Bert Model, i.e., Bidirectional Encoder Representations from Transformers, it uses transformers that finds the relationship between the words. Transformers has two main functioning first is to read the input that is done by encoder and other is to predict the words which is done by decoder.

It uses semi supervised learning. BERT is an encoder stack of transformer architecture with following properties that are:

- 12 Attention Heads
- 12 layers
- 768 hidden nodes
- 110 M parameters.

Bert Model works in two steps first step is to apply Bert Model on data for pre-training to calculate pre-training loss. Second step is to apply Bert Model with extra layers on data for fine tuning to calculate fine tuning loss.

Fine tuning of data is done by adding single layer on the top of the core model.

5.3 Masked Language Model (MLM)

This model is used to masked words that is to convert words into tokens before sending them to Bert model. It is a better way to train a language model in a self-supervised learning. This model can then be used for fine tuning to accomplish various NLP tasks. MLM converts 15% of words from sequence into tokens. Final predicted values are shown to the user using these models.

6 Implementation

The implementation for the project has been done by making the use of machine learning models.

The programming language that has been used for the implementation is Python 3. The back-end technology that has been used is machine learning. The front-end technology that has been used is Flask for GUI.

The implementation is done for three main components that are mentioned above. The first two components are implemented using machine learning technology and last component is implemented using front-end technology that is Flask.

The implementation of project is done in various steps. Below flow chart explains all the steps in which implementation is done.

The proper implementation for the project can be explained in a series of steps which have been described in a flowchart which is the process of understanding (Fig. 2).

Flowchart explains the how the flow of implementation of system goes.

Firstly, we collect the data, then to obtained the cleaned data cleaning and normali zation of data is done. For cleaned data, tokenization and vocabulary generation is done. Tokenization is the process to convert words into tokens as model accept to kens. Vocabulary generation is the process of generating vocabulary of unique words from converted tokens. Example of cleaned data is shown below in Fig. 3.

Secondly, the data set is segmented into two segments:

1. **Training Data set (80% of data set)**: This data set is used to train the system. Using this portion of data set system learns the data.
2. **Test Data set (20% of data set)**: This data set is used to test the system. Using this portion of data set validation of system is done.

Thirdly, the models are applied on the training data set. The Bert Model is trained on the custom data set and for next word prediction in Hindi hyper parameters are tuned.

The two main functions that are used to obtain desired output are “encode” and “decode.” The encoding function is used to convert the sequence of text into embeddings. Embeddings are the numerical version of tokens, i.e., tokens are converted into numerical values. This function returns two values that are input-ids and mask-ids.

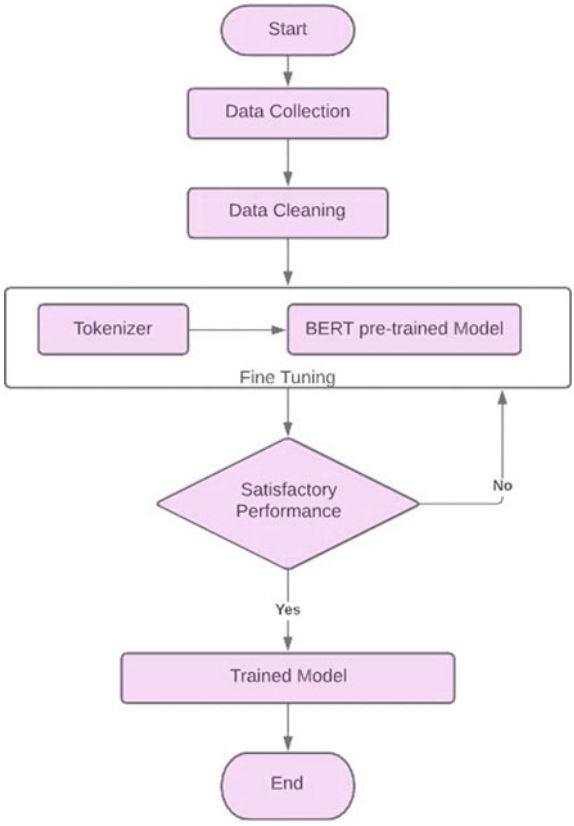


Fig. 2 Caption-process flow

Fig. 3 Cleaned data

वो वैन के पास आ रहे है
वो उसका इस्तेमाल बर्फ हटाने की बजाय
यूएनएफपीए सहस्राब्दि विकास लक्ष्यों से संबंधित मामले
ऑर्गनिक उत्पादन का प्रमाणीकरण प्रणाली आरंभ करना
हम समाज में परिपूर्णता नहीं पा सकते
चयनित रेसिपि आयात की जा रही है
और कई बार हम बहुत जतन करके
स्मार्ट डेटा को ताज़ा करने में त्रुटि
मृतजात शिशु प्रसव पूर्व मर जाता है
सामान्यतः पुरुष गंजेपन का शिकार होते हैं
मुखीय गृहिका को मुंह भी कहते हैं
चयनित संदेश को नए विंडो में खोलें

The decoding function is used to present output to the user by converting numerical values into string formats which are human understandable, and it predicts the first k predicted words.

The loss after training is:

The model is evaluated with a loss of 2.992 on validation data.

7 Conclusion and Future Scope

As in today's world, everything is moving toward digital. Writing at digital platform is a common task. Typing has become the important task and very frequent. So, to make typing more efficient and error-free in most frequent language, i.e., Hindi language we require system such as next word prediction system. In this paper, we have used machine learning technique, BERT Model and Masked Language Model is used for predicting words in this system. This ensures a good accuracy. Using these models, we train our model to predict the top k words according to current sequence of words. User can choose desired word from these predicted words (Figs. 4 and 5).

Epoch	Training Loss	Validation Loss	Runtime	Samples Per Second
1	No log	3.208789	2.027000	703.488000
2	3.889100	2.967895	2.286000	623.798000
3	3.172000	3.010731	2.243100	635.718000

Fig. 4 Training loss

```
{'epoch': 3.0,
 'eval_loss': 2.922750949859619,
 'eval_mem_cpu_alloc_delta': 0,
 'eval_mem_cpu_peaked_delta': 0,
 'eval_mem_gpu_alloc_delta': 0,
 'eval_mem_gpu_peaked_delta': 720004608,
 'eval_runtime': 2.2462,
 'eval_samples_per_second': 634.839}
```

Fig. 5 Evaluation loss

In future, work can be improved for different natural language processing applications like for automatic task compilation of poetry, story, etc.

For this system, there is a particular data set that we have used. To increase the performance of system, more words can be used and also if we focus more on Hindi mantras and symbols the system performance can increase more and scope of predicting more words will also increase.

References

1. Sharma R, Goel N, Aggarwal N, Kaur P, Prakash C (2019) Next word prediction in hindi using deep learning techniques. In: 2019 International conference on data science and engineering (ICDSE)
2. Lal Prajapati G, Saha R (2019) REEDS: relevance and enhanced entropy based Dempster Shafer approach for next word prediction using language model. J Comput Sci
3. Sharma MK, Smanta D (2014) Word prediction system for text entry in Hindi. ACM Trans Asian Lang Inform Process
4. Shah NN, Bhatt N, Ganatra A (2016) A unique word prediction system for text entry in Hindi. In: Proceedings of the second international conference on information and communication technology for competitive strategies. ACM, p 118
5. Devi R, Dua M (2016) Performance evaluation of different similarity functions and classification methods using web based Hindi language question answering system. Procedia Comput Sci 92:520–525
6. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5:157–166
7. Barman PP, Boruah A (2018) A RNN based approach for next word prediction in Assamese phonetic transcription. In: 5th International conference on advances in computing and communication
8. Devlin J, Chang M, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv.Org. https://arxiv.org/abs/1810.04805
9. Jain K, Deshpande A, Shridhar K, Laumann F, Dash A (2020) Indic-transformers: an analysis of transformer language models for Indian languages. ArXiv.Org. https://arxiv.org/abs/2011.02323
10. Kunchukuttan A, Mehta P, Bhattacharyya P (2018) The IIT Bombay english- hindi parallel corpus. In: Language resources and evaluation conference

Evolution of the Evolutionary Software Development Methods

Anurag Mishra, Shailesh Tiwari, and Ashish Sharma

Abstract In the early days of software development model, researchers expressed his views on handling various software projects of many years of experience. What they felt like handling it difficult for large software projects. So, they proposed a method that was considered as the base and standard till late 90s for handling various types of software projects. As software projects were moving towards more commercial utilities and more customer centric, there were demands that were concerned with lower or no downtime, rapid updates implementation and highly available software projects, so it was evident that changes in the methodologies were required and various development groups were creating and naming their own standards as per their own need of aspirations. Clearly, it was the era of evolutionary life cycle methods. Then almost all these evangelists met and created a common manifesto that was considered as Agile methodologies. In this document, we have tried travelling from Waterfall model to DevOps via Agile. In this study, we have collaborated the various evolution strategies and their need.

Keywords Evolutionary · DevOps · Software engineering

1 Introduction

Around for more than 15 years, almost every text book and commercial purposes used the either waterfall model or similar model like waterfall model. But around in mid-90 s, it was started the feeling of being stagnate in the process of waterfall model, and according to Tom glib [1] two main issues with the waterfall model can be described as below. He associated the terms like unrealistic and dangerous with this model [2–6].

A. Mishra (✉) · S. Tiwari
ABES Engineering College, Ghaziabad, India
e-mail: anurag.mishra@abes.ac.in

A. Mishra · A. Sharma
GLA University Mathura, Mathura, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_11

1. First main issue is to choose one delivery date, and every planning is revolved about that one date.
2. Every detailed and minor planning and analysis is to be done before coding and testing, and it is very unrealistic.

The need of evolutionary method was clearly felt, and many of the development teams were incorporating as per need. But it was not documented and published and was waiting for to get recognized in categorical and authentic way so that fraternity could take advantage of this evolutionary delivery methods as per need.

A term EVD was coined initially which is having detailed meaning of evolutionary delivery and principally was based on following items.

a. Deliver for real user
b. Measure the values
c. Adjust the design based on feedbacks.

Used EVD Model

See Fig. 1.

Few take away from this model

a. It doesn't declare anything as best step.
b. No permanent fixes in quick manner.
c. Being work in recursive manner, every involved individual can notice its contribution and upgrade its value towards the project according to the feedback received.

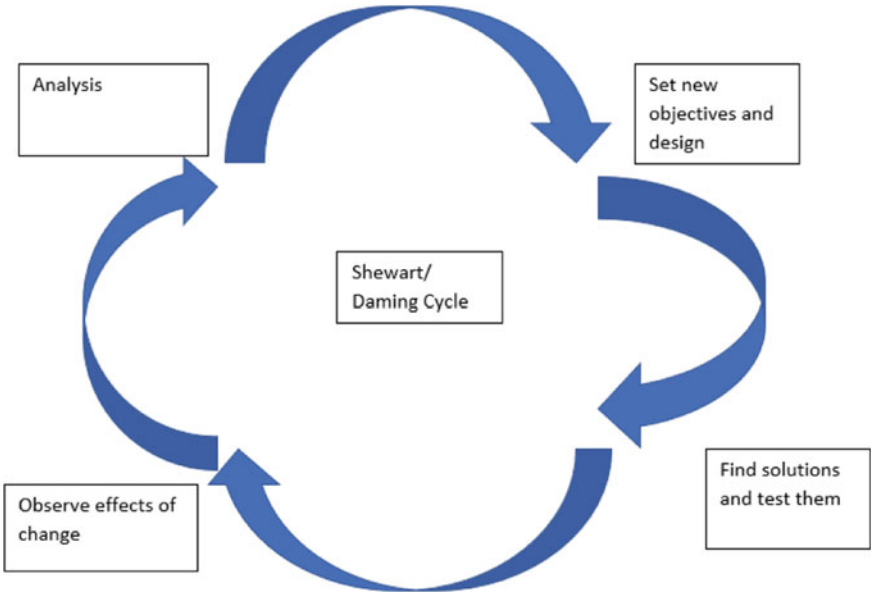


Fig. 1 A simple diagrammatic perception of EVD

Important characteristics of EVD

a. Multi-objective driven
b. Early-frequent driven
c. Analysis in each step
d. Design by feedback and objects updates
e. Focus on user need
f. Open-ended system
g. Result oriented.

2 Agile Manifesto

It was evident that there were many parallel schools, running on similar grounds sort of evolutionary computing unlike waterfall models, based on customer regular feedback and type of continuous delivery. For coming on common platform in Feb 2001, 17 technology experts belonging to different thought of schools, which were coming from Extreme programming, SCRUM, DSDM, Adaptive software development, crystal, feature-driven development, pragmatic and some other who were constantly feeling the need of a software development life cycle which can take load of large software.

What came out from this meeting of minds was termed as "Agile Manifesto", and it was beautiful in many senses and of course customer feedback was the heart of the Agile practices [7].

Principles behind the Agile Manifesto (Taken as it is described in Manifesto)

We follow these principles:

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Business people and developers must work together daily throughout the project. Build projects around motivated individuals.
5. Give them the environment and support they need, and trust them to get the job done.
6. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7. Working software is the primary measure of progress. Agile processes promote sustainable development.
8. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9. Continuous attention to technical excellence and good design enhances agility.
10. Simplicity--the art of maximizing the amount of work not done--is essential.
11. The best architectures, requirements, and designs emerge from self-organizing teams.
12. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

3 DevOps

In the exact terms of Christof Ebert DevOps can be defined as.

> DevOps is about fast, ¬ flexible development and provisioning business processes. It efficiently integrates development, delivery, and operations, thus facilitating a lean, ¬ fluid connection of these traditionally separated silos

The requirement of continuous demand and supply was getting felt in the domains of development and operations; DevOps comes with the answers handy of these situations by automating development and deployment. In a nutshell, we can say, it's a cultural mix for different groups in the organizations and imparting more collaborative deliveries for cross-functional task forces. As per expectation, DevOps practices were able to produce deliverables at faster rate and overcoming the situations which arises due to miscommunications [8–10].

DevOps empowers the mutual collaboration among quality and development departments. We can draw a representative flow of events which can elaborate different collaborations and deliverables and frequency of meetings. As tech giants like Google and Amazon have shown the world path in this manner for continuous deliveries, for example, in place of actual software deliverable, it can be replaced by corresponding service.

Tools Used in DevOps:

The main task of adopting DevOps is to adapt mind-set of using tools for automating the process. To achieve high quality of standard, it's very essential to adapt very level of automation and picking perfect tools as per requirement is in fact the task of DevOps experts [11–14].

Tools required to Automate Builds:

Code compilation and to deal with interdependencies, creating documentations, perform testing and deploying and much more tasks which need to be automated for efficient and continuous deliveries and can be judged as the crucial task for DevOps culture.

Apache Ant:

First tool which we are discussing is the Apache Ant as its open source and having specialization when you need to build the application using sources. Apart from this, it contains the virtue having simplest syntax in comparison to its counterparts, and frankly sometimes it becomes overhead for some basic tasks. Main cause of this is having used of XML.

Maven:

If we have to count on the merit of Maven, clearly it can overcome some of the issues of Ant's. Even though it has XML as backbone for creation of build process. In a summarized manner, Maven targets to define project object model (Fig. 2).

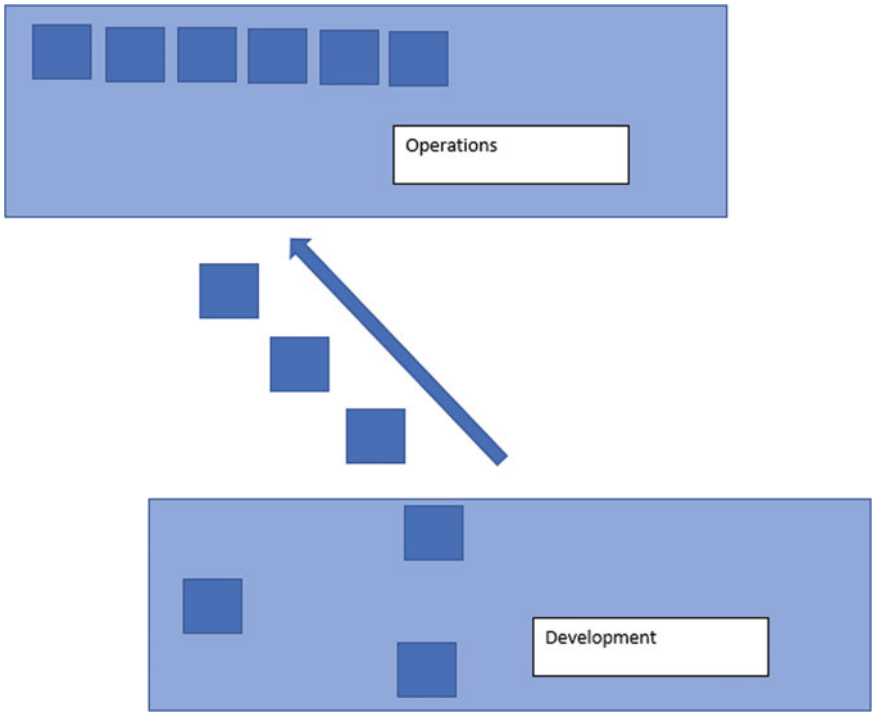


Fig. 2 Illustration of continuous development and deployment process

Rake and Gradle:

As we discussed that Ant and Maven both were using XML as their backbone documentation for automating build, Rake and Gradle tried to solve this issue by adopting a programming procedural language for the. Rake used the language Ruby, while Gradle adopted a domain-specific language based on Groovy.

Integration Tools:

In DevOps culture, we say it as Continuous Integration or as CI. Main objective of CI is to put the work of developer on testing as soon as possible to make it go for a test run.

Most used and appreciated tool for this regard is Jenkins, which is and open-source Java-based tool. It is widely used to carrying a large user base for support. Another tool is TeamCity which is also Java-based and having very good support for Java as well as for .Net.

Deployment Tools:

Deployment is the step where development needs to be in sync with continuous manner. To think in this manner deployment which is considered as administrator's

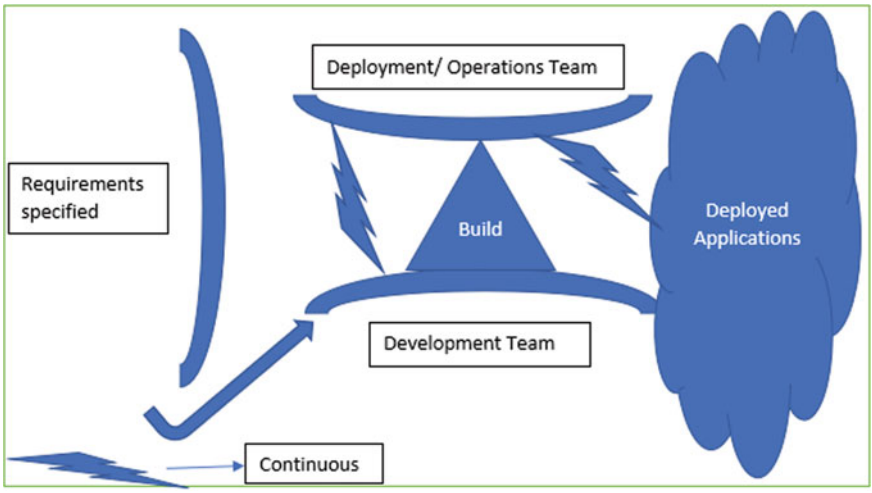


Fig. 3 Generic DevOps process-continuous development and delivery

step needs to be redesign and rethink as a task to be done as coding. So, goal is to keep on maintaining automation culture developed in development into deployment. When we achieve continuous delivery, it almost becomes synchronized process to get feedback from user after latest automated build and to get it published (Fig. 3).

Puppet is a deployment provisioning tool which is based on Ruby. It uses data-specific object model to provide client and server provisioning. Another widely used tool is Ansible, which is very easy to use because it doesn't prompt client machine to be equipped with agents and basically it uses PUSH configuration on SSH and based on Python.

Logging Tools:

When working with DevOps, it is almost necessary to log everything happens as being automated build to deployment. It's required to have traces for every shift and transitions. CI is the core of Agile, and it actually serves the purpose Agile took over the traditional software development methodologies. CI keeps on uploading the tasks of developers frequently and at a pace as fast as possible, so that it always be on testing phase and in early manner (Table 1).

Conclusion:

As being graduated slightly, it was even clear from the advent of twentieth century that it's going to be the work of continuous updates as neither user nor vendor can take overhead left for being in discontinued manner so getting update in a serial and regular level is quite a required and important move. To process automated build, deployment and monitoring a variety of tools are being listed and few important of them and have been discussed in the document. Still, we can feel the required area

Table 1 Summarization of widely adapted tools

Tool	DevOps phase	Tool type	Configuration format	Language	License
Ant	Build	Build	XML	Java	Apache
Maven	Build	Build	XML	Java	Apache
Rake	Build	Build	Ruby	Ruby MIT	MIT
Gradle	Build	Build	Based on Groovy	Java	Apache
Jenkins	Build	Configuration integration	UI	Java	MIT
TeamCity	Build	Configuration integration	UI	Java	Commercial
Bamboo	Build	Configuration integration	UI	Java	Commercial
Puppet	Deployment	Configuration integration	DSL similar to jason	Ruby	Apache
Chef	Deployment	Configuration integration	Ruby-based DSL	Ruby	Apache
Ansible	Deployment	Configuration integration	YAML	Python	GPL
Laggly	Operations	Logging		Cloud Based	Commercial
Graylog	Operations	Logging		Java	Open source
Nagios	Operations	Monitoring		C	Open source and GPL
NewRelic	Operations	Monitoring			Commercial
Cacti	Operations	Monitoring		PHP	GPL

Taken from IEEE software magazine

which DevOps culture needs to look more upon is the security channel vulnerability and fault tolerance and recovery processes.

Road Ahead:

As in this study, we have seen how from classical model developers graduated towards DevOps. Today software world is basically looking for such solution for their SDLC which can deal with the updates in a better way, and DevOps and its future researches would definitely would be in this domain only [15].

References

1. Gilb T (1985) Evolutionary delivery versus the “waterfall model”. ACM Sigsoft Softw Eng Notes
2. Rovce WW (1970) Managing the development of large software systems. IEEE, WESCON

3. BecK K (1999) Embracing change with extreme programming. IEEE-First Class Softwares
4. Schwaber K, Sutherland J (1995) The SCRUM development process. In: OOPSLA, Conference '95 in Austin, Texas
5. Oreizy et al (1999) An architecture based approach to self adaptive software. IEEE Intell Syst
6. Thomas D et al (1999) Pragmatic programmer. Addison Wesley
7. Beck K et al (2001) Agile manifesto. Evangelist Round table, Utah
8. Virmani M (2015) Understanding DevOps & bridging the gap from continuous integration to continuous delivery. In: Proceedings of 5th international conference on innovative computing technology (INTECH 15), pp 78–82
9. Spinellis D (2012) Don't install software by hand. IEEE Softw 29(4):86–87
10. Hernantes J, Gallardo G, Serrano N (2015) IT Infrastructure-monitoring tools. IEEE Softw 32(4):88–93
11. Ebert C, Gallardo G, Hernantes J, Serrano N (2016) DevOps. IEEE Softw
12. Mathieson JT, Mazzuchi T, Sarkani S (2020) The systems engineering DevOps lemniscate and model-based system operations. IEEE Syst J
13. Hasselbring W et al (2020) Industrial DevOps. IEEE-ICSA-C
14. Hamunen J (2016) Challenges in adopting a Devops approach to software development and operations. Published at Aalto University
15. Abantika Chaterjee (2018) Website Links: https://dzone.com/articles/overcoming-the-challenges-of-devops-implementation

An Application for Facial Image Beautification

Payal Bose, Samir K. Bandyopadhyay, Amiya Bhaumik, and Sandeep Poddar

Abstract In today's world, social media, virtual meetings, online demonstrations, etc. are rapidly growing in different parts of the world. With the rapid growth of the digital world, image and face beautification became one of the biggest concerns. Each human wants to look perfect and natural through the viewfinder. Therefore, appearance in an image became the most important factor. In this kind of situation, cosmetic process is not very promising all of the time and it is cost-effective. Therefore, the beautification application plays the important role in these circumstances. In this experiment, an image and face beautification application is created. Different tools and filters are created to make the input image more natural and dramatic. This application is fully in the choice of human. Every tool can be adjusted by the slider option. Finally, a user can easily beautify their image without any prior knowledge of the application.

Keywords Face application · Standalone software application · Filters · Presets · Image beautification · Face beautification · Cosmetic process

1 Introduction

Unattractive looks and poor image quality is the biggest issue for users. With the rapid growth of social media, everybody wants to look natural and beautiful. Therefore, demanding for the beautification process is much higher. At first, cosmetic process was the biggest concern for human beings for making themselves beautiful. But in today's world, the virtual conference or social media appearance, the beautification application is much more acceptable than the cosmetic one. These beautification applications motivate the users to communicate with others through virtual meetings

P. Bose (✉) · S. K. Bandyopadhyay · A. Bhaumik · S. Poddar
Lincoln University College, Kota Bharu, Kelantan, Malaysia

A. Bhaumik
e-mail: amiya@lincoln.edu.my

S. Poddar
e-mail: sandeeppoddar@lincoln.edu.my

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_12

more confidently. This also helps to enhance the image or video quality and make the image more vibrant. Image retouching not only helps to beautify the image but also helps to make the image more natural and attractive. In this application, an image and face beautification application are built up. This application not only helps to beautify a human face but also helps to beautify the background image.

2 Existing Work Details

Face and image beautification are popular topics nowadays. Because the face is an important part of the human body. It is biggest concern of a person to make his/her face beautiful. The researchers show in their research papers that in this digital era, it is possible to create a virtual mirror. This mirror can help to beautify a human face using different beautification techniques [1]. For face beautification, face retouching or face skin smoothing is one of the most important tools. The researchers propose a real-time face smoothing tool in their paper. This tool helps to reduce skin blemishes and make the skin smooth. They also compare this tool with different commercial products and got a good result [2]. Brightness and contrast adjustment are another important part of image beautification. The researchers proposed two novel histogram-based algorithms to enhance the brightness and contrast of an image. They only adjusted two factors to control the brightness and contrast to enhance the image quality [3]. Image cropping is also an important operation for beautification of facial image. The researchers proposed an automatic image cropping operation in their research paper. They studied different traditional approaches and create an extensive study in their paper [4]. Sometimes it is needed to change the color of an image to make the image more dramatic. The approaches of researchers are color printing approach employing sub-micrometer scale plasmonic pixels of a single constituent structure method. This method helps to adjust the hue, brightness, and contrast of an image and create a more natural saturated image [5].

3 Experiment and Experimental Result

In this experiment, a graphical user interface is created to beautify an input image. To beautify an image, the graphical user interface has different categories. Figure 1 shows the categories of the application.

Again all the main categories are divided into sub-categories based on the requirements.

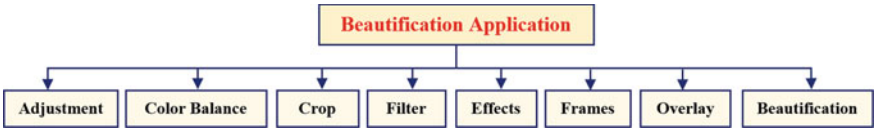


Fig. 1 All categories in the beautification application

3.1 Adjustment

The adjustment tab is used to adjust or fix the image over or low exposure, balance the contrast level, sharpen the image, or used to adjust the hue of the input image. In this experiment, the adjustment tab is divided into four sub-categories.

Brightness. The brightness tool [6] is used to adjust the exposure or the brightness level of the input image. It is a very important and the simplest tool for the image editor. This tool helps to adjust the brightness level of each pixel in the entire image. The result of this tool for different levels is shown in Fig. 2.

Contrast. The contrast tool [6] is used to remap the image intensity value within a full display range of data. A good contrast image can differentiate the different levels of white and black pixels. Different levels of image contrast for this application are shown in Fig. 3.

Sharpness. The sharpness tool [7] usually builds to enhance the low-quality image by enhancing the brightness level at the edges of an object in an image. It is used to

Fig. 2 Images with different brightness levels

Fig. 3 Images with different contrast levels

Fig. 4 Images with different sharpness levels

Fig. 5 Images with different tint levels

enhance the overall image clarity. Increasing the sharpness level gives the detailed texture of an object in an image. Figure 4 shows the different sharpness levels of the input image.

Tint. Tint tool is used to recolor the image within a certain range [8]. It is basically the change of hue of an image. The range of colors lies between 0 and 1 of an image in the HSV (hue, saturation, and value) color space image. Figure 5 shows the different tint levels of the input image.

3.2 Color Balance

The color balance tool is one of the most important tools for image beautification. It helps to adjust the intensity level of the three channels—red, green, and blue—of the input RGB image. Balancing the color gives the output image a natural look. In this experiment, the color balance tab is divided into sub-categories. They are described below.

Saturation. Color saturation [9] is also known as the intensity of color channels or the purity of color in an image. The higher saturation means the intensity of color channels is more vivid. The lower saturation means less intensity of all color channels. At the very low saturation value, the RGB image turns into a pure grayscale image. The saturation level of the input image is adjusted by using the saturation channel in HSV color space. The different levels of higher and lower saturation of the input image are shown in Fig. 6.

Fig. 6 Different levels of saturated RGB color image

Fig. 7 White balance levels **a** −ve maximum saturation with 20% white balance, **b** up to –ve 25% saturation with 20% white balance, **c** original image, **d** up to +ve 25% saturation with less than 20% white balance, **e** +ve maximum saturation with greater than 20% white balance

Fig. 8 Black balancing **a** −ve maximum saturation with 10% black balance, **b** up to –ve 25% saturation with 20% black balance, **c** original image, **d** up to +ve 15% saturation with less than 10% black balance, **e** +ve maximum saturation with 10% black balance

White Balance. To adjust the exposure of the saturated image, the white balance tool is used in this experiment. The result of different white balance levels of a saturated image is shown in Fig. 7.

Black Balance. Black balancing is the adjustment of the contrast level [10] of a saturated image. This black balancing tool helps to adjust the exposure level of the saturated image more precisely. The result of different black balance levels of a saturated image is shown in Fig. 8.

3.3 Crop

Cropping an image means trimming the input image [4] from its required edge. This tool helps to change the aspect ratio of an image. In this experiment, four

Fig. 9 Crop tool. **A** Original Image, **a1** custom cropping area, **a2** cropped area, **b1** square (1:1) cropping, **b2** cropped area of b1, **c1** (2:3) pre-defined crop area, **c2** cropped image of c1, **d1** (9:16) pre-defined crop area, **d2** d1 cropped area, **e1** (4:3) pre-defined crop area, **e2** e1 cropped area, **f1** (5:4) pre-defined crop area, **f2** f1 cropped area

types of cropping are used, (1) custom, (2) perfect square, (3) horizontal cropping, and (4) vertical cropping. For custom cropping, users can ably draw a responsive rectangle box to crop the required area. For square, horizontal, and vertical cropping, a pre-adjusted aspect ratio rectangle is created. This rectangle box is draggable. The pre-adjusted aspect ratios are 1:1 for square cropping, 2:3, 3:4, 4:5, 9:16 for vertical cropping, and vice versa for horizontal. After placing the draggable box at the required area, we need to double click on it. Then the area is cropped automatically. The results are shown in Fig. 9.

3.4 Filter

In the image beautification process, a filter is one of the most important tools. It uses to either suppress the high frequency or smoothing the image or enhancing the different levels. Filter tool can help to change the appearance of some parts of the image without changing the pixel position. It is also used to add different shades, textures, tones, and special effects to an image. In this experiment, six different preset filters and one noise filter are used.

Vintage. Vintage effect is used to create a cinematic or a nostalgic effect in an image. Here to create saturation level, it is needed to convert the input RGB image into HSV color space image. The color effect is adjusted by using the below algorithm.

```
Step: 1 Convert RGB to HSV
Step: 2 reduce the saturation level (0.3 to 0.7)
Step: 3 adjust the brightness (0.10 to 0.35) and contrast level (0.0
to 0.12)
Step: 4 adjust the red channel (0.10 to 0.40) and greed channel color
(0.00 to 0.20) of the new saturated image
```

Pastel and Pastel Grain. Pastel color filter is used to reduce the complex colors to a certain level. It is basically used to soothe the image. Here to perform the operation conversion pastel, first convert the input color image into an indexed level image. Then adjust the color map index. Again, for operation pastel grain add salt and pepper noise into the pastel conversion image. The algorithms are given below.

```
Step: 1 Convert RGB to Index mapping Image
Step: 2 adjust the index mapping level (4 to 25)
Step: 3 For Pastel grain add salt and pepper noise.
Step: 4 Adjust the dither level(10 to 20)
```

Sketch. Sketch is the rough and unfinished freehand pencil drawing. To convert an image into a sketch image, here the following steps are followed.

```
Step: 1 Convert RGB to grayscale
Step: 2 create an 8 X 8 filter matrix
Step: 3 filter the image using the filter matrix
Step: 4 subtract the filtered image from 255
Step: 5 use the gaussian filter to reduce the noise
Step: 6 sharp the edges to clearer the image
```

Soft Pink Glam. Soft glam means where the image looks soft and glowy. All the colors are in a balanced state. Not in full saturation level but no over color is used. In this application, the soft glam filter is known as a soft pink glam filter. The effect of this filter is pink-based to look a little dramatic. The algorithm is given below.

```
Step: 1 Convert RGB to HSV
Step: 2 Adjust the saturation level (0.2 to 0.6)
Step: 3 Convert HSV to RGB
Step: 4 Adjust the R, G, B channel layers to create the soft pink look.
(0.9<R<1.17, 1.12<G<1.19, 1.1<B<1.29)
```

All the results of filter tools are shown in Fig. 10.

3.5 Effects

In this experiment, the effect tool is applied only to the background of the input image. The users select the area where the effect tool is applied shown in Fig. 11b. This tool has two sub-categories.

Fig. 10 Different filters **a** original image, **b** vintage, **c** pastel, **d** pastel grain, **e** sketch, **f** soft pink glam

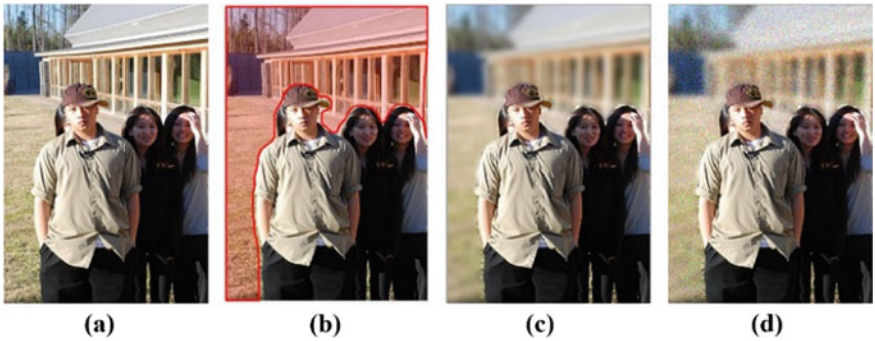


Fig. 11 **a** Original image, **b** user-selected area for using effect tools, **c** simple blur effect, **d** noise blur effect

Background Blur Effect. This effect tool uses to blur the user-selected area shown in Fig. 11c. In this experiment, the user can adjust the strength of the blurriness using the slider option.

Background Noise Blur Effect. Here a salt and pepper noise [11] is used to blur the user-selected area shown in Fig. 11d. The user also adjusts the noise blur effect by using the slider option.

3.6 Frame

The frame is a decorative edge of an image. This tool is used to decorate an image from its edge. It gives much dramatic visual effect and added depth to an image. In this experiment, different color frame as well as blur background portrait frame is created. All the frame details are shown in Fig. 12f. The user can able to adjust the frame size using the slider option. Figure 12 shows the images using different frames.

Fig. 12 Frame tool, **a** original image, **b** white border frame, **c** black border frame, **d** color border frame, **e** blur portrait frame, **f** different frames in the application

3.7 Overlay

The overlay effect [12] is used to add an extra image or texture layer to an input image. When an image is captured by a camera, then not all elements are added to the image. To add that extra layer of those elements, the overlay effect is used. This is actually the post-production process that is used to enhance the outcome of the input image. This effect creates a stunning look for an image. In this experiment, the overlay tool is used to create a stunning partial color effect to a grayscale image. The algorithm is given below.

```
Step: 1 User needs to draw a freehand masking area where the effect
is required
Step: 2 Convert RGB input image to Grayscale with 3 channels
Step: 3 Subtract the selected masking area from the grayscale image
Step: 4 Convert the overlay image into HSV space. So that only color
area will affect.
Step: 5 Adjust saturation level using slider control to blend the
color to look natural.
Step: 6 If necessary adjust brightness and contrast using adjust
tool.
```

The result is shown in Fig. 13.

Fig. 13 Overlay effect, **a** original image, **b** user selected masking area, **c** overlay image [adjusted saturation level to look more natural]

3.8 *Beautification*

Face and image beautification is used to make an image and a face more beautiful to look. It is the revolution of making an image dramatic and a face flawless. This is the most important and necessary tool for image and face beautification. In this application, the beautification tools have three main parts.

Skin smoothing. Skin smoothing [13] tool is used to make human skin more smooth and flawless. It helps to reduce the pores and unevenness of the skin and make it more natural looking. The algorithm is given below.

```
Step: 1 User need to mark the area for apply the tool
Step: 2 Create a smoothing filter using gaussian blur
Step: 3 Apply the filter on the marked area.
Step: 4 If required adjust the smoothing effect using slider option
```

Figure 14 shows the skin smoothing effect of a face.

Face Brightness and Darkness. During the photoshoot, sometimes a particular area of an image undergoes insufficient light or overexposure. To correct those areas, these tools are used. It mainly corrects the brightness level and contrast level of an image. Here user needs to select the area first, and then the brightness or darkness filter is applied. Users can adjust the amount using the slider option. The experimental result is shown in Fig. 15.

Fig. 14 Skin smoothing effect, **a** original image, **b**1, **b**2 user selected area, **c**1, **c**2 applied effect

Fig. 15 Face brightness and darkness effect, **a** original image, **b** user selected area, **c** face brightness, **d** face darkness

Fig. 16 Eye clarity effect, **a** original image, **b** selected eye areas, **c** eye clarity image

Eye Clarity. Sometimes it is also required to make the eyes more clear [14] in an image to look the face sharper. This is basically enhancing the eye brightness. In this application, users need to select the eye areas and then the filter is applied. Users also adjust the clarity by using the slider control option. The experimental result is shown in Fig. 16.

4 Discussions

This is a standalone application used to beautify an image as well as the face of a human. This is a trial version application. Here all the main tools and their sub-tools are manually adjusted by the user. The advantages of using this application are (1) easy to use, (2) no need to be an expert photo editor to use this application, (3) this application does not need a huge amount of space to install, (4) this is a standalone application applicable for PC (personal computer), and (5) like other professional photo editors like Adobe Photoshop, Google Snapseed, this application is not very complicated to understand.

The disadvantages of this application are (1) this application has a limited amount of presets and other tools; (2) the application does not have an option to install any predefined presets; and (3) also this application is not able to save any created preset for future use like Photoshop or Snapseed.

This application is applicable for commercial purpose also. This application is used for cosmetic purpose and for any image beautification for social media also.

The future scopes of this application are that it (1) needed to add different facial presets, (2) different face medication, (3) face morphing or swapping, (4) include more trending presets, and (5) want to create a real-time image and face beautification.

Sources of Images. The images used for the experiment are taken from the pratheepan database [15] and Image of groups dataset [16].

5 Conclusion

In this experiment, a face and image beautification application is created to beautify an image. This application is totally user-dependent. All the tools and sub-tools are adjusted by the slider option. Here different image filter presets are created to make an input image different looking. Also, different image color enhancement tools are created to make an image dramatic.

References

1. Jagruti B, Kasat DD (2018) A survey on face beautification techniques without cosmetic surgery. Res Rev Int J Multidisc 3:516–521
2. Velusamy S, Parihar R, Kini R, Rege A (2020) FabSoften: face beautification via dynamic skin smoothing, guided feathering, and texture restoration. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
3. Xiao B, Tang H, Jiang Y, Li W, Wang G (2018) Brightness and contrast controllable image enhancement based on histogram specification. Neurocomputing 275:2798–2809
4. Chen Y-L, Huang T-W, Chang K-H, Tsai Y-C, Chen H-T, Chen B-Y (2017) Quantitative analysis of automatic image cropping algorithms: a dataset and comparative study. In: 2017 IEEE winter conference on applications of computer vision (WACV)
5. Chen YL, Huang TW, Chang KH, Tsai YC, Chen HT, Chen BY (2017) Quantitative analysis of automatic image cropping algorithms: a dataset and comparative study. In: Proceedings of 2017 IEEE winter conference on application computer vision, WACV 2017, pp 226–234. https://doi.org/10.1109/WACV.2017.32
6. Sinha P, Crucilla S, Gandhi T, Rose D, Singh A, Ganesh S, Mathur U, Bex P (2020) Mechanisms underlying simultaneous brightness contrast: early and innate. Vision Res 173:41–49. https://doi.org/10.1016/j.visres.2020.04.012
7. Santhalakshmi M, Sukumaran S (2019) Adaptive sturdy guided filtering technique for sharpness enhancement in two-scale image fusion. J Adv Res Dyn Control Syst 11:1364–1373
8. Kinoshita Y, Kiya H (2020) Hue-correction scheme based on constant-Hue plane for deep-learning-based color-image enhancement. IEEE Access 8:9540–9550
9. Endo H, Taguchi A (2019) Color image enhancement by using hue-saturation gradient. In: 2019 International symposium on intelligent signal processing and communication systems (ISPACS)
10. Anku A, Farnand SP (2021) Preferred white balance for applications using virtual backgrounds. Color Res Appl
11. Rahim R, Ikhwan A (2017) Implementation of modified median filtering algorithm for salt & pepper noise reduction on image
12. PicMonkey. How to use overlays. https://www.picmonkey.com/blog/video-tutorial-overlay-tool
13. Qiu J, Dai S (2018) Fast facial beautification algorithm based on skin texture preserving. J Comput Aided Des Comput Graph 30:336
14. Yildiz M, Yorulmaz M (2018) Eye gaze location detection based on iris tracking with web camera. In: 2018 26th Signal processing and communications applications conference (SIU)
15. Tan WR, Chan CS, Yogarajah P, Condell J (2012) A fusion approach for efficient human skin detection. IEEE Trans Industr Inf 8:138–147
16. Gallagher AC, Chen T (2010) Understanding images of groups of people, pp 256–263. https://doi.org/10.1109/cvpr.2009.5206828

A Framework for Component Mapping Between Event-B and Python

Rahul Karmakar

Abstract Event-B is a formal modeling language that helps to prove the critical requirement properties of a software. The failure or malfunction of this software may cause a huge loss in terms of life and money. Executable code generation from Event-B makes the validation and early verification process more accurate. The proposed framework would improve requirement specification that eventually leads toward a consistent system design. This approach is very useful for industrial automation. Event-B has the extended tool support RODIN to verify the model and generate proof obligations. This paper proposes the translation framework of Event-B to Python code. The translation rules are presented that will map each component of an Event-B model into Python class.

Keywords Formal method · Event-B · RODIN · Python · Automatic code generation

1 Introduction

Guarantying the correctness of a safety-critical system is very difficult. We manage to validate a formal model to a certain extent but still, some ambiguities remain in the final system. The specification can be early verified with executable code before going for the final design. This technique helps to fix the validation bugs effectively [1, 2].

Event-B is a formal modeling language invented by J-R Abrial in 2010. It is the extension of the classical B language [3]. Event-B is based on set theory and predicate logic. Event-B allows the system to design incrementally. Context and machine are the main components of Event-B. Context defines the static part of the model, whereas the machine defines the dynamic behavior of the model.

R. Karmakar (✉)
Department of Computer Science, The University of Burdwan, Burdwan, India
e-mail: rkarmakar@cs.buruniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_13

A context can be defined as [3] CONTEXT(S, C, A, T) where

S is the set of sets.
C is the set of constants.
A is the set of axioms A(S,C) and
T is the set of theorems.

The machine can be defined as [3] MACHINE (V, I, T, Va, A, E) where

V is the set of variables to define the state of the system.
I is the set of invariants rules that defines the properties of the variable.
T is the set of theorems for the machine.
Va is the set of variants for the machine.
E is the set of events that defines the state changes of the system.

The event is guarded with some parameters, and then some actions are performed as shown in Fig. 1. A context can be extended by other contexts and seen by other machines. A machine can be refined by other machines [4–6]. The relationships between contexts and machine are represented in Fig. 2.

RODIN [7, 8] is the tool support for Event-B model verification. It is an eclipse-based open-source software that can be extended further by plugins. RODIN generates proof obligations for Event-B models.

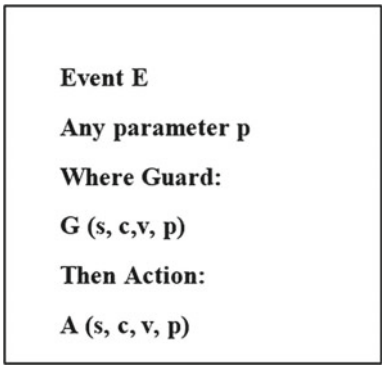


Fig. 1 Event actions and guards

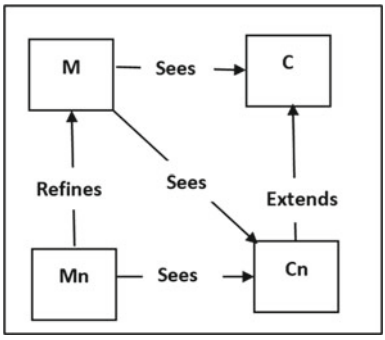


Fig. 2 Component relationships

Different plugins have been implemented by the researchers for automatic code generation of any high-level programming languages like C, Java from Event-B models. Still, there are lots of scope in this area. Automation software needs more correctness. Only a tool-dependent validation and verification is not enough for the safety-critical systems to ensure correctness. The executable codes generated from the Event-B model allow doing two-level verification. We can also add some additional features like scheduling or sequencing of events to the Event-B model for a better understanding of the system specification and design.

Python is an open-source programming language and widely accepted. It is a general-purpose language that supports both structured and objected-oriented programming . The advantage of choosing Python is its object-oriented nature, and a lot of Event-B syntax and operation are directly supported by Python. Python has the flexibility to use class, object, abstraction, encapsulation, inheritance and polymorphism features for real-world programming [9, 10].

This paper presents the translation rules for automatic code generation from Event-B to Python. The rules define how each Event-B components is mapped into Python components. After the introduction and background, in Sect. 2 the literature review has been done. The mapping rules from Event-B to Python are described in Sect. 3 followed by the conclusion in Sect. 4.

2 Related Work

We have reviewed some of the relevant works on automatic code generation from Event-B as presented below.

An interesting B2C extension tool to RODIN is presented in [11] that performs automatic code generation from Event-B to C language. The translation is done in three phases. The first phase is the rewrite phase where all the syntax mapping between Event-B and C is done.

The authors in paper [12] model the different physical measures like speed, temperature uniquely using Event-B, and generate the corresponding Java code. It addresses the issues like the heterogeneous representation of physical measure and incoherence in composition when modeling.

The work [13] presents a tool to support E2ALL of automatic code generation from Event-B to C, C++ and Java codes. It is used as a plugin in the RODIN platform. The authors present translation architecture. The Event-B code is preprocessed by converting it into target language context files. In the next step, the lexical and syntactic analysis is done for translation followed by the event scheduling and code validation.

A source-to-source mapping of the Event-B model to Eiffel programming language is presented in [14]. The authors theoretically define translation rules for Event-B components to Eiffel class. Context, machines and events are mapped into corresponding Eiffel code. An Event-B model is manually translated into Eiffel code using the proposed rules.

The Event-B to Java tool is presented in [15]. The tool automatically generates the JML annotated Java program from the Event-B model. A case study of the development of an Android application where a social event planner is designed and validated using the tool. Another case study of the tokeneer system which is a secure authentication system to enter an enclave by doing the fingerprint scanning.

Authors in [16] present a tool that combines Event-B theorem proving and SMV symbolic model checking together for better real-time verification of a safety-critical system. Event-B model to SMV model translation rules is designed first then the converter is designed accordingly.

Another work [6] presented a translation of the Event-B model into SMV.

A tool is presented in [17] that translates hardware definition language VHDL from Event-B code. The hardware system can be formalized using Event-B notation first, then the mapping between Event-B components and VHDL is done.

The JeB tool is designed in [1] for automatic simulation generation from the Event-B model. The tool has components, translators and simulators. The translator is installed as a RODIN plugin which generates the executable JavaScript

Apart from code generation, Event-B is used to design systems in different domains like verification of automatic pump controller [18], prototype modeling of irrigation system [19], formal modeling of zone routing protocol [20], graphical notation to Event-B conversion to design stop and wait ARQ protocol [21] and many more. The review shows that researchers extended the RODIN platform with plugins for automatic code generation. Different structured and object-oriented programming languages are chosen for the target language. We propose Python as the target language because it is an open-source general-purpose language that has many syntactical similarities with Event-B. Python is used as a language to convert Event-B into text [22]. But, we have not found significant works where Event-B is used as a target language by the researchers.

3 Event-B to Python Mapping Rules

We propose the translation rules as a part of a module that takes the Event-B model as input and produces Python classes. A similar kind of component mapping between Event-B and Eiffel programming language described in [14], and we take Python as a target language in this case.

Module : Event-B (total function)→Python code.

The module consists of three sub-modules to translate the Event-B model into Python code. Module (Π (Δ,$\diamond$))

Δ = It is the sub-module that maps the Event-B components (Expression and predicates) to Python code.

$\diamond$ = It is the sub-module that maps the Event-B type (variables) to Python types.

Π (Δ, $\diamond$) = It is the main module that converts the Event-B model into Python code.

The Event-B model has two main components, context and machine. We try to define the rules theoretically to map these components into the Python code. The rules are represented below.

3.1 Translation of Context

Event-B context has constants, sets and axioms. The constant and set are mapped into Python constant and set. The axiom can also map into Python axioms. All the components are directly mapped into Python code because Python supports the use of constants, sets and axioms directly. The rule is given below and the corresponding mapping is represented in Table 1.

RULE 1: Π (Axiom X(s ,c)) = (axiom X(s,c))
◇ (Constant c) = (constant c)
◇ (Set s) = (set s)

Rule 1 maps the constant and set of Event-B context directly into Python because it is supported by Python language.

3.2 Translation of Machine

The Event-B machine is mapped into Python class. The variables are directly mapped into Python class variables. The invariant rules of Event-B are mapped with Python invariant class. The initialization event of Event-B is mapped into the init function of the Python class. The mapping is represented using Rule 2, and the detailed translation is represented in Table 2.

Rule 2: ◇ (Event-B Machine) = type Δ (I(s,c,v))
=Invariant(I) Π(event e)=Python Class
Π (event initialization A(s,c,v))=init

Rule 2 maps the Event-B machine type into a Python class type. All the components of Event-B machine like variable (v) and invariant (I(Set(s),Constant(c), Variant(v))

Table 1 Mapping of context

Event-B Context	Python
Context Name C	
Extends C'	
Sets s	Set s
Constants c	Constant c
Axioms X (s,c)	Axioms(S,c)

Table 2 Mapping of machine

Event-B Machine	Python
Machine Name M sees C	Class M
Refines M'	
Variables v	Variable v
Invariants I (s,c,v)	Invariants
Variants	
Events	Functions
End	End

are mapped into Python class. The initialization event of the Event-B machine is also mapped into the init method of Python class.

3.3 Translation of Events

The events of the Event-B model are mapped into Python functions. The Guards (G (Set(s), Constant(c), Variable(v), Axioms(x))) with all the parameters are represented as the precondition operations, and the actions are represented as the postcondition operations in Python class. The mapping rule is given in Rule 3, and the detailed mapping is represented in Table 3.

Rule 3: Δ(A(s,c,v,x)) = A (for initialization event)
$\diamond$(G(s,c,v,x)) = G $\diamond$(A(s,c,v,x)) = A (for other events)

The initialization event of Event-B is translated into the init method of Python class. The event parameters are mapped into function parameters, and the action is mapped into Python statements. The other events are also represented the same way, only the guarding is done with the Python statements before the action statements are performed. The refinement relationship is not shown in Table 3 because these rules do manual translation.

3.4 Translation of Event-B Syntax

We try to summarize some of the Event-B syntaxes that can be mapped to the Python syntaxes in Table 4.

Python has different packages and classes which support the Event-B factions directly. When we go for the tool development, then we can create our packages and define different functions of Event-B like partial injection, bijection and surjection, etc., that are not directly supported by Python.

Table 3 Mapping of events

Event-B events	Python
Event_Name	Function()
Event initialization	init method(parameter)
Then A(s,c,v)	A(s,c.v)
Refines	
Event_Name E	
ANY identifier x	
Where predicates	Guard statements
Label_guard	(Preconditions statements)
With witnesses	Then
Then	Action statements
Label_action	(postcondition statements)

3.5 Manual Translation from Event-B Machine to Python Class

A simple Event-B code [3] is taken to apply the translation rules and convert the Event-B model into a corresponding Python code. A small model of a Bank that has two events Deposit and Withdraw in Event-B machine. Table 5 represents the manual conversion from Event-B machine to Python class. The Event-B context C0 has two sets ACCOUNT and PERSON. The machine name is BANK which sees the context C0. The machine has three variables act, owner and bal. The act is a subset of ACCOUNT and initialized with null. The owner belongs to act of a PERSON and initializes with null. The bal of an account is a natural number and initialize with null. These variable properties are represented by the invariant rules. The Deposit event has two parameters a and amount. The guard suggests that a is an existing account and it is a natural number. Then the action is performed where the amount is added to the balance of a. On the other hand, Withdraw has the same parameters with the additional guard checking that the withdrawal amount is less than the balance of the account.

The corresponding Python code shown in Table 5 is the manual conversion of the Event-B model. All the rules described above and with the help of Table 1 to Table 4, the Event-B machine BANK is converted into Python class Bank. The variables have the same name in the Python class. The initialization event of Event-B is translated into the init method in Python with all the actions. The Deposit and Withdraw events are also translated into the deposit and withdraw functions with the same parameters a, and amount. The event guards are translated into Python precondition statements. The actions are translated into Python postcondition statements. The overall translation architecture is not presented in this paper because we have done the component level mapping. These rules help to translate the behavior of the system into the target Python class without losing the semantics. The rules proposed in this paper will help to convert a simple Event-B model into Python code and verify the Python code successively.

Table 4 Event-B syntax mapping into Python

Event-B Notation	Event-B Syntax	Python Translation
True	$\top$	is True
False	$\bot$	is False
Conjunction	A$\wedge$B	If (a = x) and (b = y)
Disjunction	A$\vee$B	If (a = x) or (b = y)
Equivalence	A $\Longleftrightarrow$ B	No direct support
Implication	A$\Rightarrow$B	No direct support
Negation	A$\neg$	Not (a = x) and (b = y)
Universal quantifier	$\forall$	function all()
Existential quantifier	$\exists$	function any()
Equality	A = B	A == B
Inequality	A $\neq$ B	A = B
Enumeration	Enumeration A, B	using enum class
Union	A $\cup$ B	A | B
Intersection	A $\cap$ B	A and B
Difference	A\B	A $-$ B
Set membership	A $\in$ B	A in B
Not set membership	A $\ni$ *n* B	A not in B
Subset	A $\subseteq$ B	set function includes, issubset()
Not Subset	A B	Set function not includes
Proper Subset	A $\subset$ B	Set function
Not Proper Subset	A $\not\subset$ B	Set function
Finite Set	finite(s)	finite(s)
Partition of set	(S, A, B)	Set function partition()
Powerset	$\wp$(s)	Set function powerset()
Boolean True	TRUE	Type (True)
Boolean False	FALSE	Type (False)
Becomes equal to	A: =B	A = B
Becomes member of	A $\in$ B	B A
Addition	A: = B + C	A = B + C
Subtraction	A: = B $-$ C	A = B $-$ C
Quotient	A: = B/C	A = B/C
Reminder	A: =B$\div$C	A=B$\div$C
Multiplication	A: =B $*$ C	A=B $*$ C
Function	a = f(c)	a = f(c)
Less than	A $<$ B	A $<$ B
Less than equal	A $\leq$ B	A <= B
Greater than	A $>$ B	A $>$ B
Greater than equal	A $\geq$ B	A >= B

Table 5 Event-B Machine to Python Class

Event-B Machine	Python Class
Machine BANK SEES C0	class Bank
VARIABLES	def_init_(self, act, owner, bal)
act	self.bal=0
owner	self.account
bal	self.ownner
INVARIANTS	def deposit(self, amount)
iv1: act⊆ACT	Guard Statements
inv2: owner∈ act→PRS	self.amount=
inv3: bal ∈ act→N	float(input("Enter amount to be Deposited "))
EVENTS	Action Statements
INITIALISATION	self.balance+= amount
BEGIN	print("AmountDeposited",amount)
act1 : act: =∅	def withdraw(self, amount)
act2 : owner: =∅	Guard Statements
act3 : bal: =∅	self.amount =
END	float(input("Enter amount to be Withdrawn "))
Deposit	if
ANY a, amount	self.balance>=amount
WHERE	Action Statements
grd1 : a∈act	self.balance−=amount
grd2 : amount∈N	print("You Withdrew", amount)
THEN	
act1 : bal(a): =bal(a)+amount	
END	
Withdraw	
ANY a, amount	
WHERE	
grd1 : a∈act	
grd2 : amount∈N	
grd3 : amount≤bal(a)	
THEN	
act1 : bal(a): =bal(a)−amount	
END	
END	

4 Conclusion

The proposed rules mapped the components of Event-B into Python accurately. The Event-B and Python syntax mapping also helps in the automatic code generation process. The mapping shall improve the correctness of the Event-B model by validating and verifying it using the Python programming language in the early stage of the software development. We verified the Event-B model using the RODIN tool and executed the corresponding Python code to check the correctness. These rules and syntax mapping between the languages help to design the architecture of a RODIN

extended plugin that automatically translates the Event-B model into Python code. At present, we are pursuing works toward developing a tool for the purpose as future work.

Acknowledgements I sincerely thank the Department of Computer Science and Engineering, University of Calcutta, India, and Department of Computer Science, The University of Burdwan, India, for their assistance to pursue my research work.

References

1. Yang F (2013) A simulation framework for the validation of Event-B specifications. Université de Lorraine, Theses
2. Neeraj Kumar Singh (2013) Using Event-B for critical device software systems. Springer, London
3. Abrial J-R (2010) Modeling in Event-B: system and software engineering, 1st edn. Cambridge University Press, Cambridge, New York
4. Karmakar R, Biman Sarkar B, Chaki N (2019) System modeling using Event-B: an insight. SSRN Scholarly Paper ID 3511455, Social Science Research Network, Rochester, NY
5. Robinson K (2012) System modelling & design using Event-B. The University of New South Wales, School of Computer Science and Engineering
6. Sen L, Xiangyu L, Zuxi C (2017) Combining theorem proving and model checking in the safety-critical software development through translating Event-B to SMV. MATEC Web Conf 128:04004
7. Jastram M (2010) Rodin user's handbook. DEPLOY Project
8. Abrial JR, Butler M, Hallerstede S, Hoang TS, Mehta F, Voisin L (2010) Rodin: an open toolset for modelling and reasoning in Event-B. Int J Softw Tools Technol Transf 12(6):447–466
9. Kane F (2017) Hands-on data science and Python machine learning: perform data mining and machine learning efficiently using Python and Spark. Packt, Birmingham Mumbai. OCLC: 1002179793
10. Lee KD (2014) Python programming fundamentals. Springer, London
11. Wright S (2009) Automatic generation of C from event-B. In: Workshop on integration of model-based formal methods and tools
12. Gibson JP, Méry D (2018) Explicit modelling of physical measures: from Event-B to Java. In: Electronic proceedings in theoretical computer science, vol 271, pp 64–79
13. Méry D, Singh NK (2011) Automatic code generation from event-B models. In: Proceedings of the second symposium on information and communication technology—SoICT '11, Hanoi, Vietnam. ACM Press, p 179
14. Rivera V, Lee J, Mazzara M (2020) Mapping Event-B machines into eiffel programming language. In: Proceedings of 6th international conference in software engineering for defence applications, Advances in Intelligent Systems and Computing, Cham. Springer International Publishing, pp 255–264
15. Rivera V, Cataño N, Wahls T, Rueda C (2017) Code generation for Event-B. Int J Softw Tools Technol Transf 19(1):31–52
16. Hassan S, Taher M, Wahba AM (2010) Mapping SMV models to event-B models. In: 2010 5th International design and test workshop, Abu Dhabi, United Arab Emirates. IEEE, pp 161–166
17. Ostroumov S, Tsiopoulos L (2011) VHDL code generation from formal Event-B Models. In: 2011 14th Euromicro conference on digital system design, Oulu, Finland. IEEE, pp 127–134
18. Karmakar R, Sarkar BB, Chaki N (2021) Event-b based formal modeling of a controller: a case study. In: Proceedings of international conference on frontiers in computing and systems. Springer Singapore, Singapore, pp 649–658

19. Rahul Karmakar and Bidyut Biman Sarkar (2021) A prototype modeling of smart irrigation system using Event-B. SN Comput Sci 2(1):36
20. Fu C, Zheng K (2019) Formal modeling and analysis of ad hoc zone routing protocol in Event-B. Int J Softw Tools Technol Transf 21(2):165–181
21. Karmakar R, Sarkar BB, Chaki N (2020) Event ordering using graphical notation for Event-B Models. In: CISIM
22. Shchepetkov I (2020) Event-b to txt converter. GitHub Repository

A Comprehensive Investigation of Workflow Scheduling in Cloud Computing Environment

Nagresh Kumar and Sanjay Kumar Sharma

Abstract The cloud resource scalability feature allows the users' applications to meet its need during runtime or before execution and hence it required to be organized dynamically as needed. These scalable and distributed features of the cloud resources allow workflow management systems to meet the expectation of the service provider and customer. The service-level agreement (SLA) is a major concern in workflow algorithms. It also looks at the economic benefits for service providers and customers. Due to these multi-objective natures of workflow scheduling and various constraints imposed by user and cloud environment, a large number of algorithms are suggested by various researchers. There is not a single algorithm proposed by researchers which handle all known constraint imposed by user and service provider. At the end of this research paper, the authors have suggested that workflow scheduling for cloud environments is an optimization problem.

Keywords Service-level agreement (SLA) · Workflow · Multi-objective

1 Introduction

Workflow is a sequence of a process that together accomplishes a task. The workflow is responsible to distribute the task at appropriate computing resources for effective and fast execution. A large number of workflow scheduling algorithms in the cloud environment have been proposed by the researcher with various considerations like resource constraint, homogeneous, and heterogeneous clustering, task dependencies, and utility-based computing. The workflow management system allows shared resources based on demand [1]. This is done by virtualizing the resources as per the required number of storage, CPU, and memory bandwidth. The VM can be leased or released as per requirement and charged for the unit billing period.

N. Kumar (✉) · S. K. Sharma
Deparment of Computer Science, Banasthali Vidyapith, Vanasthali, Rajasthan, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_14

2 Workflow Scheduling in Cloud Computing

Now, workflow scheduling in a cloud computing environment is a mechanism that involves in managing allocation to resources and preservation of execution of an interdependent task. Workflow implementation in a cloud environment requires two mechanisms: (i) resource allocation and (ii) resource provision. Resource provisioning is the selection and provision of computational resources used to execute tasks. This means that there are heuristics that can determine how many VMs are being rented, when execution starts and stop, their type of VM, etc. The second approach is about the mapping of tasks with appropriate resources. This process is called resource provisioning.

2.1 Cloud Workflow Management System

The responsibilities of a workflow management system are to supervise, build, and manage data and task transparently between distributed resources [2]. The workflow management system consists of various components to address tasks, data, and resource concerns for customer quality of service requirements. The reference architecture shown in Fig. 1 can be classified into three parts.

(A) **User interface—**The user interface allows the user to work with workflow creation, execution, submission, and workflow monitoring. All these are web services and distributed through a web portal or standalone application at the user end. Users are responsible for defining task attributes and linking them based on their data dependencies.

(B) **Workflow engine—**The core part of the workflow management system is the workflow engine. It is responsible for managing the workflow. The parser modules of the workflow engine are responsible for translating high-level workflow descriptions for tasks and data objects. The scheduling components are responsible for implementing user-defined scheduling policies and giving space to the workflow at various stages of implementation. Works with data dispatcher and task resource interface plugins are constantly monitors tasks in the workflow.

(C) **Plug-in**—Plug-in of WFMS is responsible for communication between various resource management systems for distribution of resources. Plugins are responsible for querying work and data attributes, transferring data from and to resources and monitoring the execution status of tasks and applications to run, and recover working status, and also to measure power consumption. three parts.

The bottom part of architecture includes cloud, grid, and cluster. Further resource managers are responsible for communication with market resources, scalable application managers, and various cloud services for resource management [3].

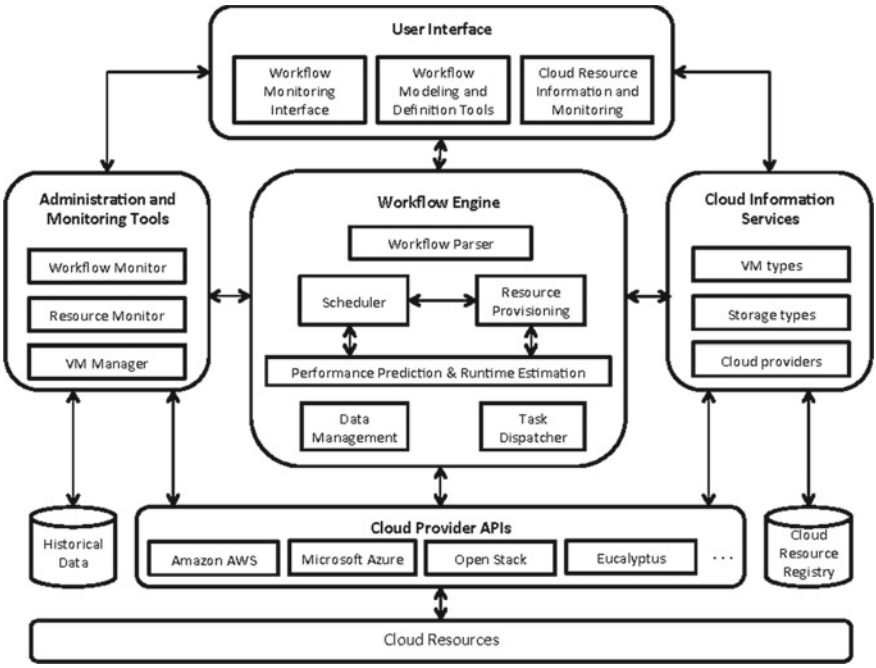


Fig. 1 Cloud workflow architecture [7]

2.2 Goals of the Workflow Scheduling Algorithm

In general, cost optimization is a common goal of workflow scheduling algorithms. Other performance metrics such as overall runtime, power consumption, safety, maximization of workload, VM utilization, awareness, security awareness, and reliability are also major concerns of workflow scheduling algorithms [4].

Resource provisioning—Allocating tasks to appropriate resources are an important challenge for workflow algorithms. The research work is shown by Michan et al. [5], and Villegas et al. [6] researched on the dependency between task and resources.

Performance variation and other sources of uncertainty—Cloud features like virtualization, resource sharing nature, and diversity of non-virtualized hardware in the cloud are the result of diversity in resource performance [8–12]. Due to uncertainty among the dependencies in a large workflow schedule cannot make an accurate decision about runtime and QoS requirements [13–15].

Utility-based pricing model—Some schedulers are interested in utility-based pricing model. For example, Amazon's EC2 Spot Instance feature is available for dynamic pricing. Spot instances vary over time based on market price and demand [16].

3 Related Works

This section has taken a detail surveys on various workflow scheduling algorithms:

A. **Deadline and energy-aware task scheduling (DEATS) in cloud—**In this paper, the authors proposed the task scheduling algorithm concern to energy-aware and deadline aware. The proposed algorithm is to reduce the makespan and reduce energy consumption in meeting the time limits. Final results in this research paper show that it is research performance over FCFS and EDF. The authors also suggest integrating other optimization methods to improve performance [17].

B. **Optimizing power consumption for expiration-constrained workflow in cloud data centers—**In this research paper, the authors take a survey about electricity consumption by datacenter. The authors proposed TMMS and WTM adopted scheduling algorithms to optimize power consumption. This proposed algorithm outperforms better in terms of the number of VM used, electricity consumed by the data center. In this proposed work, electricity cost, power consumption model, and time taken inside and outside the VM are not taken into account [18].

C. **Cost-effective fault-tolerant (CEFT) scheduling algorithms for real-time tasks in cloud systems—**In this paper, the authors proposed a novel CEFT algorithm to reduce implementation costs while trying to do more to meet the expiration requirements. The CEFT algorithm optimizes repetition and resource allocation in a repetitive manner with PSO. The rescheduling method is followed to increase the time limit guarantee ratio. Simulation results show that CEFT works better in terms of guarantee ratio and average cost, which ensures the feasibility and effectiveness of CEFT [19, 20].

D. **Cost- and energy-efficient algorithm for scientific workflow with the time limit in the clouds—**In this research paper, the authors proposed an algorithm that is cost-effective and energy-efficient. The authors proposed five steps to meet the deadline and reduce electricity consumption. All these steps effectively reduce the cost of workflow. Experimental results have shown that it performs better in terms of cost and power consumption. In future work, the author suggests taking into account the actual power cost for workflow scheduling [21, 22].

E. **An improved genetic algorithm to solve scheduled work problems with time limits on hybrid clouds—**In this research paper, the authors proposed a modified version of GA using novel crossover which identifies the world's best genes hidden in the population to inherit offspring. The authors experimentally shown that it performs better than normal GA [23].

F. **Multi-objective hybrid cloud resource scheduling method and scheduling method based on expiration and cost limits—**In this paper, the authors developed a mosaic algorithm on hybrid cloud computing by combining two single

objective optimization algorithms. These single objective optimization algorithms used improved ant colony algorithm by following entropy optimization. Proposed algorithm optimizing resource utilization subject to cost limits and time duration with the objective of resource utilization cost, QoS, task completion time, and service provider interest [24].

G. **Better budget-time-scheduling algorithms to work on constrained resources—**In this paper, the authors improve the scheduling algorithm for a variety of systems. The schedule is less focused and less complicated to meet budget and deadline limits. The algorithm is compared to three recent algorithms, such as the DBCS, BHEFT, and HBCS algorithms. Experimental result shows a high success rate when time and cost constraints are tight. In the future, author wants to minimize makespan according to the condition that the total cost of the workflow [25, 26].

H. **Multi-objective accelerated particle swarm optimization technique for scientific workflow in the IaaS cloud—**In this research paper, the authors proposed an algorithm that works in two steps. First, the algorithm follows a strategy that finds the most suitable VM which reduces the execution time. Second, the algorithm creates a proper schedule of tasks that can increase the output of the server. Through several performance analyses, the authors demonstrated the dominance of the proposed algorithm over sophisticated algorithms [27, 28].

I. **Particle swarm optimization-based resource provisioning and scheduling algorithm—**In this paper, the author proposed a particle swarm optimization-based resource provisioning cost minimization which considers resource diversity, VM performance, and elastic provisioning. Scheduling and resource provisioning are integrated to resolve PSO issues. The advantage of the global optimization technique algorithm is that it permits the creation of good quality schedules. The authors have pointed out that as unlimited resources are successfully extracted, the calculation overhead increases rapidly in the types of workflows and VMs provided by the provider [29].

J. **Budget-driven scheduling of scientific workflows in IaaS clouds with fine-grained billing periods—**In this paper; the authors present task-based budget delivery techniques to implement scientific workflows to reduce makespan. The proposed algorithm divides the workflow budget and guides the resource utilization of the divided budget of each task. The proposed algorithm implements the VM recharge policy to use idle time slots caused by a rigorous and fast billing period. Since reusable passive VMs are not available; it provides the fastest VM type on a budget when needed [30].

Fault-tolerant schedule using spot instances—Authors in this research paper proposed an algorithm which works on on-demand VM and spot VM. According to the proposed algorithm, on-demand VM will change according to the user-defined time limit. The author also proposed a bidding strategy for spot VM [31].

4 Basic Elements for Workflow Scheduling Algorithms

On the basis of a survey taken in Sect. 3, this section describes the various approaches to developing workflow algorithms in IaaS cloud environments.

4.1 Task-VM Mapping Dynamics

The classification of the scheduling algorithms generally proposed by Casavent and Kuhl [32], and categories as static and dynamic. In addition to these two cloud environments, it can combine static and dynamic environments to get the most out of it, i.e., hybrid.

Static—This type of scheduling algorithm produces advanced task-VM mapping and does not change during the run. This perspective is not allowed to be dynamic and therefore not subject to change. A small change in Task-VM mapping can lead to implementation failure and therefore affect QoS.

Dynamic—Dynamic scheduling algorithm determines task resource mapping during work. These runtime decisions are based on workflow implementation and the latest system status. Dynamic scheduling algorithm decisions are taken during the run when ready to execute. These regulations allow us to adapt to changes in the environment.

Hybrid—In cloud environments, both static node dynamic scheduling has its own significance. To take advantage of these, a new algorithm must be designed to find the difference between the compatibility and performance of the dynamic scheduling algorithm. To build a hybrid algorithm for categories into two parts: Runtime Purification and Sub-Workflow Static.

4.2 Resource Allocation Strategy

Kindly, resource allocation is categorized as static and dynamic resource allocation. In a static resource allocation, all resources are configured before implementation in the workflow. In dynamic resource allocation, all decisions are made during runtime.

Static VM pool—This strategy can be used through static resource allocation. When the VM pool is split, the resources are rented and become active during workflow execution. Its implementation resources will be released when the workflow is completed. The advantages of such an algorithm are that the algorithm mainly focuses on the task VM allocation after making resource allocation decisions.

Elastic VM pool—The workflow algorithm in this category follows static and dynamic resource allocation strategies. This mechanism permits the algorithm to change the type of VM and the number of VM during runtime. Due to flexibility on

VM, the workflow algorithm in these categories may take a decision on cost, number of VM used, and output.

4.3 Optimization Strategy Classification

The most important part of the optimization strategy algorithm is the mathematical model of problem and solution space. The optimization technique is classified into two categories, i.e. optimal and suboptimal, in which suboptimal is classified into three categories.

Heuristic—Heuristic is the property of rules for the solution of a particular problem. These types of rules are specific and designed for a specific problem to find solution within an acceptable time. The main purpose of heuristic-based scheduling algorithms is to find satisfactory solutions in a timely manner.

Meta-heuristic—Meta-heuristic algorithms are commonly used for optimal solutions to a problem. In general meta-heuristic algorithms are more computationally intensive and take longer to solve workflows using meta-heuristic algorithms, which is a challenging problem due to the virtualization of an unlimited number of resources and dependencies [33].

Hybrid-Meta heuristic algorithms are used to solve a set of tasks in the workflow and heuristic algorithms are used to solve the task in a small simplified workflow. In such an approach, we get better optimization than heuristic and meta-heuristic algorithms.

4.4 Resource Model Classification

This section presents a resource model based on the assumption of the classification algorithm. Classification is done with services provided by providers like VM pricing model and data transfer cost. Resource model is classified as follows-

4.4.1 VM Leasing Model

This model is maintained by assuming that the classification service providers provide restricted or unrestricted leased VMs for a given customer.

Restricted—Limited VM leasing model service provider has a high limit on the number of virtual machines that a customer can lease. Algorithms that use this model work with a limited number of VM resources.

Unrestricted—In these VMs leasing model algorithm assumes that there is no limit on the number of VM to lease to a customer.

4.4.2 Deployment Model

The workflow algorithm under the deployment model is classified on the basis of the datacenter and resources leased by the service provider.

Single service provider—The algorithm in this approach considers a single cloud service provider for IaaS. There is no need to consider inside and outside data transfer cost.

Multiple service providers—In this category, algorithms consider different cloud service providers to schedule tasks. The different service providers may have different service level agreements and different resource provisioning price offerings. Now it's the scheduler's choice to choose suitable offers.

Single data center—These taxonomic algorithms choose to provide VMs in a single data center. This approach reduces latency and faster transfer time and costs because most providers do not charge for transfers made in the data center.

Multiple Data Centers—In this classification of workflow algorithms, service provider offers VM available in different datacenter which is geographically distributed. It is most suitable for the application whose inputs are from various geographical locations. It reduces the latency and data transfer cost.

4.4.3 Data Sharing Model

Data sharing in the workflow process affects the performance of the scheduling algorithm because it affects scaling, cost, makespan, etc. There are two common approaches are used to improve the performance of the scheduling algorithm-(i) peer-to-peer and (ii) shared storage.

P2P—In this process, data are transferred directly from the VM running child process to the VM running parent process. This mechanism leads to synchronize the communication. This process leads to higher cost because lease time increased but increase the scalability.

Shared storage—This process reduces the VM leading cost and improves the resource utilization by storing the output in globally shared storage systems like AWS simple storage system (S3) or AWS elastic file system (EFS). There are many advantages of such a system. First, data integrity is maintained and recovered data in case of failure. Second, it permits asynchronous computing because data can be released to a VM running an actual function while continuing in the storage system.

4.4.4 Data Transfer Cost Awareness

The data transfer pricing plan is different in a different scenario. Trans bound and inbound data are usually independent, and transferring data from a cloud provider is charged and expensive. In many service providers, algorithms that schedule workflows have to relate to this cost because data transfer may belong to different service provider.

4.4.5 Storage Cost Awareness

In this strategy, an algorithm decides the cost of storage based on the type of operation stored, as well as the amount of data stored. The algorithm for storage, therefore, agrees with the cost and generally estimates the data size and data based on a fixed price per unit.

4.4.6 VM Price Model

In this category, four different price models are proposed as static, dynamic, subscription based, and time unit.

4.4.7 VMcore Calculation

This classification refers to algorithms that schedule multiple tasks simultaneously about multiple VMs.

Single core—In this mechanism, algorithm assumes that the VM has only one core. It simplifies the scheduling process because it resolves resource conflicts.

Multi-core—In the multicore facility, IaaS provides multiple cores. Workflow scheduling takes advantage of this task to execute multiple tasks simultaneously on a single VM, possibly avoiding the time, expense, and intermediate data transfers.

5 Analysis

Workflow scheduling in cloud computing is a very challenging task due to various constraints and multi-objective nature. So there is a need for an optimization technique that works on multi-objective nature and constraints imposed by the user and cloud environment.

Let T_n is the task with n objective then we need to optimize the objective function-$T_n(X)$, where X is a variable and $X = 1, 2, 3, \ldots, N$, constraints to.

$F_m(X)$ where $m = 1, 2, \ldots, M$

F is denoting constraints function, m is the number of constraints, and X is the domain of feasible solution. Suppose problem A is decision problem, and problem B is optimization problem; then it is possible that A α B. Hence, we can say that such type of problem can be solved only by multi-objective optimization technique. A diagram given in Fig. 2 completely described that how a multi-objective workflow scheduling algorithm can be incorporated.

On the basis of various taxonomy in Sect. 4 and survey in Sect. 3, we have found that such scenario can be handled by a multi-objective optimization technique

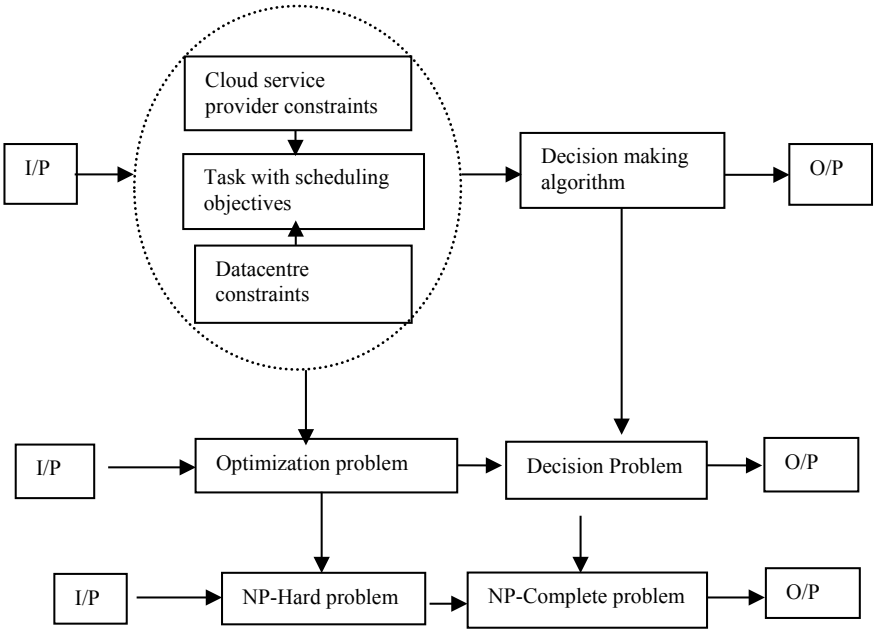


Fig. 2 Workflow is an NP-problem

like genetic algorithm, ant colony optimization algorithm, particle swarm optimization algorithm, etc. Genetic algorithm and ant colony optimization technique are mathematically well proof than PSO but PSO works with very few parameters [34].

6 Conclusion

Workflow scheduling in cloud computing is a very challenging task due to various constraints imposed during service-level agreements (SLAs) by users and the cloud environment. These various constraints make workflow scheduling as a multi-objective nature. A multi-objective problem can be solved only with optimization techniques with a decision problem. Since it is an NP-hard problem so there is a need for a meta-heuristic optimization technique that works suitably in such type of environment. As a future scope, multi-objective optimization techniques which address a maximum number of constraints should be implemented and compared.

References

1. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Generat Comput Syst 25(6):599–616. ISSN 0167-739X
2. Yu J, Buyya R (2004) A novel architecture for realizing grid workflow using tuple spaces. In: Proceedings of the fifth IEEE/ACM international workshop on grid computing. IEEE, Pittsburgh, USA, pp 119–128
3. Vöckler J, Juve G, Deelman E, Rynge M, Berriman BG (2011) Experiences using cloud computing for a scientific workflow application. In: Proceedings of 2nd workshop on scientific cloud computing (ScienceCloud 2011)
4. Zeng L, Veeravalli B, Li X (2015) Saba: a security-aware and budget-awareworkflow scheduling strategy in clouds. J Parallel Distrib Comput 75:141–151
5. Michon E, Gossa J, Genaud S et al (2012) Free elasticity and free CPU power for scientific workloads on IaaS clouds. In: Proceedings of the eighteen IEEE international conference on parallel and distributed systems (ICPADS). IEEE, Singapore, pp 85–92
6. Villegas D, Antoniou A, Sadjadi SM, Iosup A (2012) An analysis of provisioning and allocation policies for infrastructure-as-a-service clouds. In: 12th IEEE/ACM international symposium on cluster, cloud and grid computing, pp 612–619, 13–16 May 2012. ISBN: 978-1-4673-1395-7
7. Rodríguez M, Buyya R (2016) A taxonomy and survey on scheduling algorithms for scientific workflows in IaaS cloud computing environments: workflow scheduling algorithms for clouds. Concurrency Comput Pract Exp 29(8). ISSN: 1532-0626
8. Schad J, Dittrich J, Quiané-Ruiz JA (2010) Runtime measurements in the cloud: observing, analyzing, and reducing variance. In: 36th International conference on very large data bases, Singapore. Proc VLDB Endowment 3(1–2):460–471, 13–17 Sept 2010
9. Ostermann S, Iosup A, Yigitbasi N, Prodan R, Fahringer T, Epema D (2010) A performance analysis of EC2 cloud computing services for scientific computing. In: Cloud Computing. Springer, Munich, Germany, pp 115–131
10. Gupta A, Milojicic D (2011) Evaluation of HPC applications on cloud. In: Open cirrus summit (OCS), 2011 Sixth, Atlanta, Georgia, pp 22–26
11. Iosup A, Ostermann S, Yigitbasi MN, Prodan R, Fahringer T, Epema D (2011) Performance analysis of cloud computing services for many-tasks scientific computing. IEEE Trans Parallel Distrib Syst 22(6):931–945
12. Jackson KR et al (2010) Performance analysis of high performance computing applications on the amazon web services cloud. In: 2010 IEEE second international conference on cloud computing technology and science, pp 159–168. ISBN: 978-1-4244-9405-7
13. Nabrzyski J, Schopf JM, Weglarz J (2012) Grid resource management: state of the art and future trends, vol 64. Springer Science & BusinessMedia, Berlin, Germany
14. Valentin C, Ciprian D, Corina S, Florin P, Alexandru C (2010) Large-scale distributed computing and applications: models and trends
15. Berman F, Fox G, Hey Anthony JG (2003) Grid computing: making the global infrastructure a reality, vol 2. : John Wiley and sons, Hoboken, New Jersey, United States
16. Amazon EC2 Spot Instances. Available on: https://aws.amazon.com/ec2/spot/?cards.sort-by=item.additionalFields.startDateTime&cards.sort-order=asc/. Accessed on 15 Oct 2020
17. Hicham BE, Said BE, Touhafi A, Ezzati A (2018) Deadline and energy aware task scheduling in cloud computing. In: 4th International conference on cloud computing technologies and applications (Cloudtech), 26–28 Nov 2018. ISBN: 978-1–7281-1637-2
18. Al-Dulaimy A, Itani W, Zekri A et al (2016) Power management in virtualized data centers: state of the art. J Cloud Comput 5:6
19. Guo P, Liu M, Wu J, Xue Z, He X (2018) Energy-efficient fault-tolerant scheduling algorithm for real-time tasks in cloud-based 5G networks. IEEE Access, pp 1–1. https://doi.org/10.1109/ACCESS.2018.2871821
20. Wu L, Ding R, Jia Z, Li X (2020) Cost-effective resource provisioning for real-time workflow in cloud. Complexity 2020(Article ID 1467274):15

21. Peng K, Zhao B, Xue S, Huang Q (2020) Energy- and resource-aware computation offloading for complex tasks in edge environment. Complexity 2020(Article ID 9548262):14
22. Zhu M, Wu Q, Zhao Y (2012) A cost-effective scheduling algorithm for scientific workflows in cloud. In: Proceedings of 31st IEEE international performance computing and communications conference
23. Yassa S, Sublime J, Chelouah R, Kadima H, Jo GS, Granado B (2013) A genetic algorithm for multi-objective optimisation in workflow scheduling with hard constraints. Int J Metaheuristics 2(4):415–433. https://doi.org/10.1504/IJMHEUR.2013.058475
24. Zuo L, Shu L, Dong S, Chen Y, Yan L (2017) A multi-objective hybrid cloud resource scheduling method based on deadline and cost constraints. IEEE Access 5:22067–22080
25. Sun T, Xiao C, Xu X, Tian G (2017) An improved budget-deadline constrained workflow scheduling algorithm on heterogeneous resources. In: 2017 IEEE 4th international conference on cyber security and cloud computing (CSCloud), New York, NY, pp 40–45. https://doi.org/10.1109/CSCloud.2017.8
26. Kaur N, Singh S (2016) A budget-constrained time and reliability optimization BAT algorithm for scheduling workflow applications in clouds. Procedia Comput Sci 98:199–204. ISSN 1877-0509
27. Konjaang JK, Xu L (2021) Multi-objective workflow optimization strategy (MOWOS) for cloud computing. J Cloud Comput 10:11
28. Adhikari M, Amgoth T (2018) Multi-objective accelerated particle swarm optimization technique for scientific workflows in IaaS cloud. In: 2018 International conference on advances in computing, communications and informatics (ICACCI), Bangalore, India, pp 1448–1454
29. Gill SS, Buyya R, Chana I et al (2018) BULLET: particle swarm optimization based scheduling technique for provisioned cloud resources. J Netw Syst Manage 26:361–400
30. Rodriguez MA, Buyya R (2017) Budget-driven scheduling of scientific workflows in IaaS clouds with fine-grained billing periods. ACM Trans Auton Adapt Syst 12(2, Article 5):22
31. Poola D, Ramamohana Rao K, Buyya R (2014) Fault-tolerant workflow scheduling using spot instances on clouds. Procedia Comput Sci 29:523–533
32. Casavant TL, Kuhl JG (1998) Taxonomy of scheduling in general-purpose distributed computing systems. IEEE Trans Softw Eng 14(2):141–154
33. Talbi EG (2009) Metaheuristics: from design to implementation, vol 74. Wiley & Sons, Hoboken, New Jersey, United States
34. Kumar N, Kumar Sharma S (2018) Inertia weight controlled PSO for task scheduling in cloud computing. In: 2018 International conference on computing, power and communication technologies (GUCON), pp 155–160. https://doi.org/10.1109/GUCON.2018.867499.

Personalized Food Recommendation—State of Art and Review

Anjali Jain and Alka Singhal

Abstract In the present scenario, people are becoming more health-conscious. There are various factors like age, gender, and genetic history on which diet matters. The internet is flooded with a tremendous amount of information and data related to healthcare. An enormous amount of data related to medical and diet on websites makes it difficult for the user to find the correct and valuable information, and it is a high demand of society. This problem of finding the proper diet plan needs a recommender system in healthcare is required. The recommender system helps end-users and medical professionals to see the results in a better manner. As the number of available food items increases exponentially, it is impossible to take all of them. In such a scenario, a recommendation system is required to consider user preferences, maximize the number of healthy compounds, and minimize the unhealthy ones in food. The proposed paper has collectively studied, reviewed, categorized, and summarized different techniques and algorithms to suggest the diet plan and identify the issues and gaps in existing algorithms and food/diet recommendation strategies. The review includes basic categories like a recommendation based on user preferences for item features, based on user behavior, hybrid filtering (H.F.), and knowledge-based filtering (KBF). Additional state-of-art for existing food recommendations is analyzed in this paper.

Keywords Recommendation system · Healthcare recommendation · Food recommendation · Diet recommendation

A. Jain (✉) · A. Singhal
Jaypee Institute Of Information Technology, Noida, India

A. Singhal
e-mail: alka.singhal@mail.jiit.ac.in

A. Jain
Kiet Group of Institutions, Ghaziabad, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_15

1 Introduction

Nowadays, health is one of the major concerns in our daily life. Many factors influence an individual's health, such as physical exercise, sleep, nutrition, heredity and pollution. Being nutritious is one of the most significant modifiable factors in our lives; small changes can have a considerable impact. Healthy eating has always been an essential aspect of life as "health is wealth." A healthy mind lives in a healthy body. People used to take medicine and diet recommendations from various sources. With the changes occurred in the environment, every other person is looking for healthy nutritious diet as the life has become confined to specific routine. In present scenario, health related searches is one of the most searched topics on the internet. As the outbreak of pandemic, deadly coronavirus has ruined our daily life style and people are falling sick frequently. Virus has spread over a large geographic area and every other person is looking for medicines and remedies to boost immunity and to be safe from this deadly virus. A lot of research has been done in this area using different techniques.

The techniques on which research has been done includes similar user behavior, user preferences for item features, Hybrid filtering (H.F.), knowledge-based filtering (KBF) and various other techniques including innovative algorithm like ant colony algorithm, particle swarm optimization, simulated annealing, classification, clustering, decision tree, logistic regression, natural language processing, inductive logic programming, ontologies, sparse canonical correlation, support vector machines, semantic technologies, multi-criteria decision making, graph-based recommendations, context-aware recommendation, sensor smart plate, artificial intelligence and matrix factorization. These are the basic recommendation techniques applied in the various healthcare domain. The authors have discussed pros and cons of these filtering techniques (Table 1).

2 Literature Survey

Eating right and healthy diet is serious concern. Over nutrition and under nutrition diet should be monitored and at major consideration. The foods one should take daily should be balanced in terms of nutritional needs. Recommendation system has become primary source for searching something in electronic commerce and information access [1] reviewed the pros and cons of traditional techniques that are used in e-commerce recommendation and expressed the main challenges faced and relevant solutions for them. The author used rich data set and showed the comparison and evaluation of each of the proposed mechanism with significant advantages and disadvantages to point out the essential issues to improve their accuracy and functionality for further studies. Different types of recommendation systems are available like E-commerce, healthcare, banking, tourism, e-learning, etc. Generally, people are more biased toward the attractive items, to solve the problem [2] introduced a novel hybrid

Table 1 Pros and cons of basic filtering techniques

Technique	Pros	Cons	Example
Similar user behavior	Able to recommend new and complex items. No domain knowledge needed	Cold start, sparsity and scalability issues	Amazon, Google
User preferences for item features	No need of user data. Just item information is required. Able to recommend new and unpopular item	Limited recommendation scope. Problem for new user	Movies, web news etc
Hybrid recommender system	Effective and accurate results over C.F. and C.B. Resolved cold start and sparsity issues	Suggestion ability is better	Netflix
Knowledge-based recommender system	Recommendation do not depend on user ratings, there is no ramp-up problem	Noisy environment, validation, transplantation and general problems	Taxi routing apps

PMF model called DUPIA which can simultaneously differentiate user's preference and item's attractiveness. The two real-world datasets for recommendation systems is used, i.e., MovieLens 1 and Amazon 2. The original MovieLens dataset consists of 72,000 users, 10,681 movies and 10,000,054 rating scores ranging from 1 to 5. The author has made predictions using both users' auxiliary information and items' textual comments simultaneously. But the descriptive information which attracts persons verified partially. Kluver et al. [3] and Mansoury et al. [4] also introduced algorithm to solve the cold start problem and the problem faced for new item and new user recommendation. As the huge amount of data related to the medical information and if anyone need it for wellbeing and improvement faces lot of struggles to find useful information.

To overcome the above-mentioned problem [5] has given a systematic survey for scenarios and approaches in healthcare recommendation like drug recommendation, food recommendation, health status prediction recommendation, health care service recommendation and healthcare professional recommendation. The work included the food recommendation as well. Sak et al. [6] used artificial neural network for recommendation of dietary foods. In this paper, we are providing an overview and comparative study of existing techniques and approaches of recommendation system in Healthcare domain. In health care recommendation, there are food recommendation, drug recommendation, healthcare service recommendation and healthcare professionals. Here, authors are talking about food recommendation using different techniques and algorithm for the benefits of end user. Issues and challenges in food recommendation and its future scope.

3 Various Approaches Followed by Healthcare Recommendation

A key procedure to achieve balanced nutritional intake is dieting. Generally, people restrict their diet in account of maintaining the nutrients required for the body or for medical reasons. But these type of diets and restriction against the desire to eat some specific food for that someone is craving for may destroy mental, physical and social health. So, one must eat food that they desire along with recipe with of nutritious ingredients. For the purpose of healthy recipe and diet recommendation system works well and serves the purpose. Below mentioned are filtering techniques which are used to recommend food/diet and recipe for healthy and nutritious diet.

3.1 Similar User Behavior

Similar user behavior uses similarities between items and users together; it recommends items of user based on the interest and tastes of similar user. If a person has similar health concern or disease, they would have recommended similar type of food with same ingredients [5]. Similar user behavior can be combined with other techniques to give accurate results and perform better [7] highlighted the similar user behavior and user preferences for item features for the user's recommendations where system takes the user item descriptors with user's preferences of the items to recommend the results using association rules and inverse filtering methods. Another author [8] used K-means clustering along with collaborative filtering is used for diet recommendation. The higher recommended recipes included in the cluster which meets the user's nutritional balance. The method has proved to be effective with accuracy more than 70% (Fig. 1).

3.2 User Preferences for Item Features

In user preference for item features, item features are used to recommend other similar items similar to the likes of other user observing their actions and feedback of previous activities. In this if a product or service is similar to the previous product used by a particular person then it is recommended for the person [5]. To predict recipe rating using derived features of ingredient network [9], proposed recipe recommendation using ingredient network based on content-based filtering as each ingredient present in food affects health of a person. As the hypertension is one of the major non-communicable diseases among the youths Sookrah et al. proposed a recommendation engine that is DASH, i.e., dietary approaches to stop hypertension based on content-based filtering and machine learning for hypertensive patients. Also, the app was developed and this system also helped to monitor and control B.P. level [10] (Fig. 2).

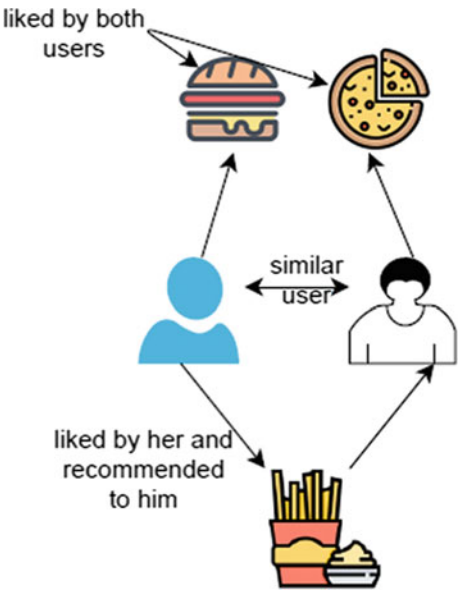


Fig. 1 Similar user behavior

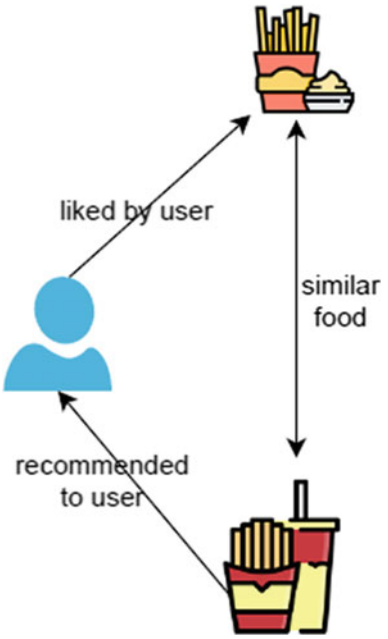


Fig. 2 User preferences for item features

3.3 Hybrid Filtering

Generally, when items added to the catalog have either none or very little interactions this leads to cold start and data sparsity issues. Hybrid filtering is the solution to overcome such type of problems, and it is the combination of similar user behavior and user preference for item features. Both of the techniques are combined to overcome the limitations of each other. Basically, it is used to fix the cold start problem, i.e., when a new item is added to the system and has no rating. Chang Kim et al. [11] proposed knowledge-based hybrid decision model which considers both physical and mental health to solve the cold start and sparsity problem. In this paper author used huge amount of data set which consists of 80% of learning and 20% of collected data as test data and evaluation was done using regression and correlation. The proposed method was the food recommendation model to help users that made the dietary nutrition-related decisions on a health platform. Li et al. [12] proposed a nutritional combo model by analyzing the user demands. As a lot of people is moving toward online ordering and compromise food taste and nutrition. The model is based on the functional of similar user behavior and user preference for item features and overcomes cold start problem and gives high performance.

3.4 Knowledge-Based Filtering

This is applied when the quantity of item rating is limited. Knowledge-based recommender system is based on the knowledge of specific items or we can say queries made by user. Knowledge-based recommender system can be used where it is not possible to use the recommender system. Nature plays an important role in daily life and lots of creatures has some technique and pattern in their life style that can be boon to human beings when observed and followed. Rehman et al. [13] proposed cloud-based food recommendation where ant colony optimization a nature inspired technique is used for generating optimal food list and food recommendation after viewing the user's pathological reports and attained a accuracy by increased number of ants. Another author [14] has designed sensor based smart plate for the children aged 3–6 years to measure the weight of food before and after the meals. The proposed methods are designed to observe the nutritional deficiencies among children and growth index. On the basis of results and analysis of sufficient nutrition required the nutritious menus were recommended.

4 Methodology Used

The methodology used for the literature review process is contains a systematic process suggested by (Kitchenham and Brereton 2013). The review process brings out the clear picture of food recommendation.

A. **Formation of research Question**:

The aim of the research question formation is to collect all the credible information of work done by food recommendation which helps to identify and explore techniques, issues and critical points of selected food recommendation.

Research questions are:

- Food recommendation system.
- Techniques used in food recommendation.
- Diet recommendation + collaborative filtering.
- Approaches to food recommendation.
- Issues in food recommendation.
- How different authors conducted study.

B. **Process for paper selection**:

There are two major parts: First is the used Google scholar to identify the selected papers, secondly the relevant keywords with required data combination were used are: Recommender system, Food recommendation, Diet recommendation, Food recommendation + collaborative filtering, Filtering techniques + food recommendation.

- Applied "allintitle" in Google scholar.
- Search included papers from 2011 to 2020.
- Title of most of the papers directly related to the topic while few papers are not directly related. Total number of detected papers is approx 353. Then, we applied our filtering method according to our problem and review 15 papers.

C. **Review of selected papers**:

In the review, it's been observed that mostly papers are solving the information overload problem related to the healthcare. As we are reviewing the basic models and approaches of food recommendation in healthcare domain. Nag et al. [15] mainly introduced novel techniques and model using filtering techniques and solve the problem of cold start, sparsity and latency.

5 Strength's, Weakness and Opportunities Analysis

Paper	Strength	Weakness	Scope
Ueta et al. [16]	Huge dataset of 80,000 dishes and trial was done on twenty persons	Authors have not considered existing system and hospitals	The system would be more accurate and efficient if existing system and hospitals diet plan can be considered
Yuen et al. [9]	Shown how different data can be obtained and used for construct network	Does not shown cooking methods for recipes	Cooking recipes along with ingredients can be included to generate region specific and diet specific results
Osadchiy [7]	Compared the three algorithms for 20,000 randomly sampled meals	Compared the result with two other systems but didn't justify the improvement in accuracy and efficiency	Accuracy and efficiency can be taken for improvement
Min et al. [17]	Incorporate the various domains in food recommendation, multidisciplinary research that combines nutrition, food science, psychology, biology, anthropology and sociology	Accuracy and efficiency of the proposed system is not highlighted	Accuracy and efficiency can be included
Chang Kim [11]	Model is well implemented on users and the feedback of users was also mentioned in the paper	Satisfaction level of the end users is not so good as the recommendation system claimed less than 80%	Some parameters of the model can be modified to increase customer satisfaction
Yuan [8]	Accuracy rate is almost above 0.70, and the average accuracy rate is 0.77, which indicates that the collaborative filtering algorithm based on user preference is stable	Algorithm does not focus on the nutritional content of the recommended food. It should also be considered	Nutritional content should be included

(continued)

(continued)

Paper	Strength	Weakness	Scope
Jiang et al. [18]	Given the detailed information of the optimum intake of different nutrition	Ranking based experiment is performed against a three-category classification so multi classification is not tried so far	Deep learning models such as graph neural network for the recommendation can also be included
Bodnar Lisa [19]	Ensemble model in machine learning is a challenging task and higher accuracy proves the novel work done by authors	1. Work is limited to data taken from an FFQ, which are affected by systematic measurement error to a greater extent than other self-report methods such as 24-h recalls	2. Future research should explore other cut-offs as well as the use of continuous measures of dietary intake
Tran [5]	Discussed multiple techniques used in different health recommendation	Author has not covered upcoming challenges in health recommendation	Cold start and data sparsity problem can be tackled using some other algorithms
Sookrah [10]	Recommended dishes also helps normal people to prevent from health complications	No provision to track the progress of user	User eating pattern can also be identified using DASH
Zhang [2]	Explained all the strategy in single paper while using collaborative filtering	Does considered the performance metrics	Experiment can be performed on large data set. As it is difficult to work with large data set
Namgung [14]	The plate is bale to guess how much nutrition child take and how much is required	Experimental data is collected through survey. It was not possible to build plate for home	The data set was collected through survey. In future real data sets can be collected and used for the results
Rehman et al. [13]	Used heuristic information to maximize food diversity	Doesn't considered diet for particular meal of the day	Group food recommendation can be worked upon in future

(continued)

(continued)

Paper	Strength	Weakness	Scope
Chena et al. [20]	The framework considers both ingredients user-defined and relevant for particular diet with quantity. Experimental results are able to prove better healthiness of recipes	Precompiled nutritional recipes are not compared and performance is not measured	Work can be done on performance improvement along with healthy nutrient recipe
Li et al. [12]	High feasibility combos are designed. Solved the problem of cold start and achieved higher performance	Doesn't solved the data sparsity problem	Combos can be recommended according to age, gender and disease
Lores et al. [21]	Doesn't rely on specific user and item. Better performance and solved the persistent problem of cold start, latency and wrong recommendation	In proposed approach time is not considered	Proposed algorithm can be applied for social tagging environment, and managing tags in folksonomy
Sak et al. [6]	Review includes all the application areas of AI	Such type of research creates diverse spectrum of problems. Range of areas doesn't consider by the authors	Integration on use of medical robots
Turkmenoglu et al. [22]	Worked on the real-world data set of Istanbul technical university to make results applicable and realistic	Considered limited food groups. Division of daily like breakfast, lunch, snacks and dinner can be considered can for diversity meal plan	In future food groups can be added by maximizing food diversity

6 Conclusion

Food recommendation is one of the emerging fields on the internet as it makes easy for the professional, health care worker and individual to follow and adapt the healthy and nutritious food. The main aim of the paper is to analyze and review the various issues in food or diet recommendation that have not been presented carefully. Provided a collaborative study by summarizing work done by different authors and provides a novel overview table for strength and weakness of each identified gap. Similar user behavior is more efficient and accurate for finding the results. The

reviewed papers include nutritional factors in recommendation process as well as suggesting better personalized recommendation using various techniques. Although many authors have tried to solve the problem of cold start, sparsity and latency by applying different algorithm and techniques but there still exist some challenges which needs to fixed in food recommendation.

References

1. Alamdari PM, Navimipour N, Hosseinzadeh M, Safaei A, Darwesh AA (2020) A systematic study on the recommender systems in the E-commerce. IEEE, pp 115694–115716
2. Zhang X, Liu H, Chen X, Zhong J, Wang D (2016) A novel hybrid deep recommendation system to differentiate user's preference and item's attractiveness. Inform Sci 306–316
3. Kluver D, Ekstrand M, Konstan J (2018) Rating-based collaborative filtering: algorithms and evaluation. Social Inform Access 344–390
4. Mansoury M, Shajari M (2016) Improving recommender system's performance on cold-start users and controversial items by a new similarity model. Int J Web Inform Syst 1744–0084
5. Tran T, Felfernig A, Trattner C, Holzinger A (2020) Recommender systems in the healthcare domain: state-of-the-art and research issues. J Intell Inform Syst
6. Sak J, Suchodolska M (2021) Artificial intelligence in nutrients science research: a review. Nutrients 13:322
7. Osadchiy T, Poliakov I, Olivier P, Rowland M, Foster E (2018) Recommender system based on pairwise association rules. Exp Syst Appl
8. Yuan Z, Luo F (2019) Personalized diet recommendation based on K-means and collaborative filtering algorithm. J Phys
9. Yuen C, Ru Y, Lin LA (2012) Adamic, Recipe recommendation using ingredient networks. In: 4th annual ACM web science conference, pp 298–307
10. Sookrah R, Devee Dhowtal J, Devi Nagowah S (2019) A DASH diet recommendation system for hypertensive patients using machine learning, pp 1–6. https://doi.org/10.1109/ICoICT.2019.8835323
11. Chang Kim J, Chung K (2019) Knowledge-based hybrid decision model using neural network for nutrition management. Inform Technol Manage
12. Li C, Yang C (2016) The research and design of recommendation system for nutritional combo. In: International conference on computer and communications (ICCC), pp 830–837
13. Rehman F, Khalid O, Haq N, Khan A, Bilal K, Madani S (2017) Diet-right: a smart food recommendation system. KSII Trans Internet Inform Syst 11(6)
14. Namgung K, Hwan Kim T, Hong Y (2019) Menu recommendation system using smart plates for well-balanced diet habits of young children. Wirel Commun Mob Comput
15. Nag N, Pandey V, Jain R (2017) Live personalized nutrition recommendation engine. In: Proceedings of the 2nd international workshop on multimedia for personal health and health care, pp 61–68
16. Ueta T, Iwakami M, Tto (2011) A recipe recommendation system based on automatic nutrition information extraction, pp 79–90
17. Min W, Jiang S, Jain R (2019) Food recommendation: framework, existing. IEEE Trans Multimed 2659–2671
18. Jiang KLZ, Wang H, Li X (2020) Healthy diet recommendation via food-nutrition-recipe graph mining, pp 1–7
19. Bodnar Lisa L, Cartus A, Kirkpatrick S, Himes K, Kennedy E, Simhan H, Grobman W (2020) Machine learning as a strategy to account for dietary synergy: an illustration based on dietary intake and adverse pregnancy outcomes. Am J Clin Nutrit 1235–1243

20. Chena M, Jiaa X, Gorbonosb E, Hongb CT, Yuc X, Liub Y (2019) Eating healthier: exploring nutrition information for healthier recipe recommendation. Inform Process Manage 102051
21. Lores ML, Farnandez YB, Arias JP, Solla AG (2012) Property-based collaborative filtering for health-aware recommender systems. Exp Syst Appl 39:7451–7457
22. Turkmenoglu C, Uyar A, Kiraz B (2021) Recommending healthy meal plans by optimizing nature-inspired many-objective diet problem. Health Inform J 1–10

Breast Cancer Detection in the IoT Cloud-based Healthcare Environment Using Fuzzy Cluster Segmentation and SVM Classifier

Umesh Kumar Lilhore, Sarita Simaiya, Himanshu Pandey, Vinay Gautam, Atul Garg, and Pinaki Ghosh

Abstract Early-stage detection of breast cancer disease and its accurate diagnosis have been always challenging for the healthcare professional. An IoT healthcare system can play a vital role in this field. The existing diagnosis technique does not efficiently identify breast cancer in the beginning phases, and many of the women struggled from such a deadly illness. In this research paper, we introduce IoT cloud-based predictive analytics mainly based on fuzzy cluster-focused augmentation and optimal SVM classification for forecasting breast cancer infection via regular inspection and enhancing the health services by giving healthcare guidelines. In the proposed model, the fuzzy clustering algorithm is being used for efficient image segmentation that mainly focused on transition region filtration. Besides that, fuzzy C-means clustering and optimal SVM are also applied to characterize the transition period region attributes and feature extraction. The experimental phase is divided into three parts: parameter optimization testing, feature selection, as well as optimal SVM. The experimental phase results reveal that the proposed enhanced SVM method performed outstandingly in terms of feature selection, precision, TPR, FPR, F1-score over existing machine learning methods, i.e., logistic regression, decision tree, and SVM method.

U. K. Lilhore · S. Simaiya (✉) · A. Garg · P. Ghosh
Institute of Engineering and Technology, Chitkara University, Punjab, India
e-mail: sarita.simaiya@chitkara.edu.in

U. K. Lilhore
e-mail: umesh.lilhore@chitkara.edu.in

A. Garg
e-mail: atul.garg@chitkara.edu.in

P. Ghosh
e-mail: pinaki.ghosh@chitkara.edu.in

H. Pandey
Faculty of Engineering & Technology, University of Lucknow, Lucknow, U.P., India

V. Gautam
School of Computing, DIT, Dehradun, India
e-mail: vinay.gautam@dituniversity.edu

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_16

Keywords SVM · Fuzzy C-means · IoT · Machine learning · Breast cancer · Cloud computing

1 Introduction

In a woman, breast cancer is a form of tumor cells that arises in the breast tissues. This is the most widespread type of disease spotted in women and girls as well as this is among the major causes of demise in females. There are numerous artificial intelligence-based methods, i.e., machine learning, artificial neural networks which are used for the forecasting of cancer. Discovering the most relevant and efficient method for both the forecasting as well as diagnosis of cancer is one of the ambitious goals of all medical researchers. Breast cancer defines one of the other cancers which creates a high amount of fatalities annually. This is the most frequent source of all malignant tumors as well as the primary cause of female deaths annually [1].

Breast disease is the most frequent cancer experienced by women all across the world and accounts for around 31–37% of all forms of cancer in women in metropolitan areas [2]. This is extremely alerting to monitor that age range by which an individual grows cancer has moved from 45–55 years to 20–30 years; as well as forms of cancer in the youths are eager to be more confrontational. Machine learning and statistical techniques are an efficient way of categorizing statistics. Especially in the healthcare profession, in which those measures are commonly included in treatment and diagnosis to take action. Regrettably, sometimes it is just uncommon to reveal cancer signs, so initial detection is complicated. This is essential to utilize mammograms as well as self-breast examinations to identify such initial abnormalities well before cancer can get developed. Figure 1 is showing the breast cancer in a woman's body.

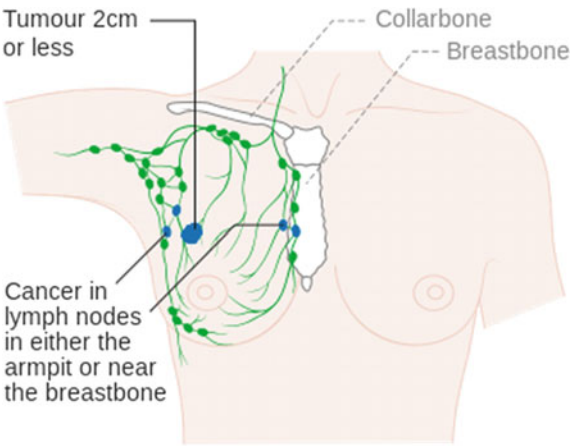


Fig. 1 Breast cancer in body [2]

The IoT has considerably improved healthcare. This innovation facilitates cancer patients and also health professionals for enabling a wide assortment of infections accurately as well as diagnosis and treatment of such infections according to the outcome measures. An IoT technology can be integrated with a machine learning-based system. A significant supervised training method is a support vector machine (SVM) [3]. The progressing sections of the article are structured as follows. Section 2 covers related work in the field of breast cancer detection, and in Sect. 3, it covers research material and methods. Section 5 mainly covers implementation and result from analysis and later covers the result comparison. Section 6 mainly covers conclusions and future improvement in the proposed work.

2 Related Works

With the advancement of clinical research, innumerable methodologies have been introduced for breast cancer diagnosis. The studies correlated with this area are described concisely as follows.

Gopal et al. [1] worked on [4] "feature selection and classification in breast cancer prediction using IoT and machine learning." This research focuses on a procedure to carry out early cancer diagnosis using the IoT as well as machine learning. This research mainly introduced machine learning (ML) approaches toward forecasting cancer disease by using embedded applications. Classification accuracy led to an increase in precision 98%, recall 97%, F-measure 96%, and accuracy 98%, respectively. Iotti et al. [2] worked on [5] "accuracy and reproducibility of contrast-enhanced mammography in the assessment of response to neo-adjuvant chemotherapy in breast cancer patients with calcifications in the tumor bed." This research emphasizes on to analyze precision and reliability in the diagnosis of detected tumor cells upon "neo-adjuvant chemotherapy in breast cancer patients to metastases, utilizing clinical insect path physiology as the reference.

Golagani et al. [3] worked on [6] "supervised learning breast cancer data set analysis in MATLAB using novel SVM classifier." In this research, a conceptual model clinical trial for the classification of the tumor but also cancerous tumor is designed using SVM machine learning techniques. Once choosing a feature, designers retrieve the outposts set across the kernel capacity and then enlarge the edge-construct component restraint scheme. Rahmani and Katouli [4] worked on [7] "breast cancer detection improvement by grasshopper optimization algorithm and classification SVM." The objective of the research work is to develop a novel process to overcome the limitation and minimize the process of diagnosis of breast cancer by utilizing "grasshopper optimization method." The proposed GOA mainly complies with three major sections, and after obtaining, the details the very first stage are to preprocess the data by normalization.

Kavitha et al. [5] worked on [8] "enhanced SVM classifier for breast cancer diagnosis. Breast tumor is a foremost chronic condition to cause fatal especially for

women." The researchers proposed a method, just to a division of data into two categories, one is cancer, and another is non-breast cancer by utilizing machine learning (i.e., supervised). A method applies to select all the essential features which are mainly derived in the initial phases. Salvi and Kadam et al. [6] worked on [9] "breast cancer detection using deep learning and IoT technologies." In this work, researcher proposed a machine learning-based classification tool that accurately categorizes the malignant or benign individuals throughout the surroundings of IoT.

Savitha et al. [7] worked on [10] "a distributed key authentication and OKM-ANFIS scheme based breast cancer prediction system in the IoT environment." This research proposed an efficient key validation concerning "OKM-ANFIS"-based secure breast cancer prediction application based on IoT ecosystem to outclass these drawbacks and also utilized a GA for predicting the mixed layer frameworks. At first, the validation is done with the help of the healthcare provider. Ghasemzadeh et al. [8] worked on [11] "breast cancer detection based on Gabor-wavelet transform and machine learning methods." This work presented an effective system for classifying as well as identifying the public in mammograms and also used machine learning as judgment phase and accomplished average precision more than 93.9%, mean sensitivities 94.11%, and finally the mean specificities higher 91.21%.

3 Research Materials and Methods

This section covers the research material and methods used in this research.

3.1 Proposed System Background

Throughout the following sub-sections, the aspects of the proposed technique are described in depth.

3.1.1 Preprocessing and Optimization of Parameters

The first phase of any classification method is data preprocessing. The dataset lowered the computation period of the classification algorithm as well as enhanced the accuracy, efficiency of the classifier. Preprocessing of a dataset is mainly used to integrate null values, identify or outliers, as well as fix self-contradiction. A method Min–Max [12] vector mainly transforms the information in such a way that almost all characteristics will be varied around 0 or 1.

- **Feature selection**—The main feature selection method goal is to achieve appropriate prediction performance precision with certain possible variables. A prototype of novel walker-assisted gait has been already projected previously that integrates SVM and GA-based attribute selection technique to make a distinction between many non-assisted as well as assisted gait mostly with forearm assistance that has been prosperous at feature score constraint [13].
- **Support vector machine (SVM)**—A support vector machine classifier is a supervised or teacher-based learning method that is focused on the numerical proposition. This was first recommended by Cortes and Vapnik (1995). SVM simply describes the associating of communication corresponding to the characteristics and the tedious attributes. SVM divides the data source into two vector pairs under the "n" spatial dimensions vector. The SVM technique mainly creates a hyperplane ecosystem as each component has to be especially in comparison with the divided normal distribution [14].

A few of the kernel functions are:

1. A polynomial kernel function with the degree value can be defined as:

$$d\,S\,(X_i,\ Yi)\ =\ (X_i * T\,Y\ +\ 1)\,d \tag{1}$$

2. The width value with a radial function kernel can be defined as [18]:

$$\sigma\,S * (X_i,\ Y_i) = \text{exponential}\ [(-\|(X_i - Y_i)\|_2/(2 * \sigma_2))] \tag{2}$$

   - It uses a strongly connected linked with the kernel function of ANN (artificial neural networks).
   - In Eq. (2), a feature vector variable has a limitless value.

3. It is also uses a sigmoid function. This function mainly contains the properties (θ) as well as (.). These properties can be defined by following equation:

$$\begin{aligned} &S\,(X_i,\ Y_i)\ =\ \tan\left(S_x * T_y+\right)\,S\,(X_i,\ Y_i)\ =\ \tan\left(S_x * T_y+\right)\,S\,(X_i,\ Y_i) \\ &=\ \tan\left(S_x *\ T_y+\right)\,S\,(X_i,\ Y_i)\ =\ tan\,(S * x_i *\ T\,y_i+)\,S_i \end{aligned} \tag{3}$$

   where the user-chosen attributes are d, θ, and σ.

   - **Gray wolf's optimization algorithm**—GWO technique is based upon this hunting method of organizational hierarchy. A gray wolf's method is ranked as the important level large predators and that it generally resides mostly in counts of 10–12. Relying on the hunting habits of wolves, this is classified into various kinds including such α, β, δ, and ω. A subsequent pattern of gray wolves is β which are subordinate wolves that assist α to take action [15]. A β wolf is also supposed as α wolf in scenarios whenever α wolf end up dead. The lowest possible scoring gray wolf seems to be ω wolf that performs the aspect of scape goat. Gray wolf encircles natural prey even during the

search. Also to scientifically replicate the surrounding behavior patterns of gray wolves, the model equation is formulated:

$$D_G = \left|C.X_p(T) - X_{wolf}(T)\right|, X_{wolf}(T+1) = [X_p(T) - (A * D)] \tag{4}$$

where variable (T) represents the iteration value, A, C is coefficient, X_p is the value for prey vector, and X_{wolf} is representing the position value of the wolf. The value of A and C can be obtained as:

$$A = [r_1 * 2a_i - a_i] \text{ and } C = [2 * r_2] \tag{5}$$

where variable a can be from 0 to 2, r_1, r_2 *are random vectors, and the value can be from* 0 *to* 1.

A hunt process [20] is primarily governed by key variables alpha (α), beta (β), and delta (δ) according to Eqs. 6–12.

$$D_\alpha = |X_\alpha * C_1 - - X| \tag{6}$$

$$D_\beta = \left|X_\beta * C_2 - - X\right| \tag{7}$$

$$D_\delta = |X_\delta * C_3 - - X| \tag{8}$$

$$X_1 = [X_\alpha - (D\alpha * A_1)] \tag{9}$$

$$X_2 = [X_\beta - (D_\beta * A_2)] \tag{10}$$

$$X_3 = [X_\delta - (D_\delta * A_3)] \tag{11}$$

$$X_{(n+1)} = [(X_1 + X_2) + (X_3)/3] \tag{12}$$

With these formulas, a search assistant upgrades the position as according alpha, beta, as well as delta inside an n-dimensional state space.

- **Fuzzy C-mean**—The FCM technique [16] is a fuzzy clustering method in the perspective of reductions of a quadratic paradigm in which clusters are recognized through their separate hubs. The FCM technique was introduced by Dunn in 1973 and further enhanced by researcher Bezdek. For a working group of C-data patterns $X = \{X_1, X_2, \ldots, X_n\}$, each method enables to sequence the information area, by finding the distribution centers with classes and the subscription matrix. FCM methodology is mainly used to enhance

the precision of cluster analysis of a noise dataset. FCM procedure is based on optimization of the feature:

$$Fj = \sum_{i=1}^{n} \sum_{j=1}^{C} \left(\begin{matrix} m \\ U_{ij} \end{matrix} \right) |\mathrm{X}_i - \mathrm{C}_j| \mathrm{pow2}, 1 <= m < \infty \tag{13}$$

where the U_{ij} is the degree of membership of an element (X_i) in the cluster set j and the squared element is showing the Euclidian distance among the two centers in the cluster (i and j). After each of the iteration, the updated value of the center cluster (C_j) and the membership function is estimated as follows:

$$U_{ij} = 1/ \sum_{(j=1)}^{C} ([X_i - C_j \,\mathrm{pow}(2-1)/m]/|X_i - C_k|) \tag{14}$$

Fuzzy systems set and membership functions, as well as fuzzy rules, that form the compositional elements of a probabilistic reasoning decision-making system are calculated separately.

4 Proposed Method

Throughout this research paper, we introduce an Internet of Things-based predictive analytics just by using fuzzy cluster-focused augmentation as well as optimal SVM classification for forecasting the breast cancer infection via regular inspection as well as to enhance the health services by giving healthcare guidelines. The working of the proposed method is described in Fig. 2.

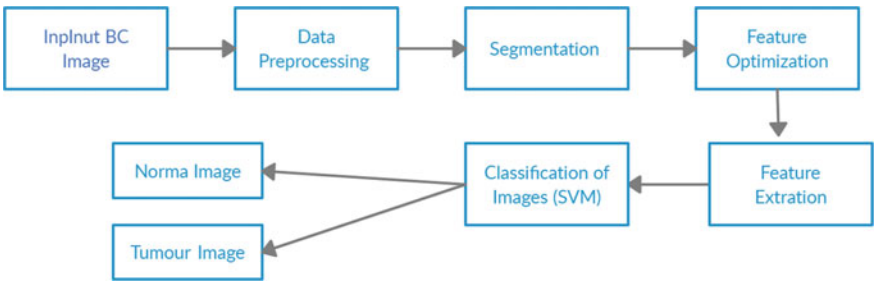


Fig. 2 Proposed enhanced model

4.1 Performance Comparison Parameters

The performance of proposed and existing methods is analyzed with relation to the few measurement criteria. The confusion matrix along with parameters TP, FP, TN, and FN for overall data is to predict data and formulated the parameters [17], where TP = true positive, FP = false positive, TN = true negative, FN = false negative.

- **Accuracy**: Accuracy can be defined as "the ratio of accurately classified data samples and the total number of samples."

$$\text{Accuracy} = [(\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})] \tag{15}$$

- **Sensitivity or Recall**: Sensitivity can be defined as "the ratio of perceived positive data and the number of total positive entries in dataset."

$$\text{Sensitivity} = [(\text{TP})/(\text{TP} + \text{FN})] \tag{16}$$

- **Specificity**: It can be defined as "the relationship between the observed negative entries with all the negative entries in the dataset."

$$\text{Specificity} = [(\text{TN})/(\text{TN} + \text{FP})] \tag{17}$$

- **Precision**: It can be defined as "the ratio of positively selected in between all the predicted positive datasets."

$$\text{Precision} = [(\text{TP})/(\text{TP} + \text{FP})] \tag{18}$$

- **F1-score**: It can be defined as "the F1-score is the harmonic mean of the precision value with sensitivity in the dataset."

$$\text{F1 - score} = [(2 * \text{TP})/(2 * \text{TP} + \text{FP} + \text{FN})] \tag{19}$$

- **Matthews's correlation coefficient (MCC)**: It is mainly used to measure the quality of a binary classification. The range of MCC can be [+1 (best performance) to −1 (worst performance)].

$$\begin{aligned}\text{MCC} = [(\text{TN} * \text{TP}) - \text{FP} * \text{FN})] /[(\text{FP} + \text{TP}) \\ * [(\text{TP} + \text{FN}) * (\text{TN} + \text{FP}) * (\text{TN} + \text{FN})]1/2\end{aligned} \tag{20}$$

5 Experimental Outcomes and Result Analysis, Discussion

This section of the research mainly covers the experimental implementation, result from the analysis of various breast cancer dataset analysis, and diagnosis forecasting by existing and proposed methods. The proposed method and existing methods, i.e., SVM, logistic regression, decision tree, and SVM are implemented and tested on the UCI breast cancer dataset.

5.1 Dataset Description

A breast cancer database was retrieved from the UCI repository (Wisconsin Hospital) [18]. Figure 3 is showing the breast cancer image dataset classification into two categories. One is a normal mammogram, another is a benign cyst (non-cancer), and the next is a breast cancer image. The experimental study utilizes two different datasets with different scenarios.

- **The first dataset** [19] contains 800 data cases in this set, in which the occurrences are malignant or benign. This set contains instances, 560 (70%) seem to be benign, and 240 (30%) seem to be malignant. The first dataset falls into 10 groups includes 250, 100, 50, 125, 75, 135, 65, 60, 40, and 100 data samples. This dataset contains 29 attributes for cancer-bacteria cell which includes: [mitosis, cell thickness, cell size uniformity, nude nucleus, single epithelial cell size, normal nucleus, cell shape uniformity, marginal adhesion, and chromatin]
- **The second data** contains 669 data samples collected from the UCI breast cancer dataset Wisconsin Hospital. This dataset contains a mixer of 467 benign and 312

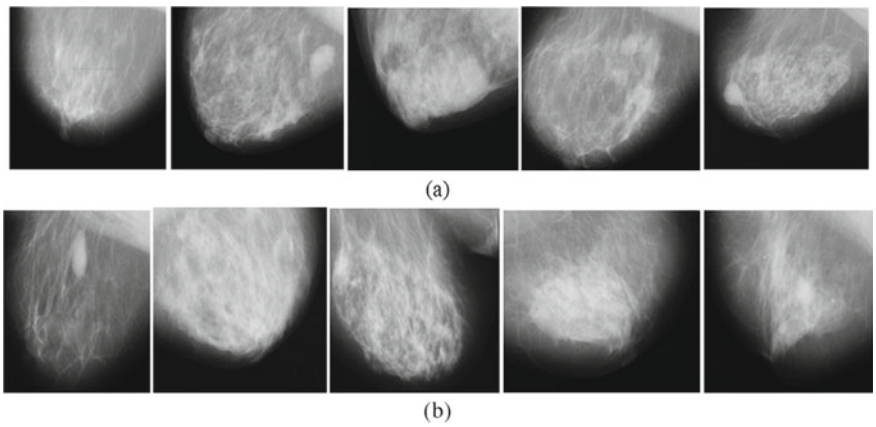


Fig. 3 Breast cancers dataset: **a** Breast cancer initial stage, **b** Breast cancer at the medium stage, and

cancer patients. The dataset mainly contains the following 32 attributes. **A few of them are as follows**:

[Patient ID number, the diagnosis/treatment details (M or B) where M = malignant and B = benign cell nucleus data feature, distance from center to perimeter (radius), the grayscale value standard deviation data (texture), data for perimeter, data for the area, data for the radius in local variation, data for compactness (where compactness= [perimeter^2/area – 1]), data concavity, data point concavity, data for symmetry, and data for attribute dimension].

- The third dataset contains: 498 entries collected from UCI official site Wisconsin data. This dataset contains thirty-four attributes. Figure 4 is showing the detailed description of the dataset. This dataset can be classified into two categories of recurring data and non-recur data. A few of them are as follows:

  [Patient ID, result (R, N, where R for recurring and N for non-recur), recurrence time details (2 for R and 4 for N), cell nucleus data feature, distance from center to the perimeter (radius), the grayscale value standard deviation data (texture), distance from center to perimeter (radius), data for total area, details of smoothness (radius lengths), details of compactness (where compactness= [perimeter pow 2/area – 1], details for concavity, details for concave points, details for symmetry, data for fractal dimension for each data] (Table 1).

Feature ranking result for dataset—1, 2, and 3 by using the proposed method is as follows (Figs. 5 and 6).

A confusion matrix [11] is constructed for 115 instances of Dataset 1. The Confusion matrix result on dataset1 on 115 instances, by proposed method and existing methods (i.e. Logistic Regression, SVM, Decision Tree) (Figs. 7, 8 and 9).

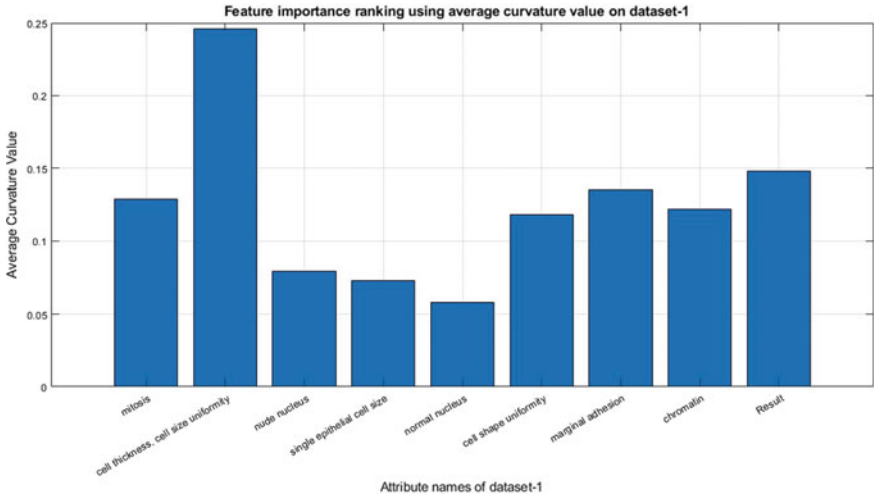


Fig. 4 Feature ranking for the attribute (Dataset 1)

Table 1 Dataset description

Breast cancer dataset	Total attributes inset	Total number of instance in the dataset	Total number of class
Dataset 1	29	800	2
Dataset 2	32	669	2
Dataset 3	34	498	2

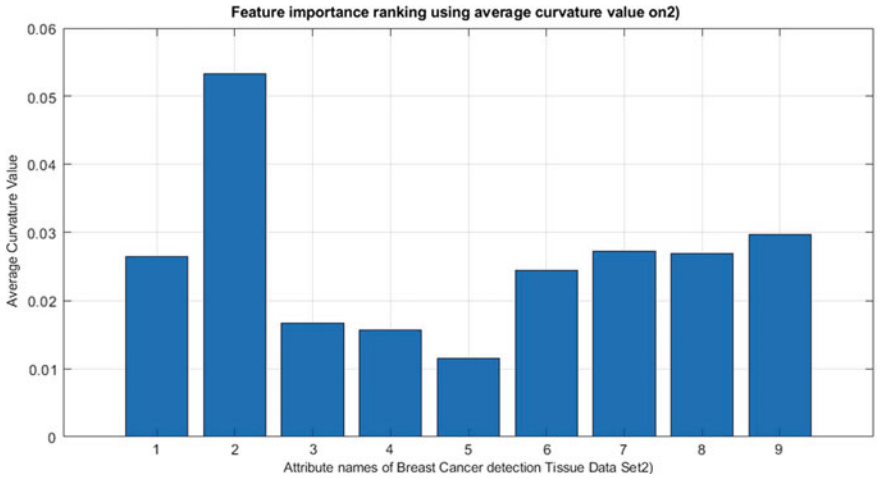


Fig. 5 Feature ranking for the attribute (Dataset 2)

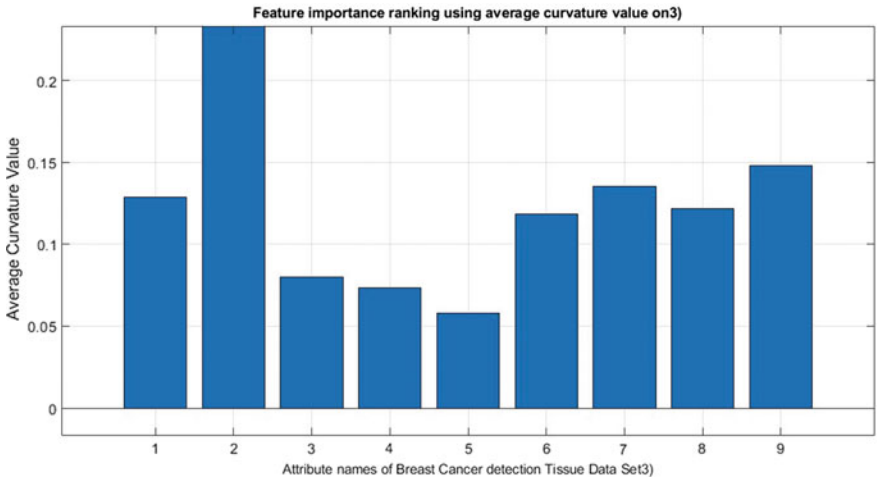


Fig. 6 Feature ranking for the attribute (Dataset 3)

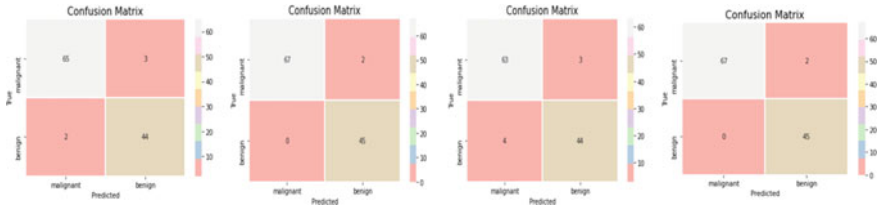


Fig. 7 Confusion matrix result for Dataset 1 by existing and proposed (IV) methods

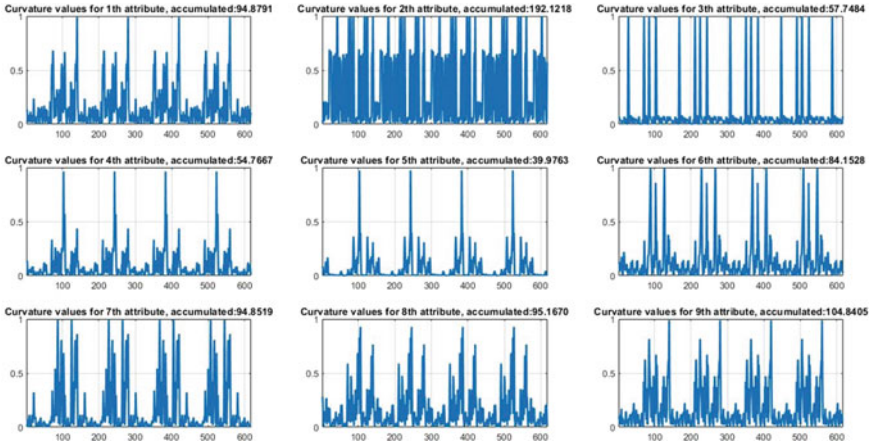


Fig. 8 Curvature value for a various attribute of Dataset 1 using the proposed method

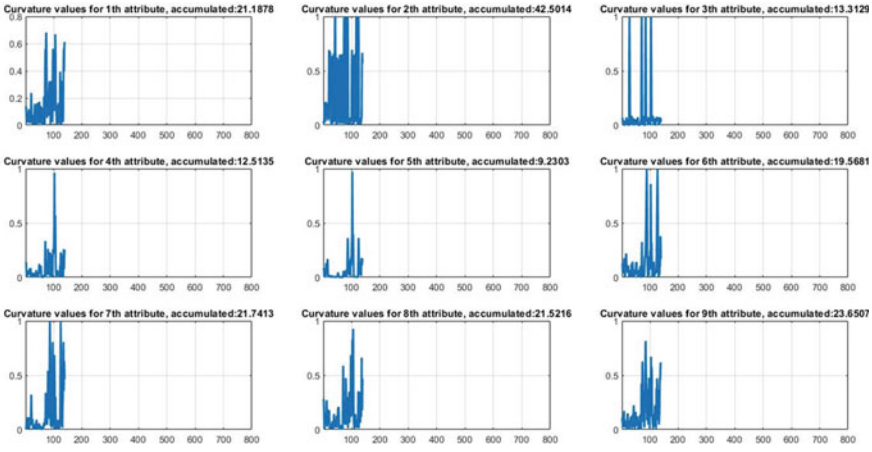


Fig. 9 Curvature value for a various attribute of Dataset 2 using the proposed method

Breast cancer curvature value for a various dataset using the proposed method is as follows.

Table 2 is showing the experimental results for three datasets (Dataset 1, Dataset 2, and Dataset 3) for various methods existing methods (i.e., logistic regression [13], SVM, decision tree [14]) and proposed (enhanced SVM). For Dataset 1, the logistic regression method is showing precision value 87.8%, sensitivity value 90.98%, and specificity value 92.8%, whereas SVM method is showing precision value 95.87%, sensitivity 96.89%, and specificity value 97.99%, whereas decision tree is showing 95.8% precision value, 96.87% sensitivity value, and 97.87% for specificity value, and proposed enhanced SVM method is showing precision 96.26%, sensitivity 97.15%, and specificity 98.5%. Similar for Dataset 2, existing logistic regression method is showing 84.78, 87.88, and 91.28% results for precision, sensitivity, and specificity, whereas the existing SVM method is showing 94.91, 93.74, and 96.32 for precision, sensitivity, and specificity, and the existing decision tree is showing 93.82, 94.74, and 96.89% value for precision, sensitivity, and specificity. Similarly for Dataset 3, existing logistic regression method is showing 85.44, 88.96, 90.45% results for precision, sensitivity, and specificity, existing SVM is showing 92.89, 93.45, and 95.45% results for precision, sensitivity, and specificity, and existing decision tree is showing 91.25, 94.12, and 95.6% results for precision, sensitivity, and specificity. Based on the experimental results on Dataset 3, we can say the proposed method is performing outstanding over existing methods.

Table 2 Experimental result analysis for proposed versus existing method

Dataset	Method	Precision %	Sensitivity %	Specificity %
Dataset 1	Logistic regression	87.8	90.98	92.8
	SVM	95.87	96.89	97.99
	Decision tree	95.8	96.87	97.87
	Proposed method	96.26	97.15	98.5
Dataset 2	Logistic regression	84.78	87.88	91.28
	SVM	94.91	93.74	96.32
	Decision tree	93.82	94.74	96.89
	Proposed method	97.06	97.37	96.25
Dataset 3	Logistic regression	85.44	88.96	90.45
	SVM	92.89	93.45	95.45
	Decision tree	91.25	94.12	95.6
	Proposed method	96.89	95.68	97.88

6 Conclusions and Future Work

Now, the world's more common cancer in females is breast cancer. Accurate intervention at the initial stage can significantly enhance the likelihood of sustainability. The major obstacle against it in the recognition will be how to categorize cancer cells into the cancerous (malignant) as well as (non-cancerous (benign). The main purpose of this investigation was to formulate an efficient classifier for the diagnosis of breast cancer. In this research, we introduced a novel IoT-based predictive analytics just by using fuzzy cluster-focused augmentation as well as optimal SVM classification for forecasting the breast cancer infection via regular inspection as well as to enhance the health services by giving healthcare guidelines. The experimental analysis takes part with various three UCI breast cancer datasets (Datasets 1, 2, and 3) for three existing methods (i.e., logistic regression, SVM, decision tree) and proposed enhanced SVM method. In the future work, we can improve the proposed method by applying the inclusion of the deep learning (DL) algorithm on more datasets.

Conflicts of Interest We declare that no conflict of interest.

References

1. Gopal VN, Al-Turjman F, Kumar R, Anand L, Rajesh M (2021) Feature selection and classification in breast cancer prediction using IoT and machine learning. Measurement (Lond.) no. 109442:109442
2. Iotti V et al (2021) Accuracy and reproducibility of contrast-enhanced mammography in the assessment of response to neoadjuvant chemotherapy in breast cancer patients with calcifications in the tumor bed. Diagnostics (Basel) 11(3):435
3. Golagani PP, Mahalakshmi TS, Beebi SK (2021) Supervised learning breast cancer data set analysis in MATLAB using novel SVM classifier. In: Advances in intelligent systems and computing. Springer, Singapore, pp 255–263
4. Rahmani A, Katouli M (2020) Breast cancer detection improvement by grasshopper optimization algorithm and classification SVM. Rev d Intell Artif 34(2):195–202
5. Kavitha M, Lavanya G, Janani J, Balaji (2020) Enhanced SVM classifier for breast cancer diagnosis. Int J Eng Tech Mgmt Res 5(3):67–74
6. Salvi S, Kadam A (2021) Breast cancer detection using deep learning and IoT technologies. J Phys Conf Ser 1831(1):012030
7. Savitha V, Karthikeyan N, Karthik S, Sabitha R (2021) A distributed key authentication and OKM-ANFIS scheme based breast cancer prediction system in the IoT environment. J Ambient Intell Humaniz Comput 12(2):1757–1769
8. Ghasemzadeh A, Sarbazi Azad S, Esmaeili E (2019) Breast cancer detection based on Gabor-wavelet transform and machine learning methods. Int J Mach Learn Cybern 10(7):1603–1612
9. Arya N, Saha S (2021) Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowl Based Syst 221(106965):106965
10. Silva Araújo V, Guimarães A, de Campos Souza P, Silva Rezende T, Souza Araújo V (2019) Using resistin, glucose, age and BMI and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer. Mach Learn Knowl Extr 1(1):466–482
11. Huo L et al (2021) Machine learning models to improve the differentiation between benign and malignant breast lesions on ultrasound: a multicenter external validation study. Cancer Manag Res 13:3367–3379

12. Fan J et al (2016) Structure-leveraged methods in breast cancer risk prediction. J Mach Learn Res 17
13. Ribelles N et al (2021) Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients. Eur J Cancer 144:224–231
14. Panagopoulou M, Karaglani M, Manolopoulos VG, Iliopoulos I, Tsamardinos I, Chatzaki E (2021) Deciphering the methylation landscape in breast cancer: diagnostic and prognostic biosignatures through automated machine learning. Cancers (Basel) 13(7):1677
15. Patil RS, Biradar N (2020) Improved region growing segmentation for breast cancer detection: progression of optimized fuzzy classifier. Int J Intell Comput Cybern 13(2):181–205
16. Sudhakar S, Logesh R, Saravanan P (2020) A fuzzy-based high-resolution multi-view deep CNN for breast cancer diagnosis through SVM classifier on visual analysis. J Intell Fuzzy Syst 39(6):8573–8586
17. Shen T, Wang J, Gou C, Wang F-Y (2020) Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis. IEEE Trans Fuzzy Syst 28(12):3204–3218
18. Venkatesh V, Anishin Raj MM, Mohamed Sajith K, Anushiadevi R, Suriya Praba T (2020) A precision-based diagnostic model ADOBE-accurate detection of breast cancer using logistic regression approach. J Intell Fuzzy Syst 39(6):8419–8426
19. Ramesh Dhanaseelan F, Jeyasutha M (2021) A novel fuzzy frequent itemsets mining approach for the detection of breast cancer. Int J Inf Retr Res 11(1):36–53

An Improved Deep Neural Network-Based Predictive Model for Traffic Accident's Severity Prediction

Sarita Simaiya, Umesh Kumar Lilhore, Himanshu Pandey, Naresh Kumar Trivedi, Abhineet Anand, and Jasminder Sandhu

Abstract In recent decades, a vehicle crash has become a worldwide issue and labeled it the twelfth prominent reason for death worldwide. Artificial intelligence-based techniques, i.e., machine learning and deep learning play a vital role in various aspects of modern society and accidental predication is one of them. This research presents an accidental prediction model based on an improved deep neural network (IDNN). A DNN model mainly contains various hidden layers toward nodes. The proposed IDNN model includes two modules. The first model is based on an unsupervised feature learning interface to recognize operational networks and correlation factors. The second model is based on a supervised optimization subsystem with an extended negative binomial distribution which helps in forecasting road crashes. The proposed model was simulated over traffic datasets collected from the online data sources, and various performance measuring parameters have been calculated, i.e., precision, accidental detection. The experimental results demonstrate strengthen of the proposed model over the existing model.

Keywords Improved deep neural network extended binomial distribution · Road safety · Accident prediction · AI

S. Simaiya · U. K. Lilhore (✉) · N. K. Trivedi · A. Anand · J. Sandhu
Institute of Engineering and Technology, Chitkara University, Punjab, India
e-mail: umesh.lilhore@chitkara.edu.in

S. Simaiya
e-mail: sarita.simaiya@chitkara.edu.in

N. K. Trivedi
e-mail: nareshk.trivedi@chitkara.edu.in

A. Anand
e-mail: abhineet.anand@chitkara.edu.in

J. Sandhu
e-mail: jasminder.sandhu@chitkara.edu.in

H. Pandey
Faculty of Engineering & Technology, University of Lucknow, Lucknow, U.P., India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_17

1 Introduction

Road safety is still an extremely important issue in an efficient transport implementation. Road accidents will get some negative, effects on individuals, which include fatalities, traffic congestion, as well as air pollution, which are not amenable to the stable and efficient advancement of the public transport system. Also with a rapid enhancement of the standard of automatic vehicle information management, in previous years, a few other agencies and road transport manufacturing industries have been devoted to the success of smart transport solutions to enable the sustainability of road transport [1].

Road crash forecasting has become a complex and important problem throughout the domain of smart traffic safe operation governance systems, and it is of particular importance for evaluating the potential development pattern of road accidents and incorporating strategic preventive actions under existing road traffic patterns. To enhance road safety supervision and operation, this is vital to obtain consistent and appropriate techniques for identifying road accident intensity [2]. In this research paper, we are presenting an accidental prediction model and road safety enhancement model by utilizing an improved deep neural network (IDNN). The complete article is divided into various sections which cover introduction, theoretical background, proposed model, experimental results, and result comparison, and finally covers conclusion.

2 Theoretical Background

Heavy traffic has been incorporated further into the complicated framework in positions of managing and coordinating mostly because of a growing number of people. Following research papers have been analyzed in literature survey work.

In researcher [3] worked on, "Leak flow prediction during the loss of coolant accidents using deep fuzzy neural networks." Throughout this research, leak flow all through LOCAs has been anticipated by using a deep fuzzy neural network (DFNN) framework. A DFNN system is predicated on fuzzy neural network (FNN) components and also has a framework in which the FNN components are concurrently linked. In research [4] worked on, "Method of highway risk assessment and accident quantity prediction based on multi-source heterogeneous data and deep neural network." Throughout this article, by using multi-source statistics, integrated with a deep learning model, autonomous open learning for highway threat as well as collision percentage prediction was indeed established.

In research [5] worked on, "Bounded memory-based frequent pattern growth approach with deep neural network and decision tree for road accident prediction." An analysis of analytical outcomes confirmed that the proposed "Bounded FP-growth-Deep Neural Network with J48 classifier" (BFP-growth-DNNJ48) accomplished higher accuracy (94%) when especially in comparison to the conventional

methods such as Ada-Boost-SO as well as TASP-CNN. In research [6] worked on, "Predicting injury severity of road traffic accidents using a hybrid extreme gradient boosting and deep neural network approach." Technological tools like data analytics as well as simulation methods are crucial to predict potential events of traffic safety.

In research [7] worked on, "Deep neural-network-based predictive modeling of road accidents." The whole paper proposed using the deep neural networks (DNN) method for forecasting accidents. DNN comprises two or more hidden units mostly with a massive quantity of nodes. Accident statistics of non-urban segments of eight expressways were obtained from official data, and the sample size consisted entirely of several 2680 fatalities. In research [8] worked on, "FPGA based human fatigue and drowsiness detection system using deep neural network for vehicle drivers in road accident avoidance system." This research is intended purely for the development of extreme fatigue and drowsiness detector by using a deep neural network in FPGA. Within the proposed model, an image is filtered utilizing median filtering as well as Viola-Jones face detection method for further extracting the facial features. The established Vision-based system has become a convenient method for real-time control of vehicle constant vigilance [9].

3 Proposed IDNN Model

In this research paper, we are presenting an accidental prediction model and road safety enhancement model by utilizing an improved deep neural network (IDNN). To resolve the road accident problem discussed in the previous section, the proposed method efforts to only anticipate the middle frame which is deleted from the sequentially frames as just the direct input, though sometimes it can be recognized as an interpolation of the deleted frame. A DNN model [10] mainly contains various hidden layers toward a variety of nodes. The proposed IDNN model includes two systems, an unsupervised feature learning interface to recognize the operational network among the correlation factors as well as the attribute representations and a supervised optimizing subsystem to accomplish road crashes forecasting by using the extended negative binomial distribution (ENBD).

In an attempt to enhance the existing DNN efficiency, we analyzed all use of input standard normalization, dropout, and cross-validation.

- **Over-fitting with Dropout**: even in relatively large neural networks, this becomes challenging to initiate with over-fitting issues with features and data. An ideal solution to fix this issue, without expanding its data size here is to implement the dropouts' strategy that also partially disables components throughout the training process.
- **Over-fitting with Cross-Validation**: If the multi-layer neural network experimental test itself is not stopped somewhere at the right point, over-fitting can happen. To prevent this issue, the cross-validation technique can be applied also with a training sample, i.e., within a week of completing the assigned series

of training iterations, training is stopped, and also the development sequence is circulated to calculate the corresponding data inaccuracy.

3.1 Unsupervised Feature Learning (UFL)

Deep Neural Networks on image file processes take in Pixel Vectors of its shape (height, width, channels) as well as process individuals into one of lesser characteristics via a sequence of multivariate features. Supervised and unsupervised activities both aspire to understand a contextually better understanding of characteristics from the original [11]. Training Deep Supervised Learning techniques necessitate a large quantity of data in the shape of labeled (x, y) sets. An unsupervised algorithm does not require correlating labels (y), the other most widely accepted instance of these was the auto-encoders method. An auto-encoder process takes parameter x as input, moves it via a sequence of nested to constrict the dimensions as well as is then excoriated about how well individuals can reshape x. The auto-encoders learn quickly a selection of characteristics that will define a content x; even then, such characteristics were also probable to be not very beneficial for predefined classes as well as exclusionary activities [12]. A prediction model for perceptron can be calculated by using Eq. 1.

$$yP = f\,(Wx + b) \text{Where}: \ \mathrm{W} = \text{perceptron weight}, \ \mathrm{b} = \text{output layer bias value}, \tag{1}$$

A perceptron is an exclusionary linear classification method that is primarily used to rectify the simple multiclass classification difficulties of AND, NOT, and OR. The perceptron has quite a single-static neuron that is not very smart enough to learn and therefore has no way to deal with nonlinear problems.

3.2 Supervised Optimizing Subsystem (SOS)

The SOS model involves a regression layer on the surface of the possibly resulting hidden participation layers to evaluate the possibility of collision incidences. The generated transmitted feature interpretations first from the unlabeled data module are being used as an activation function for the SOS module [13]. To collectively approximate the incidence possibility for yet more than one form of accidents concurrently as well as resolve the prospective variability problems in the interlinked accident statistics, a "multivariate negative binomial system" (MNBS) has been used in the SOS module to evaluate the predictive the severe injuries all over accident occurrences.

4 Experimental Results and Discussion

Accident statistics of urban regain of ten expressways [14] have been obtained from official databases, and a set of data consists mainly of several 4800 accidents. The statistics of explanatory variables linked to road geometry, traffic congestion atmosphere have been obtained from public documents as well as through research projects. Another difficult aspect of this investigation is accumulating sample sets of data from the organization known. The initial dataset accumulated from the authority was managed individually.

4.1 Road Accidental Dataset

The statistics utilized for the whole research were mostly gathered out over a particular sequence of ten expressways in India. The raw data for traffic characteristics as well as traffic conditions were obtained via site visits together across various expressways. Traffic density statistics for varying highways have been collected from Toll booths, various detailed project reports (DPRs) designed by the National Highway Authority of India (NHAI), and also from vehicular registers of the India highway department [15]. The statistics of particular spot rates of speed were obtained from DPRs as well as incomplete information through using speed gun [16]. Mostly on premise of the gathered data, road segments showing equal traffic, traffic volume, mean average length, and maximum acceleration was identified [17]. The road network length of various local highways discussed in this research has been more than 400 km, and then was partitioned in 90 consistent segments of different lengths as can be seen in Table 1 (Fig. 1).

In the above images, the red circle is showing the possible accidental zone identified by using the improved deep learning method [18]. This zone is based on the previous history and number of accidents in a particular location. The confusion matrix mainly shows the performance of machine learning methods. Classification accuracy itself can't be observed in the results in multiple classes. A confusion matrix can be defined as "Is a matrix which is frequently utilized to characterize the

Table 1 Road accident data (year wise)

Year	Road accidents in thousand	Persons injured (in thousands) %	Persons killed (in Nos.)	No. of vehicles (In Thousands)
2019	447.9	493.5	1,54,732	2,53,311
2018	449.8	489.65	1,52,780	2,53,311
2017	458.9	469.98	1,50,093	2,53,311
2016	478.5	479.99	1,50,093	2,30,031
2015	467.5	469.8	1,48,707	2,10,023

Fig. 1 Accidental images used for training

Table 2 Confusion matrix training and testing data for proposed IDNN

	Actual (critical injuries) 1	Actual (non critical injuries) 0
Training data		
Predicated (critical injuries) 1	3763	2073
Predicated (non critical injuries) 0	2088	3755
Testing data		
Predicated (critical injuries) 1	2215	1622
Predicated (non critical injuries) 0	1613	2330

effectiveness of the algorithm on a set of test data values is computed." Table 2 is showing the experimental result for confusion matrix training and testing data for the proposed IDNN. In the confusion matrix, 1 is representing critical injured and 0 is representing non critical injuries [19]. In the proposed method training data is showing a total of 3763 critical injuries and 2073 non critical injuries and for testing 2215 critical injuries and 1622 for non critical injuries. Similarly, Table 3 is showing experimental results for confusion matrix training and testing data for the existing

Table 3 Confusion matrix training and testing data for existing DNN

	Actual (critical injuries) 1	Actual (non critical injuries) 0
Training data		
Predicated (critical injuries) 1	1940	1350
Predicated (non critical injuries) 0	1319	1891
Testing data		
Predicated (critical injuries) 1	3949	1859
Predicated (non critical injuries) 0	1791	2900

DNN method. The existing method is showing 3949 critical injuries and 1791 non critical injuries for training and 1940 (1) and 1319 (0) for testing.

The precision is calculated as the quantity of accurate classification divided by the total of observations. The precision seems to be the percentage of accurate evidence throughout the list among all retrieved search queries. Equation 2 is showing the precision formula [20]—

$$\text{Precision} = [\text{TP} / (\text{TP} + \text{FP})] \tag{2}$$

where, TP = True positive, FP = False positive. Figures 2 and 3 are showing the precision results for the proposed and existing method for accidental and non-accidental datasets. In Fig. 2, existing DNN method is showing precision value for accidental related data 80.28, 78.8, 77.9, 77.5, and 77.2% and the proposed IDNN method is showing precision value 81.89, 81.45, 81. 3, 81.2, and 81.1% for number of feature 10, 20, 30, 40, and 50.

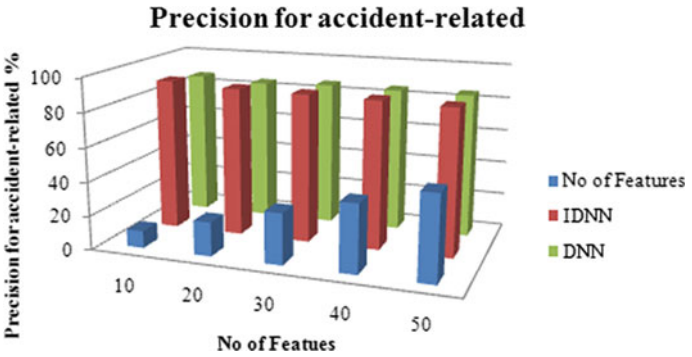


Fig. 2 Results for precision value accident-related data IDNN versus DNN

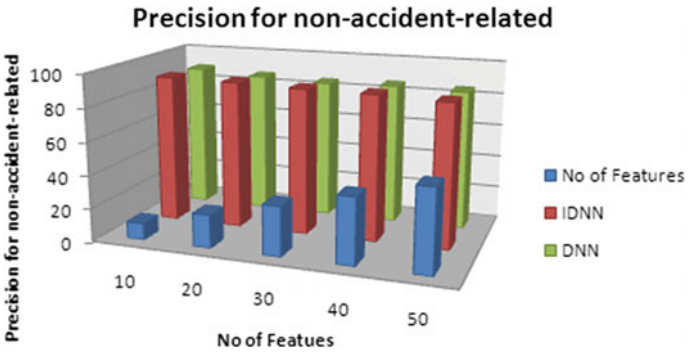


Fig. 3 Results for precision value non-accident-related IDNN versus DNN

In Fig. 3, existing DNN [21, 22] method is showing precision value for non-accidental related data 83.11, 83.10, 82.57, 81.2.5, and 80.2.2% and the proposed IDNN method is showing precision value 85.19, 85.05, 84.73, 84.02, and 82.21% for number of feature 10, 20, 30, 40, and 50. The above simulation results are clearly showing the strength of the proposed IDNN method over the existing DNN method.

5 Conclusions

Detection of crashes will prevent possible accidental deaths, and therefore will encourage authorities to make the exterior framework usable for traffic yet again. Throughout this research, we have shown that heavy traffic activity can always be evaluated utilizing vehicular locations as well as the speed of operation and unusual actions on the highway could also be taken into consideration potential hazards for such road users who already are nearer to the incident region. To achieve accidental detection in a better way, this research presented an improved deep neural network model for accidental prediction.

The experimental results demonstrated that the proposed IDNN model is showing better precision value for accidental and non-accidental data over the existing DNN model. The proposed IDNN method shows 88.9% precision for accidental images and 89.1% for non-accidental images where DNN shows 84.5% for accidental and 86.65% for non-accidental images. The methodology measured in this research can be utilized to reduce road accidents in such an actual circumstance that mostly potentially results in better emergency services. More precise features can be configured shortly by a joint group effort in developing a classifier so that it can be frequently used during data analysis. The experimental results are clearly showing the proposed IDNN method is showing a better precision value (more than 2 or 3% higher) for accidental and non-accidental datasets over the existing DNN method for various features.

Conflicts of Interest We declare that no conflict of interest.

Funding Not received any funding from any of the agencies for the research.

References

1. Park JH, An YJ, Yoo KH, Na MG (2021) Leak flow prediction during loss of coolant accidents using deep fuzzy neural networks. Nucl Eng Technol 7(11):1–9. https://doi.org/10.1016/j.net.2021.01.040
2. Zhang X, Huang C, Chen Y (2020) Method of highway risk assessment and accident quantity prediction based on multi-source heterogeneous data and deep neural network. In: Resilience and sustainable transportation systems. American Society of Civil Engineers, Reston, VA, pp 118–125. https://doi.org/10.1061/9780784482902.014

3. Prasath (2020) Bounded memory-based frequent pattern growth approach with deep neural network and decision tree for road accident prediction. Indian J Comput Sci Eng 11(5):623–633. https://doi.org/10.21817/indjcse/2020/v11i5/201105189
4. Singh G, Pal M, Yadav Y, Singla T (2020) Deep neural network-based predictive modeling of road accidents. Neural Comput Appl 32:12417–12426. https://doi.org/10.1007/s00521-019-04695-8
5. Selvathi D (2020) FPGA based human fatigue and drowsiness detection system using deep neural network for vehicle drivers in road accident avoidance system. In: Human, behaviour analysis using intelligent systems, vol 6. Springer, Cham, pp 69–91. https://doi.org/10.1007/978-3-030-35139-7_4
6. Pradhan B, Sameen MI (2020) Laser scanning systems in highway and safety assessment. Springer International Publishing, pp 119–127. https://link.springer.com/chapter/https://doi.org/10.1007/978-3-030-10374-3_1
7. Harith SH, Mahmud N (2020). Trend on Drink drive and road accident across Asian region: a review study. Indian J Publ Health Res Devel 11(1):1789–1796. Retrieved from https://medicopublication.com/index.php/ijphrd/article/view/1349
8. Kumar MK (2020) Pattern of facial bone fractures in road traffic accidents riding on two-wheelers in a tertiary hospital in Bangalore. Int J Sci Res 9(8):1–3. https://doi.org/10.36106/ijsr/3526783
9. Suman S (2020) Deep neural network-based prediction of burst parameters for Zircaloy-4 fuel cladding during loss-of-coolant accident. Nucl Eng Technol 52(11):2565–2571. https://doi.org/10.1016/j.net.2020.04.025
10. Afghari AP, Haque MM, Washington S (2020) Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes. Accident Anal Prev 144:105615. https://doi.org/10.1016/j.aap.2020.105615
11. Vingilis E, Beirness D, Boase P, Byrne P, Johnson J, Jonah B, Wiesenthal DL (2020) Coronavirus disease 2019: what could be the effects on road safety? Accident Anal Prev 144:105687. https://doi.org/10.1016/j.aap.2020.105687
12. Lilhore UK, Simaiya S, Guleria K, Prasad D (2020) An efficient load balancing method by using machine learning-based VM distribution and dynamic resource mapping. J Comput Theor Nanosci 17(6):2545–2551. https://doi.org/10.1166/jctn.2020.8928
13. Tang J, Zheng L, Han C, Yin W, Zhang Y, Zou Y, Huang H (2020) Statistical and machine-learning methods for clearance time prediction of road incidents: a methodology review. Anal Methods Accident Res 27:100123. https://doi.org/10.1016/j.amar.2020.100123
14. Høye AK, Hesjevoll IS (2020) Traffic volume and crashes and how crash and road characteristics affect their relationship—a meta-analysis. Accident Anal Prev 145:105668. https://doi.org/10.1016/j.aap.2020.105668
15. Yadav AK, Velaga NR (2020) Alcohol-impaired driving in rural and urban road environments: effect on speeding behaviour and crash probabilities. Accident Anal Prev 140:105512. https://doi.org/10.1016/j.aap.2020.105512
16. McIlroy RC, Kokwaro GO, Wu J, Jikyong U, Nam VH, Hoque MS, Stanton NA (2020) How do fatalistic beliefs affect the attitudes and pedestrian behaviours of road users in different countries? A cross-cultural study. Accident Anal Prev 139:105491. https://doi.org/10.1016/j.aap.2020.105491
17. Zhu L, Li T, Du S (2019) TA-STAN: a deep spatial-temporal attention learning framework for regional traffic accident risk prediction. In: 2019 IEEE international joint conference on neural networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN.2019.8852212
18. Tarko AP (2020) Analyzing road near departures as failure-caused events. Accident Anal Prev 142:105536. https://doi.org/10.1016/j.aap.2020.105536
19. Akkayasamy VS, Panneer S (2020) Situational analysis of road traffic accidents—a case of Madurai district rural areas, Tamil Nadu. Indian J Community Health 32(3). https://doi.org/10.47203/IJCH.2020.v32i03.015
20. Jain S (2020) Road accidents in India: constraints and challenges. Gedrag En Organisatie 33(03). https://doi.org/10.37896/gor33.03/412

21. Jaidev B, Garg S, Makkar S (2019) Artificial intelligence to prevent road accidents. Int J Mach Learn Netw Collaborative Eng 3(01):35–45. https://www.mlnce.net/index.php/Home/article/view/72
22. Do Koo Y, An YJ, Kim CH, Na MG (2019) Nuclear reactor vessel water level prediction during severe accidents using deep neural networks. Nucl Eng Technol 51(3):723–730. https://doi.org/10.1016/j.net.2018.12.019

Grading of Apples Using Multiple Features

Smrithy R. Sunil, H. B. Anita, and P. Renupriya

Abstract Apple is the most demanding food product that has the utmost importance when it comes to drupes. Food is the very basic necessity for our survival. Every new day brings a change, and the demand for a better quality is no greed. Quality food benefits the health of the living beings, and thus, it increases the economic growth of our country. There is a huge possibility that identifying the different varieties of apples is quite a tedious job for these traders and time consuming. Generally, identification is done manually by the very three basic senses: sight, hearing and smell. In the proposed work, an image processing technique is used to differentiate between the varieties of apples such that the manual process can be eliminated. Commercially available seven varieties of apple with various size, shape and color are considered to create database. Apples are purchased from different places across Karnataka, India to create the database. Various spatial and frequency domain based features are extracted from the images of apple. Naive Bayes, Random Forest and Multilayer perceptron (MLP) classifiers are used and got motivating results. An average accuracy of 78.47% is obtained using methods like Fourier Transform and Discrete Cosine Transform.

Keywords Image processing · MLP · Naive Bayes · Random Forest · Fast Fourier transform · Discrete cosine transform · Apple · Drupes · MATLAB · WEKA

S. R. Sunil (✉) · H. B. Anita · P. Renupriya
Department of Computer Science, CHRIST (Deemed To Be University), Bengaluru, Karnataka, India
e-mail: smrithy.sunil@cs.christuniversity.in

H. B. Anita
e-mail: anita.hb@christuniversity.in

P. Renupriya
e-mail: p.renupriya@cs.christuniversity.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_18

1 Introduction

Life of a person depends on their lifestyle. A healthy lifestyle will lead to a long happy life. Three most important factors that guard a human being are food and water, clothing and shelter. Each of these play a significant role in our survival but the most significant one among these are food and water. Human intuition plays a very important role for identification purposes. The basic human senses are a part of these intuition. Normally, people identify apples just by the color but identifying it by its variety is quite difficult. Human brain tries to process these colors and relate it to the corresponding variety but human intuition can go wrong at this stage, and this is applicable to the traders as well.

Image based techniques are used in various fields such as astronomical analysis, medical field [1], automation, industry and military. Agriculture is the start and end of human evolution. Apples are the most significant of all. Apples have a lot of importance in the food community. Apples are of various types, and each type has various nutritional levels. Apples are scientifically known MalusDomestica, are rich in 2.4 g of fiber and vitamin C. Apples contain a 52 nutritional caloric value. It contains 86% of water content, 0.3 g of protein, 13.8 g of carbs, 0.2 g of fat and 10.4 g of sugar. These are the minimum compound present in any type of apple. Apples contain potassium in excess, and it can benefit heart health. Apples prevent type 2 diabetes and takes a control on cholesterol. It lowers the risk of the most dangerous disease cancer and that is why, it is said to consume at-least one apple a day.

Images get audience attention quickly and derive a significant conclusion from it. Human brains reciprocate to images and make assumptions. Color components are one of the major factors that contributes to this field of study. Human brains have the capability to make correct comparisons between these visuals, and this can be automated in order to make it an efficient product. In the proposed work, 13 features were extracted in order to classify 7 prominent varieties of apples. Spatial and frequency domain based features are extracted. Images are captured in various angles and constant distance from the camera. The main aim of the proposed work is to automate the classifying of different types of apple for trading. This helps not only the vendors but also benefits the consumers.

2 Literature Survey

There is very high scope in automation of agriculture production in India. Very less research work is done in the area of agriculture automation.

Oo and Aung [2] did classification of strawberry based on shape and size. The classification was done using an image processing algorithm such that it can estimate the size and shape of strawberry. It was achieved using geometrical properties called “right kite” and “simple kite.” The classification accuracy is between 94 and 97%

and the average processing time for one strawberry is below 0.45–0.5 s. Wang [3] discussed on how to sort mushrooms. Mushrooms are one of the most commonly used products but these pileus are quite difficult to be sold in the market. An algorithm was created to sort these white button mushrooms. This automatic sorting system will reduce the labor fatigue. Various principles and algorithms were developed to sort out mushrooms of various price. The accuracy of grading was 97.42%. The system achieves machine automatic sorting and grading of fresh white button mushrooms with minimal destruction. This process was involved in the hardware systems such that sorting can be done. Tu et al. [4] aim at identifying passion drupes, and their maturity of using RGB depth images. A dataset of 4000 passion drupe images was created and applied Convolutional Neural Networks and Dense scale invariant features with locality-constrained linear coding. These machine learning algorithms were used in feature extraction process and then the SVM classifier was used in the work and obtained an accuracy of 92.71% for detection and an accuracy of 91.52% for maturity classification. Utai et al. [5] aim at estimating the mass of mango drupe using image processing technique and artificial neural network. 61 randomly selected mango drupes were targeted as the dataset. A mathematical tool such as MATLAB which incorporated image processing algorithms was used in extracting the 2D geometrical features, and Artificial Neural Network was used to detect the mass. An accuracy of 97% and 0.99 highest coefficient of efficiency using two inputs or three inputs were obtained. Hossain [6] proposed a classification accuracy to identify infected date drupe. A dataset of 75 date samples was collected, and 64 principle component features were extracted. Using LDA, an accuracy of 91.5, 91 and 99% was obtained for 2,4 and pairwise model and QDA an accuracy of 92.4, 100 and 99.6% was obtained for the same model. Tan et al. [7] took nearly 40 samples of apple to identify the bruising degree. Random forest, Kennard-Stone algorithm and standard-normal variate was used in feature extraction process after which SNV-SPA with SVM based on GS-SVM, was used in identification process. An accuracy of 95% was obtained in identifying the bruising degree on apple. Kuang et al. [8] created 1778 images dataset and 500 object proposals. Algorithms like HOG, LBP and Gabor LBP were applied on the multi-class drupes. Using feature fusion method, a 98.5% accuracy was obtained. Beyaz et al. [9] used image processing techniques to identify Spanish olives. Around 4400 Spanish olive images were taken. Applying Artificial Neural Network, it was noticed that 90% accuracy was obtained to identify Spanish olives. Philip and Anita [10] created the dataset of nine varieties of rice grains and extracted features based on spatial and frequency domain for the classification. A flat-bed scanning techniques applied to capture the images of rice grains. Classification done using NB and SMO classifier. An accuracy of 95.78% was obtained. Wajid et al. [11] used 335 images of orange drupes. RGB features were extracted. Different ripening stages were analyzed using decision tree classifier. A precision rate of 93.45%, sensitivity of 93.24% and accuracy of 93.13% was obtained from this model. Pariente et al. [12] used partial least square and artificial neural network to distinguish the quality of 4 types of olives. Around 190 images were captured to do the analysis. Mean value of each color channel was calculated for the classification.

Linear models and non-linear models were used to evaluate and got 70–75% and 93–95% of accuracy, respectively.

The survey led to a conclusion that five research scholars have applied SVM classifiers for various classification and identification purpose. An aggregate of 95% accuracy can be obtained using SVM classifier. Four scholars used LDA and QDA classifiers for different methods and obtained about 90–100% of accuracy. Some of the researchers claimed Random Forest is better than K-Nearest Neighbor and Support Vector Machine.

3 Method Description

The proposed work adopts image processing techniques for classifying apples based on its varieties. Varieties of apple are purchased from the local market to create a database. The images are captured under a set of significant parameters. All the images are captured and saved in JPEG format. Various image processing algorithms are applied on the dataset to extract features. The spatial and frequency domain features are extracted from the images. Machine learning algorithm such as Naive Bayes, Random Forest and MLP classifier is used to classify them.

An automated process always holds advantages over the manual process. The manual process of classifying is time consuming, and the intuitive judgments made can be wrong or varied from person-to-person.

3.1 Data Collection

The source of the dataset was collected from the regular markets in Bangalore, Karnataka. 7 varieties of apples were purchased. Each of these apples have various price per kilogram. The most commonly available apples were taken into consideration like: Apple Newzealand (ANZ), Apple Australian (AUST), Ooty Apple (AOO), Green Apple (AGN), Australian Small Apples (AUSTS), Apple USA (AUS) and Apple Indian (AIN).

Various vendors were approached in order to get different varieties of apples. Since some of the varieties are imported from other places it is important to consider them for creating the dataset. In order to differentiate between the apples, appropriate code names were created for every variety. Each of these categories contains quality apples to check if there is a correct classification between them and hence 100 images were captured in total from these apples. The code names are used such that it can differentiate between the apples during classification. The sample dataset is given in Table 1.

To capture these images a web camera with a tripod stand is required. Digital images are captured under the environmental conditions with a set of constraints.

Table 1 Sample dataset (variety and image sample)

ANZ	AUST	AUS	AIN

Images are captured under such a controlled environmental setup. The following measures are considered:

- Constant and noise-free background is maintained.
- Sustained and balanced light intensity is maintained around the object throughout the session. Constant light is given through a panel of LED lights.
- The distance between the camera and the object was persistent.
- The base height of the object from the ground was also constant throughout the procedure.
- Every sample is captured in three different angles.
- Same web camera is used throughout the session.

These parameters were considered such that all the images are captured under the same environmental conditions. External parameters are controlled significantly such that all images of apple have the same properties.

3.2 Feature Extraction

A feature is an important piece of information that can solve computational task for any application. Image processing has a refined collection of features. A total of 100 apple images are captured to create the database. The following methods are adopted to extract 13 features from the images which includes regional, boundary, statistical and frequency descriptors.

In the proposed method, the following features are extracted from the apple images: Convert the image to binary images and calculate the on pixel density and mean of the same. Images are converted to grayscale images to obtain mean, standard deviation entropy and compactness. FFT and DCT standard deviation of the Fourier transformed grayscale image and discrete cosine grayscale image is measured. Red color component is extracted from the apple images. Statistical descriptors such as mean, standard deviation and entropy were calculated.

Methods used for extracting features is mentioned in the below algorithm.

Input: Image

Output: Total number of 13 features

Method

- For each original images, the following features are extracted.

  I. Mean of R-channel.
  II. Standard Deviation of R-channel.
  III. Entropy of R-channel.

  A total of 3 features were extracted from the images.

  The original images are converted to grayscale and binary images in order to extract more features.

- For every grayscale Images, the following features are extracted.

  I. Mean of grayscale image.
  II. Standard deviation of grayscale image.
  III. FFT is applied on the grayscale image and the standard deviation is computed from the Fourier transformed image.
  IV. DCT is applied on the grayscale image, and the standard deviation is computed from the discrete cosine image.
  V. The first pixel value of the DCT of the grayscale image is measured.
  VI. Entropy of the grayscale image is calculated.
  VII. Compactness of the grayscale image is measured.

  A total of 7 features were extracted from the grayscale image.

- For each binary images, the following feature descriptors are extracted.

  I. The on pixel density of the binary image is calculated.
  II. Mean of the binary images are computed.

  A total of 3 features are extracted from the binary image.

MATLAB is used for creating the algorithms in order to extract features of spatial and frequency domain based. The extracted features are stored in an excel sheet and is normalized, and WEKA tool is used for classifying these 7 varieties of apples. WEKA is trained and tested against Naive Bayes, Random Forest and MLP, and persuading results were obtained. The procedure to perform the experiment is given in Fig. 1.

3.3 Classifier

The extracted features are experimented with different classifiers using WEKA tool. Naive Bayes Classifier, Random Forest and MLP gave encouraging result. Naive Bayes Classifier uses the concept of the popular theorem Bayes theorem. Naive Bayes classifier determines to find the probability of a sample that belongs to its associated class. The class with the highest probability is considered as the most probable class. MLP are efficiently used to solve non-linearly separable problems. MLP trains a part of the dataset and gathers information regarding the dependencies between the input and output. The training phase in MLP guides one to adjust the parameters in order to minimize error. Random forest uses the concept of decision tree. The main aim of random forest is to achieve low-prediction error.

The created database contains 100 images in total. 60 images and 40 images are considered for training and testing, respectively. The size of the database is less, so K-fold method is adopted for experimenting. K-fold value is set to 10. It is noticed that in the proposed model MLP gives a better accuracy when compared to Naive Bayes and Random Forest Classifiers. Table 2 gives the details of the experimental results.

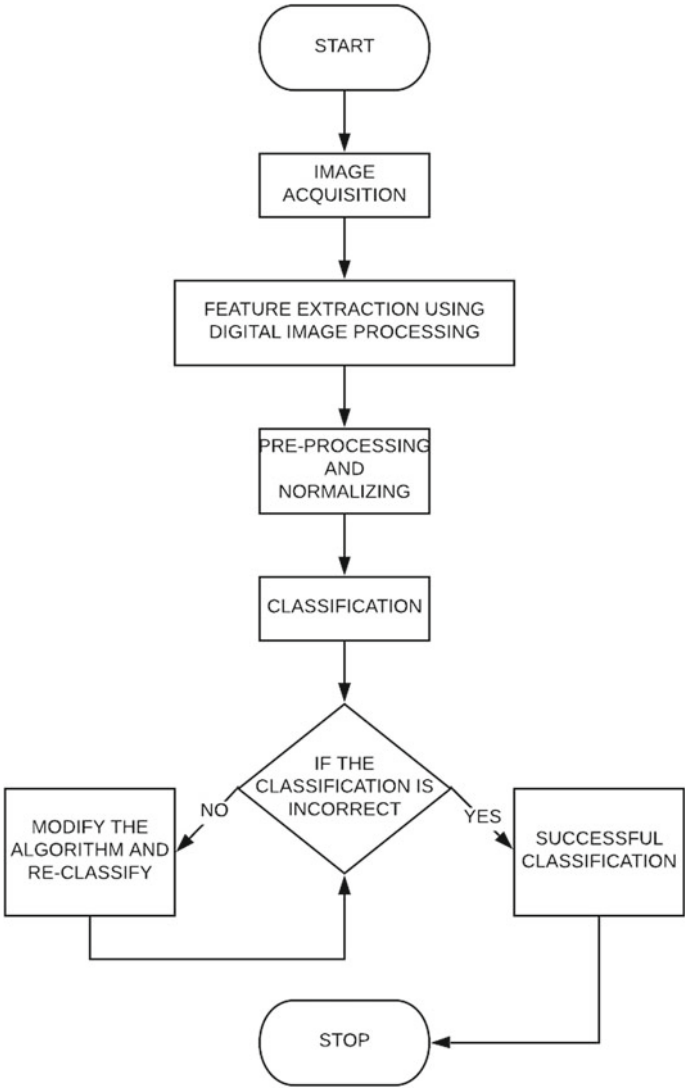


Fig. 1 System work flow

Table 2 Result

Naive Bayes (%)	MLP (%)	Random Forest (%)
65.62	94.79	75

4 Experimental Results

Total number of 100 images is captured and stored to create a database. Both spatial and frequency domain features are extracted and for each image 13 features are considered. The features are normalized and stored in an excel sheet. WEKA tool is used to classify the images. Multiple classifiers are evaluated to check the best classifier empirically. Table 2 shows the result analysis of various classifier. It is seen that MLP is able to classify varieties of apples with the highest accuracy of 94.79% when compared to Naive Bayes Classifier and Random Forest which has 65.62% and 75% of accuracy, respectively.

5 Conclusion

The anticipated work aimed at the grading of apples using image processing. In order to produce an efficient model of automation for the classification of apples, the proposed work extracted relevant features from the images. A total of 100 apple images were captured to create the database. The images were captured with a controlled environment setup. Various regional, boundary, statistical and frequency-based features are extracted. Feature extraction is done using many algorithms using MATLAB software. WEKA tool is used for classification. MLP gives the highest accuracy when compared to Naive Bayes and Random Forest Classifiers. In future, the dataset size can be increased to improve the training and classification accuracy. By approaching different vendors, a larger database can be created to accommodate various varieties of apple. This technique can also be applied to other drupes like bananas, Grapes, etc. This could be better informatics for the consumers.

References

1. Britto CJ, Anita HB (2019) Classification of radiolucency in dental X-ray image. Res J Pharm Technol 12(5):2171–2174
2. Oo LM, Aung NZ (2018) A simple and efficient method for automatic strawberry shape and size estimation and classification. Biosys Eng 170:96–107
3. Wang F, Zheng J, Tian X, Wang J, Niu L, Feng W (2018) An automatic sorting system for fresh white button mushrooms based on image processing. Comput Electron Agric 151:416–425
4. Tu S, Xue Y, Zheng C, Qi Y, Wan H, Mao L (2018) Detection of passion fruits and maturity classification using red-green-blue depth images. Biosys Eng 175:156–167
5. Utai K, Nagle M, Hämmerle S, Spreer W, Mahayothee B, Müller J (2019) Mass estimation of mango fruits (Mangiferaindica L., cv.'NamDokmai') by linking image processing and artificial neural network. Eng Agric Environ Food 12(1):103–110
6. Hossain MS, Muhammad G, Amin SU (2018) Improving consumer satisfaction in smart cities using edge computing and caching: a case study of date fruits classification. Futur Gener Comput Syst 88:333–341

7. Tan W, Sun L, Yang F, Che W, Ye D, Zhang D, Zou B (2018) Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM. Optik 154:581–592
8. Kuang H, Liu C, Chan LLH, Yan H (2018) Multi-class fruit detection based on image region selection and improved object proposals. Neurocomputing 283:241–255
9. Beyaz A, Özkaya MT, İçen D (2017) Identification of some spanish olive cultivars using image processing techniques. Scientiahorticulturae 225:286–292
10. Philip TM, Anita HB (2017) Rice grain classification using Fourier transform and morphological features. Indian J Sci Technol 10(14):1–6
11. Wajid A, Singh NK, Junjun P, Mughal MA (2018) Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification. In: 2018 international conference on computing, mathematics and engineering technologies (iCoMET). IEEE, pp 1–4
12. Pariente ES, Cancilla JC, Wierzchos K, Torrecilla JS (2018) On-site images taken and processed to classify olives according to quality—the foundation of a high-grade olive oil. Postharvest Biol Technol 140:60–66

Technical Invoice Data Extraction System: State of the Art, Research Challenges and Countermeasures

Oshi Varma, Samarth Srivastava, and M. Gayathri

Abstract In the world of E-commerce, recording sales play a significant role. To achieve it, companies use invoices where each seller generates an invoice for each transaction. As tons of invoices are generated by different sellers in different formats, it becomes difficult for the E-commerce industries to track trade and transactions manually. With exponential increment in the invoices, E-commerce industries need a solution to replace the manual intervention and adapt a completely automated data extraction system. Invoice data extraction systems are widely used in the E-commerce industries to remove the need of middleman for manual data extraction and automate the information retrieval process. Conventionally, an optical character recognition is used to salvage data from the invoice document, followed by a regress text analysis, which is an expensive and impractical solution. With the advancement in this field, pattern matching became a popular methodology, but the absence of specified patterns restricted the accuracy. Usage of recurrent neural network and graph neural network tackles most of these issues conveniently. This paper aims to give a glance of the researches in this domain and categorize the literature survey based on the adapted methodology, dataset involved, year of publication, and precision in data extraction. Furthermore, we have discussed the limitations and improvement in the research. We have also stated some applications of the invoice data extraction system and drawn the inference from the available researches.

Keywords Invoice data extraction · Optical character recognition · Recurrent neural network · E-commerce

O. Varma (✉) · S. Srivastava · M. Gayathri
SRM Institute of Science and Technology, Chennai, India
e-mail: or1447@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_19

1 Introduction

The core of invoicing involves three broad steps which are customer and seller information, product information and currency amount involved. Invoicing is extremely vital for all businesses for their evaluation and growth. Therefore, correct and efficient management of invoices is critical. The best way to achieve this is to automate the process of invoice generation. There are many methods introduced in the industry to achieve consistency and reliability in the process of data extraction from different invoice structures. Figure 1 represents variable invoice structures.

Study in [1] suggests the usage of morphological features on the data extracted by OCR and then mapping the data on the features to retrieve structured information. However, it involved many issues like structure intolerance as it worked well with tabular invoices but not with non-tabular invoices. The next major study presented was about using recurrent neural network in [2]; in this, the data was extracted using OCR and then converted to N-grammar to form features; these features were then assigned values which signified the probability of them being a defined field. As this method only gave a probabilistic output, it was not reliable. Since identifying the position of named entity seemed to be the problem, so the next research in [3] indicated using template matching to extract information using OCR. Template matching worked by determining to which pre annotated invoice template does the given invoice match. It uses the position of the annotated tags in the pre defined template to extract data from the same position in the given invoice. This method required the user to maintain a large dictionary of invoice templates. It also used techniques like contour extraction and secondary rotation. Any identification issues during contouring or rotation can increase its complexity and make the entire process error prone. More recent advancements in this field involved the use of graph convolutional network and artificial intelligence in [4] and [5]. It uses tree data structure with the invoice being the root of the tree. It decomposes the invoice into nodes with

Fig. 1 Different structures of invoices

each node representing a word. The bunch of nodes forms graphs. These graphs given to the graph classifier which groups them together from left to right. This method too requires a large collection of datasets and is considered time consuming and costly.

We will comprehensively analyze these techniques subsequently to understand what these methods aimed to accomplish, what challenges they overcome and what their flaws were.

2 Invoice System

An invoice system is used to automatically generate structured invoice from bills or receipt of provided services or products. The suppliers and vendors upload a copy of the bill, which is fed to the system, which converts them into consistent structured format. The system can be extended to process wider range of invoices involving different currencies and multiple structures. It immensely improves the productivity and efficiency of the organization.

3 State of the Art

This section provides a comprehensive survey of the state of the art in the field of invoice data extraction. Table 1 provides a complete overview based on adapted methodology and problem. Traditionally, industrial invoice data extraction takes place manually which causes low efficiency and high error rate. With the advancement in the field of computer intelligence, manual intervention has reduced and efficiency has increased. E-commerce industry is widely implementing invoice data extraction solution. Tradeshift's solution is detailed by Palm et al. in [2]. It extract and validate invoices easily and quickly. Palm et al. [2] use recurrent neural network's long short-term memory (LSTM) and provide high accuracy of extraction. Andreas et al. briefed his implementation in [6]. It is a solution developed by DFKI spin-off INSIDERS and extends its application in the medical industry. Schuster et al. proposed implementation in [7] is used by a document management software company, DocuWare which uses rule-based system for data extraction. The available methodology used in invoice data extraction system is described in detail for you.

3.1 Natural Language Processing (NLP)

A subfield of artificial intelligence and linguistics, NLP, allows synergistic communication between humans and computer. With NLP, computers get the power to analyze and process large volume of natural language as data. Belaïd and Belaïd [1] use part

Table 1 Overview of the state of the art on invoice data extraction

References	Adapted methodology	Problems
[1]	Part of speech tagging	Under segmentation
[10]	Pattern matching and rule-based system	Works only for tabular structure
[2]	Recurrent neural network	Requires manual ordering of long short-term memory (LSTM)
[3]	Template matching	Time consuming and issues due to noise
[12]	Deep learning	Limited range of invoices can be processed
[4]	Graph convolution network and supervised learning	More concentration on table detection than data extraction
[5]	Graph convolution network and artificial intelligence	Issues due to noise
[13]	Case-based reasoning	Increased solving time due poor database indexing
[11]	OCR and FRESCO model (knowledge base)	No data extraction in absence of the keyword
[7]	Rule-based system	Need fixed weights
[8]	Computer vision	Image size affect results
[9]	Computer vision	Not suitable for handwritten invoices
[6]	Machine learning	Requires human intervention
[14]	Machine learning	Works only for tabular structure

of speech (PoS), a field of NLP which assumes that a part of natural language contains syntactic structure.

3.2 Computer Vision

It allows computer to develop high-level knowledge and understanding from images and videos. Computer vision involves acquiring of image followed by processing, analyzing, transforming and extracting data from it. Zhang et al. [8] use hough transform and [9] use canny edge detection algorithm for line and edge detection in image. Further, segmentation allows extracting information.

3.3 Graph Convolution Network

Graph Convolution Network is also referred as graph neural network. It extends the usage of convolution on graphs. It allows filtration and classification of graph nodes and edges. Graph neural network provides an additional feature of retaining information about the neighborhood with some arbitrary depth. In [4] and [5], nodes contain words, and edges represent relationship between the nodes. Repetitive structure of invoice forms graphlets or subgraphs because of their similarities.

3.4 Recurrent Neural Network (RNN)

A subfield of artificial neural networks forms a directed graph to represent node connections. It offers usage of variable length of input as it processes internal state. In [2], long short-term memory is used for invoice data extraction. Classification is followed by IOB labeling to output final extraction.

3.5 Pattern Matching

A subfield of rule-based expert system allows automatic recognition of repetitive structures or patterns in data. With this algorithm, the system looks in the input for an exact pattern defined in model. Kosiba and Kasturi [10] and Bayer and Mogg-Schneider [11] work with pattern matching to classify documents as seen or unseen. For the seen documents, patterns are matched to obtain data from invoices in minimal time. Table 1 shows a comparative study of the mentioned methodologies in the domain of invoice data extraction. An intelligent invoice system can use any of the methodology followed by a verification step by users to rectify any error in data extraction.

4 Challenges

In this section, we will highlight the challenges present in the available methodology. Table 2 shows the disadvantages of the methodology in invoice data extraction. Further explanation of the challenges is listed below.

Table 2 Comparative study of the methodology used in invoice data extraction

S. No.	Methodology	Advantages	Disadvantages
1	Natural language processing	With NLP, you do not need any prior knowledge of invoice structure	With weak redundancy, OCR quality decreases. Also, as OCR is applied on entire text, it becomes an expensive process
2	Computer vision	Images are segmented, smoothened, and sharpened which results higher efficiency	Image size affects recognition rate
3	Graph convolution network	It makes the model language independent and preserves privacy of data	Complexity increases with constraints
4	Recurrent neural network	Automatic learning of new words	It does not consider image feature and it can have discrepancies between output document and input
5	Pattern matching	Identification of region of interest (ROI) requires minimal time	Errors in case of no prior knowledge of a pattern

4.1 Under Segmentation

When natural language processing is used in invoice data extraction system, PoS tagging is used, and words are segmented to get a clear meaning of it. Belaïd and Belaïd [1] use OCR to read the text from the image and then assign a label to it. But, with a noisy image, OCR can read erroneous data and lead to incorrect labeling and segmentation.

4.2 Tabular Structure Dependence

Kosiba and Kasturi [10] and Ming et al. [14] are the systems dependent on the tabular structure in invoices. As there is no standardized invoice structure available, these systems will lack the flexibility feature. Figure 1 depicts variance in the invoice structure. Kosiba and Kasturi [10] and Ming et al. [14] will fail in the absence of the tabular structure and will provide incorrect output. Also, in case of unwanted lines, the image can lead to incorrect detection of the structure.

4.3 Noisy Images

Noise refers to presence of disturbance in the image that affects the image quality by hiding the region of interest. Invoices received by the industries are images or scanned PDFs. These images often contain noise such as tilts, stamp marks, rotations and unwanted lines. In [1, 2, 13], and [7], OCR is used to extract data without any corrections in the images and PDFs. It results into incorrect extraction or recognition of characters.

5 Countermeasures

Automatic invoice data extraction is tempting for industries as it reduces manual efforts and reduces cost involved. However, the system needs to offer high accuracy with minimal human intervention. Palm et al. [2] and Schuster et al. [7] allow user to check the identified information before forming the final output. Most of the systems suffer incorrect recognition because of the lack of smoothing, tilt correction, rotation, etc. and result lower accuracy. Some countermeasures for the stated problems are as follows.

5.1 Image Correction

With the implementation of hough transform and canny edge detection, you can solve incorrect recognition by OCR as they identify the invoice region correctly. Usage of contour detection allows easy shape detection and analysis. All the images must undergo rotation and tilt correct before proceeding with data extraction. Conversion of image to a gray scale and using color channel allows you to deal with stamp removal.

5.2 Structure Independent

Segmenting your image on the basic of the tabular structure will reduce the flexibility of the system. Implementing a solution in NLP or graph convolution network will allow data extraction for dynamic invoice structure.

6 Applications

Invoice systems can be used in various fields. Its variant applicability makes it an important choice in the E-commerce, medical, transportation, manufacturing, etc. industry. Some of its applications are listed for you.

6.1 Consistent Invoice Generation

The invoice system is used to generate exactly same invoice each time from the input bill or receipt, which plays a critical role in maintaining proper track of the sales.

6.2 Predefined Templates

The system can be used to generate invoices in any predefined template seamlessly. The required template is fed in the system, and it converts any given bill to that format; this makes the system extensible and very easy to use.

6.3 Multiple Currency Setting

The system can also be used to process invoices with different currencies without any functionality change. The system can be set to work for a specific currency or it can also detect the currency form the input.

7 Inference from the Survey

The carried out work in this field discusses about the invoice data extraction system and its application in detail. We have also looked at the state of the art in this domain and further explained the challenges with respect to the available solutions. A broader glimpse of the challenges in the available solution was followed by the counter-measures for it. Invoice data extraction system is a widely used solution of the E-commerce industries, who are dealing with 10,000 invoices each day. In such cases, accuracy and time consumption take the front seat. Table 3 has discussed the accuracy of all the available methodology to get you a clear picture of the best available solution. It was observed that NLP and pattern matching offer high accuracy but involve vast usage of OCR, leading to privacy issues. Computer vision allows

Table 3 Accuracy of the adapted methodology

References	Methodology	Accuracy >90%
[1]	Natural language processing	Yes
[3, 10]	Pattern matching	Yes
[2]	Recurrent neural network	No
[4, 5]	Graph convolution network	No
[8, 9]	Computer vision	Yes

preprocessing of the image before extracting information with highest accuracy and minimal time, but its recognition rate is affected by image size.

8 Conclusions

The above study identifies the potential in invoice data extraction system. It can be further extended in data extraction from documents, web pages, medical bills, etc. Its interoperable and extensive architecture helps to extend the usage of these system in multiple domains. Invoice systems particularly in E-commerce will be the next step forward. Equipped with cutting edge deep learning tools, the invoice systems will allow extreme flexibility and accuracy in processing invoices. This study shows systematically how advancements in this field were made incorporating and developing various technologies and methodologies. The proposed solution studies the previous research and tackles the drawbacks of previous solutions providing better results.

References

1. Belaïd Y, Belaïd A (2004) Morphological tagging approach in document analysis of invoices. In: 17th international conference on pattern recognition, vol 1
2. Palm RB, Winther O, Laws F (2017) CloudScan—a configuration-free invoice analysis system using recurrent neural networks. In: International conference on document analysis and recognition (ICDAR), vol 1
3. Sun Y, Mao X, Hong S, Xu W, Gui G (2019) Template matching-based method for intelligent invoice information identification. In: IEEE access, vol 7
4. Riba P, Dutta A, Goldmann L, Fornes A, Ramos O, Llados J (2020) Table detection in invoice documents by graph neural networks. In: International conference on document analysis and recognition (ICDAR)
5. Blanchard J, Belaïd Y, Belaïd A (2019) Automatic generation of a custom corpora for invoice analysis and recognition. In: International conference on document analysis and recognition workshops (ICDARW), vol 7
6. Andreas R, Dengel, Klein B (2002) smartFIX: a requirements-driven system for document analysis and understanding. In: 5th international workshop on document analysis systems, vol 5, pp 433–444

7. Schuster D, Muthmann K, Esser D, Schill A, Berger M, Weidling C, Aliyev K, Hofmeier A (2013) Intellix—end-user trained information extraction for document archiving. In: 12th international conference on document analysis and recognition
8. Zhang J, Ren F, Ni H, Zhang Z, Wang K (2019) Research on information recognition of VAT invoice based on computer vision. In: 6th international conference on cloud computing and intelligence systems (CCIS)
9. Sidhwa H, Kulshrestha S, Malhotra S, Virmani S (2018) Text extraction from bills and invoices. In: International conference on advances in computing, communication control and networking (ICACCCN)
10. Kosiba DA, Kasturi R (1996) Automatic invoice interpretation: invoice structure analysis. In: 13th international conference on pattern recognition, vol 3
11. Bayer TA, Mogg-Schneider HU (1997) A generic system for processing invoices. In: Fourth international conference on document analysis and recognition, vol 1
12. Sun Y, Zhang J, Meng Y, Yang J, Gui G (2019) Smart phone based intelligent invoice classification method using deep learning. IEEE access, vol 7
13. Hamza H, Belaïd Y, Belaïd A (2007) A case-based reasoning approach for invoice structure extraction. In: Ninth international conference on document analysis and recognition (ICDAR), vol 1
14. Ming D, Liu J, Tian J (2002) The design and implementation of a chinese financial invoice recognition system. In: International symposium on VIPromCom video/image processing and multimedia communications

Spectrum Sensing and Radio Resource Allocation in Cognitive Radio Network System

C. Rajesh Babu, Shruti Garg, and Upasana Chakraborty

Abstract The rapid advancement of radio communication technology, as well as the growing use of IoT devices, has prompted research into the fifth generation (5G) terrestrial telecommunications. This poses a range of difficulties; as a result, we suggest cognitive radio technology, which shares the permitted spectrum of primary users (PUs) with a variety of unlicensed users, also known as secondary consumers, in an opportunistic manner. This continues to alleviate the underutilization of bandwidth and the lack of cellular applications. This boosts spectrum use as well as major bandwidth growth and multiplexing gains by remote transmission and distribution. The authors of this paper propose an energy-efficient resource allocation method for Long Term Evolution (LTE) systems that considers both energy efficiency and the quality of service (QoS) to users. The proposed model uses traffic-based resource and energy allocation algorithm to incorporate orthogonal frequency division multiple access (OFDMA) technology in conjunction with MIMO (multiple-input multiple-output) in a network with radio access. Our suggested algorithm significantly increases energy efficiency, while still meeting users' requirements in heavy data traffic, according to simulation results. The primary idea is to focus on relay architectures of LTE-A (multiple and hop based relay transmission). Pros and cons of developed transmission schemes together with relay selection pairing scheme are also reviewed. Experimental results verify that OFDM technology works best when paired with MIMO technology.

Keywords Cognitive radio network · Spectrum utilization · Orthogonal frequency division multiple access

C. Rajesh Babu · S. Garg · U. Chakraborty (✉)
CSE Department, SRM Institute of Science and Technology, Chennai, India
e-mail: ub1223@srmist.edu.in

C. Rajesh Babu
e-mail: rajeshbc@srmist.edu.in

S. Garg
e-mail: sd9468@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_20

1 Introduction

Spectrum is the frequency range between 3 kHz and 300 GHz that can be used for radio transmission. Spectrum management is the process of regulating how this finite bandwidth is used in order to maximize performance. Due to the exponential increase in demand caused by the proliferation of IoT or "smart" devices, spectrum management has become a requirement [1]. Cognitive radio (CR) is a technology that uses several accessible channels in the spectrum band. [2] suggests cooperative communications, in which many relays assist the source to the destination, increases CR efficiency and spectrum usage even more. The channels in a spectrum band which are not being used or in an idle state can be accessed dynamically and this process of using the available unused spectrum band is known as dynamic spectrum access (DSA). This helps to reduce the number of unused spectrum bands. Effective spectrum sharing is a very vital requirement to guarantee CR users can make complete use of free resources without creating any interference for other customers. However, a reduction can be seen in detection performance due to reasons like—(I) multipath fading, (ii) shadowing and (iii) hidden primary users, which is why CSS makes use of spatial diversity to avoid these.

OFDM allows a big number of orthogonal, concurrent narrowband sub-channels to be transmitted in parallel by splitting the usable transmission bandwidth. The way OFDM handles multipath interference at the receiver end is its key selling point. The property of orthogonality allows a large number of sub-carriers to be transmitted simultaneously in a small frequency space without interfering with each other. As a result, OFDM is increasingly becoming the modulation technique of choice for wireless communication. OFDM utilizes resource units to allocate resources to different channels in the signal. With OFDM, different transmit powers may be applied to different RUs. With the aid of OFDM, it is possible to achieve adequate robustness in order to provide high data rates to radio channel impairments [3]. The way OFDM handles multipath interference at the receiver end is its key selling point. The BW waste generated by the cyclic prefix is reduced using wavelet-based OFDM. The transmission power is decreased due to the use of wavelet transform. OFDM in combination with MIMO technology offers the most favorable outcomes [4]. Multiple-input and multiple-output systems involve placing antennas at both the transmitter and receiver end. MIMO plus OFDM is a popular pairing often employed in WLAN, WMAN and 4G cellular standards.

2 Background Work

According to [5, 6], dynamic spectrum management (DSM), often called dynamic spectrum access (DSA), has its foundation in different theories such as game theory and network information theory. Traditionally, DSM is hugely dependent on either correct information about the network or on assumptions that we make, however, this

can become difficult and tedious as the user pool grows. Another potential method of implementation of DSM is through cognitive radio. It is a form of wire-free communication that has emerged as an effective solution to spectrum utilization issues where the system can rationally identify which communication channels are occupied and which are not.

In [7], every primary customer contributes a part of radio capital and guarantees a certain level of service quality (QoS). The suggested method is a two-phase arrangement of resource allocation. It solves the resource allocation problem in complex systems using Auction and Stackelberg Game in Spectrum Sharing (TAGS). In the first step, the spectrum allocation is calculated using a spatially restricted version of an auction that do not take spectrum recall into account. A Stackelberg game is created at the end of the process so that all customers can make a decision.

Heterogeneous secondary users consider a spectrum auction scheme with numerous quality-of-service parameters and a primary base station in [8]. It suggests a recall-based spectrum auction (RSSA) algorithm with one winner which eventually extends to allow multiple winners for better utilization of the spectrum. Following the formulation of the auction model, the Vickrey–Clarke–Groves (VCG) approach is used in the payment function. The proposed algorithm increases spectrum efficiency, while maintaining QoS for secondary users, according to simulation results.

Classical algorithms are known to control power and beamforming in cognitive radio (CR); however, this may lead to the degradation of performance in a more practical system. In [9], the issue of power control as well as joint beamforming in underlay CR networks with non-ideal channel state is investigated. The main purpose is to reach the maximum value of spectrum utilization for secondary users (SUs) under the interference power constraints of primary users (PUs). Conclusively, it is shown through simulation that the suggested scheme can be optimized locally and converges to a pair of beamforming vectors and transmission power levels in the presence of channel uncertainty.

The ways in which power can be allocated for the mechanism of sharing of spectrum for multi-band cognitive radio networks is discussed in [10]. The primary users allow secondary users to transmit at the same time as their own signal. It employs a conventional ascending clock auction-based power allocation scheme. According to the announcement, secondary users repeatedly apply their optimum power demand to the primary consumer.

[11] consider show secondary users efficiently employ D2D communications in a network of the cognitive radio environment. In this network, the transmissions by the primary users are done through a base station (BS mode), while secondary users (SUs) may use a variety of transmission modes. A base station (BS mode) transmission is the first option and D2D connectivity is the second option (D2D mode). Inside the party, the SUs will communicate via BS or via D2D directly. Other than this machine, only BS mode is available. Based on this model, it is proposed to create a distributed protocol for SUs to converge automatically into ESS within a D2D community. The arithmetical outcomes demonstrate that the anticipated protocol is both effective and reliable.

3 Proposed Algorithm

A cognitive radio is a smart radio with complex programming and configuration capabilities. Its transmitter is set to be using the region's current best wireless networks. This form of radio detects usable channels in the wireless spectrum and changes its propagation or reception parameters to allow more data transmission in a given spectrum band to take place at the same time in a single location.

3.1 A. Genetic Algorithm

We suggest that the channel allocation problem be solved using a genetic algorithm. The genetic algorithm simulates the biological process of evolution and can be used to solve both constrained and unconstrained optimization problems. It replicates Charles Darwin's theory of "Survival of the Fittest." The genetic algorithm is an optimization algorithm for finding a close solution to an objective function without recognizing any gradient-related details or the objective function's derivatives. The significant idea behind the genetic algorithm is to first choose a set of practicable judgment variable values, then design real ideas that improve the goal function based on previous collection. In reality, the understanding of channel access is defined by a chromosome. This could look like a chromosome with 2 PU channels and 4 SU-TXs:

$$\text{Chromosome} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The search space is defined as Φ, including every likely channel allocation, and has the maximum dimension of $(N + 1)^K$. The process of implementing GA is explained further. At first, ϑ chromosomes are chosen at random from called $R^{(i)}(W, \varphi)$, the initial generation, $^{(0)}$.

At first, two chromosomes, P1 and P2, are selected randomly from the two groups $G^{(g)}_{best}$ and $G^{(g)}_{luckies}$ with P1 and P2 being called the parents.

Following this, the crossover operation is started between the two chosen parent chromosomes and starts generating two new children, namely TempCH1 and TempCH2. In the end, two randomly selected entries x^{n1}_k,x^{n2}_k, swapped on an arbitrary basis, this process is called mutation. It is important that during the selection process, either of the entries should have a value that one. To prevent GA from converging into a value that yields local optima, mutation is very essential.

In this paper, we show how the system reacts in different scenarios of SISO, MISO, SIMO and MIMO and what is the spectrum strength using cognitive radio networks by the application of a genetic algorithm.

3.2 B. System Model

The system model for this experiment is described as follows:

- The parameter-based system has multiple base station (3), 6 destinations, and 1 relay path.
- Every channel is processed on the random variable of signals (channels) [rand (3, 64)].

Each channel route must be examined using a chromosome set loop—
3 Base Stations to 1 Relay:
The matrix is defined to display the base stations to relay

Station 1 to relay: [1 0 0]

Station 2 to relay: [0 1 0]

Station 3 to relay: [0 0 1].

The next phase is to choose the best path depending on the corresponding; the worst path using a descending order condition. Following the discovery of the route, we must transmit the highest possible throughput rate using the crossover condition:

P1 ← select ($G^{(g)}_{best}$, 1, "Random")

P2 ← select ($G^{(g)}_{luckies}$, 1, "Random")

[Temp CH1, Temp CH2] ← Crossover (P1, P2)

[CH1, CH2] ← Mutation (TempCH1, TempCH2)

Optimal channel allocation analysis on the relay path to the destination is implemented.

4 Results and Discussion

Figure 1 shows different trends in the increase in capacity of SISO, MISO, SIMO and MIMO with a higher SNR. These are different variants of antennae propagation algorithms and hence have different capacities and spectrum strength for a certain value of the signal-to-noise ratio which portrays the performance of each scheme. These algorithms are compared and the results are calculated in the form of spectrum strength from the graph of signal-to-noise ratio (in dB) versus capacity (in Hz) plotted through simulation.

The overall spectrum strength is calculated for different values of capacity-SNR pairs as shown in Tables 1, 2, 3 and 4. Table 4 verifies that multiple input and multiple output (MIMO) is the best way to the implement OFDM in a cognitive radio network.

In reality, there is no such thing as a noiseless channel; it is always noisy. To calculate the theoretical highest data rate for a noisy channel, Shannon capacity is used:

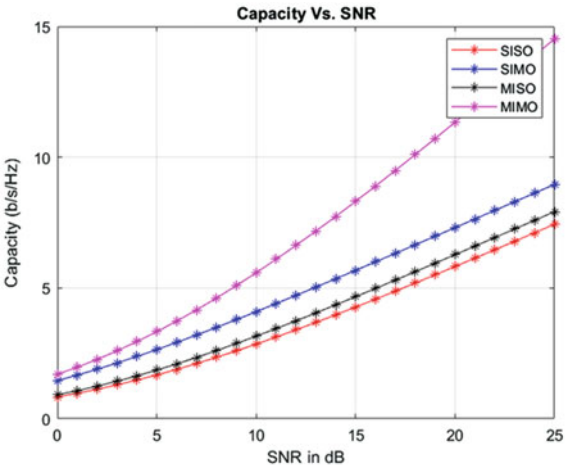


Fig. 1 Capacity (in Hz) versus SNR (in dB)

Table 1 Increase in capacity with increase in SNR for SISO

SNR (in dB)	Capacity (in Hz)	Spectrum strength
0	0.826022	0
5	1.66259	8.31295
10	2.8395	28.395
15	4.25534	63.8301
20	5.80467	116.0934
26	7.41841	185.46025

Table 2 Increase in capacity with increase in SNR for MISO

SNR (in dB)	Capacity (in Hz)	Spectrum strength
0	0.91239	0
5	1.85768	9.2884
10	3.15739	31.5739
15	4.6687	70.0305
20	6.2755	125.51
25	7.91831	197.95775

Table 3 Increase in capacity with increase in SNR for SIMO

SNR (in dB)	Capacity (in Hz)	Spectrum strength
0	1.44632	0
5	2.62235	13.11175
10	4.06676	40.6676
15	5.64466	84.669
20	7.27648	145.5296
25	8.92773	223.19325

Table 4 Rapid increase in capacity in case of MIMO

SNR (in dB)	Capacity (in Hz)	Spectrum strength
0	1.66637	0
5	3.27827	16.39135
10	5.51266	55.1266
15	8.22081	123.31215
20	11.2305	224.61
25	14.4056	360.14

$$\text{Capacity} = \text{bandwidth} * \log2(1 + \text{SNR}) \tag{1}$$

Bandwidth is defined as the channel's capacity in bits per second, SNR is the signal-to-noise ratio, and capacity is the channel's capacity in bits per second in the above equation. It is impossible to change bandwidth since it is a fixed resource. As a consequence, channel capacity is proportional to signal power in the following way:

$$\text{SNR} = ((\text{Powerofsignal})/(\text{powerofnoise})) \tag{2}$$

The signal-to-noise ratio (S/N) is generally expressed in decibels (dB), as follows:

$$10 * \log10(\text{S/N}) \tag{3}$$

The graph between outage probability and signal to noise ratio is shown in Fig. 2. The probability that the information rate falls below the required threshold information rate is known as outage probability.

In Fig. 3, the same information parameters are plotted just changing the level of probabilities of secondary users. In Fig. 2, the letters L = 1, 2, 3, 4 denote the

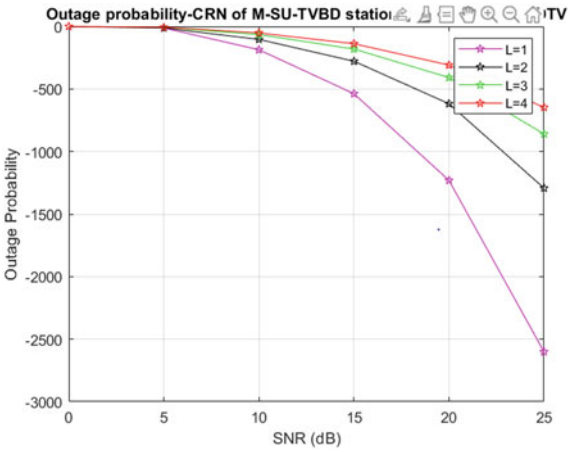


Fig. 2 Outage probability versus SNR (in dB) in 4 different levels of secondary user

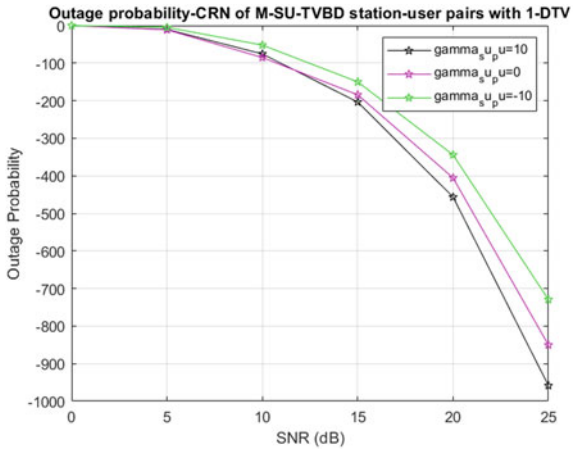


Fig. 3 Outage probability versus SNR (in dB)

levels. By holding all parameters constant but changing probabilities, the outage likelihood is determined with respect to SNR in dB. Gamma values are set to −10, 0, 10, respectively, and each is plotted accordingly. It is thus evident that MIMO technology provides the most favorable outcomes when coupled with the OFDM system.

5 Limitations

This section discusses the limitations of the proposed system:

- Due to issues such as channel loss, receiver confusion for future decision making, and so on, sensing data quickly becomes obsolete.
- Requires prior information of the primary user in certain operations such as matched filtering, this is difficult to achieve especially in tactical environments.
- The interference of the channel is high because of the nature of the wireless medium especially for information generated from resources with high sensitivity.

6 Conclusion

In this paper, we have successfully verified that MIMO (multiple-inputs, multiple-outputs) technology provides the best signal strength when paired with OFDM (orthogonal frequency division multiplexing). This algorithm helps propagate the signal to large distances with minimal noise. We have utilized genetic algorithm to find the best solution. The paper also discusses the limitations of this method which provides scope for future research works. These limitations can be countered through some research in the areas of machine learning algorithms such as semi-supervised

learning to perform dynamic spectrum management. Semi-supervised learning eliminates the need for a large amount of prior information. Machine learning algorithms help make intelligent decisions and thus yield better results. Thus, there is great future scope for research works in this area.

References

1. Axell E, Leus G, Larsson EG (2010) Overview of spectrum sensing for cognitive radio. In: 2010 2nd international workshop on cognitive information processing, Elba, Italy, pp 322–327. https://doi.org/10.1109/CIP.2010.5604136
2. Nandakumar S et al (2019) Efficient spectrum management techniques for cognitive radio networks for proximity service. IEEE Access 7:43795–43805. https://doi.org/10.1109/ACCESS.2019.2906469
3. Jadav NK (2018) A survey on OFDM interference challenge to improve its BER. In: 2018 second international conference on electronics, communication and aerospace technology (ICECA). Coimbatore, India, pp 1052–1058. https://doi.org/10.1109/ICECA.2018.8474748
4. Patil P, Patil MR, Itraj S, Bomble UL (2017) A review on MIMO OFDM technology basics and more. In: 2017 international conference on current trends in computer, electrical, electronics and communication (CTCEEC). Mysore, India, pp 119–124. https://doi.org/10.1109/CTCEEC.2017.8455114
5. Yi C, Cai J (2015) Multi-item spectrum auction for recall-based cognitive radio networks with multiple heterogeneous secondary users. IEEE Trans Veh Technol 64:781–792.
6. Zhao F, Li B, Chen H, Lv X (2013) Joint beamforming and power allocation for cognitive MIMO systems under imperfect CSI based on game theory. Wireless Pers Commun 73:679–694. https://doi.org/10.1007/978-3-642-31869-6_14
7. Wu Q, Zou J, Zhu K (2013) Power allocation in primary user-assisted multi-channel cognitive radio networks. Commun Netw 5:238–244. https://doi.org/10.4236/cn.2013.53B2044
8. Cheng P, Deng L, Yu H, Xu Y, Wang H (2012) Resource allocation for cognitive networks with D2D communication: an evolutionary approach. IEEE Wireless Commun Netw Conf WCNC. 2671-2676. https://doi.org/10.1109/WCNC.2012.6214252
9. Zhao F, Li B, Chen H, Lv X (2013) Joint beamforming and power allocation for cognitive MIMO systems under imperfect CSI based on game theory. Wireless Pers Commun 73:679–694. https://doi.org/10.1007/978-3-642-31869-6_14
10. Wu Q, Zou J, Zhu K (2013) Power allocation in primary user-assisted multi-channel cognitive radio networks. Commun Netw 5:23–8244. https://doi.org/10.4236/cn.2013.53B2044
11. Cheng P, Deng L, Yu H, Xu Y, Wang H (2012) Resource allocation for cognitive networks with D2D communication: an evolutionary approach. In: IEEE wireless communications and networking conference, WCNC, pp 267–12676. https://doi.org/10.1109/WCNC.2012.6214252

Analysis of Energy Efficiency in Hierarchically Structured Homogeneous Routing Protocols for Green WSN

Akanksha Singh and Anand A. Bhaskar

Abstract The world of IoT has witness huge expansion in the past few decades, which has brought our focus on the energy conservation aspect of wireless sensor networks. WSN constitutes of small sensing units which are responsible to sense the event, compute its sensed data and communicate it to the receiving base station for further processing and decision making. Sensors are extensively used for various applications ranging from environmental monitoring to health care and to surveillance. Green WSN focuses to optimize the amount of energy used by these tiny sensors. These sensors sustain in the network for long period of time on limited power supply of battery. Therefore, managing energy decay of sensors for increasing network lifespan is an underlying matter of concern in these sensor networks. Efficient routing is a prominent technique that targets on developing a Green WSN by reducing energy used in communication. Furthermore, routing protocols which are based on network structure are majorly categorized into flat, hierarchical and location-oriented routing protocols, among which hierarchically structured homogeneous routing protocols are best-suited solution for reducing energy utilization of WSN, as it avoids unnecessary and repetitive data transmission. In this paper, we have analyzed and created a Green WSN using Leach, Heed, Pegasis and tree-based clustering protocols that deals in energy efficiency and simulation is done in MATLAB 2019b. The results exhibit that TBC protocol surpasses Leach, Heed and Pegasis protocol with respect to the number of packets addressed to the receiving base station, average residual energy and the round through which the first node dies, hereby increasing the lifespan of the network.

Keywords Green wireless sensor network · Clustering · Routing protocols · Energy efficiency · Network lifespan

A. Singh (✉) · A. A. Bhaskar
Sir Padampat Singhania University, Bhatewar, Rajasthan, India
e-mail: akanksha.singh17@spsu.ac.in

A. A. Bhaskar
e-mail: anand.bhaskar@spsu.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_21

1 Introduction

Owing to the rapid growth in the IoT, the popularity of wireless sensor network has expanded in all dimensions. Wireless sensor network is made up of small tiny devices called sensors, which are responsible to sense, process and transfer the sensed set of data values to the adjoining base station for further and advance computations. A base station (BS) is a transceiver that is connected to a number of other nodes or to a wider area [1–3].

These sensors are used to monitor and analyze a wide variety of events ranging from environmental parameters such as temperature, humidity, seismic activities to healthcare, wildlife habitat and surveillance. Sensors are mostly used where human intervention may not be a feasible solution [4]. Sensors operate on a finite energy support, and since they are established in unfavorable environmental conditions, their battery cannot be changed or recharged after deployment [5, 6]. Also, the major part of sensor's energy is used in communication, so if the node dies, the WSN has to be created again, which would turn into a costly affair [7].

Today, in the era of information technology, our main focus is energy conservation and exploring renewable sources of energy which primarily deals in 3R's—Reduce, Renew and Recycle.

Green wireless sensor network aims to reduce this energy consumption factor of communication. Routing of data is one of the suitable mechanisms to preserve the energy of these tiny devices and thereby improving the life span of the network [8]. The characteristics of Green WSN are—periodic data reporting, no time stamp synchronization, and most important of them all is that sensors can be customized as per the requirement of the site [9]. Routing protocols which are classified on the basis of the structure aspect of the network are mainly classified as flat, hierarchical and location-oriented protocol, among which hierarchically structured routing protocol with homogenous configuration of nodes is considered to be the best suited for creating green wireless sensor network (GWSN). Here, clustering gets a larger credit in delaying the time for first node dead in the network, and aggregation process on the other hand helps in delaying the last node dead [10, 11].

Leach, Heed, Pegasis and TBC are some of the homogeneous hierarchical clustering methods that deal in energy efficiency.

Low energy adaptive clustering hierarchical routing (Leach) is widely studied as an energy-efficient protocol. It creates clusters with nodes who are in close vicinity and every cluster elects their cluster head (CH) through a random function and to whom every participating nodes sends their sensed set of data. All the CH then communicates the data to the BS. In every round, a new CH is elected. In Leach although load balancing is achieved, minimum energy of the CH is maintained as well as collision in transmission is prevented, but Leach has few disadvantages also, like unsystematic selection of CH, asymmetric distribution of clusters and one hop communication between CH and BS, due to which its network lifetime is low [12–14].

Hybrid energy-efficient distributed routing (Heed) elects the cluster head who has the highest residual energy and requires the minimum distance for communication.

Here, Intra-cluster communication cost, which is defined as the energy consumed by the sensor to sense and communicate its data to CH within the cluster, is calculated. The Intra-cluster communication cost is a function of cluster properties, that is, cluster size, and whether or not variable power levels are permissible for such communication. This again leads to overload in CH selection because of which the network dies sooner [15].

Power-efficient gathering in sensor information system (Pegasis) is another homogeneous routing protocol which creates chain-based network structure. Here, nodes will sense the data and transmit it to the closest neighboring node. The consecutive node takes up the data, aggregates with its own sensed data and forwards it to the next closest node forming a chain structure. Finally, the last node in the chain transfers the set of sensed data to the BS. Although Pegasis very efficiently creates a balance of energy utilization, between CH election and sensed data transmission but since there is only one node acting as head, it may lead to network congestion and delay in data processing [16].

Tree-based clustering routing protocol (TBC) is used to solve the energy utilization issue for the CH's which are situated at a larger distance from the BS. Through tree-based clustering (TBC), inter-cluster distance is calculated and a shortest path routing table is created among cluster head as parent and child node, and base station as a root node, where their distance is calculated for every round and the table is updated [17, 18]. Transmitting sensed data through multiple hops between CH to finally BS is an energy-efficient method for establishing green WSN. The CH instead of using direct communication method for data transmission to the BS chooses to send the sensed set of data to the nearby CH [19, 20].

In our paper, we have analyzed and simulated Leach, Heed, Pegasis and TBC routing protocol using MATLAB 2019b. The performance of the network is evaluated based on network throughput, average residual energy and network lifetime parameters.

The simulation results show that to construct a green wireless sensor network, tree-based clustering hierarchical routing protocol outperforms the remaining three protocols with regards to energy efficiency and extending the time period of a network and can be further studied to evaluate intra-cluster routing using TBC as well as to calculate the optimal level of tree within a cluster.

2 Experimental Section

2.1 Hierarchical-Based Homogeneous Routing Protocols

2.1.1 Performance of a Network Depends up on the Subsequent Factors

- Network Throughput—It is the average data rate of all positively acknowledged set of data transmitted over the particular communication channel. It is measured in bps (bits per second).
- Network lifetime—It is the time when the first node of the network gets completely exhausted and dies.
- Average residual energy of the network—It is the optimal amount of energy of all alive nodes after completing all rounds [21, 22].

2.1.2 Network Model and Radio Communication Energy Model

It is supported by the subsequent assumptions:

- Sensor nodes in a square area A (a ∗ a) are randomly distributed.
- At fixed location within this region (A), is a unique base station (BS).
- Nodes are stationary after deployment. All nodes have a same capability, equal standing and same amount of energy.
- Transmission power is manageable and controlled so that nodes can adjust with reference to the transmission distance.
- The energy model follows multipath fading model.

2.1.3 Radio Communication Energy Model

Total energy consumed can be calculated by the following formula:

$$E_{total} = E_{send} + E_{receive}$$

i.e., when sending k-bit of data to a dist—distance, the nodes will consume the energy as (Fig. 1):

$$\text{ETX (kbit, dist)} = \text{ETX} - \text{elect(kbit)} + \text{ETX} - \text{Eamp (kbit, dist)}$$

$$\text{ERX (kbit, dist)} = \text{ERX} - \text{elect(kbit)}$$

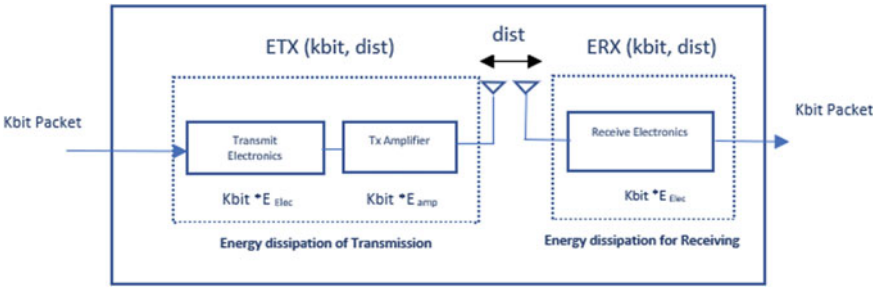


Fig. 1 Radio energy consumption model for WSN

2.1.4 Network Simulation Tool-MATLAB 2019b

See Table 1.

Table 1 Simulation set-up for Leach, Heed, Pegasis and TBC protocol

Type	Parameter	Value
Topology of node	Size of network	300 m × 300 m
	Number of nodes	100
	Location of base station	100,75
	Node distribution	Random
	Channel	Wireless
	Type of network	Fixed
Setting of node	Initial energy (E0)	0.5 J
	Sleep and idle energy	~0 J (ignored)
Radio model	ETX = ERX	50 nano Joules per bit
	Efs	10 Pico Joules per bit sq. mt
	Emp	0.0013 Pico Joules/bit/m4
	EDA	5 nano Joules per bit
	Packet header size	25 bytes
	Size of data packet	4000 Bytes

2.2 Leach Protocol

Low energy adaptive clustering hierarchical routing (Leach) is a MAC protocol. The network is divided into group of nodes called as cluster. Every cluster has its own head node which is responsible to communicate with the base station. All the nodes in the network continuously sense the information and transmit it to its own CH. Leach routing protocol operates in two phases; one is set-up phase and second is the steady state phase. In the first phase, every cluster elects its head node (CH) based on a random number (RN) generated by all nodes, in the range of 0 and 1. If the random value is less than the preset threshold value, then that particular node is elected as CH. Clusters are formed, and time division multiple access scheduling is executed.

The threshold value is calculated as,

$$T\ (\text{node}) = \text{prob}/(1 - \text{prob}\ (r \bmod (1/\text{prob}))) \text{ if node} \in G,$$

If (RN < T (node)), current node is a new CH; else, current node is a normal node. prob = probability or chance to be a CH, r = number of the round, n = total number of participating nodes of the network, G = group of all those nodes who have not been a CH since (1/prob) round. In the second phase, the sensed information from all the remaining normal nodes is received and sent to the CH and from CH to the fixed BS using the time division multiple access scheduling algorithm. Hence, the BS receives only the aggregated information from all the CH thereby, reducing the energy utilization of the network. After a fixed interval, both the phases repeats themselves to elect a new CH. Leach protocol is simulated in a virtual environment to analyze the network lifespan, throughput and average residual energy of the network.

Advantages of Leach

This algorithm does not require information of the complete system and is totally distributed. Every node has good probability to become a cluster head. Fault tolerance is attained. All load is not on one cluster head, so its energy does not get drained. Minimum energy in all nodes is maintained. Load balancing is efficient, and collision is prevented.

Disadvantage of Leach

Fit for regions with smaller geographical area only. Since the cluster head selection is predicted by a random function, energy is not utilized for selection of cluster head that ends up as energy overhead.

2.3 Heed Protocol

Hybrid energy-efficient distributed clustering is another homogeneous cluster-based topology. Heed calculates intra-cluster communication cost. This cost determines the

feasibility to allow variable levels of power within a cluster for establishing communication and determine the size of cluster. If the level of power for each cluster is kept fixed, then intra-cluster communication cost will be dependent on either: n degree, if we want to evenly distribute the load of energy consumption on all the CH.

Heed Protocol Works in 3 Phases

In the initial or starting phase, at each T_n seconds, we initiate a clustering mechanism so as to select new CH where T_n is the summation of the time required to make a cluster and the time gap between any two consecutive cluster formations. In every round, all node set their probability to be a CH as CHprobability,

$$\text{CHprobability} = \text{Cprobability} * \text{Eres}/\text{Em},$$

here,

Cprobability = It is the percentage of CH for all n nodes,

Eres = Approximate residual energy for the current round in the node,

Em = Highest amount of energy of a node.

In the repetition phase, all the sensor nodes undergo many iterations in search of cluster head that may use minimum power for transmission. If the result is zero and finds no nodes satisfying this criterion, then the sensor declares itself as CH and announces it to all the nodes of the cluster.

In this situation, all the remaining nodes double their CH probability value and undergoes subsequent iteration process until its CH probability becomes 1.

In final stage, every sensor finalizes its status that whether it is a normal node within a cluster range or a cluster head. Heed protocol terminates after a fixed set of iteration, and it works irrespective of the network field area.

Advantages of Heed

Distributed clustering method where CH is uniform supports long range of communication and minimizes communication cost.

Disadvantages of Heed

It causes overhead in cluster head selection and expires shortly because of energy overload on CH.

2.4 Pegasis Protocol

Power-efficient gathering in sensor information system is one of the suitable protocols for data transmission in a small network. Pegasis is a homogeneous protocol, where every node communicates only with the nearby neighboring node and takes turn to

send the sensed data to the BS, thereby reducing the amount of energy utilized in every round. The main objective in routing with Pegasis is to create a chain-like structure where every node can send or receive the sensed data from its predecessor or successor node. All the node gathers the sensed data, assembles it and forwards it to the BS using a direct link. All the nodes get an equal chance to transmit the sensed data, thereby decreasing the factor of energy utilization per round. If the node gets exhausted and dies, in that case, a new chain will be create, hereby handling the fault tolerance issue at its best.

Advantages of Pegasis

In Pegasis, every node is connected to its predecessor and successor node; hence, every node has two links and CH gets the aggregated message from all nodes, which is less as compared to each protocol. Since every node is elected only once, energy dissipation is balanced between sensor nodes. Decreased burden in dynamic cluster formation .CH is uniform and has balanced node formation.

Disadvantages of Pegasis

The parameter of distance among CH and BS is not taken into account after cluster head node is elected. As the CH is elected, the energy level factor is also not considered. As there is only one head node, there might be a traffic congestion problem and might cause delay in communication. Data processing is considerably delayed.

2.5 TBC Protocol

Tree-based clustering routing protocol is used to create a balance on energy consumption problem for the head nodes within the cluster which are located at remote and larger distance from BS. Every cluster constitutes of a cluster head (CH). In single hop sensor networks, the sensed information is taken from the respective CH where the information is assembled and passed on to the receiving BS. As the cluster head uses direct communication method, where every CH sends the data directly to the Base Station, irrespective of its distance from it. Consequently, such cluster heads die sooner as compared to the nearer ones. Hence, the network lifespan is decreased [22]. In tree-based clustering (TBC), inter-cluster distance is calculated, and a shortest path routing table is created among cluster heads as parent and child node and BS as the root node, where their distance is calculated for every round and the table is updated. Sending the information through multi-hop routing method is more effective and energy-efficient scheme for large scale network. Therefore, it is an optimal method where cluster heads can coordinate together and allow passing of sensed data from one to other and finally to the receiving base station [23].

Advantages of TBC

Energy is balanced among cluster heads to a considerable extend.

Cluster heads will use their energy for data aggregation and communication and will not get drained due to its large distance from BS.

TBC is fault tolerant as at any level if the cluster head dies, there are alternative path in the hierarchical structure to reach the base station.

Disadvantages of TBC

CH close to the receiving BS will get exhausted early as all the data from parent CH and their child cluster head will send aggregated data to it.

Data processing is much delayed.

Routing table needs to be updated for every round with the shortest path to base station, which again is an overhead.

3 Results and Discussion

Simulation Comparison of Leach, Heed, Pegasis and TBC Protocol

A GWSN environment is created where Leach, Heed, Pegasis and TBC protocols are simulated using MATLAB 2019b for analyzing network lifetime, average residual energy and throughput of the network. It has been observed that with the analysis of number of alive nodes to the number of rounds, using TBC protocol, the first node died at round 168 (Fig. 5a) as compared to Leach, Heed, Pegasis, where their first node died at round 113 (Fig. 2a), 42 (Fig. 3a), 97 (Fig. 4a) respectively, thereby increasing the network lifespan. Secondly, the average residual energy of the network calculated after all rounds using TBC protocol is found to be 0.0712 nJ (Fig. 5b) which is less as compared to Leach, Heed and Pegasis where their average residual energy is found to be 1.159 nJ (Fig. 2b), 1.066 nJ (Fig. 3b), 0.128 nJ (Fig. 4b) respectively, as the CH who are close to base station, besides sending their own data also transmit the aggregated data sensed by other CHs who are in the lower level of the tree structure. Therefore, a large amount of energy is consumed in this process. Furthermore, on comparing the throughput of TBC and pegasis protocol, the throughput of TBC, that is 7585 packets is considerably better as compared with Pegasis protocol which delivers 6651 data packets at the end of all rounds (Table 2 and Fig. 6).

4 Conclusion and Future Extension of the Work

This paper and research work exhibits the fundamental framework for different types of hierarchically structured homogeneous routing protocol for GWSN. Leach, Heed, Pegasis and TBC routing algorithm is simulated by creating a virtual environment in MATLAB 2019b and its comparative studied in carried to analyze its energy efficiency through various parameters. We conclude that performance of tree-based clustering protocol tends to conserve energy aspect and enhance the efficiency factor

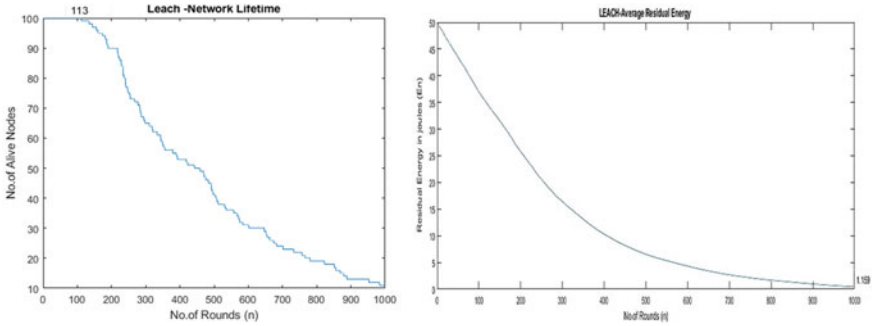


Fig. 2 **a** Leach network lifetime. **b** Leach average residual energy of the network

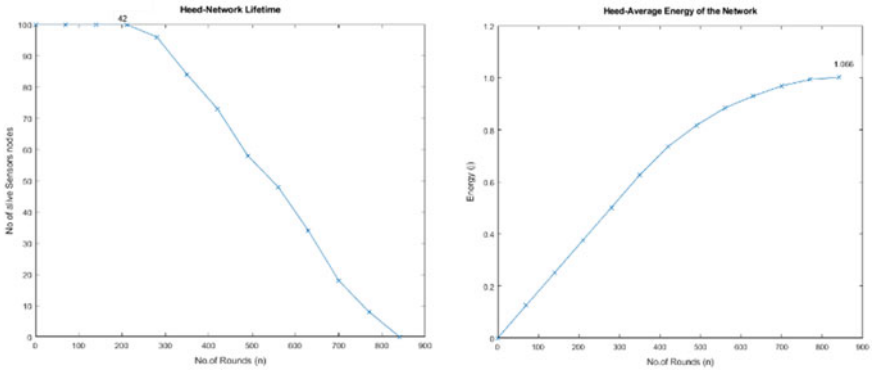


Fig.3 **a** Heed network lifetime. **b** Heed average residual energy of WSN

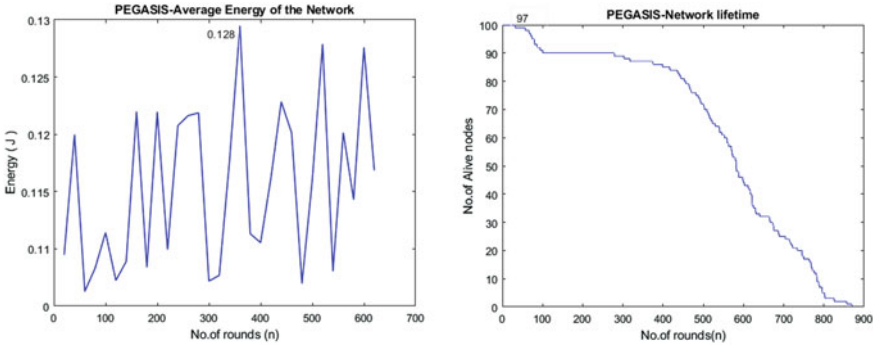


Fig.4 **a** Pegasis network lifetime. **b** Average residual energy of network

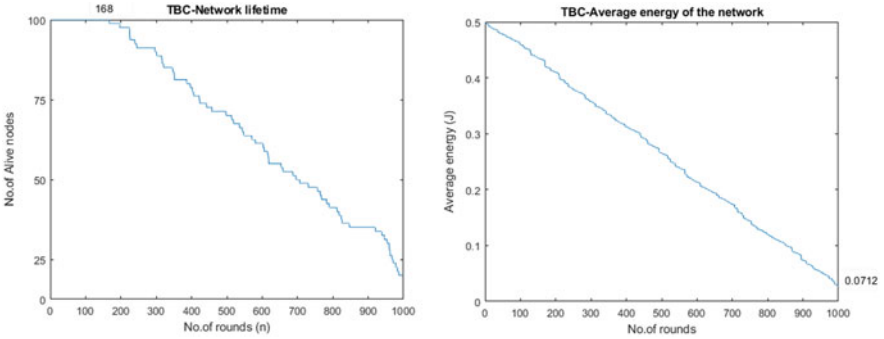


Fig.5 **a** TBC network lifetime. **b** TBC average residual energy of network

Table 2 Simulation results

S. No.	Protocol	Network lifetime (round number)	Average residual energy (En) (nJoules)	Throughput of the network (no. of successfully delivered data packets)
1	Leach	113	1.159	16,738
2	Heed	42	1.066	16,054
3	Pegasis	97	0.128	6651
4	TBC	168	0.0712	7585

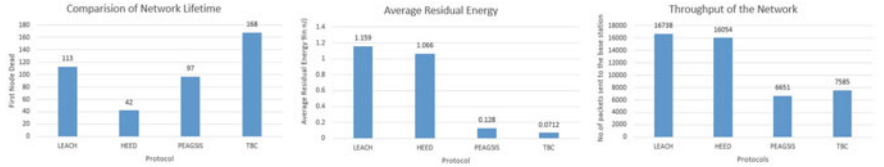


Fig. 6 **a** Network lifetime. **b** Average residual energy for the network at the end of all rounds. **c** Throughput of the network of hierarchical protocols

in network performance. This research can be extended to design and simulate intra-cluster tree-based clustering routing for hierarchical homogeneous clustering routing to increase energy efficiency in green wireless sensor network as well as an optimal number of levels in the tree structure can be analyzed to maximize throughput of the sensor network.

References

1. Anand S, Gowthami N (2020) International conference on emerging trends in information technology and engineering, pp 1–7. https://doi.org/10.1109/ic-ETITE47903.2020.485
2. Almasri AK, Darabkh KA (2020) Fifth international conference on fog and mobile edge computing (FMEC). Paris, France, pp 263–269. https://doi.org/10.1109/FMEC49853.2020.9144765
3. Shanmukhi M, Ramanaiah OBV (2016) Int J Appl Eng Res 11(10):6990–7002
4. Shetty NP, Kamath VK, Sen S, Mallaya V (2017) In: Proceedings of IEEE second international conference on electrical, computer and communication technologies (ICECCT). Coimbatore, pp 1–6
5. Altaharwa R (2018) J Softw Netw 1:1–26
6. Gupta N, Samyal VK (2018) Int J Future Revol Comput Sci Commun Eng 4(4):368–372
7. Hassan AAA, Shah WM, Iskandar MF, Mohammad AAJ (2017) Int J Appl Eng Res 12(21):11350–11360
8. Alhalafi LS, Naous R, Shihada B (2016) In: Proceedings IEEE conference on computer communications workshops (INFOCOM WKSHPS), pp 136–143
9. Fernado X, Latif S (2016) J Electrical Electron Syst 5(4)
10. He W (2019) IEEE access, p 1. https://doi.org/10.1109/ACCESS.2019.2956068
11. Jorio BE, Casablanca (2018) pp 1–6. https://doi.org/10.1109/REPSGIE.2018.8488835
12. Sinha S, Parsai MP (2020) Coimbatore, India, pp 486–489. https://doi.org/10.1109/ICCES48766.2020.9137980
13. Jambli MN, Bandan MI, Pillay KS, Suhaili SM (2018) IEEE conference on wireless sensors (ICWiSe). Langkawi, Malaysia, pp 44–49. https://doi.org/10.1109/ICWISE.2018.8633291
14. Qubbaj N, Taleb AA, Salameh W pp 414–419. https://doi.org/10.1109/ICICS49469.2020.239516
15. Seyyedabbasi GD, Kiani F (2020) IET wireless sensor systems, vol 10, no 3, pp 130–136, 6. https://doi.org/10.1049/iet-wss.2019.0153
16. Ali N, Mardini W, Abuein QQ, Khair M (2020) In: 11th international conference on information and communication systems (ICICS). Irbid, Jordan, pp 355–360. https://doi.org/10.1109/ICICS49469.2020.239543
17. Gandhi M, Patil B, Daflapurkar PM (2017) In: Proceedings IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI). Chennai, Sept 21–22, pp 2450–2456
18. Dolly VK, Arora R (2018) Sawhney. Int J Comput Appl 180(25):32–36
19. Jatothu R, Singh RP (2017) Int J Innov Adv Comput Sci (IJIACS) 6(12). ISSN 2347-8616
20. Singh DK, Diksha MA, Hasan HA (2018) Int J Adv Res Ideas Innov Technol 4(4)
21. Mittal N, Singh U, Salgotra R (2019) Proc Wireless Pers Commun 108:473–492
22. Sajwan M, Gosain D, Sharma AK (2018) Springer science business media, Springer Nature
23. Sodairi AS, Ouni R (2018) Sustainable computing: informatics and systems, 20:1–13

Resume Classification Using Bidirectional LSTM and Attention Mechanism

Swayami Bera, Biraj Ghosh, and D. Vanusha

Abstract Document classification using Deep Learning encompasses handling documents and sorting them on the basis of the subjects that they belong to. Due to the increasing growth in online recruitment, job portals receive a mammoth amount of resumes from candidates with different fields of expertise. This has rendered the traditional hiring methods inefficient and given rise to the necessity of an automatic resume classification system that efficiently directs applications to their corresponding occupational categories. For building the training and evaluation dataset, we have scraped data from websites to obtain resumes in readable text format to form a real-world dataset. Here, we implement a Word2Vec model to form the word embeddings and a bidirectional LSTM paired with a custom attention layer as our proposed model to carry out the resume classification task. We also compare our proposed model against other algorithms such as SVM, KNN, etc., to demonstrate its effectiveness.

Keywords Classification · Bidirectional LSTM · Word embedding · Attention · Accuracy

1 Introduction

Data is largely distributed on the internet as we produce so much information every moment from our application, uploading files in different formats across different web pages. Information in different forms of posts, documents and blogs on a variety of issues, videos are all uploaded on the World Wide Web. This magnitude of data can be used more effectively deriving valuable insights from it.

With the advent of the digital age, communication through the internet has opened up job opportunities for people and remotely being able to share information to a

S. Bera (✉) · B. Ghosh
Computer Science Engineering, SRMIST, Chennai, Kattankulathur, India

D. Vanusha
Department of CSE, SRMIST, Chennai, Kattankulathur, India
e-mail: vanushad@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_22

potential employer is the more efficient way to interact as opposed to the archaic hard copy submissions. This makes for a great way to hasten the process of sieving through candidates without having to deal with tedious printed paperwork. This has given birth to a lot more avenues of increasing the efficiency involved in the resume handling process. The task, of automatically classifying resumes based on the job one is likely to apply for, is a boon for companies getting thousands of applications and the automation process saves an enormous amount of time when compared to manually reading through each and every resume. This also removes the chances of human error as we humans are invariably prone to making mistakes, while skimming through data by ourselves.

This piece of work mainly focuses on text documents as the prime form of data; resumes, to be specific. The idea is to classify each resume based on a particular label. When it comes to resumes, it translates to classifying a resume based on the type of job that the person is likely to apply for. Resume classification is a machine learning task, and can fall under supervised, or even unsupervised learning category. In case of unsupervised learning problems, there is no label attached to the resumes, and the idea is to find clusters of documents that belong to a similar category, without explicitly identifying the particular category. However, in a supervised learning problem, there are labels attached to every data item and a machine learning model is used to learn to understand the sentiment of every document so that it can better 'predict' which occupational category(s) is a resume likely to belong to.

In this paper, we propose a bidirectional LSTM network along with an attention layer to carry out the resume classification task. The bidirectional LSTM preserves backward as well as forward information about the features at every step. This helps the network understand the context in long documents and works very well on this dataset where the average size of the resumes is around 1200 words. While the aforementioned layer preserves the context of the sequences, the attention layer endeavors to selectively focus on a few relevant tokens, while ignoring others in the input sequence. This provides immense help in correctly classifying resumes that belong to closely related labels such as Data Scientist and Business Analyst which might include multiple common aspects such as educational qualifications, skills, projects completed, certifications, etc.

2 Literature Survey

There has been some great work in the field of analysis and classification of text documents in recent times with rapid advances in the field of Natural Language Processing. Shuo Xu, Yan Li and Wang Zheng present a Bayesian version of the multinomial NB classifier where every word in the document contributes in determining the label that should be assigned to an unseen document [1]. But when the NB classifier is used for text classification it is often affected by data scarcity because a likelihood value has to be estimated for each and every feature. The authors in [2] implement a KNN model to classify a news dataset that shows better performance

for fewer features but its performance severely degrades with the increase in the dimensions of the data and the presence of noise. Earlier, BoW has often been used for text data representation but it completely ignores the semantics of the document. The authors in [3] propose a bag-of-concepts (BoC) depiction of documents instead of the commonly used bag-of-words (BoW) with a SVM model to carry out the classification task. Although the BoC method works better in case of documents with extensive descriptions and explanations of the subject, it negatively affects the performance when the document contains a lot of abbreviations, numerical measures, and other words or symbols that cannot be translated into concepts. Manh Hung Nguyen introduces a new feature called label-oriented score for each term which represents its importance with respect to all the terms assigned or unassigned to a certain label. For the testing data, the sum of the LO score of all terms in a text will be used to determine whether that document should be categorized into that certain label or not [4]. The authors in [5] use a neural network on the vector representation of the text using the TF-IDF values of the tokens to perform classification of resumes. The proposed neural network model shows better performance than KNN and SVM models.

However, none of the aforementioned works take into account the context in which the words have been used in the respective documents, i.e., the semantic information. This limitation is overcome by Yoon Kim who implements a model that gives importance to the local information that is essential for the classification task, using the Word2Vec embedding, followed by a single-layer CNN, and achieves remarkable results [6]. The design proposed by Abdalraouf Hassan and Ausif Mahmood is a framework of RNN and CNN combined, where they replace multiple pooling layers with an LSTM layer in order to retain maximum details in neighboring text, capture the long-term dependencies proficiently and simultaneously decrease the amount of parameters [7]. The attention mechanism was initially proposed as a layer implemented in between the encoder and decoder modules used for language translation to give more importance to words that were more relevant to the context [8].

3 System Architecture

A. *System Design*
See Fig. 1.

B. *Data Acquisition*
Data in the form of resumes is scraped from a number of websites and search strings used to find the required data are in accordance with the labels that we want to classify the data into.

A dataset is then prepared as a csv file with each resume in the form of a long string of characters and another column with its corresponding label (Fig. 2).

C. *Data Preprocessing*
Lowercase: The resume string is converted into lowercase letter so that the data is no more case sensitive for obvious reasons.

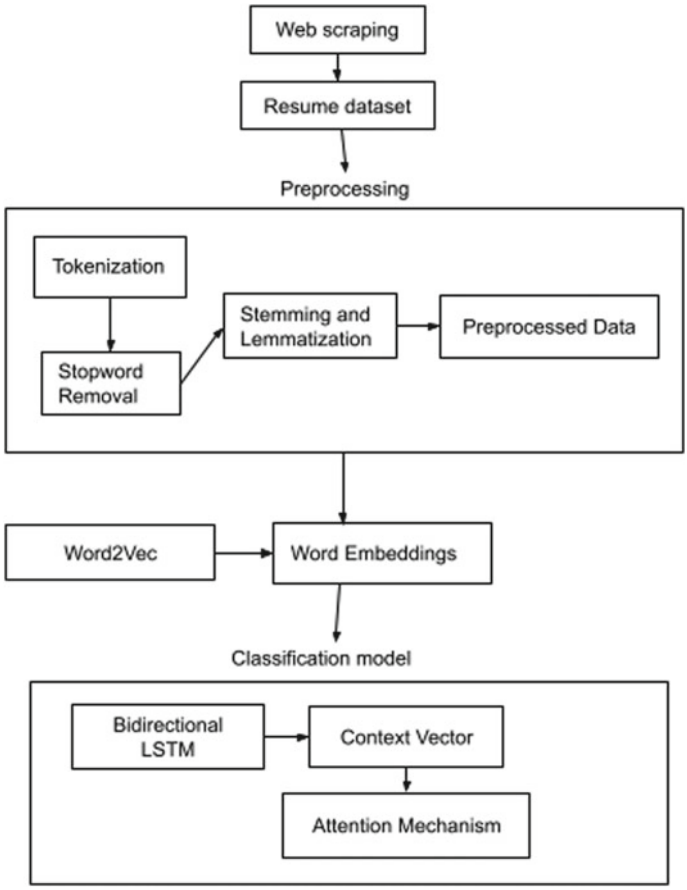


Fig. 1 Proposed system architecture

Fig. 2 Sample of the dataset used

Category	Resume
HR	JGSKILLSMicrosoft Office Medical Terminol
Electrical Engineering	JHONNY LOOR**** *** ** ********** **
Data Science	Jiali ZhangWaltham, MA ***** 857-***-**
Web Designing	Jianchao LiBoston MA 617-***-**** adhm
Electrical Engineering	Jianhua YingEmail: adh7cb@r.postjobfree.c
Electrical Engineering	JIBIN GEORGEKOCHUVEMBILLIL HOUSETH
Data Science	Jichong Chaiade5p2@r.postjobfree.com 2C
Data Science	Jiekai Maadh7t1@r.postjobfree.com +1-91
Web Designing	Jignesh Babubhai PrajapatiGhatlodiya, Ahm
Web Designing	JIGYASA MESHRAMComputer Science & En
Data Science	Jijun "David" DuSan Francisco, CA 628-***
HR	Jillian WeishaarPRESENT ADDRESS: PERMA
Web Designing	Jinali MistryFront-End Angular Developerac

Tokenization: Tokenization is a process by which a string is split into a token of words. For this dataset, all words which are alphanumeric in construction have been retained, and the other tokens have been discarded.

Stopword Removal: Stopwords are commonly used words in a corpus which do not lend any valuable context and sentiment to the data item and can be safely dropped for reducing the redundancies in the corpus.

Stemming: It is how a word is transformed to its stem by dropping the attached suffixes or prefixes. Stemmed words help the model understand the similarities among seemingly different word forms of the same word.

Lemmatization: It eliminates endings of a word which are inflected, by performing a morphological analysis on them. This facilitates obtaining the root form of a word, called the lemma. For example, 'better' and 'good' both upon lemmatization, become 'good.'

D. *Word Embeddings*

Word embedding: It represents a word in vector form which can be plotted in Euclidean space.

$$\text{Cosine Similarity}(\alpha, \beta) = \frac{\alpha \cdot \beta}{\|\alpha\|\|\beta\|} = \cos(\theta)$$

where $\alpha.\beta$ here represents the dot product of two vectors (which in turn are words), $\|\alpha\|$ stands for the length (commonly called norm) of the vector α, and θ is the angle between α and β. The similarity of two wordsis subject to the angle formed between word vectors α and β. The closer in meaning the words are, the more their cosine similarity tends to 1; the more dissimilar they are, the lower is their cosine similarity value.

Embedding dimension is an important metric because the higher the dimension, the better the representation of a word. Although, the drawback of a higher dimension is that it increases the time taken for obtaining word vectors as well as for training the model.

Word2Vec model is a neural network model used to find embeddings of a word. Two ways of doing this are CBOW (Continuous Bag of Words) or Skipgrams. TensorFlow does provide a default Embeddings layer but we replace the weights in that embedding layer with the weights (word vectors) obtained from Word2Vec.

E. *Proposed Model*

The neural network model that is used to train the aforementioned data makes use of a few layers. The algorithms implemented in them are as follows:

Bidirectional LSTMs: In sentences, the relevance of the current word can be determined better when both the past and the future context are taken into account. Every LSTM cell has two outputs, a new candidate value and an activation value. Bidirectional LSTMs provide two activations for every input in the sequence, and these are used to calculate the encoding of the corresponding input (Fig. 3).

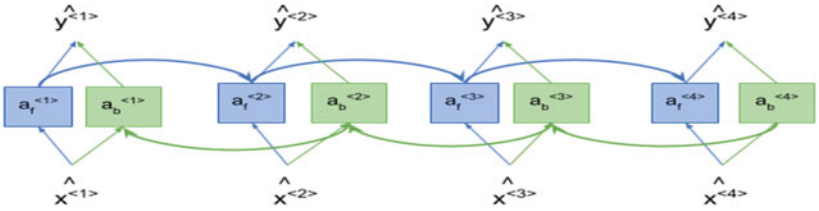


Fig. 3 Working of the bidirectional LSTM

Attention Layer: The attention layer endeavors to selectively focus or give 'attention' to a few relevant tokens while ignoring others in the input sequence. The attention mechanism considers all the hidden states of the previous bidirectional LSTM layer but pays attention to those parts or tokens which have a strong association with the specific label.

Instead of taking the final hidden state of the bidirectional LSTM as an input to the first feedforward dense layer, we utilize the individual hidden states of all the words as produced from the bidirectional LSTM layer. The intention is to pay attention to the individual hidden states and conclude which words hold more importance in predicting the instance label, than others. We form a custom attention layer for the same. If the overall word count of each resume is N, then the bidirectional LSTM creates a vector with dimension d for each of the N words.

First, we find the 'attention score,' one for each word in the resume (α_i where $i \in [0, N]$) as follows:

$$\alpha_i = \text{softmax}\left(\tanh\left(q^T w_i + b_j\right)\right)$$

Here, q is the weight vector ($d \times 1$) and b_j ($j \in [0, d]$) is the bias—these are optimized during the training of this layer. w_i is the word vector. The softmax function is used to normalize the attention score.

Next, we obtain the attention adjusted word vectors by multiplying the attention scores with the corresponding word embeddings. All these vectors are summed up to obtain the final output vector that represents the resume.

$$y = \Sigma \alpha_i w_i$$

The attention adjusted word vectors are then passed through the first dense layers containing 32 neurons, followed by the dropout layer to avoid overfitting. Finally, there is the output layer with 8 neurons corresponding to the different labels.

Table 1 Data sample size

Label	Resumes
HR	1000
Web designing	1000
Business analyst	1000
Electrical engineering	1000
Java developer	1000
Mechanical engineer	991
Data science	800
Python developer	475

4 Experimentation

A. *Experimental Setup*

For the setup, the first task of data collection is done by scraping the internet for resumes belonging to the required labels. The data is then preprocessed using standard techniques such as stopword removal, stemming and lemmatization.

After that the word embedding matrix is obtained using the Word2Vec model. The feature vectors are then fed to the model architecture to train the neural network model. The results are then compared with other machine learning algorithms.

TensorFlow and Scikit-learn are used for deep learning and machine learning algorithms, respectively. Jupyter notebooks and Google colaboratory are used as the training and testing environment.

B. *Evaluation Results*

Unique as well as similar labels were scraped to test the efficiency of the attention mechanism. 75% of the data is randomly chosen for training, and the rest is kept for validation. The size and specifications of the dataset are as follows (Table 1 and Fig. 4).

The accuracy of the neural network model is tabulated with the accuracies of Support Vector Classification and KNN algorithms (Table 2 and Fig. 5).

Inference: From the above results, it is clear that a neural network model outperforms classical machine learning algorithms such as SVMs and KNNs. It is also noteworthy that on adding the attention mechanism we see an increase in overall accuracy of classification.

5 Conclusion

Through our work, we have introduced an efficient and accurate way to classify resumes. We proposea model which is a combination of a bidirectional LSTM, followed by an attention layer applied on the word vectors obtained after using

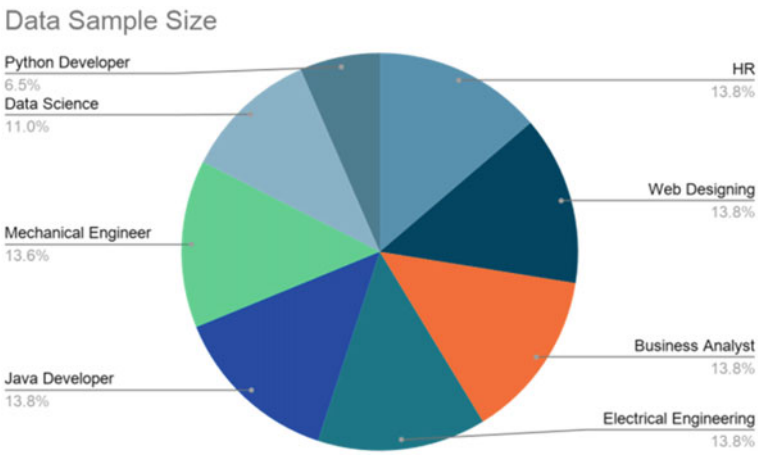


Fig. 4 Distribution of the dataset

Table 2 Algorithm accuracy

Algorithm	Training accuracy (%)	Validation accuracy (%)
KNN	76.65	75.3
SVM	88.40	87.23
NN (Bidirectional LSTM)	90.75	85.9
NN (Bidirectional LSTM + Attention Mechanism)	93.67	90.17

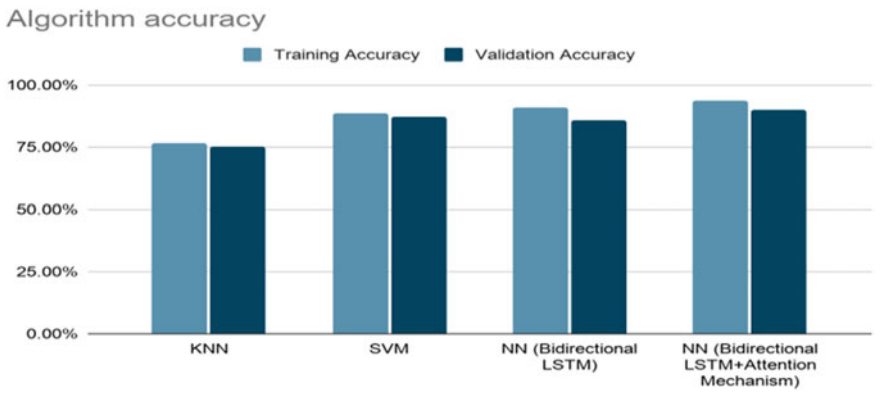


Fig. 5 Comparing performance obtained with different algorithms using the accuracy metrics

the Word2Vec model. This model overcomes the shortcomings of traditional resume classification algorithms by preventing the loss of context, while improving accuracy of prediction. Evaluation metrics such as accuracy of a neural network clearly justifies how it is superior to the other standard classification algorithms when it comes to real-world data. Training and testing costs were not too expensive, as epoch durations were short and little space was required for text data which is inherently small in size.

The only limitation which the model possesses is its bias to a limited number of labels. Due to its scalable nature, we will be adding more labels to our dataset in future and will focus more on hyperparameter tuning. We will also aim for a threefold increase in the size of the dataset.

References

1. Xu S, Li Y, Wang Z (2017) Bayesian multinomial Naïve Bayes classifier to text classification. Advanced multimedia and ubiquitous engineering. Springer, Singapore, p 448
2. Jodha R, Gaur Sanjay BC, Chowdhary KR, Mishra A,Text classification using KNN with different features selection methods. Int J Res Publ 8(1)
3. Khan A, Baharudin B, Lee LH, Khan K (Feb 2010)A review of machine learning algorithms for text-documents classification. J Advan Inf Technol 1
4. Nguyen MH (Oct 2020)A label-oriented approach for text classification. Int J Innovative Comput, Inf Control 16(5)
5. Pimparkar K, Lulla R, Rathod P, Anirudh V, Dedgaonkar SG (2019)Document management using artificial neural network. In: International conference on communication and electronics systems (ICCES). Coimbatore
6. Kim Y (2014) Convolutional neural networks for sentence classification. In: Conference on empirical methods in natural language processing (EMNLP)
7. Hassan A, Mahmood A (2018) Convolutional recurrent deep learning model for sentence classification. IEEE Access 6
8. Bahdanau D, Cho K, Bengio Y (2015)Neural machine translation by jointly learning to align and translate. In: International conference on learning representations (ICLR)

An Empirical Analysis Data Mining Frameworks—An Overview

K. Sivakumar, S. Kalaivani, D. Venkatesan, and V. Vetrivel

Abstract In recent years, a massive breakthrough has attracted interest, dramatically altering the way companies produce, conduct, and operate data and analytics structures. Hadoop, NoSQL, and the cloud have ushered in a new age of scale-out, flexible, and real-time computation, resulting in the development of new data processing, collection, and analytical knowledge to enable advanced machine learning and optimization techniques. Over the last decade, many developments in data processing and development have occurred, including big data systems, cloud services, data lakes, optimization, self-service, data collation, machine learning, and many others. Today's data management systems imitate the old data warehousing and BI systems on the outskirts, but with new data management concepts and updated processes. The need to modernize data processing architecture is widespread, as shown by the amount of architecture consultancy inquiries. As a result, this research explores a simple path to massive data handling framework as well as various popular file system architectures for handling big data for data mining. Furthermore, major frameworks for data mining are investigated and compared.

Keywords Amazon Web Services · Big data · Database architectures · Google cloud · Hadoop · Hazure

1 Introduction

Data processing is widely more testing than simply finding, perceiving, comprehension, and referring to data. For fruitful tremendous extension assessment, the sum of this necessities to happen in a completely modernized manner. This requires contrasts in data plan and semantics to be conveyed in structures that are PC sensible, and a

K. Sivakumar (✉) · D. Venkatesan · V. Vetrivel
Department of Mathematics, Saveetha School of Engineering, SIMATS, Thandalam, Chennai, Tamil Nadu, India

S. Kalaivani
Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_23

short time later "precisely" resolvable. There is a strong assortment of work in data blend, arranging and changes. In any case, amazing additional work is expected to achieve robotized botch free qualification objective.

Methods for questioning and mining enormous information are in a general sense unique in relation to conventional factual examination on little examples. Information administrations give coordinated proficiently open information. All the while, data mining itself can moreover be used to help with chipping away at the quality and dependability of the data, grasp its semantics, and give cunning addressing limits.

Regardless of how massive or unstructured the data is, effective big data management allows an organization to easily find valuable information. The data was collected from different sources, including call logs, machine logs, and social media platforms. All businesses must make decisions on how to interact and use their data. Companies who take a systematic approach to implementing an enterprise-grade strategic strategy are best positioned to maximize their innovation investment while lowering costs. Companies can no longer continue to disregard big data's immense potential in light of the current situation. Information system plans must be designed for a wide variety of potential and possible concerns [1, 2]. The most obvious flaw is human height. Due to the current influx of information that fills the planet on a daily basis, divine discoveries within a given time period are exceedingly difficult for a human data team. As previously stated, big data applies to every business sector, and staying on top of new trends in this field pays off—a credential in this vein will go a long way toward improving a professional profile and ensuring that individuals and businesses gain a significant competitive advantage. Modern computing strategies include the speed, energy, and versatility needed for accessing large quantities and types of big data effectively [3]. Data management, data integrity, data storage and processing, data preparation for evaluation, and secure access will all be covered by corporations. Although some information can be stored on-site in a traditional data center, cloud systems, unstructured data, and Hadoop are all versatile, low-cost alternatives for storing and handling large amounts of data [4].

A broad data architecture is intended to enable data collection, processing, and interpretation for data sets that are too big and complicated for conventional database systems. Batch processing, real-time storage, digital exploration, predictive analytics, and machine learning are all common tasks in big data strategies. The general data mining architecture for managing unstructured data is depicted in Fig. 1 [5].

At least one information sources are the establishment of all enormous information examination. Bunch handling information is typically saved in a different document store that can oblige enormous quantities of records in various arrangements. Since the informational indexes are so enormous, they will be overseen by a major information framework that will utilize long-running bunch responsibilities to sort, examine, and arrange the information for the examination. Perusing, checking on, and moving yield to new records are regular undertakings for these occupations. Except if the arrangement utilizes solid realities, the execution should consider the assortment and capacity of constant messages through stream processing. It might be a basic information stockpiling framework, with approaching messages being saved into a capacity organizer. The arrangement will sort, dissect, and in any case

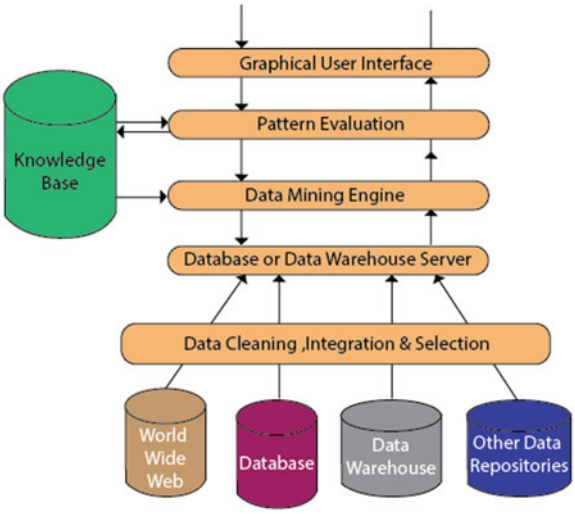


Fig. 1 General data mining architecture

orchestrate the information for examination in the wake of assembling continuous messages. The information got from the stream is moved to a channel in the yield. Numerous huge information frameworks pre-measure information prior to showing it in a normalized design that can be gotten to utilizing testing strategies. The point of most huge information arrangements is to follow and dissect information to give bits of knowledge into it. Work processes that convert source information, move information between various assortment and reusing measures, load prepared information into a functional information store, or push moment couriers to a report or entry make up most of enormous information advances [6–8]. With this in-depth introduction to big data handling design for data mining, Sect. 2 outlines the related research, Sect. 3 analyzes various data mining frameworks with comparisons, and Sect. 4 concludes the work.

2 Big Data: Opportunities and Challenges

There are challenges [l] even in big data, counting how to catch, move, store, clean, examine, channel, search, share, get, and imagine information. In any event, putting away and recovering large information is one of the primary issues around here. Large information issues require making a few compromises among wanted adaptability, accessibility, execution and security. Big data starts with tremendous volume, heterogeneous, free sources with scattered and decentralized control and attempts to examine mind boggling and creating relationship among data. These characteristics make it a crazy test for discovering significant data from enormous data. According to an artless point of view, we can imagine that different outwardly disabled men

are endeavoring to assess a goliath elephant, which will be large information in this setting [2].

The goal of each outwardly debilitated man is to draw a picture of the elephant according to the piece of information he accumulates during the cooperation. Since, each individual's view is confined to his close by region. Exploring the Colossal Data in the present circumstance is indistinguishable from accumulating heterogeneous information from different sources (daze men) to help with drawing a most ideal picture to uncover the authentic badge of the elephant in a continuous style. Without a doubt, this endeavor is not essentially pretty much as fundamental as mentioning that each outwardly weakened man portray his feelings about the elephant. Then getting an expert to draw one single picture with joined view stressed that each picture individual may impart in a substitute language and they may even have security stresses over the messages they intentional in the information exchange measure.

Indeed, the effective mining of huge information empowers to work on the upper hand and to add an incentive for some areas. Enormous information is principally characterized by its 3 versus essential attributes. The 3 versus incorporate speed (information are developing and changing quickly), assortment (information come in various and numerous arrangements), and volume (immense measure of information is created each second). Examination people group from various areas have been attempting to foster a new, quick, dynamic, and easy to use innovation for enormous information.

Big data concern enormous volume, awesome, creating instructive lists with various, independent sources. With the fast improvement of frameworks organization, data amassing and the data grouping limit, Huge Data are presently rapidly filling in all science and planning spaces, including physical, regular and biomedical sciences.

Wu et al. [2] planned a HACE hypothesis that portrays the provisions of huge information transformation and proposed a major information handling model from the information mining point of view. This model incorporates request driven conglomeration of data sources, mining, and examination. In general, it plainly investigations testing issues in the information driven model. However, it does not have the testing and execution assessment regarding any informational indexes.

Nowadays, various endeavors and government can exploit tremendous data to isolate huge agreement. Such information can assist bosses with working on their frameworks and upgrade their courses of action. It helps the relationship with procuring advantage and gives expanded the worth of various monetary and social regions. Benjelloun et al. [4] presented a couple of significant data projects, openings, models a great deal in various regions like clinical consideration, business, the movement business and authoritative issues. It surmises that immense data disturbance adds to improve the unmistakable coherent fields by allowing advanced complex assessment across different sources.

MapReduce is a structure for taking care of and supervising enormous degree enlightening assortments in a passed on bundle, which has been used for applications, for instance, creating search documents, record gathering, access log assessment, and various sorts of data examination. Li et al. [9] give a total review of a wide extent of

proposals and systems that bright lights on scattered data the leaders and planning using the MapReduce structure. This is a promising and best resource for extra experimentation and assessment in the enormous data dealing with.

The advancement of unstructured and somewhat coordinated data in online media, land information, normal data and other web bases applications present an open test to the cloud informational collection neighborhood. Subsequently, an approach to manage Colossal Data examination which joins coordinated and unstructured coordination has gotten essential nowadays. Data recuperation using the request getting ready is one of the fundamental troubles in Huge Data. Mathew et al. [10] arranged new requesting approaches explicitly Llndex and Hlndex to offer an assistance to list in Hadoop Dispersed Report System and MapReduce systems without changing the current Hadoop structure. These techniques give a degree for investigates to grow the work on composite inquiries in a more prominent educational assortment up.

3 Literature Study

In the field of innovation, stockpiling limit has consistently been a worry. Momentarily after the online stage's hypothetical structure of the 1930s, the emphasis was on the most proficient method to deal with the interminable correspondence measure. The information stockpiling strategies must be modified in files as well as in all information combination organizations because of this assessment. In the start of the 1950s, the idea of virtual memory was proposed by Fritz-Rudolph Güntsch, who respected last stockpiling to be interminable. The advancements in information stockpiling have seen persistent improvement, which is basically "more on less," with most in-house facades. The advancement of information stockpiling innovations has seen consistent improvement, for all intents and purposes "all the more less," with many in-house administrations and cloud administrations being provoked [11]. According to Gartner, interest in the arising advances of enormous information has developed continuously in the course of recent years, and in 2018 and 2019, organizations are proceeding to put more in IT, with an emphasis on IoT, blockchain and large information. Organizations like Instagram, one of the world's most online media stages, started to develop from 1.5 million to 1 billion new clients, posting 95 million every day photographs and recordings, separately [12].

In the big data sector, numerous assessments were completed. A significant number of them manage advancements, instruments, qualities, and shortcomings. They attempt to explain the large information industry more and present its benefits and weaknesses. They evaluate the specialized hardships and the developing energy in high parcels for every one of the standard handling ventures from information age to appraisal. Some exploration approaches and resources for huge information examination were additionally conveyed out. Existing executions have gotten broadly archived over the long haul. Three of the best ones are the models of Lambda and Kappa, the most elevated piece of in progress contemplates [9].

An incredible arrangement has effectively been done according to a modern perspective to clarify the utilization of large information supply better consideration and upgrade productivity in different fields. A major information execution manual cannot be utilized for choosing from among the most oftentimes utilized enormous information systems, which have an application market, existing programming configuration, financial plan for the acquisition of new modules, and the issues to be settled in the app. Intake the board, information assortment, representation and examination are likewise planned with huge or too convoluted huge information designs for the treatment of conventional instruments. Hundreds or many terabytes starting with one association then onto the next can include [10]. Inside this unique situation, the most famous enormous scope information the board structures portrayed in Segment III is AWS, Cloud, and Azure.

4 Data Mining Frameworks

An architecture in data mining governs the collection, storage, analysis, and visualization of data. We also discuss how organized, unorganized, and semi-structured data for tracking and analysis can be converted. In this segment, we discuss three of the most famous data mining frameworks that have become known to the industry over the years [13].

Big data gives parcel of freedoms to a wide scope of use regions. In the brilliant item applications like customized medical services which empowers patients to screen their own surroundings, for example, the organization of custom far off help and portable wellbeing arrangements. This offers the capacity to send cautions, anticipate potential irregularities progressively. These sort of enormous information gathered may help future investigates to further develop the treatment strategies. Jara et al. [14] introduced an information procurement engineering, which coordinates and moves improved information from clinical gadgets to the information based data frameworks. The design has been executed and tried on the sensor information. This paper has the chance for the explores to think on taking on the design to the next various areas in Huge Information.

Data acquisition is the essential advance in the information handling in huge information. The presentation of information obtaining contributes the significant job in the exhibition of the information handling framework. Nie et al. [15] carried out information obtaining dependent on LabVIEW correspondence framework with continuous transmission and simple upkeep. The strategies utilized can be utilized as an asset to the analysts for the enormous information obtaining frameworks in more extensive spaces.

4.1 Azure

In 2010, Microsoft acquainted Purplish blue with offer cloud benefits that permit its executions in Microsoft's server farms to be created, checked, carried out, and oversaw. Such server farms have been appropriated in 54 locales around the world. Microsoft offers a scope of administrations covering a few distinct fields like PC, stockpiling, sharing of data, correspondence, and so forth. With reference to Purplish blue has the greatest portion of the overall industry. Many individuals have effectively worked on this by utilizing AWS and the AWS work market is presently clogged. All things considered, you will not be useful in the event that you pick Google Cloud in light of the fact that most organizations in the Fortune 500 are not really at Google Cloud. The quantity of Google Cloud workers accessible is in this way generally little. Remaining in the center would be a deliberately decent approach to get a Sky blue Proficient certificate [16]. Figure 2 shows that Azure data mining architecture.

AppFabric is the framework that contains capacity, recovery, and correspondence abilities. Purplish blue is open from the Microsoft Sky blue Administration Entrance and Microsoft's server farms. At the build stage, Purplish blue gives items that depict the usefulness of a stage based application. The organizations permit organizations to safely connection to their public cloud utilizing Purplish blue ExpressRoute. It is additionally utilized for overseeing virtual private organizations and building different virtual organizations. Store more proficient frameworks that can produce monstrous executions and scale up, if important, and balance traffic-based information naturally. Sky blue information base is a predictable and versatile connection

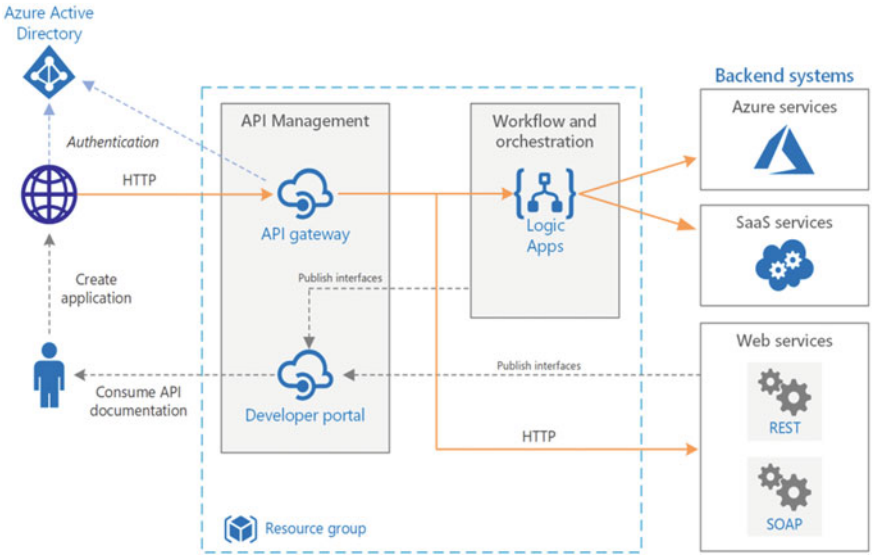


Fig. 2 Azure data mining architecture

by-administration data set that conveys superior without stressing over any facilities. Since information are not put away on a solitary actual gadget and are distinctive in various topographical regions, it would be a lot more secure on account of a calamity. Regardless of whether one spot is exposed to certain cataclysmic events, different areas are not hurt, and information are subsequently secure [17]. Azure is continually refreshed to guarantee its security. The Security Advancement Lifecycle guarantees free from any and all harm each phase of improvement. Purplish blue has likewise as of late been refreshed on its upgraded virtual machine limit adaptability. In these reservations, distributed computing costs will be decreased. Actual equipment implies purging the pockets to send gadgets. To instruct individuals how to manage,, it requires some investment and energy. In this quick clearing innovative world, there is huge contest. This load of issues are settled by Sky blue, which makes equipment and labor modest [18].

4.2 Aws

Amazon Web Services, or AWS, is a distributed computing network that most undertakings, new businesses, and government offices use. This guides organizations in adapting to high traffic and putting away huge volumes of video and information. In principle, the AWS EC2 stage gives various arrangements dependent on the prerequisites for clients of numerous virtual machines. EC2 represents Versatile Register Cloud, which gives a scope of installment choices, characterization alternatives, and planning, among different things. The straightforward AWs or AWS EC2 engineering is the thing that it is called. As such, EC2 can likewise be called a Versatile PC Cloud, which permits clients and different clients in their own framework or task to use various provisions whenever required. There are likewise unique extraordinary decisions like estimating, singular worker planning, the executives of arrangement and so forth S3 is called Basic Stockpiling Administrations. Utilized in the AWS engineering. Utilizing this application interface, clients can helpfully store or recall information by means of different information types. The administrations will likewise have no PC viewpoint [19]. Figure 3 shows that AWS data mining achitecture.

The load balance aspect helps the database and server performance to be improved appropriately throughout the AWS architecture. This load balancer is a common network device used in AWS architecture to achieve skill in conventional web application frameworks. As per AWS architecture data flow chart, it can be used. The Elastic Load Balancing System is also provided, and AWS spreads the traffic to EC2 instances via different information published. The traffic is also delegated to adaptive add-ons, and Amazon EC2 hosts load balance deletions. The Amazon CloudFront service is widely used to distribute information for web delivery. In addition to using global network locations, the service can be used in content types such as static, interactive, and subsystems. The content is automatically demanded from the user's

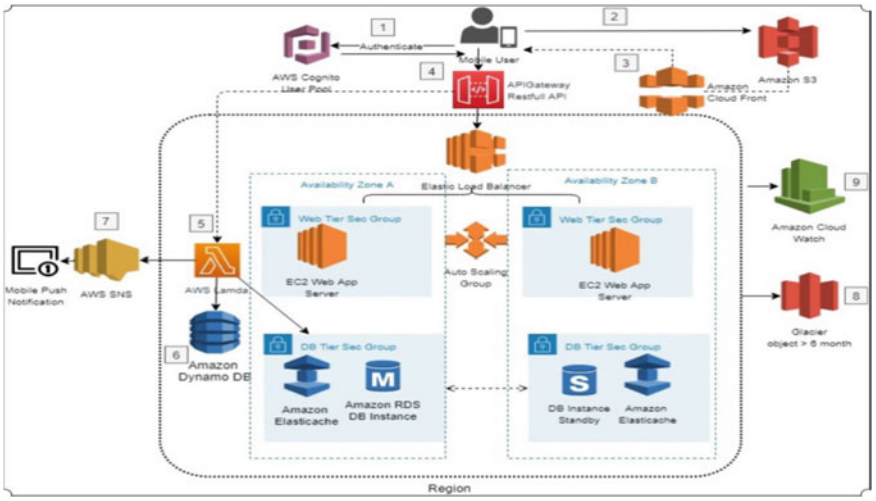


Fig. 3 AWS data mining architecture

end based on their position and also displays the various impacts on the results that will enhance correctly.

It also ensures that a security feature, known as shared folders, is enabled. It will also conflict with the inbound firewall network, necessitating the specification of ports, processes, and even source IP fields for all EC2 instances that can be accessed. Protective groups can be changed, and access to EC2 instances can be limited, using different subnets or other IP addresses. It provides similar access to database systems such as MySql, Microsoft SQL Server, and Microsoft SQL. These programs, inquiries, and applications will all benefit from Amazon RDS. This often aids in the avoidance of postulating control and the attainment of higher economic values that can easily adapt from low to high prices. It will easily increase strength and endurance, reducing the amount of time it takes to complete a task. This will also assist in preventing the use of conjectural force [14, 15].

4.3 Google Cloud

Google Cloud Platform (GCP) is a public cloud platform from Google that competes with Microsoft Azure and Amazon Web Services. The difference is that GCP is built on Google's massive, cutting-edge infrastructure, which manages all of Google's traffic and workload (Fig. 4).

For process monitoring, logging is essential. While the reports are the foundation for assessing performance, logs provide valuable information needed to test, analyze and enforce policies. Google Cloud Platform gives the automatic logging service for archiving, search, tracking and alarming the log data and events in Google Cloud.

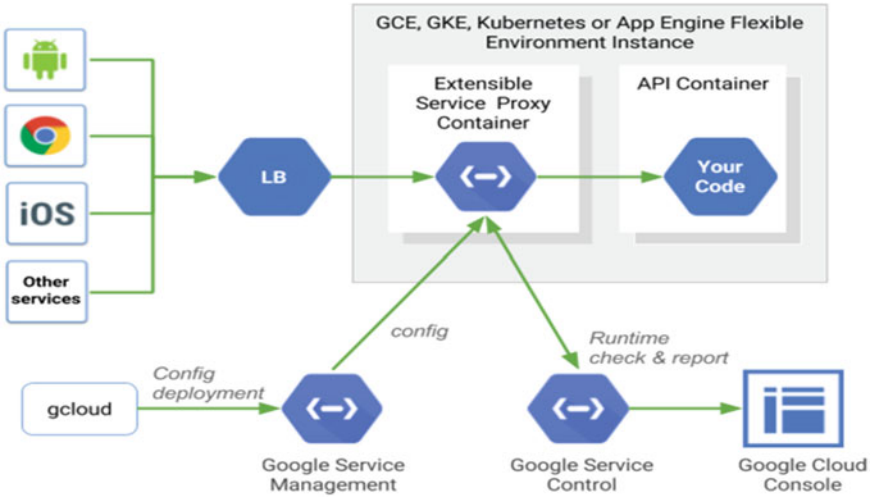


Fig. 4 Google Cloud data mining architecture

The Data Log collects instant data from the facilities of Google Cloud. Monitoring a challenging process is a key technical task. Google Cloud is a programming model in the Google Cloud Operations Suite that offers cloud management. In Google Cloud, cloud surveillance is used to track providers and metrics, and cloud analytics provides an API for third-party monitoring tools to be integrated. Cloud monitoring centralizes network measurements, reports, and occurrences to provide a rich collection of measurable signals to developers and operators to enable root causes analysis to be speeded up and time to be resolved to a minimum.

The alerting device instantly maps to the surveillance system's four golden signals, providing all the information the on-call application requires to take urgent action. Finding out how alarms work should not take too many taps. Also, it aims to reduce the number of jobs you have, such as by deleting or automating error adjustments. No servers are automatically run, maintained or updated, according to the load. Tracking, logging and debugging are referred as appropriate terms. Built-in task and functional protection based on the lowest principle of privilege [20, 21]. Table 1

Table 1 Comparison azure versus AWS versus google cloud

Parameter	Microsoft Azure	AWS	Google Cloud Platform
Age scale	7	12	6
Backup	Azure	Amazon S3	No backup
Structure	Enterprise-grade hybrid cloud storage	Virtual tape infrastructure	Partners like egnyte
Market status	Second largest provider	Dominant market position	New
Price	Price/minute	Price/hour	Price/minute

shows comparison azure versus AWS versus google cloud.

Companies who primarily base their pricing decisions on their cloud providers would need to evaluate each project on an individual basis in order to get the best price. As suppliers lower their prices on a regular basis, these calculations will need to be reviewed on a regular basis. Finally, the best public cloud provider is determined by the demands and workloads.

5 Conclusion

Big data refers to massive amounts of heterogeneous data from both traditional and new sources that are growing at an exponential pace. Designing structures for centrally handling and efficiently analyzing vast quantities of data that are both internal and external to an enterprise is difficult due to their high complexity. The blueprint for a system that manages large data volumes during collection, processing, analysis, and visualization is specified by a big data architecture. Many architectures belonging to different categories have been proposed by academia and industry, but the sector still lacks benchmarks. The arrival of the big data era necessitates the creation of more effective and efficient data discovery and analysis methods. Different clustering algorithms for parallel processing distributed systems and software platforms must be advanced in this regard. When complex clustering algorithms are required, however, implementing them in a fully scalable manner on such platforms poses a number of technical challenges that evolve as the algorithms become more complex. As a result, architectures that were designed for sequential processing must constantly be redesigned to allow effective use of distributed computational resources. As a result, clustering algorithms and big data handling architectures must be examined in future.

References

1. Azarmi B (2016) Scalable big data architecture. A practitioner's guide to choosing relevant big data architecture. Apress, Berkeley. https://doi.org/10.1007/978-1-4842-1326-1
2. Burys J, Awan AJ, Heinis T (2018) Large-scale clustering using MPI-based canopy. https://doi.org/10.13140/RG.2.2.26139.11049
3. Chattopadhyay A, Chang C-H, Yu H (2017) Emerging technology and architecture for big-data analytics. https://doi.org/10.1007/978-3-319-54840-1
4. Fahmideh M, Beydoun G (2018) Big data analytics architecture design—an application in manufacturing systems. Comput Ind Eng. https://doi.org/10.1016/j.cie.2018.08.004
5. Godson K, Behera R (2019) Big data architectures: a detailed and application oriented review
6. Komal MS (2018) A review paper on big data analytics tools
7. Rafsanjani K, Asghari M, Emami Z, Nasibeh (2012) A survey of hierarchical clustering algorithms. J Math Comput Sci 5:229–240 https://doi.org/10.22436/jmcs.05.03.11
8. Kumar A, Ingle Y, Pande A, Dhule P (2014) Canopy clustering: a review on pre-clustering approach to k-means clustering

9. Mukherjee S (2019) Benefits of AWS in modern cloud. https://doi.org/10.5281/zenodo.2587217
10. Patil P, Karthikeyan A (2020) A survey on k-means clustering for analyzing variation in data. https://doi.org/10.1007/978-981-15-0146-3_29
11. Memon M, Soomro S, Jumani A, Kartio M (2017) Big data analytics and its applications. Ann Emerg Technol Comput 1. https://doi.org/10.33166/AETiC.2017.01.006
12. Mijwil M (2018) Microsoft azure what is it and where does microsoft bet with its cloud?
13. Kalipe GK, Behera RK (2019) Big data architectures: a detailed and application oriented review
14. Mijwil M (2018) Microsoft azure what is it and where does microsoft bet with its cloud?
15. Mukherjee S (2019) Benefits of AWS in modern cloud
16. Komal MS (2018) A review paper on big data analytics tools
17. Rafsanjani MK, Asghari Z, Emami N (2012) A survey of hierarchical clustering algorithms. J Mathe Comput Sci 5:229–240. https://doi.org/10.22436/jmcs.05.03.11
18. Kumar A, Ingle YS, Pande A, Dhule P (2014) Canopy clustering: a review on pre-clustering approach to K-Means clustering
19. Memon MA, Soomro S, Jumani AK, Kartio MA (2017) Big data analytics and its applications. Annals Emerge Technol Comput 1. https://doi.org/10.33166/AETiC.2017.01.006.
20. Nielsen F (2016) Hierarchical clustering. https://doi.org/10.1007/978-3-319-21903-5_8.
21. Ramesh B (2015) Big data architecture

An IoT-Based Automated Aqua Nurturing System

R. H. Aswathy, G. Dhanush Eswar, S. Arnold Jerome, and R. Abishek

Abstract Internet of Things (IoT) extended its features to all technologies, which enables the connection of devices to the Internet and helps exchange information to enhance communication using devices. IoT is a leading technology for automation that provides unique identifiers and can share data over the Internet without any human interactions. As we are in the modern world, we need to engage with work and several other responsibilities. The loveable pets cannot be feed at the regular time by owners. Regular feeding is one of the problems in fish maintenance, and owners often forget to feed their pets because of their work. Our proposed work is to design an interactive mechanism for pet owners who have an aquarium. The prototype of an automated pet feeder builds that will be suitable for feeding pets remotely using a smartphone. The pet feeder connected to a cloud server through home Wi-Fi, and the pet owner controls feeding the pet from their mobile phone, using a predefined dashboard. An intelligent pet feeder circuit is built on a microcontroller-based NodeMCU board interfaced with cloud. The servomotor with several sensors is for notifying the owners to refill the food container and change the fish tank's water. Owners can monitor their pets on their mobile phones. This intelligent pet feeder system is involved in the IoT's ideas and satisfies pet owners' needs, who are out for work without any trouble.

Keywords Raspberry Pi · Python · Cloud · Internet of Things (IoT) · Microcontroller · NodeMCU

1 Introduction

Automation is a technological revolution in recent years and plays an integral part in our daily life gadgets and industrial levels. Consumers are more attracted to the automated devices. These devices are mainly used for time saving. Automation works with programmed software and electronic devices; with such machines, automation

R. H. Aswathy (✉) · G. Dhanush Eswar · S. Arnold Jerome · R. Abishek
Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_24

becomes more convenient for the consumer. Most pet owners love to spend time with their pets, but some of them only have time and patience to feed their pets, and someone do not have the time to care on pets. The IoT and automation come in handy to develop a device that fulfils pet owners' work without causing any harm to the pets. These pet feeders work to give a fully personalized experience and programme the schedule for feeding their pets with exact amount of food dispensed.

The IoT provides methodologies and sets of standards to associate objects in real life. Pets require care and nourishment to keep them safe and healthy, but due to the busy routine of working or studying until late hours, it is not easy to look after their pets for some owners. More often, the pets get stressed because of not getting food properly or not getting proper amounts of food. Sometimes the owner may not be available for an extended period due to a family trip or any other important work. The owners make a terrible decision to feed their pet with more food at once till their return. Overfeeding is a bad practice of feeding a pet. With a pet feeder's help, the owner can feed their pet from anywhere they are and can follow the food chart for their pet. This pet feeder comes with different ways of controlling remotely using the phone by triggering the device using an app or using the Google Assistant to trigger the feeder through a voice command. The sensors make this device more efficient in many ways. Feeding can be done in timely basis which can set up what time using the phone, the command will be triggered automatically and notified with a message. The owner can also control the amount of food served at a time. The pet feeder will be more beneficial for the pet owners who have a busy schedule and leave home for an extended period. The pet owners are comfortable leaving their pets at home because they can monitor and can feed their pets wherever they are through their smartphone.

2 Literature Review

The wireless technology called "IoT" gained widespread attention in academic, research and industry. It is a unique technique where each object provided with a unique identifier where objects communicate, correlate and exchange data. The IoT permits the user to optimize the electronic devices remotely [1] with various techniques and applications. Communications technology growth, like WiMax, Wi-Fi, IPV4 and IPV6, is the rudimentary facet of IoT [2]. The IoT is a ubiquitous technology that works based on a set of standards and protocols that integrate the objects in the real world [3]. IoT is used in many applications in the automation industry, education, logistics, health care and home applications. An automation technology connects local and global network using embedded boards and sensors. The author [4] discussed the protocols and algorithms in the Internet of Things. The author [5] designed the intelligent pet feeder that feeds the food remotely, and the remaining food in the container is intimated to the user. Various researches [6, 7] discussed that the aquarium conditions are not appropriately maintained. Few kinds of research give an intelligent mechanism for maintaining the aquarium in an efficient way [8],

which control the aquarium through the online, status of food container and owner to customize the time of feed. The author [9–12] records the timing for feeding the fish, and it is monitor through UI-based environment.

3 Proposed Methodology

3.1 Workflow of the Proposed Methodology

The workflow of the pet feeder starts with the user. When the user logs into the Blynk app using a mobile phone, the API call is sent to the Blynk API. The API will authenticate using the token ID given to the user at the time of installation and get the NodeMCU status, pH value of the water and available food quantity in the container using the photoelectric sensor. The pH sensor will calculate the water's pH value and notify the user when the water needs the change. The photoelectric sensor calculates the available quantity of the food in the container. It uses the photoelectric light to pass on the food inside the container and calculate the amount of food available using the reflection of the light from the food matter. The available amount of food is calculated and displayed through the application. The Blynk app will send an alert to the user if the food quantity reaches the threshold value. Figure 1 shows the workflow of proposed method.

The API would return an "Up" signal if the NodeMCU is connected to the Wi-Fi network. Else it will return "Down" as the status. The user provides two components: a slider that allows the user to set the feeding quantity and a trigger button that sends an API call to the Blynk server, which is available over the cloud network. The NodeMCU will wait for the signal from the API after the connection establishment.

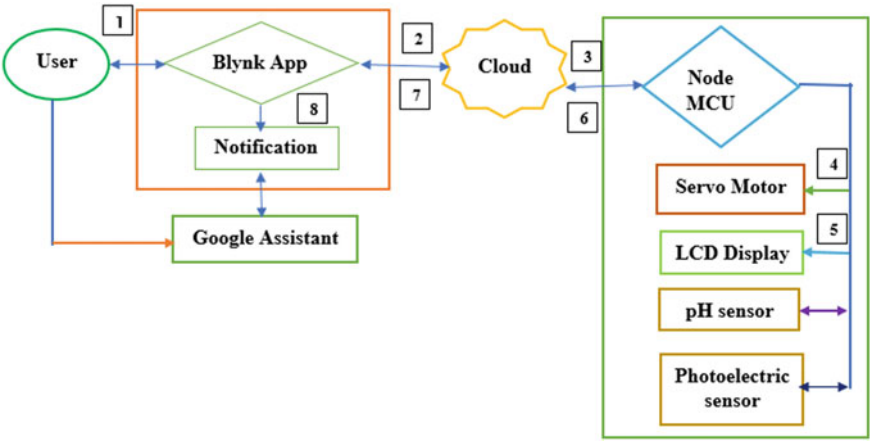


Fig. 1 Workflow of the proposed system

Whenever the user triggers the button, the API call is sent to the Blynk server, and the API will trigger the NodeMCU, which is connected to the Wi-Fi network. The Google Voice Assistant can be used instead of triggering a button, and the proper command will be assigned to it using an open-sourced platform called IFTTT. Whenever the user wants to feed the pet, the user will be saying the commands to the Google Assistant, and it will trigger the API call to the Blynk API, and the API will trigger the NodeMCU. Whenever the NodeMCU receives the signal, it will send the signal and the value for the motor's number of rotations. The LCD will display the number of the rotations and the last feed time. The pH sensor and the photoelectric sensor continuously return the user's value when connected to the Internet.

3.2 Architecture of the Proposed Methodology

The proposed system of the pet feeder relies on multiple sensors and the microcontroller. The microcontroller NodeMCU is connected to the various sensors that help the user feed the pet endlessly, and the user can monitor the pet feeder's requirements. The pH sensor senses the water's pH value, and the photoelectric sensor will sense the amount of food available in the container. The servomotor will serve food to the pet, and the LCD will show the last feed time to the user. The entire set-up is connected to the Internet through the open-source cloud platform called Blynk. The user will have the app created by Blynk and trigger the button to feed the pet. The notification slide in the app shows the food's status in the container, and the water has a monitored pH value. The Google Assistant can be used instead of triggering a button, the user can say some command, and the user can feed the pet.

The proposed methodology's architecture runs with diverse phases such as measurement unit, actuator unit and software unit, and the architecture is shown in Fig. 2.

3.2.1 Measurement Unit

The measurement unit is divided into two phases: weight scale and food container level indicator. The weight scale indicates the weight of the food available in the food container, whereas the level indicator displays the food availability in the container. The container's quantity is measured using an ultrasonic sensor and indicated to the user through the Blynk app. The pH level of the water can be measured using the pH sensor based on water conditions. The tab in the Blynk application shows the level of both measurements.

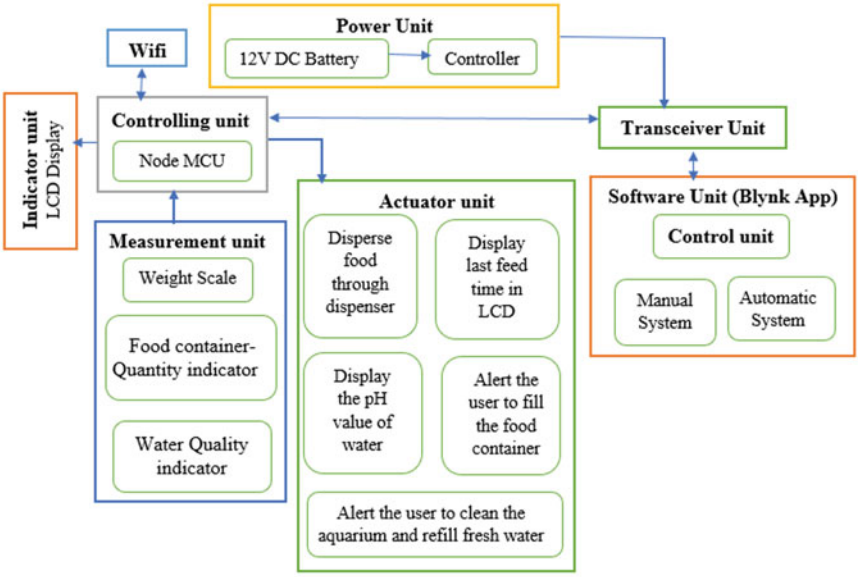


Fig. 2 Architecture of aqua nurturing system

3.2.2 Actuator Unit

The taxonomy of actuator units is to disperse food, display last feed through LCD, display the pH value of water, alert the owner to fill the food container and alert the user to clean the aquarium and refill the water.

3.2.3 Food Dispenser Mechanism

The feed is based on the size, number and quantity required for fishes present in Aquarium. Table 1 shows the consumption chart of the fishes.

Table 1 Food requirements for aquarium fishes

Types of fishes	Average weight (pounds)	Required energy per day (DE/HE)	Food weight (grams)
Gold fish	0.4	11.72 DE	0.362
Guppy	0.1	13.10 HE	0.090
Angel fish	0.3	12.97 DE	0.272
Red head cichlid	0.5	1.55 DE	0.453

Table 2 Threshold indicators for food container

Level (%)	Intimation
0 < 25	Fill the food immediately
25 < 50	Fill the food
50 < 100	Food is sufficient

Disperse Food

The user has to set the quantity of food to be dispersed in the aquarium depending upon the fishes' number.

Display the Last Feed Time

The user can view the last feed time through their mobile device and the LCD placed at the tank's top. So, the feed performed either remotely or physically.

Display of pH Level

The user can view the water's pH level indicator in the aquarium through their mobile device through the Blynk application. This application alerts the user to change the water in the aquarium depending on the pH level.

Food Container Alert

By using the ultrasonic sensor, the level of the food in the container is measured. The level in the food container fixed using threshold values is shown in Table 2.

Here the threshold value is fixed at three levels from 0 to 25, and the user has to fill the container with the food immediately. If the value lies between 25 and 50 percentage, the system will alert every 12 h to fill the container. If the threshold value is between 50 and 100, food is sufficient, and no alert is given to the user.

pH Value Indicator

The pH sensor used to display the pH value of the water in the aquarium. The average pH value of different fishes in water is shown in Table 3. Water hardness shows the pH value. The convenient water level lies between 6.5 and 7.5

The protons movement stops in the fish body when the pH value is less than 5.5. The ammonia level in water is 2.0 ppm if the aquarium water gets affected by the increase in temperature and pH value. The ammonia level also affects the aqua bodies. The pH value for different fishes is shown in Table 3.

Table 3 pH value for different fishes

pH range	Fishes
6.5–7.0	Silver dollar fish, hatchet fish, angel fish
6.0–6.5	Tiger barb, harlequin, clown fish
5.8–6.2	Neon tetra
5.0–7.0	Plecostomus
6.5–7.0	Zebra danio
7.0–7.5	Goldfish

Alerts the User to Clean the Aquarium and Refill Freshwater

The pH sensor in the fish tank alerts the user to clean the water. The water level indicator sensor refills the water when the water level is low due to evaporation. The aqua bodies will be affected by poor water conditions, leading to disease and a short life span. The actuators' automatic refill of water is when the model is connected with the water tank/pipes. The water level sensors monitor the level of water in the fish tank. Intimation is given to the user during the reduction of water, and it automatically refills the tank.

3.2.4 Control Unit

The food dispenser has a control unit. The control unit act as a scheduler, that schedule schedules the time for feeding the fish, and quantity adjuster frequency of the food dispensing and quantity of flowing food are also defined and managed.

3.2.5 Software Unit

The software used in the proposed work is the Blynk application which shows the food level, last feed time pH value of water, water evaporation rate, the water level in the aquarium and auto-refill of water. The food dispenser is designed with the feeder threshold value, and once the threshold value gets changed, the intimation is received through the application. The user can also view the last feed time and number of feeds in a day. The pH indicator sends the alert message to the user when there is a variation in water's pH level. The water level sensor is used to maintain the water level, which alerts the user of high evaporation rates. The user can also view the level of water and refill the water through the Blynk application.

3.2.6 Historical Database

The real-time chart is used to analyse the amount of food dispersed and the current pH level of water. This measurement helps the user to monitor the remaining food

in the container and the current pH level of water. This intimation helps the user to manage their aquarium in a better way.

4 Performance Evaluation of the Proposed System

4.1 Flow Rate of Food

The quantity of food released to the aquarium tank is calculated using flow rate. The flow rate is denoted as f

$$f = A * V$$

The servomotor attached in the dispenser Ga is rotated from 0° to 180°, a complete rotation. The area (A) is defined as

$A = \pi * r^2$, and we know that the radius (r), $r = 0.8$ cm

$$A = 3.14 * (0.9)^2 = 3.14 * 0.81 = 2.34 \text{ cm}^2$$

Manning's equation is used to calculate the gravitational flow rate of the container. The flow rate is defined as

$$F = A * V \quad \text{where} \quad v1/n * R^{(2/3)} * \sqrt{3}$$
$$F = A * V = \left(A * 1/n * R^{(2/3)} * \sqrt{3}\right)$$

where f is the flowrate, V is the velocity, A is an area, n is the manning's constant, R is the radius, and S is the slope of the channel.

Manning's coefficient is 0.012 cm, and the radius is calculated as

$$R = A/P,\ R = 2.54/18 = 0.14$$

The radius (R) is 0.14. The area is 2.54, radius is 0.14, slope value is 0.012, and Manning's coefficient is the flow rate of small outlet which is 5.712 cm^2.

4.2 Comparison with Existing Methods

Components and functionalities are shown in Table 4

Table 4 Components and functionalities used in other related schemes

Components and functionality	Chen et al. [11]	Dolan et al. [12]	Encinas et al. [13]	Lin et al. [14]	Raju et al. [15]	Dupont et al. [16]	Aqua nurturing system
Sensors	pH, DO, temperature	pH, CO_2, O_2, NH_3	pH, DO, temp	pH, DO, TDS, temp	DO, Salt, NH_3	pH, DO, temp	pH, photoelectric, ultrasonic
Actuators/effectors	Feeder, heater, light	NO	Feeder, fan, light, air	Feeder, fan, light, air	Light	Air pump	Servomotor LCD display
Control manual/automatic	Threshold values	NO	No	Manual	Manual	Manual	Automatic by fixed threshold values
Feeder automatic/manual	Manual	No	No	Automatic	No	No	Automatic feeder and food filler
Monitoring remote/manual	No	No	No	Yes	No	No	Remote and manual
Control boards and components	HSP430	N/A	Arduino UNO	Arduino UNO	Raspberry Pi	Raspberry Pi	NodeMCU ESP8266 microcontroller, sensors

5 Results and Discussion

This section shows the system's result, which is connected to the Internet via the Blynk server. The pet feeder is enabled using the NodeMCU and Blynk cloud server. Blynk app, which the owner uses, can control and monitor the amount of food dispensed and when the food is dispensed. When the owner clicks the Blynk app button, it triggers the installed code on the NodeMCU. The code works and signals the NodeMCU to operate the servomotor, and the container's food is dispensed to the pet.

5.1 Blynk Application

The data is updated whenever the food is dispensed, and these data are collected and uploaded in the Blynk cloud server and get uploaded in the Blynk app. The data sheet contains the amount of the food dispensed, how many times the servomotor rotates, and the number of times the food is dispensed. With sensors' help, the quantity in the food container also gets measured and updated in the cloud and displayed in the Blynk App which is shown in Fig. 3.

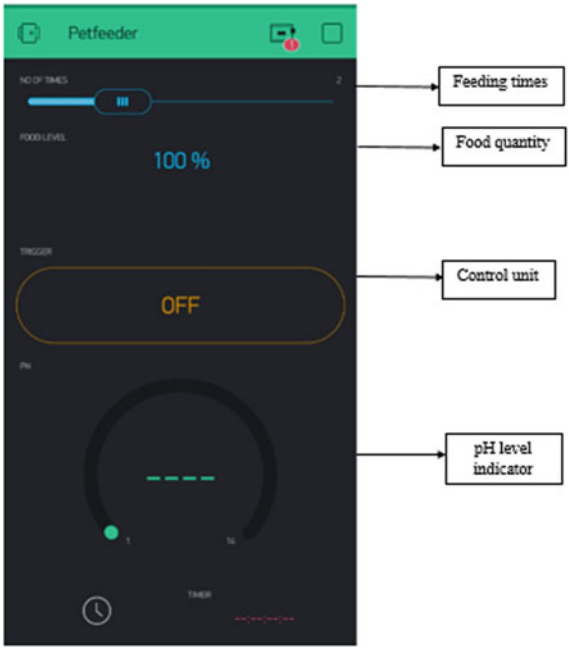


Fig. 3 Output of aqua nurturing system

Fig. 4 LCD output of last feed time

5.2 Google Assistant

The Google Assistant is also a way the pet owner can feed their pet. The pet owner can use Google Assistant to give a command through voice message or text message to feed their pet, the Google Assistant connected to the server triggers the code in the NodeMCU, and it operates the servomotor to feed the pet. After dispensing the food, the NodeMCU sends a reply message to the Google Assistant.

5.3 LCD Display

The LCD screen is fixed over the fish tank which display the last feed time. The LCD output of last feed is shown in Fig. 4.

6 Conclusion

The usage of the Blynk cloud and sensors with the pet feeder is new to implement. The measurement unit is divided into two phases: weight scale and food container level indicator. The weight scale indicates the weight of the food available in the food container, whereas the level indicator displays the food availability in the container. The measurement unit measures the feed level, and it requires a lot of research and reading. The display unit shows the last feed through LCD, display the pH value of water, alert the owner to fill the food container and alert the user to clean the aquarium and refill the water. This IoT-based aqua nurturing system implemented using Arduino provided low cost, which used to fulfil the customer's requirement with the desired result.

Acknowledgements We acknowledge all referees who have given useful suggestion to enhance this work.

References

1. Suresh P, Daniel JV, Parthasarathy V, Aswathy RH (Nov 2014) A state of the art review on the internet of things (IoT) history, technology and fields of deployment. In: 2014 International conference on science engineering and management research (ICSEMR). IEEE, pp 1–8
2. Qu Y, Chan P (April 2016) Assessing vulnerabilities in bluetooth low energy (BLE) wireless network based IoT systems. In: 2016 IEEE 2nd international conference on big data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing (HPSC), and IEEE international conference on intelligent data and security (IDS). IEEE, pp 42–48
3. Dong HJ, Abdulla R, Selvaperumal SK, Duraikannan S, Lakshmanan R, Abbas MK (2019) Interactive on smart classroom system using beacon technology. Int J Electr Comput Eng (2088–8708) 9(5)
4. Suresh P, Koteeswaran S, Malarvizhi N, Aswathy RH (2018) Internet of things (IoT): a study on key elements, protocols, application, research challenges, and fog computing. In: Handbook of research on cloud and fog computing infrastructures for data science. IGI Global, pp 124–148
5. Adriansyah A, Wibowo MA, Ihsanto E (2016) Design of pet feeder using web server as internet of things application. In: International conference on electrical engineering and informatics. pp 151–156
6. Salim TI, Haiyunnisa T, Alam HS (2016) Design and implementation of water quality monitoring for eel fish aquaculture an examination of microbubble aeration. Proc Int Symp Electron Smart Devices (ISESD) 29–30
7. Encinas C, Ruiz E, Cortez J, Espinoza A (April 2017) Design and implementation of a distributed IoT system for the monitoring of water quality in aquaculture. In: 2017 Wireless telecommunications symposium (WTS). pp 1–7
8. (Feb 2019) Int Res J Eng Technol (IRJET) 6(2). e-ISSN: 2395-0056. www.irjet.net p-ISSN: 2395-0072
9. Meshram S, Meshram G, Rokde B, Kapse R, Hedaoo O, Mandhata C. Fish feeder using internet of things
10. Xingqiao L, Jiao G, Feng J, Dean Z (July 2008) Using MATLAB image processing to monitor the health of fish in aquiculture. In: 2008 27th Chinese control conference. IEEE, pp 677–680
11. Chen JH, Sung WT, Lin GY (Oct 2015) Automated monitoring system for the fish farm aquaculture environment. In: 2015 IEEE international conference on systems, man, and cybernetics. IEEE, pp 1161–1166
12. Dolan A (2015) The effects of aquarium size and temperature on color vibrancy size and physical activity in bettasplendens. Maryville College, Maryville, TN, USA, Tech. Rep, 53309811
13. Encinas C, Ruiz E, Cortez J, Espinoza A (April 2017) Design and implementation of a distributed IoT system for the monitoring of water quality in aquaculture. In: 2017 Wireless telecommunications symposium (WTS). IEEE, pp 1–7
14. Lin Y-B, Tseng H-C (2019) FishTalk: an IoT-based mini aquarium system. IEEE Access, pp 1–1. https://doi.org/10.1109/ACCESS.2019.2905017
15. Raju KRSR, Varma GHK (Jan 2017) Knowledge based real time monitoring system for aquaculture using IoT. In: 2017 IEEE 7th international advance computing conference (IACC). IEEE, pp 318–321
16. Dupont C, Cousin P, Dupont S (June 2018) IoT for aquaculture 4.0 smart and easy-to-deploy real-time water monitoring with IoT. In: 2018 global internet of things summit (GIoTS). IEEE, pp 1–5
17. R Abdulla SK Selvaperumal A Al-Adilee T Maythem K Ahmed 2020 IOT based pet feeder Test Eng Manag 83 269 279
18. Kondapalli KJ, Sanepu VR, Kothapalli BS, Peketi SPR (April 2019) Automatic pet feeder using internet of things. J Emerg Technol Innovative Res 2019 JETIR 6(4)

Advanced Automated Crop Predictive Irrigation System

P. Suresh, K. Sujith, G. Thilak, and K. Sakthivel

Abstract In today's agriculture, water management is an essential requirement in arid and semiarid regions. The water irrigation system designed to decrease water wastage and dispatch the proper water amount. This irrigation method manages the water flow based upon plants' species, increasing the yield. The proposed system will help to create a real-time monitoring and low-cost irrigation system offering stable remote access from anywhere, anytime. This result leads to a more efficient method of agriculture. Live time sensors collect data such as dampness, soil moisture, soil temperature, air temperature, and sufficient data for the irrigation intervals as per the plant's requirements. This method allows the farmers with efficient water management and also to produce healthy crops with better yields. Our irrigation system is adaptable to any irrigation mode. It provides the best and more efficient technique to grow a plant of different kinds with different water requirements. This work mainly focuses on plants' consistent growth in the field concerning water requirements and crop monitoring. This project aims to mitigate water scarcity and increase the growth and yield of crops. Approximately, 70% of the total volume of water used in the world is used for irrigation. The water requirements in arid and semiarid regions are too high, so that irrigation management is a primary model for sustainable and economically profitable crops.

Keywords Automation · Control systems · Measurement · Portable radio communication · Sensors · Irrigation · Master–slave model

1 Introduction

Agriculture is the rudimentary source of income for the largest population in India. Agriculture contributes a significant part to the Indian economy. However, technological enhancement and their adaptability of the technology in agriculture needs to improve in India. The integration of new technology like the Internet of things will

P. Suresh (✉) · K. Sujith · G. Thilak · K. Sakthivel
Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_25

help make it easy to manage and supervise from any place at any time. Our future is moving toward the scarcity of water for drinking, agriculture. It is essential to create a system to reduce water wastage in agriculture because we use more than 78% of freshwater. As there are many soil types and regions with different temperatures, it requires adaptable irrigation. Interestingly, around 57% of the water meant to use for irrigation is lost due to various reasons such as evaporation, transpiration, land runoff, or simply inefficient primitive usage methods.

The initiatives to improve the agriculture sector are taken by the government because of the need and development of the country's economy; the government is providing help to farms by online notification and mobile messaging alert services to farmers regarding the agricultural queries, agriculture vendor's information. Moreover, implementing new schemes for the welfare of farmers, it found based on the survey done by the government of India that agriculture contributes 27% of the country's GDP and also provides an employment rate of 70% of the Indian population.

The agriculture field overcomes water deficiencies for irrigation and restricted availability of lands and has to meet expanding consumption needs for the world population. With the help of innovation in IoT with agriculture, applications are overcoming these issues to increase the quality, quantity, sustainability, and cost efficiency of agricultural production. In this rapidly growing growth of the global population, agriculture has become the more important sector to meet human race needs. Irrigation is considered one of the essential requirements for agriculture. Every year, the consumption of water is increasing rapidly than rainfall. It becomes difficult for farmers to conserve freshwater for achieving increasing demand with high yield. With the tremendous growth of the technologies in the field of agriculture, the farmers are provided with new options to implement irrigation [1] in their agriculture field like manual control of irrigation process in which they irrigate the land by manual control or based on a regular interval.

According to a statistics survey done by the researchers, agriculture uses about 85% of the available freshwater resource globally. The amount of water consumption will increase continue to meet the population consumption with increased demand for food. To overcome this issue, we have to enhance our technology in the field of irrigation and agriculture. The advanced automated predictive irrigation system is based on the Internet of things. When deployed to the farm, the soil moisture sensor will start to collect all the readings along with temperature and humidity from the farm and sent them to the server for further analysis. The moisture identification sensor shows the potential difference between the two roads [2]. Based on the data, the server analysis and provides an efficient decision to the main microcontroller, which further starts to initialize the irrigation process by locating the position of the field to be irrigated and setting the connection of the pipe toward the particular location and starts pumping water toward the filed. By using the Internet of things and sensor network technology, we can minimize the wastage of water up to 70% and maximize the yield and productivity of the crops. Hence, it can significantly improve the utilization of water and can increase water productivity.

The enhancement in the field of the Internet of Things has led to advancement in the field of agriculture; by using smartphones and laptop devices, users can monitor

and also control from any part of the world. The Internet of things is primarily focused with interconnected communication devices from any location in the world. Internet of things is a type of communication, where the sensors gather the data and sent it to a microcontroller or microprocessor, and these processors connected to the Internet through integrated or external communication modules, which in turns helps to establish a communication the channel between the microprocessor or microcontroller with the user devices to manage and perform a set of functionalities. To establish a communication the channel between the processing unit and the user device, both of the device should have a communication module along with their devices. The communication modules are mostly WI-FI modules, Bluetooth modules, or any other wireless communication systems [3].

2 Literature Survey

The literature review begins with gathering background resources in the field of precision agriculture in order to accomplish the proposed work. The sensors like soil moisture, humidity, and temperature placed near the root zone of plants and the sides of the field with some space. The data transmitted to the server is further analyzed and stored in the database. The irrigation method is used to monitor and predict the water requirement for upcoming days and also to characterize the variability of the farm field [2, 4, 5], and this viewed as well as controlled using user interfaces like Android or web app platforms. The microprocessor is programmed to set the threshold of soil moisture sensor to limit the water flow, when there is an issue with connectivity to the Internet, based on the crop, soil type, and also the climatic conditions and also can be manually configured through the UI if needed.

The UI application displays climatic factors such as temperature, humidity, soil moisture, water consumption,water threshold, and water level in the tank, making the automated irrigation process much more efficient and precise. The remote monitoring of agricultural fields is done using a wireless sensor which interconnected to all the slave modules and master modules. All the data that the slave modules have gathered have transmitted to master module, which classifies and consolidates the data so that this data can be transferred to the cloud-based server for further analyses and monitoring purpose. Based on this analysis, a threshold has set by the server and sent back to the master module, which further set to the slave modules, which are present in that particular field. The initial investment may be higher when we integrate with the agricultural field. However, when we consider the amount of freshwater consumed, monitoring, and labour costs, the investment will quickly return in the long run.

3 Architecture of the Proposed Work

The model is based on the master–slave architecture so that every module can work efficiently. The master module works as the central processor with more computing power than slave modules. Only the master modules have access to the Internet and configured with the cloud-based server. The slave modules are being designed with a low configuration microprocessor to reduce implementation costs [6]. The main work for slave modules is to gather the data regarding soil moisture level, temperature and humidity from the sensor associated with the salve microprocessor. The gathered data, along with slave module ID, is transferred using the wireless communication modules to master modules. The master module gathers all the data from the slave modules, which are then classified based on the slave ID and transmitted to a cloud-based server. Based on the data gathered by the server, the analyses done on the server. The server responds with commands stating the status of moisture content in the field and the slave ID. The slave ID is a unique identification code used to identify the location of the slave module in the field. Once the master module has received the data, it initiates the irrigation system to irrigate at that particular field location [7] with a required amount of water is released based on the crop.

3.1 Slave Module Architecture

The slave module made using a lower power microprocessor to reduce the cost of implementation in real-time monitoring [8]. The moisture sensor, humidity sensor, temperature sensor and wireless communication modules [7] connected to the microprocessor. The power supplied to the microprocessor uses the solar panel to reduce current wires' overlaying over the field. The power supply is regulated and supplied directly to the slave microprocessor unit. From there, the power is distributed to the sensor modules, respectively. The slave microprocessor is allocated with a unique identification code to identify the microprocessor superlatively.

Architecture diagram in Fig. 1 shows the slave system which will be deployed to each and every part of field to gather the required information from soil. The data are monitored and collected continuously with a specific time delay. This collected

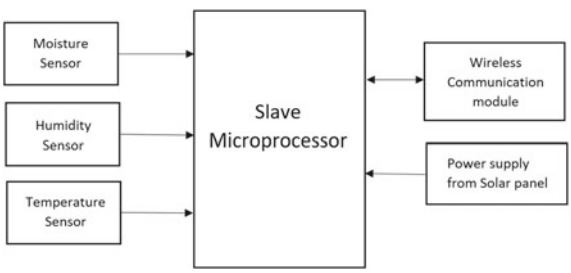


Fig. 1 Architecture of slave system

information regarding temperature, moisture, and humidity has sent to the master module for further processing through the wireless communication channel. The slave microprocessor collects the processed data. The power supplied to the microprocessor from the solar is regulated using the voltage regulator based on the process used. If the wireless communication module draws more power than the processor's capability, it can also be connected externally to the solar supply unit.

3.2 Master Module Architecture

Architecture of master system in Fig. 2 is deployed in field to fetch data gather by the slave modules and do the necessary decisions. The master module consists of a powerful master microprocessor to communicate with the salve modules through an optimal wireless communication module [9] to gather all the data from the respective slave modules. The following modules include a wireless communication module, Internet module, water diverter module and water pump in the irrigation system connected with the master module. Only the master microprocessor is provided with the Internet to communicate to the cloud server. The wireless communication module used to communicate internally to the slave modules without the Internet. The water diverter module is a stepper motor connected to the water pipe channel pump to water to a specific pipe controlled by the master processor. Once the location identified, the water pump initiated to pump the water to a specific location.

The data gathered by the slave are fetched and classified based on slave ID, further after classification. The data are converted into a meaningful format for precision farming [10, 11] and send to the cloud server for further analysis. Based on the server's response, the master microprocessor decides to select the pipe channel based on the slave ID, the location tracked, and the channel established with the water diverter module's help [12]. After this, the pump initiated to start pumping the water to that particular field [13] with a specific amount of water with analysis.

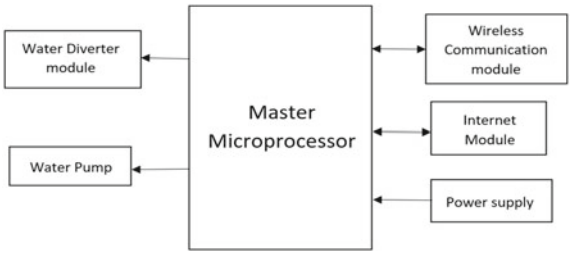


Fig. 2 Architecture of master system

3.3 Master Module Connection with Slave Module Architecture

To achieve efficiency and get more control over the field, we have to split the farmland into different pieces so that the slave modules can be fixed in the system by every part of the land. The range is being defined based on the modules used to gather the field's data, which may range based on the sensors' capability integrated into the module based on the user's budget and the field size. Slave made of microprocessors with low computing power to reduce deployment cost. The slaves are interconnected to measure the soil moisture sensor, humidity sensor, temperature sensor, and wireless communication module. The slaves allotted with a unique id to locate in the field, which interns help run the custom commands for each slave to prove the farmers' access to grow different crops in each area separately. The need for custom configuration is to grow other crops. Based on the type of crop configured, the master processor verifies the water threshold with field and the calculated point based on the crop variety grown, soil type, and climatic conditions. The area is divided into a two-dimensional matrix so that the slaves can be placed in each matrix places [14]. The slaves and master modules connected through wireless connection [15]; the redundant wired connection can be dismissed. A slave configured separately to give the farmers the flexibility to grow a different crop variety in the same agriculture using single master–slave configurations. The power supply to the slave modules provided using the solar panels to avoid unnecessary current wire running throughout the field and use the renewable energy source to provide an eco-friendlier system. The configuration mode from the user interface provided using the web app or Android application. The design validated by the server and provides suggestion to the farmers if required. Once the server approves the structure, the user is compiled with specific commands and transferred to the appropriate master microprocessor. Once the designs are set and required information provided, the automated irrigation process has initiated.

Architecture of master and slave system connected with other modules to automate the irrigations process for more efficiency is shown in Fig. 3. Each slave is a separate unit placed in a field, and the field is divided into a two-dimensional matrix so that the slaves can be placed in each matrix place [14]. The slaves and master modules are connected through wireless connection [15], so that unwanted wired connection dismissed. Each slave configured separately to give the farmers the flexibility to grow various crops in the same agriculture using single master–slave configurations. The power supply to the slave modules provided using the solar panels to avoid current wire running throughout the field. The configuration from the user interface provided using the web app or using the Android application. When the configurations set and required information provided, the automated irrigation process initiated.

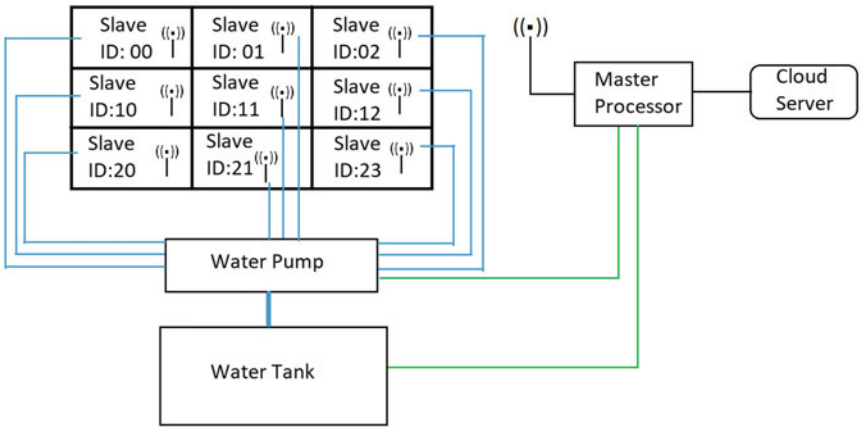


Fig. 3 Architecture of master and slave system connected with modules

3.4 Abstract Data Flow

Abstract architecture of the system data flow from the agriculture field to the end users and vice-versa is shown in Fig. 4. The data are collected from the agriculture field using the sensor modules with the help of slave processors. Then they are forwarded to the master microcontroller to classify the data from each slave module. Further, data are forwarded to the cloud-based server to analyze the data, and a response is provided to the microcontroller to initiate the process based on the server's analytical response, respectively. During this process, these data are stored in the server to monitor water consumption. Based on this data, farmers can monitor the readings, analyze their crop, and customize the settings and manually run the water pump from the user interface.

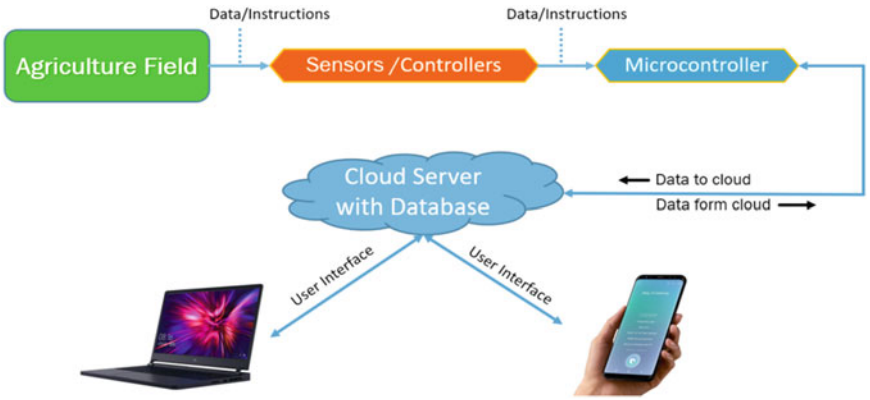


Fig. 4 Abstract architecture of the system data flow

3.5 Monitoring

Monitoring is considered the most important aspect of analyzing the data and getting the most out of the project. The gathered data are investigated based on the water consumption, temperature, moisture level, and crop verification and growth, which are the most critical parameters affecting crop growth. This data used to predict the water threshold value to be set by the server for processing. Based on the change in temperature, moisture, and humidity, the water threshold level has varied to provide the system's efficient outcome. It also creates space for the users to view the water consumption. The author [16] suggested the precision irrigation based on the agriculture land variations.

4 Experimental Setup with Results and Discussion

An experiment conducted to understand and analyze the difference between the simple automated irrigation model and the proposed system in a small field. During this experiment, the temperature is noted between 37 °C and 40 °C by the temperature sensor. Both systems started at the same time. The system is provided with a tank of 10 L separately to individually analyze the amount of water consumption. The experiment is conducted for 10 h to calculate the efficiency of both systems.

4.1 Simple Automated Irrigation Model

For the research process, we have created a modal prototype to check whether the system can provide the required efficiency as described in this paper. For this, we have deployed the system in a small field, and for the first trial, we used the simple automated system for irrigation, which starts when the threshold reaches 20% in the soil moisture sensor in the field and following are the analysis graph:

There is no water in the initial state field, so the system starts and pumps the water until all the soil moisture sensor threshold is satisfied. We can see that the amount of water consumption is primarily near the graph's initial value. If we are irrigating from one end of the field and soil moisture sensors at the other end, then the motor will pump until the last soil moisture sensor threshold is satisfied. The ten litters consumed before 10 h of the experiment (Fig. 5).

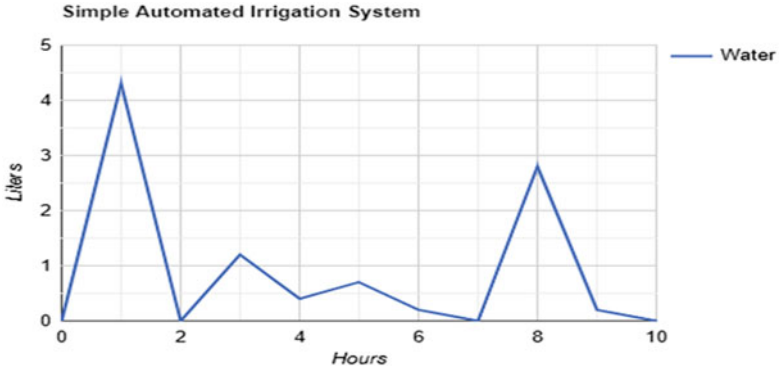


Fig. 5 Analyses of water consumed by simple automated irrigation system for a period of 10 h and temperature ranges from 37 °C to 40 °C

4.2 Proposed Model

We deployed our proposed model and the analysis made in the same agricultural field depicted in Fig. 6.

The line graph represents the overall process of the system, the initial stage reading start with 0, and by turning on the system, the water amount reaches high for 5 minutes before returning to normal. When a specific part of the field's soil moisture sensor detects a water level threshold of less than 20% for five minutes, it automatically begins to pump water to a specific field in a specific amount based on the crops' needs. We can see that only less amount water consumed for each initialization of the process. In the process, which ran for 10 h, approximately, 4.6 L of water consumed, respectively.

Based on both the analyses, we can see that the water consumption calculated to 10 L for the simple automated system, and the proposed automated irrigation system has consumed 4.6 L for 10 h. So, when we calculate the percentage of water

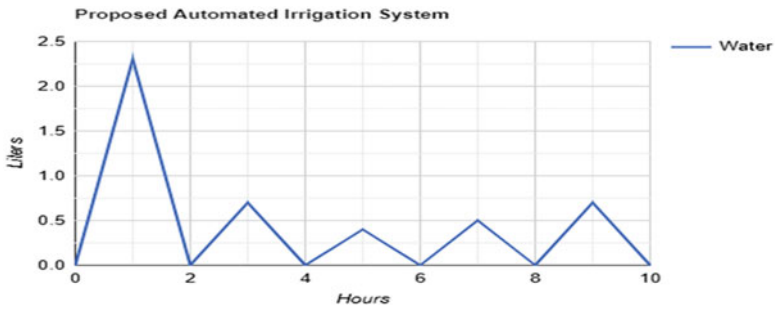


Fig. 6 Analyses of water consumed by proposed predictive automated irrigation system for a period of 10 h and temperature ranges from 37 °C to 40 °C

consumption of the proposed model with a simple automated irrigation system: (4.6/10) *100 = 46% of water usage, and when we calculate the efficiency of the proposed system, we get 100 − 46 = 54%. The proposed system is more efficient by 54% more than the simple automated system.

5 Conclusion

The proposed system designed to automate irrigation which mitigates water wastage in the agriculture field and provide more yield with high productivity. Since the traditional agriculture method consumes freshwater quantity, we need a flexible system to reduce the water wastage to cultivate crops on the farm. The analysis shows that the amount of water consumed by the existing system is 54% more than the proposed system. Implementing the proposed system assured that we could have an efficient irrigation system with precision cultivation of crops to provide a highly efficient and high yield to significantly impact the minimum use of water resources with an efficient and much more stable form of agriculture production.

References

1. Perry CD, Dukes MD, Harrison KA (2004) Effects of variable-rate sprinkler cycling on irrigation uniformity. 2004 ASAE annual meeting. Am Soc Agric Biol Eng
2. Farahani HJ, Buchleiter GW (2004) Temporal stability of soil electrical conductivity in irrigated sandy fields in Colorado. Trans ASAE 47(1):79
3. Zhang Z (2004) Investigation of wireless sensor networks for precision agriculture. 2004 ASAE Annual Meeting. Am Soc Agric Biol Eng
4. Drummond PE, Christy CD, Lund ED (2000) Using an automated penetrometer and soil EC probe to characterize the rooting zone. In: Proceedings of the 5th international conference on precision agriculture, Bloomington, Minnesota, USA, 16–19 July, 2000. Am Soc Agron
5. Jabro JD, et al (2006) Characterization of spatial variability of soil electrical conductivity and cone index using coulter and penetrometer-type sensors. Soil Sci 171.8:627–637
6. Wall RW, King BA (2004) Incorporating plug and play technology into measurement and control systems for irrigation management. 2004 ASAE annual meeting. Am Soc Agric Biol Eng
7. Kim Y, Evans RG, Iversen WM (2008) Remote sensing and control of an irrigation system using a distributed wireless sensor network. IEEE Trans Instrum Meas 57(7):1379–1387
8. Wei Z, Jinying Y, Feng Y, Rupeng L (2011) Study on agricultural distance monitoring and diagnosing integration platform based on XMPP [J]. Chin Agric Sci Bull 11
9. Kim Y, Evans RG, Jabro JD (2005) Optimal site-specific configuration for wireless in-field sensor-based irrigation. In: 26th annual irrigation association international irrigation show. Phoenix, AZ
10. Irmak A, Jones JW, Batchelor WD, Paz JO (2002) Linking multiple layers of information for diagnosing causes of spatial yield variability in soybean. Trans ASAE 45(3):839
11. Ahmad IS, Reid JF, Noguchi N, Hansen AC (July 1999) Nitrogen sensing for precision agriculture using chlorophyll maps. In: ASAE meeting presentation. pp 18–21
12. Evans RG, Iversen WM (Nov 2005) Combined LEPA and MESA irrigation on a site specific linear move system. In: Proceedings 26th annual international irrigation show. pp 6–8

13. Li J, Tu W, Fu J, Wang Y (2010) A handheld testing system for irrigation system management. In: International symposium on information and automation. Springer, Berlin, Heidelberg, pp 418–423
14. de Miranda FR (2003) A site-specific irrigation control system. 2003 ASAE annual meeting. Am Soc Agric Biol Eng
15. Kumar Sahoo P, Chiang M-J, Wu S-L (2016) An efficient distributed coverage hole detection protocol for wireless sensor networks. Sensors 16.3:386
16. Al-Sammarraie M, Ahmed A (2021) New irrigation techniques for precision agriculture: a review. Plant Archives. 21:972–5210. https://doi.org/10.51470/PLANTARCHIVES.2021.v21.S1.275

Applying Reinforcement Learning in SDN: A Packet Re-transmission Policy

Vasantharaj Karunakaran and Angelina Geetha

Abstract The main purpose of working in open flow and is its specification is very important in software-defined networks. Normally in flow control and error control protocol, the main role will be forwarding the packets from source to destination, but it seems to be vulnerable. In open flow specification, the logical switch contains at least one stream table or gathering table that performs queries, bundles sending to an outside Controller. Open flow particularly advances the bundle from the source to the external controller and then to the destination. Due to many reasons, the packets may be rejected or dropped. The rejected or dropped packets will be stored in the port. In this paper, reinforcement learning is applied for the re-transmission of the lost packets. Q-Learning formula is applied to the dropped packets to study the reason for the dropping and the same will be re-transmitted.

Keywords Open flow · Software-defined networks · Reinforcement learning · Q-formula

1 Introduction

SDN Controllers and changes include the execution of open flow information with the goal of understanding the basic language of open flow messages. The SDN controller pushes switches to power switches with the intention that they will be able to choose when system traffic reaches them. SDN is a very famous and powerful concept [1] which decouples the data plane and control plane [2], where data plane is used as the forwarding mechanism and control plane takes the responsibility of control decisions.

V. Karunakaran (✉) · A. Geetha
Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Kelambakkam 603103, India
e-mail: rs.vk0819@hindustanuniv.ac.in

A. Geetha
e-mail: angelinag@hindustanuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_26

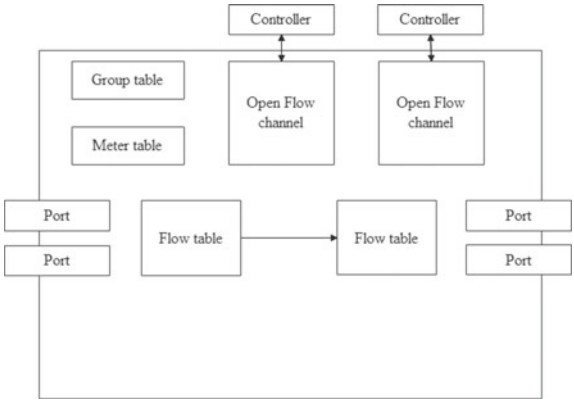


Fig. 1 Components of open flow switch

There are lot of protocols available in SDN to swap the information, but open flow is the most powerful [3, 4] in information exchange. Open flow permits a change to inform the controller of an approaching bundle for which there are no rules for forwarding that has been traced in the stream table. Thus, a controller might transfer control message to a control layer called switch informing to include a recent or alter a current sending regulation in the flow table to assist the approaching bundles. The procedure expands organizing assets and acquaints inertness with the parcel sending process. The open flow specification shields the fundamental mechanisms and fundamental tasks of the switch and the open flow switch procedure to handle an open flow switch from an isolated open flow controller (Fig. 1).

The switch protocol or open flow channel communicates with controller which is in the top and the controller accomplishes the switch through open flow switch procedure. The open flow switch components contains flow table which has the set of flow entries. Every match field contains a conventional of flow guidelines to smear to check the packets. The next step will be finding the match of the packet if any. If a similar entry is found, the instruction linked through precise flow entry will be accomplished. It will suit the flow entries in the priority order. The packet always matches from the highest priority to the lowest priority. If there is no matching found in the flow table, the following possibilities may happen.

(a) The packet might be promoted to the controller via open flow channel.
(b) The packets might be dropped or rejected or it may continue in the iteration.

The purpose of the group table will be guiding packet forwarding and packet modification. Usually, the port which follows will be 6653.

Open flow channel available in the right side of the component acts as an edge that links every open flow logical switch to the open flow controller, via controller which aligns and accomplishes the switch, collects the instruction from the switch and drives the packet out of the switch. The dropping of packet rate should be minimized. There is a possibility for a rejected packet to continue in the next iteration. This can be obtained by using trial and error method; Q-Learning is the suitable for the above proposed method.

2 Related Works

In Ting et al. [4], SDN-based methodologies can effectively assign organizing assets for such streams, the overhead because of system reconfiguration can be critical. They also proposed RL model algorithm which uses emulation results in which RL algorithm shows 60% progress in reducing the long-term plane and about 14% progress in the table-hit ration.

When related to north-bound interface, south-bound interface plays a major role which was implemented by open flow specification. Wolfgang and Michael [5] exemplified numerous SDN applications in diverse ranges that progresses network administration and process and also malignant issues and inter-domain routing. The dropped packets will be routed back and been transferred to its destination. The SDN structure must be adaptable enough to deal with a wide range of network devices, instead of one API for each device type. This implies managing both equipment and programming devices, straightforward sending devices and devices with rich and complex conduct.

As per Jean et al. [6] in the concern of flow entries, lot of previous standards had been explained, but mostly all these in very gross platform. The packet forwarding or packet processing should be decoupled from the protocol definitions. The perfect solution was adopted by open flow specification, the main concept of flow entry. Hence, the flow entries deserve to select related packets with great flexibility and access protocol layers. The dropped packets will be in the port, and then the packet may be re-forwarded again to open flow channel.

Utilizing the open flow convention, the controller can build, update and eradicate stream sections in-stream tables, both responsively (because of packets) and proactively. Each flow table in the switch contains a bunch of flow sections; each flow passage comprises of match fields, counters and a bunch of directions to apply to coordinating with matching packets as clarified by [7].

Lara et al. [8] proposed a survey that there is always a feasibility for the researchers to test the innovative ideas and novel applications; this additionally incorporates software-based traffic investigation, centralized control, dynamic refreshing of sending rules and flow reflection. They measured the open flow networks through modeling and experimentation; also, they surveyed the challenges, while deploying open flow-based networks.

Another survey provided by Hu et al. [9] conducted a thorough overview of the significant points in SDN/open flow execution, counting the essential idea, applications, semantic deliberation, controller, virtualization, quality of administration (QoS), security, just as its coordination with remote and optical systems. Various plans of merits and demerits will be analyzed and future exploration slants in this energizing region will be discussed.

Fang et al. [10] described the objective of software-defined networks is to give exposed, client-controlled administration of the sending equipment in a system. SDN misuses the capacity to part the information plane from the control plane in switches.

Zhang et al. [10] stated that SDN has responsibility to provide a centralized view; for that, SDN controller with SDN should be synchronized each other. Sometimes, there may be an issue in performance of the controllers with respect to SDN. To overcome this issue, we formulate the controller using Markov decision process (MDP), agent will be equipped in an environment to train a smooth, accessible and fine grained controller synchronization called multi-armed controller synchronization (MACS).

Yao et al. [11] have illustrated that the process of segregating the control plane from the forwarding plane is very mandatory for wide deployment of SDN. We have to manually configure the forwarding planes. This issue can be overcome by reinforcement learning getting employed to vigorously produce control strategies and produces an ideal solution.

Maintaining a balanced load while segregating the planes as mentioned earlier, SDN controller mechanism is very much essential from Zhuo et al. [12] utilized to choose the out-migration space, the in-migration area and the change to move, as consolidated into one switch migration triplet. By applying support learning, the heap adjusting system picks a lot of switch migration lines that can accomplish worldwide enhancement from neighborhood improvement which will be the primary objective of this paper. The test results clarify that the component can utilize the controllers' assets, rapidly balance the heap between controllers, lessen pointless movement overhead and get a quicker reaction pace of the parcel in demand.

5G networks are used for re-transmission for the fastest recovery of the packets. To prove the re-transmission, Chen et al. [13] state that a two-way handshaking mechanism can also be implemented. SDN-based data transmission protocol (SDTP) is presented for 5G networks which improves the data transmission and packet re-transmission. 5G network communication is also supported by SDN-based adaptive transmission protocol [14].

3 Reinforcement Learning

The reinforcement learning is applied in the SDN open flow, which the entire scenario called environment. The source node acts as an agent, which waits the instructions from the destination that is SDN controller. Reinforcement learning can be generally classified into two types.

(i) Markov decision process
(ii) Q-Learning.

Markov decision process (MDP) utilized for creating decisions whether the packets transferred travel the correct path or not. The following terminologies are very important before going to Q-Learning. MDP consist of, **State (S)**: SDN is ready to transfer the packets in open flow specification to the SDN controllers. Packets are considered as set of states. So we have five states. There is a possibility of rejecting or dropping the packets in which will be stored in port. This will be described in Q-Learning.

1. **Action (A)**: Possible set of activities available from state S.
2. **Reward** R(s'ls, a): It is the reward received after overcoming the failure.
3. **Discount (Y)**: It is the discount factor, which exemplifies the transformation in future and present rewards.

So, Markov process can be perceived as an assortment of state S for certain activities A potential from each state with some likelihood P. Each such activity will prompt some prize R. In the event that the likelihood and prizes are obscure, the issue is of reinforcement learning. A new learning methodology called re-transmit algorithm before that a brief study SDN open flow gives clarity about how the packets are transferred and how is mentioned in the open flow entry which is recognized by the packet's match fields and its priority.

4 Controller Architecture

Figure 2 represents the RYU controller architecture; the top of the architecture contains north-bound plug-in which enables the user to write their own application. REST API is responsible for that. Bottom of the architecture contains south-bound plug-in which enables the open flow switches. All the versions of the open flow switches will be supported by RYU. Inside the RYU architecture contains RYU library files. The main purpose of the library files is that the flow rules are installed in the RYU controller, so the path between source host and destination host will be established. The controller ensures all the paths possible between the switches and the hosts. Pathless switches will be rejected. Also, there is possibility of finding the nearest host, which may also act as destination host.

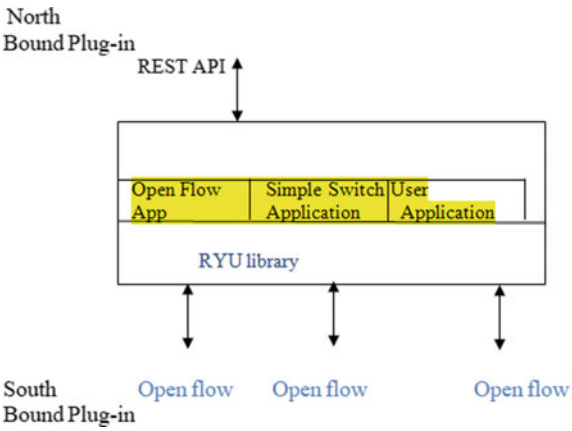


Fig. 2 RYU controller architecture

5 Proposed Model

The main purpose of this work is to apply the reinforcement learning in software-defined networks, while transferring the packets. If the packet gets stuck in the middle, then we use reinforcement learning to resume the dropped packets and re-transfer it again. A new algorithm, re-transmit algorithm, checks the circuit whether there is any packet to re-transfer. This allows to check if there is any failure of packets in the initial transferring of the packets. This can be elaborated later in the circuit diagram.

First way: The first is to utilize the q-table as a source of perspective and view every single imaginable activity for a given state. The operator at that point chooses the activity dependent on the maximum estimation of those activities. This is known as exploiting since we utilize the data we have accessible to us to settle on a choice.

Second Way: The subsequent method to make a move is to act haphazardly. This is called investigating. Rather than choosing activities dependent on the maximum potential compensation, we select an activity at irregular. Acting haphazardly is significant in light of the fact that it permits the specialist to investigate and find new expresses that in any case may not be chosen during the abuse procedure. You can adjust exploration/exploitation utilizing epsilon (ε) and setting the estimation of how regularly you need to explore versus exploit. Here is some unpleasant code that will rely upon how the state and activity space is the arrangement.

Updating Q-Table:

Updates happen each step ends when an action is done. The agent comes to a point by reaching the endpoint of the game or completing the task; in the next step,the agent comes to the conclusion that it explored enough steps.

There are three basic steps:

(i) The controller transfers all the packets initially.
(ii) The flow rules installed in the controller check whether every packet had been successfully transmitted.
(iii) If any packet failed in the middle, then the failed packets are applied for re-transmission by using the re-transmit algorithm.

The above Fig. 4 shows the respective dropped packets 0,1,2,3,4 which lays in the port. It is a chance by applying a Q-Learning algorithm to re-transfer the dropped packets from the port. Q-Learning normally plays a role of getting failed at initial phase, in which it is called as observing phase. The same methodology applied in SDN open flow packet transferring, while the dropped packets in the port will be falling under the term observation by agent. Proposed re-transmit algorithm plays a similar role like Q-Learning [15]. The failure packets are re-transmitted similarly like Q-Learning algorithm (Fig. 3).

From the port applied the Q-Learning algorithm, the packet tends to find the destination host. There may be one node in the middle or may be two node in the middle, between source and destination hosts. This provides the accuracy rate of 90%; hence, we cannot assure 100% for all the dropped packets.

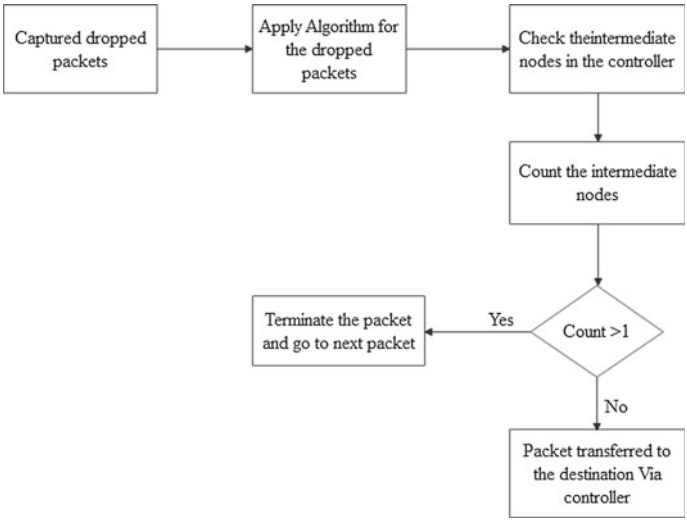
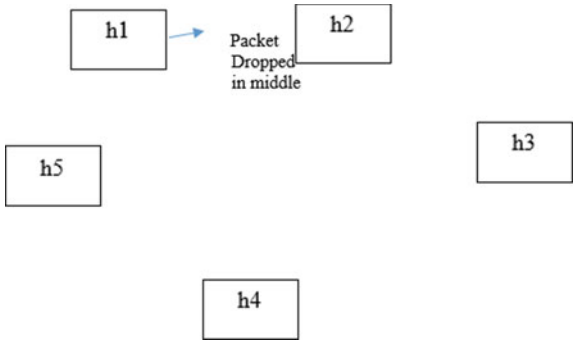


Fig. 3 Applying Q-Learning in open flow specification packets

Fig. 4 Dropping of the packets



5.1 Proposed Re-transmit Algorithm

1. *Src* - > new *Pckt*;
2. *Send(Pckt)* - > *InPort*;
3. **If** *Pckt.Header* ϵ *Flow_Table* **then**
4. Update_Counters();
5. Send(Pckt) - > Output.Port
6. Success
7. **else**
8. *re-transmit(Pckt_In)* - > *Controller;*
9. *Add Flow rule to Switches;*
10. *Src* - > dropped *Pckt*;

11. *Send(Pckt) - > InPort*;
12. **If** *Pckt.Header* ϵ *Flow_Table* **then**
13. Update_Counters();
14. Send(Pckt) - > Output.Port
15. endif
16. *Output_Port - > Dst;*

The algorithm states that how many packets we are to re-transfer, those packets are constructed in the matrix format. If we say there are four packets, then it is considered as 4*4 matrix. From the above algorithm, the first step will be setting the gamma value. That should be less than the value 1. Now, the user should take the value near to 1, may be 0.8 or 0.9. Initial input matrix will be 0, whatever the size of the matrix. Applying the formula in the eight step, the packets will be re-transferred, and some rewards will be obtained. The packet starts traveling from the source and checks the valid destination; SDN open flow controller will be routing the packet to its proper destination, by overcoming the timing issues.There should be not more intermediate nodes, which makes the architecture more complex. But in feature, there are feasibilities for this type of study.Retransmission is also applicable for UDP-based transmission [16].

The below architecture shows that how the dropped packets are transferred from source to destination (Fig. 5).

Assume 0,1,2,3,4 are the areas in which the dropped packets happened in the main flow specification packet transferring; now, this has to be re-transfer again. Take 0th packet and check the possibility of the packet to reach the destination; as explained in the architecture, intermediate node should be less than or equal to 1 or it should not be greater than 2. Now in the area 5, assume the SDN open flow controller switch is present. This controller has to be passed to the destination. If we want to transfer 0th packet from any of the source node, we say the 5th area is the open flow switch, so the role of switch will forward to any of the destination node as shown in the Fig. 6. The traveling of the packet and its drop of the packet is shown in Fig. 7.

The packet is dropped packet to RYU controller. Fig. 4 shows the number of hosts connected, connected to a switch. Five switches and five hosts are designed. The

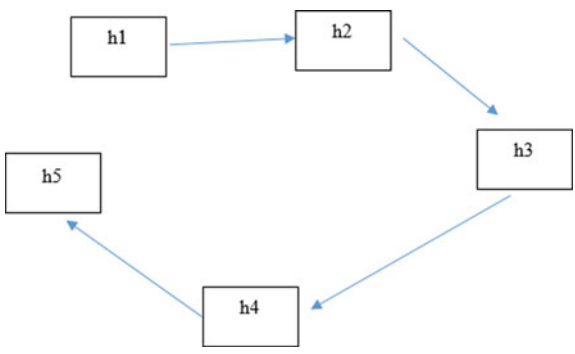


Fig. 5 Re-transmission of the packets

hosts are named as h1, h2, h3, h4 and h5, respectively. h1 is considered as source host, and a time slot fixed as t0. Inside the switch, SDN flow rules are installed. This is mainly used for Packet matching. If a packet transferred either from Host 1 or Host 2, it is transferred to switch A. From switch A fix a time t1, and then packet travels to SDN open flow controller; with the help of flow rules installed, the switch A forwards the packet to the controller; the open flow controller then checks whether the packet is matched. If the packet is matched, then it forwards the packet to switch B. If there is a mismatch, then the controller rejects the packets and then sends back to the switch. The concept of Q-Learning arises here. Now, the rejected packet should be re-transferred. The flow table in which flow rules are available will re-transmit the packet again to the SDN controller. Now, if the packet is matched, controller forwards the packet to the switch B and then it reaches the destination host.

Figure 5 shows the successful re-transmission of the packets implement the above said architecture; SDN controller is needed. There are lot of SDN controllers available; RYU controllers had lot of provisions. RYU is a segment-based SDN outline. RYU gives programming segments a very much characterized outline that makes it simple for engineers to make new organized executives and control applications. The primary motivation behind utilizing RYU is that it upholds different conventions for overseeing organizing gadgets, for example, open flow, BGP and so on When coming to open flow switches, RYU underpins a wide range of switches. A brief outline of RYU controller architecture will be discussed.

6 Results and discussion

There are more possibilities of packets getting affected or packet failure as similar in traditional networking. In similar cases, the packets may be getting rejected. RYU keeps track of the rejected packet, as the flow rules are installed in the controller; hence, the rejected packets are re-transmitted. The packet will start its re-transmission, and the controller will find a path for the packet to reach the destination. To execute this process, we need to install mininet. The platform which mininet will be well supported is Ubuntu 20.0.4. A mininet emulator will be installed, besides a simulator tool called miniedit, a tool where we can customize our topologies. To successfully do this, we follow two procedures. To install Ubuntu 18.0 or higher version, the Oracle Virtual toolbox needed to be installed. Another procedure is we can do Ubuntu 20.0.4 as main OS or in dual boot OS.

7 Conclusion

As the concept of virtualization has reached its peak, SDN replaced the traditional network. Software-defined networking (SDN) segregates the control plane and data plane. The SDN switches in the data plane, is ready to receive the packet from

the source, and transfers the packet to the destination through control plane. The possibility of packet loss arises because of the issues like latency in the controller. RYU controller always gives a better throughput when compared to other controllers, but in case of latency, it may lead to a packet loss. So it leads to the retransmission of the packets. From this, we can conclude that packets are retransferred using the Q-Learning algorithm. The dropped packets are already stored in the port in the open flow specification of SDN. Those dropped packets are taken for retransferring again into the open flow controller and provide 90% success rate. Suppose, if any failure occurs again in the retransmission, then those packets will be terminated permanently. In the future, there may be a change in the architecture and the success rate will be increased. Also, traffic analysis will be considered in the future, to obtain better retransmission between source and destination. If traffic arises while re-transferring, then it incurs a loss again. The traffic issue will be taken care of by both the switches and controllers. Less traffic will be filtered by the switches, and more traffic will be filtered by the controllers. These issues will be taken care as well as the testbed in SDN open flow can be improved.

References

1. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260
2. Wang M-H, Chen L-W, Chi P-W, Lei C-L (2017) SDUDP: a reliable UDP-based transmission protocol over SDN. IEEE access. pp 1–1. https://doi.org/10.1109/ACCESS.2017.2693376
3. Parisotto E, Ghosh S, Yalamanchi SB, Chinnaobireddy V, Wu Y, Salakhutdinov R (March 2019) Concurrent meta reinforcement learning. arXiv:1903.02710. [Online]. Available: https://arxiv.org/abs/1903.02710
4. Mu T-Y, Ala-al F, Shuhaib K (Nov 2018) SDN flow entry management using reinforcement learning. https://doi.org/10.1145/3281032
5. Braun W, Menth M (2014) Software-defined networking using OpenFlow: protocols, applications and architectural design choices. Future Internet 6:302–336. https://doi.org/10.3390/fi6020302, www.mdpi.com/journal/futureinternet
6. Tourrilhes J, Sharma P, Banerjee S, Pettit J, SDN and OpenFlow evolution: a standards perspective, DOP: 24 Nov 2014. https://doi.org/10.1109/MC.2014.326, Publisher: IEEE, pp 22–29
7. Open networking foundation. [Online]. Available: https://www.opennetworking.org/
8. Lara A, Kolasani A, Ramamurthy B, Network innovation using OpenFlow: a survey. IEEE Commun Surv Tutorials
9. Hu F, Hao Q, Bao K, A survey on software-defined network (SDN) and OpenFlow: from concept to implementation. IEEE Commun Surv Tutorials. https://doi.org/10.1109/COMST.2014.2326417
10. Fang C, Cheng C, Tang Z, Li C (2019) Research on routing algorithm based on reinforcement learning in SDN. IOP Conf Series: J Phys: Conf Series 1284:012053. IOP Publishing. https://doi.org/10.1088/1742-6596/1284/1/012053
11. Zhang Z, May L, Poularakisz K, Leung KK, Tuckerx J, Swami A, MACS: deep reinforcement learning based SDN controller synchronization policy design. https://arxiv.org/abs/1909.09063
12. Yao L, Hong P, Zhou W, Evaluating the controller capacity in software defined networking
13. Li Z, Zhou X, Gao J, Qin Y, SDN controller load balancing based on reinforcement learning. Publisher: IEEE, pp 23–29

14. Al-Duwairi B, Al-Quraan E, AbdelQader Y (2020) ISDSDN: mitigating SYN flood attacks in software defined networks. J Netw Syst Manag 28:1366–1390. https://doi.org/10.1007/s10922-020-09540-1
15. Jang B, Kim M, Harerimana G, Kim JW (13 Sep 2019) Q-learning algorithms: a comprehensive classification and applications. IEEE
16. Chen J, Yan S, Ye Q, Quan W, Do T, Zhuang W, Shen X, Li X, Rao J (2019) An SDN-based transmission protocol with in-path packet caching and retransmission

Machine Learning Approach to Detect Depression Using Social Media Posts

Samadrita Saha, Sonal Kumari Thakur, and R. S. Ponmagal

Abstract Depression is a common cause of global impairment and can lead to suicide in the worst-case scenario. It has been observed to affect a person's language of communication as well as their facial expressions. People have taken to sharing their feelings and sentiments on social media sites since its introduction in the last decade. This has provided a unique opportunity for researchers to study their interactions on these platforms and infer whether the users display signs of depression or not. This paper utilizes the Twitter platform and analyzes text posts using bag of words, term frequency–inverse document frequency, and Naive Bayes algorithm, and profile pictures using convolutional neural networks. Additionally, profile pictures containing quotes or words are first subjected to optical character recognition (OCR) to convert them into a text format so that they can be analyzed using BOW and TF-IDF. The study yields a combined result based on the analysis of both text posts and profile pictures, which can better guide the users toward seeking professional medical counsel. Unlike current systems which either focus on text or images posted by the user, the proposed system focuses on both and thus takes into consideration both facets of social media interaction.

Keywords Depression · Social media · BOW · TF-IDF · Naive Bayes · Convolutional neural networks

S. Saha (✉) · S. K. Thakur · R. S. Ponmagal
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India
e-mail: ss6158@srmist.edu.in

S. K. Thakur
e-mail: ss8661@srmist.edu.in

R. S. Ponmagal
e-mail: ponmagas@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_27

1 Introduction

According to a study published by the World Health Organization [1], 322 million people suffer from depression globally, a number that roughly equates to 4% of the worldwide population. Despite the common public perception, depression is not the same as mood fluctuations and emotional responses that arise due to challenges faced in everyday life, and while it is difficult to define it precisely in a manner that fits all the various subtypes that have been elaborated in the past, common symptoms include a considerably changed mood that could be accompanied by a dent in self-worth, a desire to shrug off interaction with other people, sluggishness, and the inability to be productive. When depression affects someone over long periods of time, it affects their health severely and has also been linked with inducing suicidal thoughts.

Although effective methods to treat mental illnesses have been established, a large percentage of people do not receive any treatment for their ailment. This lack of diagnosis is due to several reasons: a deficiency in medical resources and adequately trained healthcare professionals, the societal stigma surrounding mental illnesses, and shame and self-stigmatization that affect the depressed individuals. Studies have shown that the latter is a much more prominent reason that deters people from reaching out for proper medical help. The aforementioned reasons that fail to diagnose depression, coupled with the popularity of the internet and social media sites, have led to affected individuals expressing their concerns and fears, sharing their struggles, and supporting each other on these online platforms [2]. This has led several researchers to study the online activity of individuals as a solution to combat the lack of diagnosis of depression. While studies in the past have either focused on text or image analysis, this paper aims to combine the results obtained from both using machine learning methods such as BOW, TF-IDF, Naive Bayes classifier, and convolutional neural networks.

This paper comprises several sections: Sect. 2 outlines the most up-to-date methods for predicting depression from social media messages; Sect. 3 describes the proposed system's methods briefly; Sect. 4 explains the system's implementation by decomposing it into two modules: one for text analysis and one for image analysis; Sect. 5 discusses the evaluation metrics of the system along with two individual instances of executing it; Sect. 6 finally concludes this study by summarizing the entire system and detailing limitations and scope for future work.

2 State of the Art

The current body of literature that deals with the prediction of depression using social media is quite extensive pertaining to the study of text posts. Twitter is a popular choice for researchers because of the large volume of data that is available on it; it has around 326 million active users and 90 million publicly available tweets [3]. De Choudhury et al. [4] used linguistic features obtained from social media

interaction to predict postpartum depression. Further, they used statistical methods such as t-tests to obtain the results. Coppersmith et al. [5] searched for a specific phrase in tweets, 'I have been diagnosed with X,' where X was representative of a particular mental health condition. This study used linguistic inquiry and word count to determine the difference between these conditions. Tsugawa et al. [6] used a strictly Japanese Twitter sample and successfully predicted depression using features based on a topic modeling. They concluded that when the data was collected over a period of 16 weeks as compared to 6 weeks, the accuracy of prediction improved. Besides Twitter, researchers have also used other social media sites like Facebook and Reddit to analyze the text posts of users and detect signs of depression. Moreno et al. [7] observed the status updates of 200 users on Facebook and looked for the phrase 'I feel hopeless' in them to conclude whether the users were depressed or not, using negative binomial regression. Further, Choudhury et al. [8] observed using statistical means that users on Reddit with poor linguistic style and less social interaction who tended to express feelings of hopelessness or anxiety showed signs of depression.

Contrary to the vast research that has been done involving text posts on social media, extremely limited studies have focused on image analysis. While images showing the facial expressions of users have still been studied to predict depression, images with embedded words, most often famous quotes from popular works in media, remain an unexplored domain. Guntuku et al. [9] observed that Twitter profile pictures that showed only the user's face and images dominated by grayscale were indicative of depression. The authors also found that pictures of users with depression more often hid positive emotions than display negative emotions. Despite images on social media platforms not being analyzed in detail to evaluate the mental state of users, significant work has been done using them outside the realm of social media sites. Alghowinem et al. [10] predicted depression in users based on their eye movement and head movement using a predefined dataset of images containing facial expressions. The subjects varied in their cultural and societal backgrounds, and it was found that the best results were obtained using cross-validation. Harati et al. [11] measured a feature called multi-scale entropy (MSE) that indicated the variation occurring across a single pixel in a video. They concluded that a low entropy level suggested the person was depressed. Zhou et al. [12] detected the severity of depression in people by generating a depression activation map (DAM) by training a deep convolutional neural network. This was combined with regression models for even better performance and to successfully identify depression regions on people's faces.

Additionally, the present body of literature does not include any studies that combine the results of both text and image analysis performed on data obtained from social media platforms, which is the goal of this paper.

3 Proposed Work

The proposed system aims to analyze both text and image posts using machine learning algorithms. For the analysis of text posts, BOW and TF-IDF are used independently to first construct a vectorized matrix. A Naive Bayes model with two classes, positive and depressive, is then fitted using the said vectorized matrix. Once the model is trained, its accuracy is verified using the validation data, following which it is used to determine whether a particular text post is positive or depressive in its tone.

Image posts can be classified into two categories: images containing words and images containing facial expressions. For the analysis of images containing words, OCR is first used to convert them into text format. This intermediate result is then analyzed using the Naive Bayes model previously trained for text data. For the analysis of images containing facial expressions, a four-layered convolutional neural network is built that can classify the expressions into two classes, positive and negative, the latter indicating that the user could be suffering from depression. The outcomes of both varieties of image analysis, and the independently acquired result of text analysis, are used to predict the state of the user's mental health; if either analysis predicts that the user suffers from depression, then the overall prediction states the same.

4 Implementation

The prediction of a user's mental state is done based on both their text and image posts. To facilitate the study of both kinds of posts, the proposed system is broken down into two modules, one for text analysis, and the other for image analysis. Fig. 1 depicts the architecture diagram of the system that is proposed. It shows the two modules of the system; the text analysis module has components for preprocessing, performing feature extraction using bag of words and term frequency–inverse document frequency, and performing classification using a Naïve Bayes Classifier; the image analysis module has components for preprocessing both images containing quotes and facial expressions, an optical character recognition component for the quotes that feeds into the text analysis module, and a convolutional neural network to analyze facial expressions.

4.1 Text Analysis

Dataset: We generate the dataset by combining positive tweets from the Sentiment140 dataset with tweets containing the hashtag 'depression' (#depression) scraped using Twitter's Developer Tools. For scraping, a developer account is made

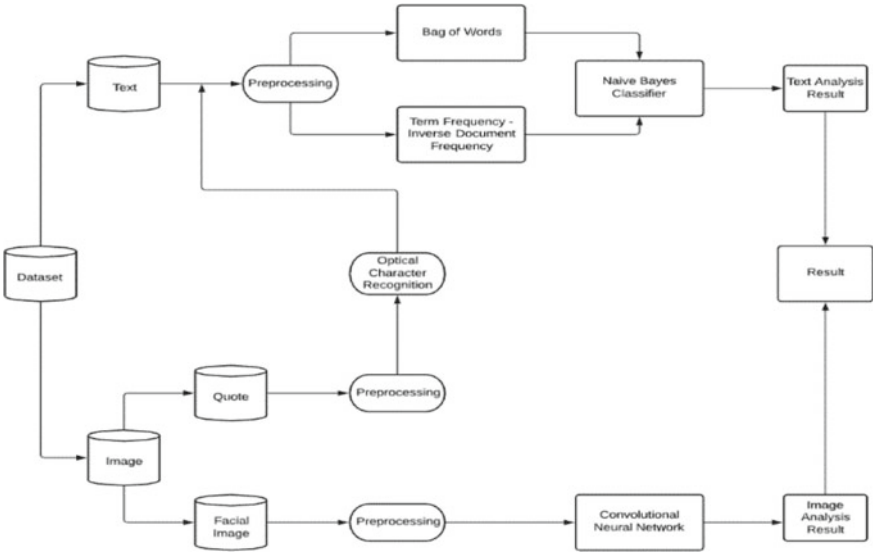


Fig. 1 Architecture diagram of the proposed system

that allows one to scrape tweets for academic and research-related purposes. The Sentiment140 dataset has 1,600,000 tweets divided into three categories labeled as follows: 0 for negative, 2 for neutral, and 4 for positive. For this study, only 8000 positive tweets are considered and are then combined with 2314 depressive tweets that are scraped from Twitter.

4.1.1 Data Preprocessing

Once the dataset is created, we use natural language processing (NLP) to preprocess it. Several algorithms are used for this purpose. To begin, we break down the sentences in the posts into words using tokenization. Next, we exclude stop terms like 'the,' 'for,' 'can,' and 'a,' which provide little insight into the user's mental state. Finally, we reduce all the words to their root or base form using stemming [13]. For example, the words 'worrying', 'worries', 'worrisome', and 'worried' are all reduced to 'worry'.

4.1.2 Feature Extraction

We perform feature extraction using two different methods, BOW and TF-IDF, and evaluate and compare the performance of the classifier using both.

(a) BOW: BOW is the simplest method to extract features in NLP. First, we create a vocabulary using all the unique words present in the corpus, following which a

vector matrix is constructed in which each row corresponds to a single tweet and each column corresponds to one of the previously determined unique words.

(b) TF-IDF: One of the major drawbacks of BOW is that it gives equal importance to every word in the text, irrespective of its impact on the tone of the message. For example, 'sad' and 'will' are both treated equally, although 'sad' is more indicative of the mental state of the user. TF-IDF rectifies this shortcoming and identifies words that are more important to the corpus, using statistical features. It works on the principle that if a particular word occurs more frequently in a single document (a tweet, in the scope of this study) and less frequently across all other documents in the corpus then that word should be given more importance because it holds more insight for classification. The TF-IDF score for every word in the corpus is calculated using the formulae mentioned in (1), (2), and (3). A high score indicates more importance of the word. We then construct a vectorized matrix using these scores.

$$tf_{t,d} = \frac{n_{t,d}}{\text{Number of terms in the document}} \tag{1}$$

$$idf_t = \log \frac{\text{Number of documents}}{\text{Number of documents with term } 't'} \tag{2}$$

$$(tf\_\text{idf})_{t,d} = tf_{t,d} * idf_t \tag{3}$$

4.1.3 Naïve Bayes Model

Finally, we train a Naïve Bayes classifier using the vectorized matrices obtained from the BOW approach and the TF-IDF approach independently. The classification algorithm Naïve Bayes is based on Bayes' Theorem and assumes that each function is independent of each other [14]. It can be expressed mathematically as (4), in terms of conditional probability, $P(x|y)$, and independent probability, $P(x)$. Once the classifier is trained, we verify its accuracy using the validation data.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4}$$

4.2 Image Analysis

4.2.1 Quotes

(a) *Dataset*: Since we analyze images containing quotes using the module for text analysis, there is no necessity for a separate train or test dataset.

(b) *Data Preprocessing*: First, we grayscale the image to convert it into a binary image. RGB images are converted to grayscale images using the formula in (5), where Y, R, G, B stand for grayscale, red, green, blue values. Next, we remove noise from the images by utilizing morphological operations like erosion and dilation.

$$Y_{\text{linear}} = 0.2126R_{\text{linear}} + 0.7152G_{\text{linear}} + 0.0722B_{\text{linear}} \tag{5}$$

(c) *Optical Character Recognition*: Finally, we convert the images containing words into machine editable and searchable text. This text is then used as an intermediate output that is analyzed using the techniques of BOW, TF-IDF, and Naive Bayes classification.

4.2.2 Facial Images

(a) *Dataset*: We use the FER2013 dataset to analyze images with facial expressions in this study. The FER2013 dataset has 35,887 images showing a variety of seven basic human emotions. We classify these emotions into two classes: positive (happiness, neutral, surprise) and depressive (anger, fear, disgust, sadness) [9].

(b) *Data Preprocessing*: Next, we perform image augmentation on the training dataset for better performance of the convolutional neural network. Multiple processing techniques like rotating, shearing, and flipping are performed to artificially generate more training images.

(c) *Convolutional Neural Network*: Finally, we create a convolutional neural network containing convolutional layers and fully connected layers. Once the model is trained using the train data, we validate its working using the validation data. Figure 2 depicts the organization of the neural network that has been used. It has four convolutional layers and two fully connected layers containing batch normalization, RELU, max pooling, and Dropout layers. The output is obtained using a SoftMax activation function.

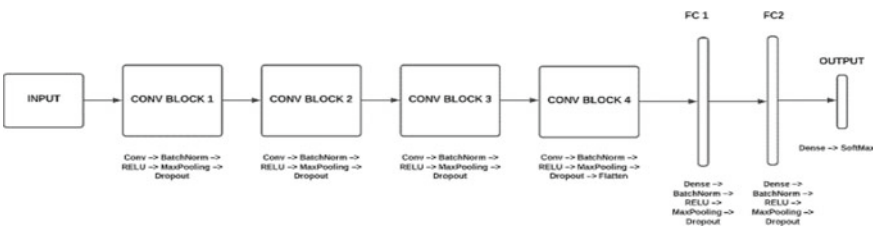


Fig. 2 Convolutional neural network architecture

5 Results Discussion

Our objective is to predict depression of the users using their text posts and profile pictures. For testing, we generate a dataset by shuffling our train dataset with scraped Twitter data; we collect the Twitter data by first scraping text posts and profile pictures of users who post using the hashtag 'depression' (#depression) and then scraping the data of users who post using the hashtag 'positivity' (#positivity). The resulting dataset thus has information about both varieties of social media interaction of users, text as well as image, and is then verified using the text and image analysis techniques outlined in Sects. 3 and 4.

We evaluate the performance of these algorithms using popular evaluation metrics. Accuracy is the percentage of true classification; precision is the number of positively identified entities that are correct; recall is the proportion of correctly identified positive samples; F1 is recall and precision's harmonic average [13]. All these metrics can be mathematically formulated as shown in (6), (7), (8), and (9), in terms of true/false positives and negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{9}$$

Our system achieves 91% accuracy, 0.94 precision, 0.89 recall, and 0.91 F1-score. Further, a single instance of working of the system is shown in Fig. 3 where the analysis of a particular user's profile picture and text post indicates that they display signs of depression. The tweet that is used for analysis here states, 'Depression

Fig. 3 Instance of working predicting the user could be depressed

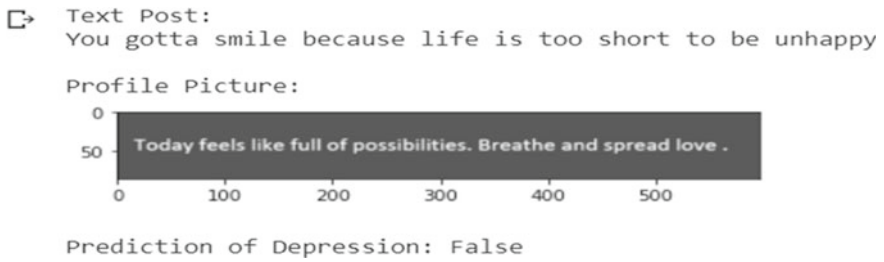


Fig. 4 Instance of working predicting the user is not depressed

isn't always sadness, it can be emptiness, anger, tiredness or loneliness, sometimes a cry for help isn't so visible,' while the profile picture depicts a woman looking away from the camera with an impassive expression on her face. In Fig. 4, another instance is observed in which the integral difference is that the user's profile image has embedded words in it. The tweet in this case states, 'You gotta smile because life is too short to be unhappy,' while the profile picture depicts the words, 'Today feels like full of possibilities. Breathe and spread love.' Here, the analysis of the text post and the profile picture indicates that the user is not displaying signs of depression. In both these examples, both the tweet scraped from the user's account and their profile picture are displayed followed by the prediction.

6 Conclusion

In this study, we observed both text and image data generated by users on Twitter and analyzed them to detect signs of depression in them. We built upon present systems that either utilize text or image data and proposed a model that took into consideration both facets of social media interaction before predicting the state of the user's mental health. For text analysis, we coupled feature extraction techniques such as BOW and TF-IDF with a Naive Bayes classifier, whereas for image analysis, we trained a convolutional neural network to study images with facial expressions and performed OCR on images with words embedded in them so that they could be studied using the previously mentioned text analysis techniques.

We found that text posts that contained words indicating feelings of sadness, loneliness, or anxiety, combined with profile pictures that either contained similar words embedded as quotes or contained facial expressions that can be classified as sad, angry, fearful, or disgusted, were signs of a user who was dealing with depression and required professional medical counsel. On the other hand, users whose text posts contained words indicating feelings of optimism, happiness, or hope, and whose profile pictures had words embodying similar sentiments or had facial expressions such as those of happiness, or surprise, were not symptomatic of signs of depression.

One of the limitations of the system is that it is unable to understand the semantic relationship between words. This can be overcome by including word embeddings along with BOW and TF-IDF. Further, the system cannot handle noise in images, especially the ones which are subjected to OCR. This can be rectified by including noise filtering algorithms that can effectively remove the noise before further analysis is done. As for future work, other modes of social media interaction such as the use of emojis, videos, and GIFs, along with the number of hours spent by a user on these platforms, can be studied for a more comprehensive result to decide whether a user displays signs of depression.

References

1. WHO (2017) Depression and other common mental disorders: global health estimates. Geneva: world health organization; 2017. license: Cc by-nc-sa 3.0 igo. http://www.who.int/en/news-room/fact-sheets/detail/depression
2. Gowen K, Deschaine M, Gruttadara D, Markey D (2012) Youngadults with mental health conditions and social networking websites: seeking tools to build community. Psychiatr Rehabil J 35(3):245–250
3. T. S. Portal Statistics and Studies (2019) Social media usage worldwide. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/
4. De Choudhury M, Counts S, Horvitz E (2013) Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors incomputing systems. ACM, pp 3267–3276
5. Coppersmith G, Dredze M, Harman C, Hollingshead K (2015) From ADHD to SAD: analyzing the language of mental health on twitter through self reported diagnoses. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality, pp 1–10
6. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H (2015) Recognizing depression from twitter activity. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM, pp 3187–3196
7. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, Becker T (2011) Feeling bad on facebook: depression disclosuresby college students on a social networking site. Depress Anxiety 28(6):447–455
8. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M (2016) Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 CHI conference on humanfactors in computing systems. ACM, pp 2098–2110
9. Chandra Guntuku S, Preotiuc-Pietro D, Eichstaedt JC, Ungar LH (2019) What twitter profile and posted images reveal about depression and anxiety. Proc Int AAAI Conf Web Soc Media 13(01):236–246
10. Alghowinem S, Goecke R, Cohn JF, Wagner M, Parker G, Breakspear M (2015) Cross-cultural detection of depression from nonverbal behavior. In: 11th IEEE international conference and workshops on automatic face and gesture recognition (FG). pp 1–8
11. Harati S, Crowell A, Mayberg H, Kong J, Nemati S (2016) Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression. In: 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp 2254–2257
12. Zhou X, Jin K, Shang Y, Guo G (2020) Visually interpretable representation learning for depression recognition from facial images. IEEE Trans Affect Comput 11(3):542–552

13. Tadesse M, Lin H, Xu B (2019) Detection of depression-related posts in reddit social media forum. IEEE Access, pp 1–1
14. Zhang H, Li D (2007) Naïve bayes text classifier. 2007 IEEE international conference on granular computing (GRC 2007). pp 708–708

Plant Leaf Disease Detection Using Mask R-CNN

P. Ranjana, Janga Pavan Krishna Reddy, Jakkamsetti Bharath Manoj, and Kalicheti Sathvika

Abstract In agriculture, disease detection is important for a productive crop yield. So many diseases will affect crop quality of tomatoes, potatoes and pepper; some of them are bacterial spot, septoria leaf spot and yellow curved leaf diseases. To classify diseases of plant by detecting symptoms of leaf disease through automatic methods, this paper presents a Mask R-convolution neural network (Mask R-CNN) algorithm for tomato leaf, pepper leaf and potato leaf. In this, we are using plant village dataset which contains more than 1000 images of potato, tomato and pepper leaves of each plant along with disease symptoms. With the help of Mask R-CNN, classification and extraction are done automatically. Mostly information of color of the leaf is used in disease detection. Based on RGB components, filters are used in our model for the three channels. The results of proposed method for the experiment will be recognized efficiently for different types of potato, tomato and pepper leaves. This technique of detecting plant leaf disease detection using Mask R-CNN will help small holder farmers for detecting diseases of plants in very efficient manner.

Keywords CNN · Image processing · Genetic algorithms · Preprocessing · Classification

1 Introduction

Day by day need of crop production is increasing. By 2050, at least 50% more production is needed. Mainly production is high in Asia and Africa as well. Without horticulture expertise, around 84% of farmers are farming. Due to this reason, 50% of loss in crop yield is happening every year. For the classification of crop diseases, so many methods are there. The traditional method is analyzing crop diseases through human visual inspection, but it is not that much of accurate and not feasible. But with the help of machine learning and computer vision, there is much more accurate and standardized solution for this issue. Once the software is designed, then it is very

P. Ranjana (✉) · J. P. K. Reddy · J. B. Manoj · K. Sathvika
School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India
e-mail: pranjana@hindustanuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_28

easy to use with the help of a smart phone with Internet connection as well as camera of that smartphone. This software can be used to build a mobile application or Web site, and with that application or Web site, we will get accurate results; already some popular commercial apps are available namely 'Plant Snap'. Smartphone usage is gradually increasing every year globally.

2 Review of Literature

Worldwide around 5 billion smartphone users are there; a part of that in India itself, 1 billion users are there and 1 billion users in Africa as well. Nowadays, artificial in intelligence apps play major role in society. In recent years, usage of plant leaf disease detection using CNN technique is gradually increased and also achieved excellent results. These techniques are very useful to reduce time as well as error rate. Due to these features, it is favorable to researchers. For grape plant leaf disease detection in 2015, ResNet was introduced. This architecture is a ground-breaking function. This technique has higher learning rate. So, we will get more accurate results, i.e., achieved a 99% accuracy rate. For model's performance data, preprocessing is very difficult to differentiate fungal infections and viral bacteria which are often sharing an overlap of symptoms. Due to pathogens difference in shape, color or function symptoms are measurable. RGB data is preferable for this complexity [1]. This results good, clarity and noise-free images for better results, and they take more amount of data to train than grayscale. Augmentation technique and transfer learning technique use smaller datasets or unvaried data that can act on model's accuracy. This technique training images not only improve model's overall performance but also reduce over-fitting and can perform actions such as rotate, adding color, zoom and contrast changes. With these smaller datasets, it is proved that it is very successful in the method of transfer learning. The quality of the dataset is also very important for getting accurate results. The photography of leaf will be unreliable when tested. So many plant leaf disease datasets are available; the most accurate and very popular dataset which we took is 'plant village' dataset. The type of disease and in which stage of disease effected by the leaf of the plant can be detected. To recognize this early disease detection, that particular image must be used. So many plant leaf diseases are not visible to naked eye, but it can be identified using this Mask R-CNN techniques. The disease detection in the leaf involves the process like image acquisition, image preprocessing, segmentation, feature extraction and classification. In the detection, it is very important for early detection of the leaf disease. Automatic methods for leaf disease classification are done by the CNN model, and learning vector quantization algorithm and plant leaf disease are classified based on the color [2]. Leaf spot is the indicative of the disease in leafs, and it is identified by integrating with the image processing and helps to inspect automatically [3]. The image segmentation techniques are used for the plant leaf disease detection using genetic algorithm which detects the disease automatically and helps the farmers by automatically motioning it

[4]. Deep learning model is used which is the best model is applied with GoogleNet architecture for plant disease detection; it proves the accuracy of 14 crop types [5].

3 Objectives of the Proposed Work

The main goal of this work is to train a model to determine the plant leaf disease. In this, we mainly focused on three plant species, i.e., potato, tomato and pepper [3]. For all the three species, models are trained to recognize which disease and how much the plant effected by that disease and also measure the healthiness of the plant.

Goals:

(i) Overall performance of the dataset with simplified training using Mask R-CNN
(ii) Model's accuracy when tested with different images with fast execution
(iii) Developing a user-friendly application.

The researches will examine once the F1-score is greater than 80%, then only it will be approved. The Mask R-CNN model is used to improve the accuracy of the prediction. The dataset has to be preprocessed before the classification using the Mask R-CNN algorithm. We also discuss the classifier creation and deployment. Three phases of classification are involved in CNN which implement different tasks. With a single machine, total work will be completed. Specifications are mentioned in Table 1.

3.1 Data Addition

Tomato, potato and pepper are represented from 'plant village dataset' which contains 54,000 images. Total 50 images will be there in a test dataset, and these images are also taken from Google as well.

Table 1 Hardware software characteristics

Hardware and software requirements	Characteristics
RAM	8 GB
Processor	Intel i5
Graphic card	NVIDIA GeForce
OS	Windows 10

3.2 Data Preprocessing

In this, 80% of training data and 20% of validation data is divided. First, settings of augmentation are enforced to the training data; each and every operation is bringing a weighted probability that is present in every epoch. So many settings are available in augmentation. Some of them are zoom with crop, reflection and flipping, etc. and can be applied to identify infected leaf in particular regions. For normalization, RGB ImageNet statistics are used.

4 Classification Using Mask R-CNN

In this phase image size effect will be investigated on model performance Resnet34 where already trained weights are downloaded. Only the last two layers are frozen. So, it contains new weights to the specific plant leaf disease classification task. The policy of the cycle is useful for the final layers training [6]. The given below table contains four images sets that are re-created by the model. To run this model, a suitable learning method will be selected.

4.1 Data Preprocessing

In this, 80% of training data and 20% of validation data is divided. First, settings of augmentation are enforced to the training data; each and every operation is bringing a weighted probability that is present in every epoch. So many settings are available in augmentation. Some of them are zoom with crop, reflection and flipping, etc. and can be applied to identify infected leaf in particular regions. For normalization, RGB ImageNet statistics are used.

4.2 Classification Using Mask R-CNN

Mask R-CNN algorithm helps to solve the problem by segmentation; it has two types of image segmentation, i.e., the sematic segmentation and instance segmentation. The Mask R-CNN takes the concept of fast R-CNN, but it will predict only the object of interest. So, it will be helpful to do the accurate prediction. The Mask R-CNN is simple to train and only a small time than the Fast-CNN (Table 2).

The first phase is image trailing phase. In this phase image size effect will be investigated on model performance Resnet34 already trained weights are downloaded. Only the last two layers are frozen. So, it contains new weights to the specific plant leaf disease classification task. The policy of one cycle is useful for the final layers

Table 2 Image size and epoch

Trail	Image size	No. of epoch	Learner rate
1	152 × 152	5	1e5 and 1e4
2	198 × 198	5	1e5 and 1e4
3	227 × 227	5	1e5 and 1e4
4	227 × 227	5	1e5 and 1e4
5	228 × 228	5	1e5 and 1e4

Table 3 Species and images

Species	Class	No. of images
Tomato	Bacterial spot	2020
Tomato	Mosaic virus	170
Tomato	Leaf mold	980
Tomato	Healthy	1000
Potato	Late blight	1000
Potato	Early blight	1000
Potato	Healthy	160
Pepper	Mosaic virus	1000
Pepper	Blight	792
Pepper	Healthy	179

training [6]. The given below table contains four images sets that are re-created by the model. To run this model, a suitable learning method will be selected (Table 3).

4.3 Model Optimization Phase

ResNet35 is optimized using most suitable images. Adding augmentation settings will improve model's performance. Isolation and training of final two layers wrapping and changing of brightness operations will be added. Isolation and training are done at default learning rate to the final two layers. With the number of epoch test, series are performed at multiple trails [7] (Fig. 1).

The Mask R-CNN model is trained by taking tomato and potato leaves which include healthy and disease leaves. We use the training dataset which contains the full images of the leaves of different types with the background (Fig. 2).

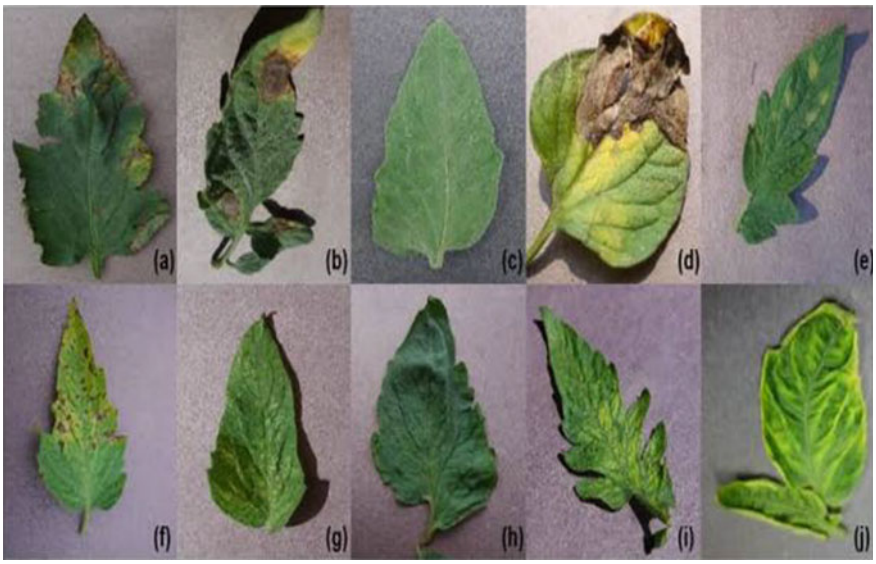


Fig. 1 Leaf types

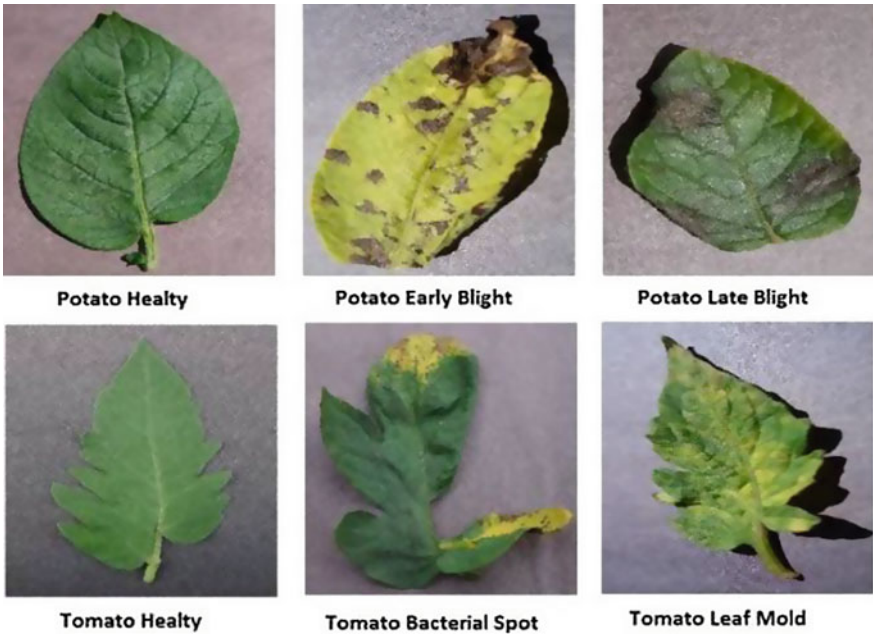


Fig. 2 Leaf types and disease

4.4 Visualizations and Plant Leaf Disease Detection

With the test datasets and validation, generation of visualizations is done; also, application will be created. All the needed files are stored in a GitHub. Git is linked to unified platforms and Render. This model is specially applied for tomato and potato leaves with ten image. The time taken to complete the work is 3 months. So many complex elements are there in this work that can be done very carefully [8]. The main complex thing is image classification and then both are unfamiliar. For the scope of the project, RAID log and Gantt chart are created. Thinking about risks or errors will be done, and taking appropriate measures to rectify these problems observing and updating the project frequently during last two weeks all the tasks including minutes' tasks will be completed. The only cost of the project is analyzing for the research because code will be stored in Google Colaboratory, free cloud storage platform, which provides 25 GB.

5 Experimental Results

5.1 Image Size Trailing Phase

In this phase, we are getting F-score of image sizes 156 × 156 to 256 × 256 which is greater than 90%. Running will be increased along with feature extraction when there is an increase in the image size. We got excellent results in this analysis with an accuracy of at least 85%. For the four epoch, each model learning rate is from 1e-04 to 1e-05. We got best results when the image size is 225. At this size, we got the highest accuracy as well as F1-score. For this research, we got 225 as image size that is best for good results [9].

5.2 Model Optimization Phase

In this phase, accuracy is increased to 0.95 as well as F1-score as 0.95. Fine-tuning is used to plot graph between logistic scale vs. loss. As learning rate increases, loss will also increase. The range of learning rate in between 1e-05 to 1e-04 gives the best results. Due to fine-tuning, here is also a little bit increase in F1-score and accuracy. Finally, after overall implementation, the accuracy will be 2.8% and F1-score as 3.8% (Fig. 3 and Table 4).

The learning dataset is divided into two halves testing and training for predicting the disease type; out of this, 80% of the data was used in the Mask R-CNN model and trained in five batches with different epochs (Fig. 4).

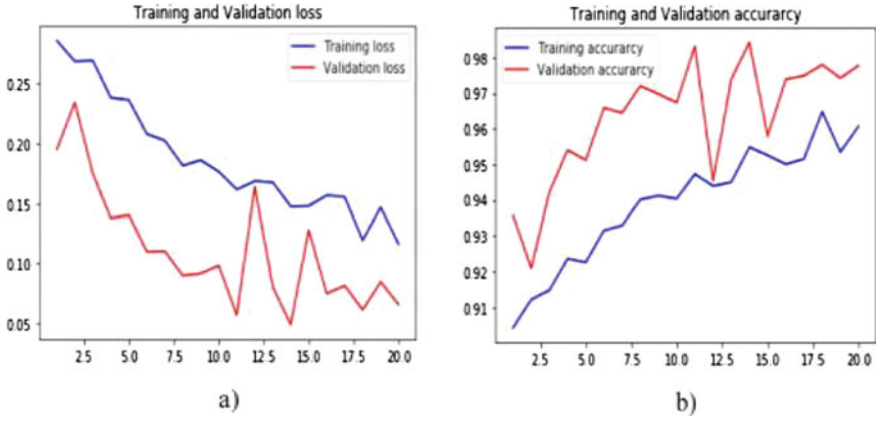
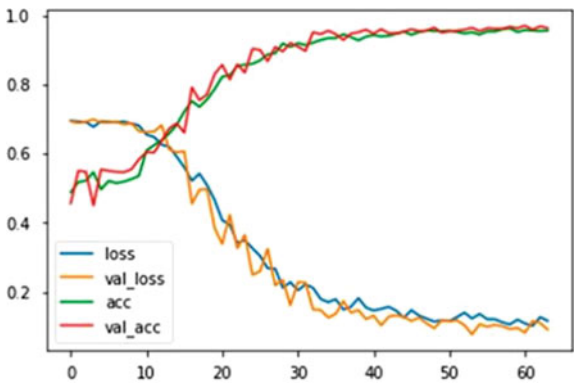


Fig. 3 **a** Training and validation loss **b** Training and accuracy loss

Table 4 Learning rate

Test	Image size	Train loss	Valid loss	Accuracy	F1-score	Time (hours)
1	156	0.17	0.1233	0.9567	0.9449	2.93
2	196	0.1568	0.1160	0.9595	0.9470	3.72
3	225	0.1788	0.1266	0.9532	0.9369	4.39
4	245	0.1320	0.1163	0.9613	0.9460	5.30
5	255	0.1617	0.1249	0.9572	0.9450	5.52

Fig. 4 Validation with loss and accuracy



Test result of the image is shown in terms of the loss and the values loss accuracy and value accuracy, and it is noticed that the detection accuracy is improved using the proposed Mask R-CNN model (Fig. 5).

The same value of loss and accuracy is calculated with different models, and the accuracy value calculated using these models is plotted for comparison (Fig. 6).

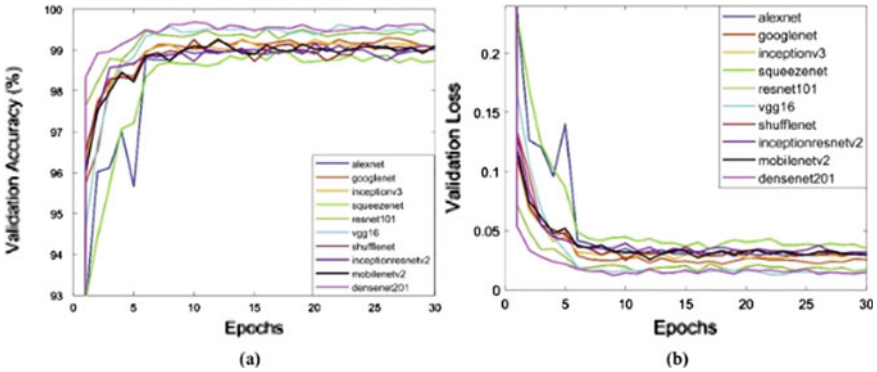


Fig. 5 Comparison of models

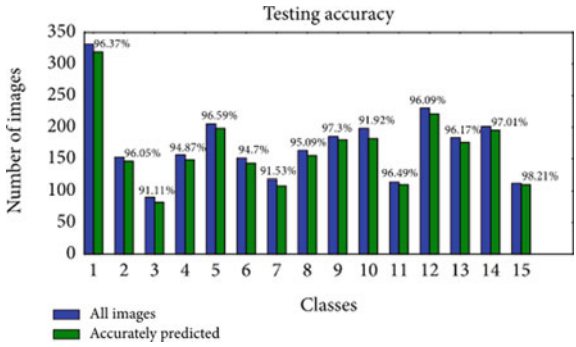


Fig. 6 No. of images versus classes

The image classification algorithm Mask R-CNN is applied on the plant leaf as shown in Fig. 7. The Mask R-CNN extracts the features of R-CNN for predicting the mask of each object. The Mask R-CNN works on selective search where each

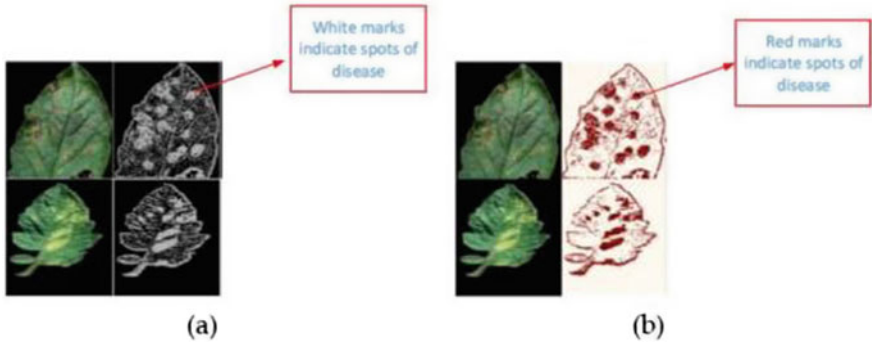


Fig. 7 Image classification using mask R-CNN

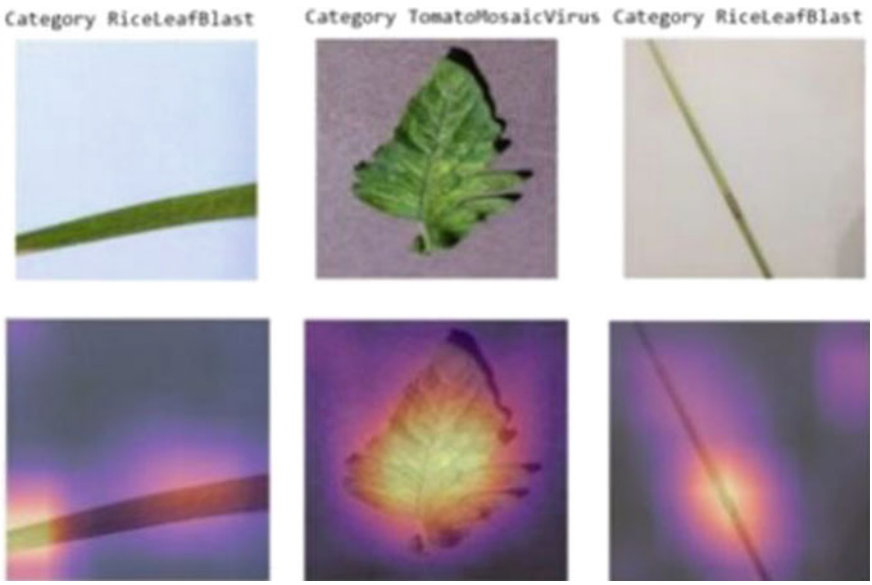


Fig. 8 Heat map for leaf types

feature is extracted using the CNN technique. And, final classification is done using the linear SVM.

For the plant leaf disease detection, heat map is very important for knowing the inner work of Mask R-CNN. To identify very similar diseases, this heat map will help us to identify the disease easily. It mainly uses RGB data for the disease detection. This work also finds the RGB images. This uses a color map as the image is represented by the 3b color component red, green and blue. The RGB color has 24 bit per pixel and takes 8 bits for each color. The heat map of the leaf is shown in Fig. 8.

5.3 RGB Image Visualizations

The image using the visualization tool heat map is shown in Fig. 9. A heat map shows the visuals way of representation and takes decision. A heat map software uses a bar graph representation to denote the height by which the images are visualized.

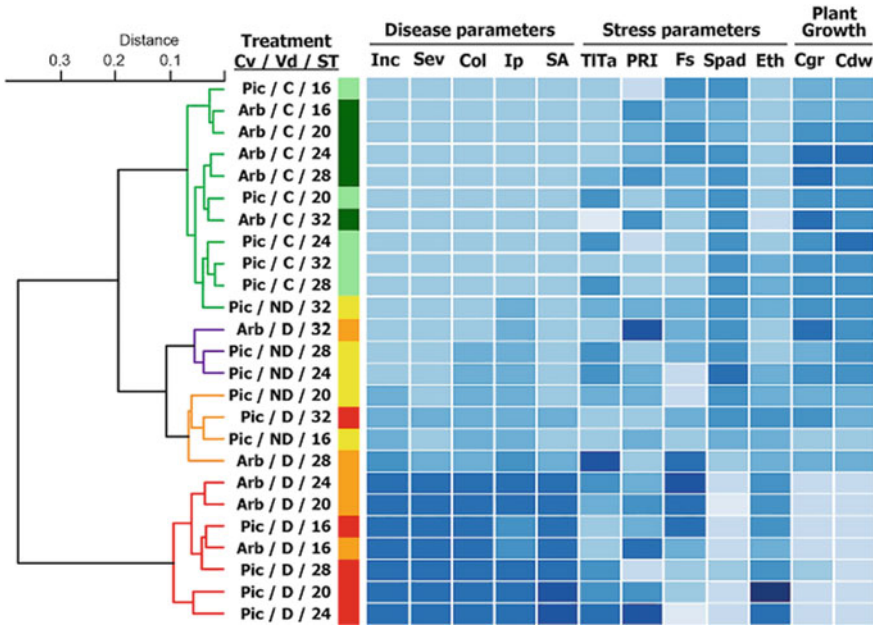


Fig. 9 Heat map results

A correlation heat map represents the correlation between two discrete dimensions using the color correlation. In Fig. 10, two colors brown and green are used to represent the correlation among the leaf length, leaf width, leaf weight, plant height, etc. to estimate the correlation matrix to determine the leaf disease.

6 Conclusion

Many small holder farmers are getting loss because of not detecting plant leaf disease at the right time. To prevent this, finally, we came with detection of plant leaf disease using Mask R-CNN, and deployment of online app will happen. The main aim of our work is to detect the disease with more efficient and accurate manner. The Mask R-CNN algorithm takes the advantage of the CNN and Fast-CNN. With this model, we are able to get 98% accuracy than the existing models [10]. The main beneficial of this model is augmentation and transfer learning. In the future, this model can be used to predict the leaf disease in many variety of leaves.

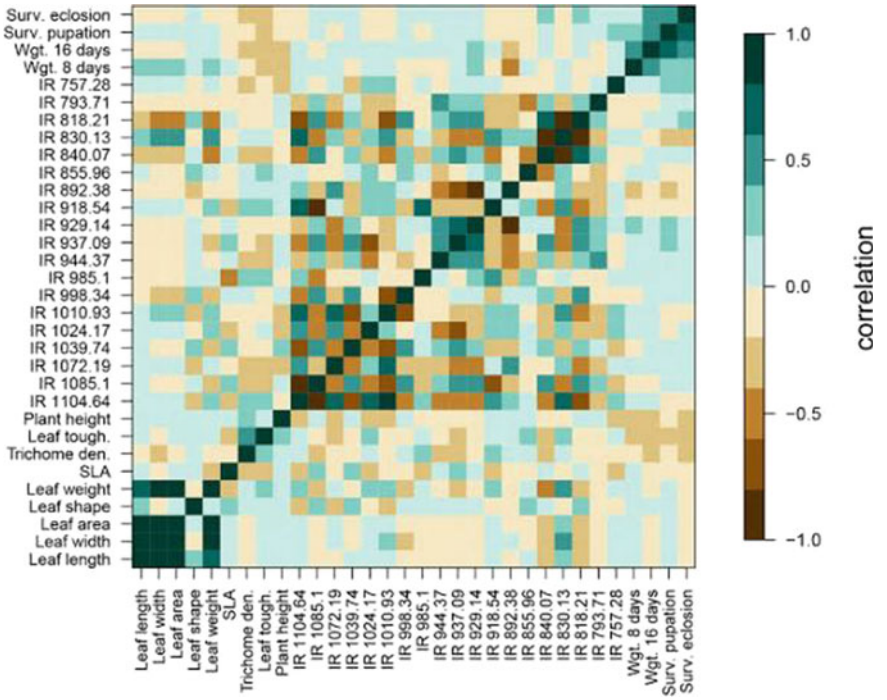


Fig. 10 Correlation map

References

1. Khirade SD, Patil AB (2015) Plant disease detection using image processing. Int Conf Comput Commun Control Autom 2015:768–771. https://doi.org/10.1109/ICCUBEA.2015.153
2. Sardogan M, Tuncer A, Ozen Y (2018) Plant leaf disease detection and classification based on CNN with LVQ algorithm. 2018 3rd international conference on computer science and engineering (UBMK), pp 382–385. https://doi.org/10.1109/UBMK.2018.8566635
3. Jagtap SB, Hambarde SM (2014) Agricultural plant leaf disease detection and diagnosis using image processing based on morphological feature extraction M. IOSR J
4. Singh V, Misra AK (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Process Agric 4(1):41–49. ISSN 2214-3173
5. Kaur S, Joshi G, Vig R (July 2019) Plant disease classification using deep learning google net model. Int J Innovative Technol Exploring Eng (IJITEE) 8(9S). ISSN: 2278-3075
6. Singh V (2019) Detection of plant leaf diseases using image segmentation and soft computing techniques information processing in agriculture
7. Gavhale KR (Nov 2019) An overview of the research on plant leaves disease detection using image processing techniques. IOSR J Comput Eng (IOSR-JCE) 16(1):10–16. e-ISSN: 2278-0661, p ISSN: 2278-8727, Ver. V (Jan 2014). www.iosrjournals.org
8. Harte E (2020) Plant disease detection using CNN, 2020, project plant leaf disease
9. Bahar H (2019) Plant leaf disease detection, a literature review. J Food Sci Technol 3(6)
10. Saleem MH (Nov 2019) Plant disease detection and classification by deep learning. Plant (Base)

An Efficient Way of Predicting Unhealthiness of Frond Using Deep Learning Algorithm

SV. Shri Bharathi, Angelina Geetha, Sarada Nidamanuri, Anil Kumar Rodda, and Suresh Reddy Bhumireddy

Abstract Farming system can improve more by using innovative technologies that will help us to improve the quality and quantity of agricultural production. Nowadays, plant leaf diseases are the major drawbacks for the crop yields, and it is reducing the production and its quality because of that we are facing major threats on food security. The proposed system will be very helpful to reduce the diseases on frond to get more effective crop by using this technique. So this system will be helpful to the farmers; for this, we are using deep learning approach with the combination of algorithms AlexNet and LeNet. A different way of approaching neuron-wise and layer-wise visualization methods was applied by using a deep learning algorithm and trained the model with a publicly available plant diseased dataset given for images. So, this proposed paper observes the neural network and can capture the colors and textures of leaves to respective diseases upon which farmer will take the decisions.

Keywords Disease detection · Deep learning · AlexNet · LeNet · TensorFlow

1 Introduction

In general, introduction of deep learning algorithm efficiently identifies the plant leaf diseases. After preprocessing and training the model (CNN), the dataset is preprocessed such as image size, shape, texture and remove noises in the images. Similarly, this is also done on the test image dataset. A dataset consisting of about 15 different plant species is obtained, out of which any image can be used as a test image for the software.

There are two types: 1. training the model and 2. testing the model. In training model, first collect the raw images, then do the normalization on the raw images and then train the model using a deep learning algorithm.

In testing the model, which is also used to collect the raw images on images, find the normalization to find leaf disease prediction using the CNN model.

SV. S. Bharathi (✉) · A. Geetha · S. Nidamanuri · A. K. Rodda · S. R. Bhumireddy
Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_29

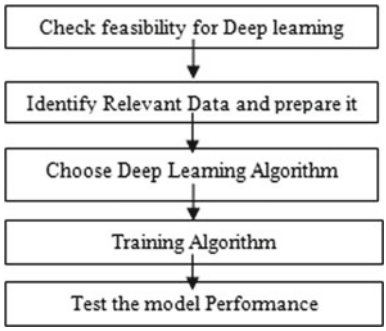


Fig. 1 Training and testing model

Figure 1 shows the training and testing model.

Our proposed system works with the help of the deep learning algorithm which is a convolutional neural network, which takes frond unhealthiest dataset and produces output as a prediction of the frond unhealthiness.

This paper is divided into five sections: Sect. 1 describes introduction, Sect. 2 explains literature work, Sect. 3 proposes system and methodology, Sect. 4 explains the module description, Sect. 5 explores the experimental study, and Sect. 5 explains conclusion.

2 Related Work

Qi et al. [1] did several types of researches to find the efficient measurement for the leaf angle distribution (LAD) from the smartphone photographs. The main challenge is to find the affected leaves and to help the farmers with which they used a structure from motion (SfM) and pyramid convolutional neural network (PCNN)-based leaf detection. By the LAD method, we assume that the leaf size is independent of its inclination angle by their 3D images. The objective of individual leaf detection is to segment individual leaves from background, and then leaf to leaf validation can be done by SfM-PCNN.

The main drawback of the process is it is valid for short focal length so it will not apply for the tall trees due to short focal length in smart phones.

Annabel et al. [2] described the usage of machine learning algorithm, which can detect the plant leaves disease and from the analysis of the data from different aspects, and they are classified into a predefined set of classes. From these processes, there are various types of plant disease, and different machine learning will be useful for finding diseases in various plant leaves. The improvement of this work is their technique is used only for detecting different bacterial, fungal and viral plant leaf diseases.

Militante et al. [3] After getting the image outline, feature extraction happens after the edge detection formula is applied to the image, and finally, horizontal and

vertical scanning methods are used to find the target area. By combining this deep learning technology with feature extraction and preprocessing of input images, it brings out a big difference in accuracy while comparing to other machine learning algorithms such as SVM and BP. The observed future work of this paper is that there are very limited datasets given for the training purpose of this system which led to the unbalance of the number of the experimental dataset.

Park et al. [4] proposed a system which makes use of IoT devices which is enabled by Hadoop and CNN algorithm used to classify the diseases found in strawberry plant leaves. This paper proposed a farm as a service-integrated system which consists of various census and IoT devices implemented in the farmland which is capable of collection of data analysis and prediction. The system registers and manages all the sensors and devices implemented, analyzes the environmental conditions where the system is implemented, and also collects the growth information of the plant integrated with a cloud service which provides an easy access to the data that is being stored.

This paper also provides a high efficiency and accuracy of results. The people search for a plant leaf disease that can be predicted much more accurately if the growth of the plant is observed, recorded and compared with the dataset and also provides specifically to control various parameters of the growth environment remotely. However, the implementation of such system in a small piece of farmland accounts to a heavy investment and maintenance which makes it economically infeasible. The system is also proposed for a control environment such as a greenhouse, a suitable place for implementation of this system.

Alajrami [5] provided an idea to use the convolutional neural network model instead of using the pre-trained models to gain much more accurate results for the real-time usage of the application.

Agarwal et al. [6] studied about the implementation of tomato leaf disease detection system which is implemented using convolutional neural network which is capable of identifying more than nine types of diseases found on the tomato leaf. In this research paper, they took the tomato leaf dataset. With the downloaded dataset, they have implemented by the convolutional neural network model to predict the images, and they have analyzed the diseases based on various parameters such as validation accuracy, testing accuracy and training accuracy.

This system was implemented on a super computer system which is having a high GPU capacity. In this system, they have used approximately around 10,000 images for training purpose, 7000 images in validation and 500 images for the testing purpose. In order to achieve 1000 images dataset per class, they have used augmenter package of Python to build similar new images by flipping, rotating and cropping the existing images. By observing their testing accuracy and training accuracy, we came to know that the proposed CNN model is producing better accurate results than the pre-trained models such as MobileNet and InceptionV3.

Mohammad et al. [7] proposed a technique for plant leaf disease detection and classification using *K*-nearest neighbor (KNN) classifier. The texture features are extracted from the leaf disease images for the classification.

Alajrami et al. [7] have implemented tea plant disease prediction system using convolutional neural network. The proposed system is capable of classifying and identifying seven types of leaf diseases that generally occur on tea plant. The preprocessing of the image input is carried out in this project which involves segmentation and enhancement of the images to reach a higher recognition accuracy of the CNN.

Here, the image segmentation and enhancement process involve renaming the dataset for the branch for reading and storing the dataset, then using the Gaussian filter is applied to smooth and remove the noise inside the image, followed by Gaussian filter, grayscale filter is used for converting the image into grayscale for getting the image outline, then feature extraction happens after the edge detection formula is applied [8] to the image, and finally, horizontal scanning and vertical scanning methods are used to locate the target area.

By combining this deep learning technology with feature extraction and preprocessing of input images, it brings out a big difference in accuracy while comparing to other machine learning algorithms such as SVM and BP. The observed disadvantage of this paper is that there are very limited datasets given for the training purpose of this system which led to the imbalance of the number of experimental dataset.

Peng et al. [9] have published article that studies about the implementation of apple leaf diseases using deep learning algorithm based on improved convolutional neural networks which is implemented using deep learning, convolutional neural networks and feature fusion. This system provides 78% accuracy; based on this, our proposed system will produce more accuracy.

Sharma et al. [10] investigate a potential solution to flaw problem in images processing such datasets with the help of segmented image data to train the convolutional neural network (CNN) models. The paper mainly deals with the comparison between the accuracies provided by the machine which is trained by full images to the machine which was trained using segmented parts of the image. Hence, the paper gives the result of the research showing that segmented images provide a higher accuracy due to more processing of the important segments of the leave when they are scanned. Ashqar et al. [11] present a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

Adekunle et al. [12] proposed a method which uses image processing techniques for the extraction of important features in order to provide the characteristics properties of the image that could be used for the identification.

3 Proposed System

To detect the plant leaf diseases, this paper planned to design deep learning technique so that a person with lesser expertise in software should also be able to use it easily. It proposed system to predicting leaf diseases.

Figure 2 shows the overview of the model. The input image must be properly cropped in the appropriate dimensions and scale before processing with the system, which raises the dependencies of a human involved. Therefore, preprocessing of

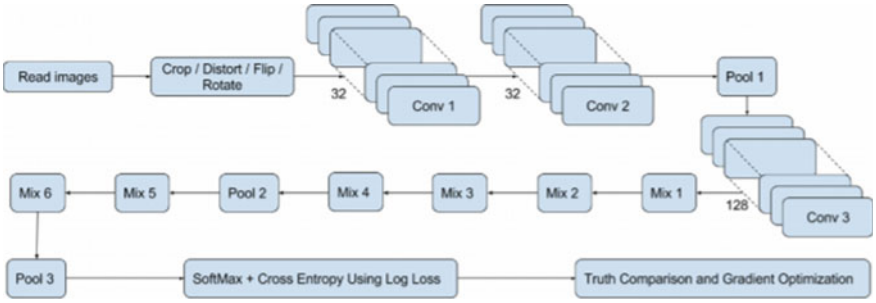


Fig. 2 Model overview

images before model training in CNN can prove invaluable to achieve high real-world performance.

By referring all the papers, the proposed system to predict the unhealthiness of frond using convolutional neural network (AlexNet, LeNet) improves the accuracy more.

Frond unhealthiness causes major production, quality and quantity of the yield were reducing continuously, many farmers are doing agriculture in the rural region, they belong to rural places so the software will not reach easily, they have to pay more money to buy the equipments and more labor, sometime they are unable to predict the diseases which the plants are having, so we planned to design project with which we can easily predict the diseases of the frond, and it will be more useful to the farmers.

Figure 3 represents the architecture of the proposed system.

Architecture Diagram:

The initial step is to collect image dataset which consists of raw images noises and blur images to remove all these things, pre-preprocessing will be done on the collected image dataset from Kaggle after preprocessing, and the dataset is divided into 7:3 ratio

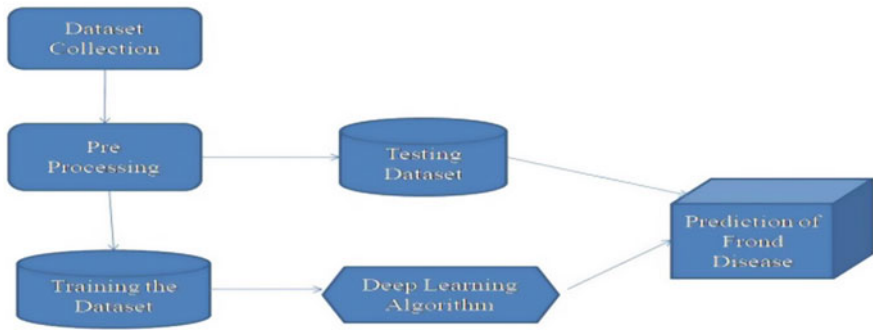


Fig. 3 Architecture diagram

of training dataset and testing dataset. After dividing, select the suitable algorithm to build the model using which algorithm unhealthiness of the frond is predicted using test image dataset.

To predict the unhealthiness of the frond, there are two algorithms used. They are AlexNet, LeNet and tkinter to the graphical user interface.

The whole system is implemented on a cloud-based workbench for data science and artificial intelligence—named as "Kaggle." It efficiently makes use of the datasets, kernel, GPU and RAM that are provided by the Kaggle community. A web interface is integrated with Kaggle's public API for accessing the actual project residing in the cloud platform.

AlexNet:

AlexNet is one of the algorithms in deep learning algorithms which is a part of CNN algorithm. In AlexNet, there are eight various types of layers in which the five first layers are used for convolutional and other three are used to fully connect. Figure 4 shows the working procedure of the system.

Working Procedure:

AlexNet:

AlexNet is one of the algorithm in deep learning algorithms which is a part of CNN algorithm. In AlexNet, there are eight various types of layers in which the five first layers are used for convolutional and other three are used to fully connect.

LeNet:

LeNet contains of seven layers.

a. Two are convolutional layers.
b. Two are average pooling layers.
c. Two are fully connected layers.

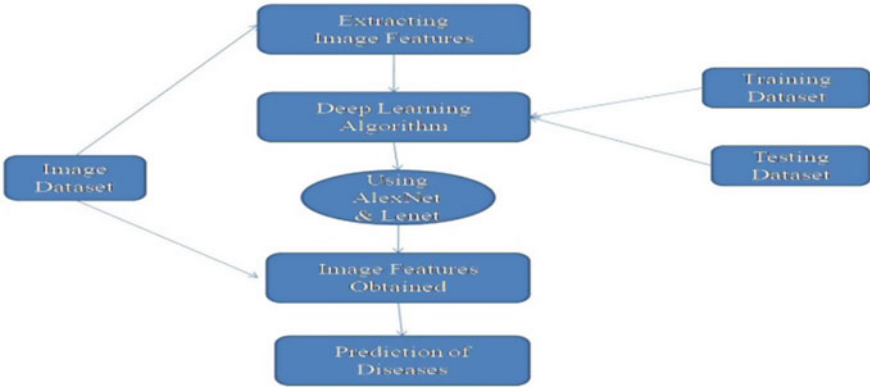


Fig. 4 Working procedure

d. The output layer with activation function.

LeNet is the base for all the convents. It is the linear operation. In LeNet, multiplying kernels or filters with the input is done by convolutional layers, and the pooling layers help to reduce the high dimensionality to curb the over fitting.

4 Experimental Analysis

There are four major modules in the proposed system starting from dataset creation and preprocessing till the final result.

Module 1: Import the given image from the dataset

The dataset is imported using Keras preprocessing image data generator function and also creates size, rescale, range, zoom range and horizontal flip. Then it is imported from image dataset from the folder through the data generator function. Here the paper sets to train, test and validate, and also we set target size, batch size and class mode from this function which we have to train. Figures 5 and 6 describe the model for AlexNet and LeNet.

Model:

Figures 5 and 6 have shown the model for AlexNet and LeNet.

Module 2: To train the module by given image dataset

To train the dataset using classifier and fit generator function also, we make training steps per epochs, then total number of epochs, validation data and validation steps, and using this data, we can train our dataset.

Using grayscale, base and edge base are applied to the image to find out the diseases on the frond, and these two terminologies will be useful to predict the unhealthiness of the frond.

Module 3: Working process of Layers in CNN model

The working process of layers in CNN model is being done to produce the final leaf disease where we will be detecting unhealthy leaves in a crop. Here the algorithm will take an input image from trained model and assign to different aspects or objects of the image different from each other. CNN gives a much lower preprocessing compared to different classification algorithms. For main methods filters, CNN has trained more and has the ability to revise filters or characteristics. Figure 4 explains the accuracy of the system.

Model Analysis: The Fig. 5 explains about the layer, Output Shape and various parameters used.

Figures 7 and 8 show the model analysis for accuracy and loss.

Figure 9 shows the accuracy of the proposed model.

Module 4: Frond disease identification

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 54, 54, 96)	34944
activation (Activation)	(None, 54, 54, 96)	0
max_pooling2d (MaxPooling2D)	(None, 27, 27, 96)	0
conv2d_1 (Conv2D)	(None, 17, 17, 256)	2973952
activation_1 (Activation)	(None, 17, 17, 256)	0
max_pooling2d_1 (MaxPooling2	(None, 8, 8, 256)	0
conv2d_2 (Conv2D)	(None, 6, 6, 384)	885120
activation_2 (Activation)	(None, 6, 6, 384)	0
conv2d_3 (Conv2D)	(None, 4, 4, 384)	1327488
activation_3 (Activation)	(None, 4, 4, 384)	0
conv2d_4 (Conv2D)	(None, 2, 2, 256)	884992
activation_4 (Activation)	(None, 2, 2, 256)	0
max_pooling2d_2 (MaxPooling2	(None, 1, 1, 256)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 4096)	1052672
activation_5 (Activation)	(None, 4096)	0
dropout (Dropout)	(None, 4096)	0
dense_1 (Dense)	(None, 4096)	16781312
activation_6 (Activation)	(None, 4096)	0
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 1000)	4097000
activation_7 (Activation)	(None, 1000)	0
dropout_2 (Dropout)	(None, 1000)	0
dense_3 (Dense)	(None, 15)	15015
activation_8 (Activation)	(None, 15)	0

Total params: 28,052,495
Trainable params: 28,052,495
Non-trainable params: 0

Fig. 5 Model for AlexNet

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 75, 75, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 37, 37, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 12, 12, 128)       36992
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 4608)              0
_________________________________________________________________
dense (Dense)                (None, 256)               1179904
_________________________________________________________________
dense_1 (Dense)              (None, 19)                4883
=================================================================
Total params: 1,222,675
Trainable params: 1,222,675
Non-trainable params: 0
```

Fig. 6 Model for LeNet

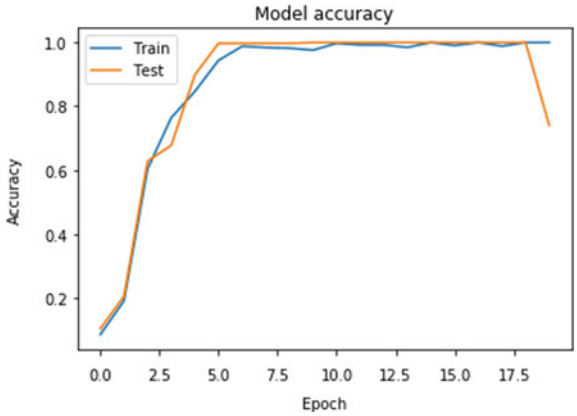


Fig. 7 Model analysis for accuracy

The input image was collected using Keras preprocessing package. That input image is converted into array value using pillow and image to array function package. From the dataset, disease of leaf is classified already. It classifies what are the plant disease leaves. Then leaves diseases are predicted using predict function. Here Fig. 10 shows with frond which is healthy or unhealthy 'for the prediction of diseases using deep learning algorithm'. Figure 10 shows the final output of the system.

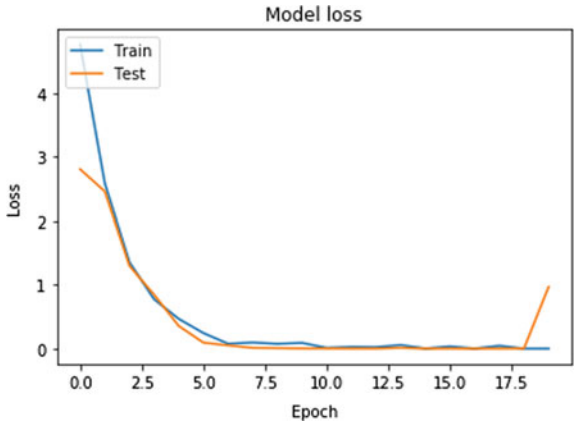


Fig. 8 Model analysis for loss

```
Train for 50 steps, validate for 12 steps
Epoch 1/20
50/50 [==============================] - 24s 481ms/step - loss: 4.7464 - accuracy: 0.0856 - val_loss: 2.8054 - val_accuracy: 0.
1042
Epoch 2/20
50/50 [==============================] - 22s 448ms/step - loss: 2.5848 - accuracy: 0.1914 - val_loss: 2.4579 - val_accuracy: 0.
2057
Epoch 3/20
50/50 [==============================] - 21s 423ms/step - loss: 1.3511 - accuracy: 0.6033 - val_loss: 1.2982 - val_accuracy: 0.
6276
Epoch 4/20
50/50 [==============================] - 20s 407ms/step - loss: 0.7720 - accuracy: 0.7639 - val_loss: 0.8459 - val_accuracy: 0.
6771
Epoch 5/20
50/50 [==============================] - 21s 419ms/step - loss: 0.4643 - accuracy: 0.8463 - val_loss: 0.3549 - val_accuracy: 0.
8984
Epoch 6/20
50/50 [==============================] - 20s 409ms/step - loss: 0.2426 - accuracy: 0.9440 - val_loss: 0.0943 - val_accuracy: 0.
9974
Epoch 7/20
50/50 [==============================] - 21s 420ms/step - loss: 0.0766 - accuracy: 0.9880 - val_loss: 0.0509 - val_accuracy: 0.
9974
Epoch 8/20
50/50 [==============================] - 21s 419ms/step - loss: 0.0951 - accuracy: 0.9843 - val_loss: 0.0122 - val_accuracy: 0.
9974
Epoch 9/20
50/50 [==============================] - 24s 472ms/step - loss: 0.0754 - accuracy: 0.9824 - val_loss: 0.0077 - val_accuracy: 0.
9974
Epoch 10/20
50/50 [==============================] - 23s 467ms/step - loss: 0.0897 - accuracy: 0.9761 - val_loss: 0.0033 - val_accuracy: 1.
0000
Epoch 11/20
50/50 [==============================] - 23s 460ms/step - loss: 0.0121 - accuracy: 0.9975 - val_loss: 0.0029 - val_accuracy: 1.
0000
Epoch 12/20
50/50 [==============================] - 23s 459ms/step - loss: 0.0261 - accuracy: 0.9924 - val_loss: 9.8704e-04 - val_accurac
y: 1.0000
Epoch 13/20
50/50 [==============================] - 21s 423ms/step - loss: 0.0232 - accuracy: 0.9924 - val_loss: 0.0011 - val_accuracy: 1.
0000
Epoch 14/20
50/50 [==============================] - 21s 415ms/step - loss: 0.0568 - accuracy: 0.9849 - val_loss: 0.0128 - val_accuracy: 1.
0000
Epoch 15/20
50/50 [==============================] - 23s 456ms/step - loss: 0.0018 - accuracy: 1.0000 - val_loss: 2.6954e-04 - val_accurac
y: 1.0000
Epoch 16/20
50/50 [==============================] - 20s 404ms/step - loss: 0.0351 - accuracy: 0.9906 - val_loss: 0.0010 - val_accuracy: 1.
0000
```

Fig. 9 Accuracy of a model

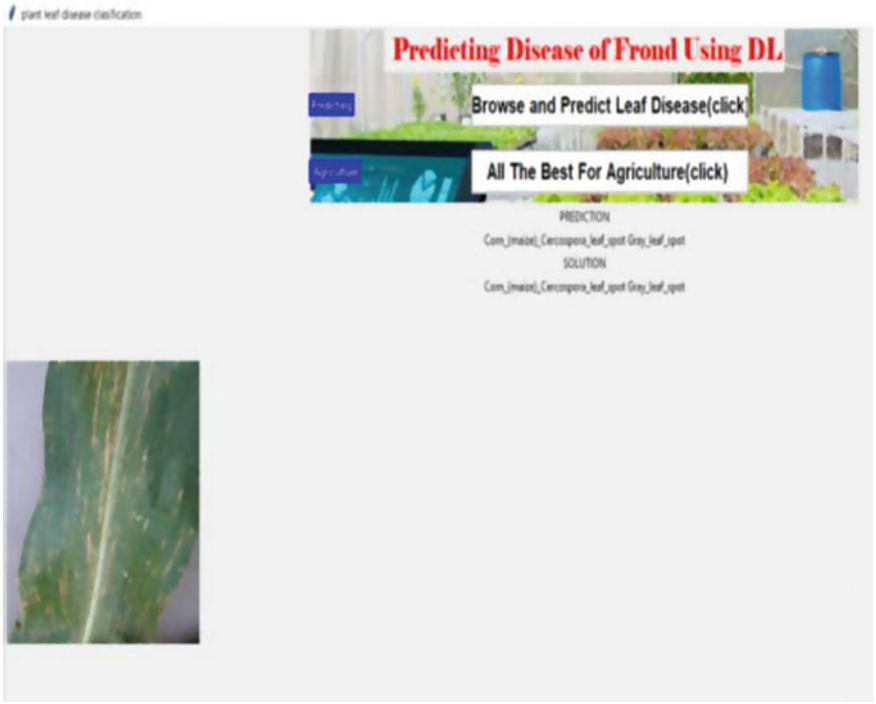


Fig. 10 Final output

5 Conclusion

The proposed paper demonstrated how convolution neural networks, AlexNet and LeNet can be used to find the best approach to detect the unhealthiness frond from the dataset (trained dataset) to get quality and quantity crop. The system worked well, and its performance is up to the mark.

In future, this system will be useful to work with identification of the diseases in flowers, roots and stems. This system is very helpful to the farmers to get more demand of the crop in market.

References

1. Qi J, Xie D (2019) Estimating leaf angle distribution from smartphone photographs. 16(8)
2. Annabel LSP, Deepalakshmi P (2019) Detecting leaf disease by using machine learning—a review
3. Militante1 SV, Gerardoij BD, Medina RP (2019) Recognizing crop sugarcane by using deep learning algorithm. IEEE explorer

4. Park H (2018) Crops disease diagnosing using image-based deep learning mechanism. (CoCoNet). IEEE explorer
5. Alajrami A (2019) Different type of tomato classification approach to detect plant leaf diseases
6. Agarwal (2020) Tomato leaf disease detection. Proc Comput Sci:293–301
7. Alajrami MA, Abu-Naser SS (2019) Type of tomato classification using deep learning. IEEE explorer
8. Sun X (2019) Tea leaf diseases based on convolutional neural network. arXi preprint arXiv: 1901.02694:465-505
9. Jiang P, Chen Y, Liu B, He D, Liang C (2019) Detection of apple leaf diseases
10. Sharma PWG (2020) Disease detection in plants using image segmentation:566–574
11. Ashqar BAM, Abu-Naser SS (2019) Image-based toato leaves diseases detection using deep learning, Int J Eng Res 2(12)
12. Adekunle IM (2020) Implementation of improved machine learning techniques for plant disease detection and classification. Int J Res Innov Appl Sci (IJRIAS) 5(6)

Emotional Recognition System Using EEG and Psycho Physiological Signals

Angelina Geetha, SV. Shri Bharathi, A. Rohan Bernard, R. Yamini Teja, and K. Pradeep

Abstract Machine learning has become the frontier for advanced development techniques, and this is very prevalent in the field of medical science and engineering. Emotion recognition using signals are directly received from the brain can be used to accurately identify and diagnose medical health and psychological problems. In this paper, EEG signals are used to predict the active mood or emotional state of the person's brain wave signals. Data is fed thorough all the given algorithms and tuned. DEAP data in the dataset that is fed into all the above algorithms and the results are observed. In the comparative testing phase, SVM is the most accurate machine learning algorithm, yielding a resulting $f1$ of about 84.73%. The results of this proposed paper shows the different grading patterns that are used to predict the various sentimental states.

Keywords EEG signals · Sentiment analysis · Emotion recognition · SVM

1 Introduction

Sentiment sorting is an analysis of adjustable factors and things considering detection usual feeling of joy or temper with changing degree. It is proven that perception of nervous facial is recognition software is difficult and costly, where those standard and strength are second nature to mankind.

EEG signals are restricted based on their relative frequency, electroencephalogram (EEG) is an extensively used neuroimaging activity which can gauge the potential changes in electric potential generated by the discharge of electric pulse of brain cell, which are found on the scalping and understanding by a device containing a structure of electrodes, where they are disunited into five major bands. Brain–computer interfaces (BCI) where EEG is used to allow the human brain to trace and intercommunicate through computing machine with no interaction. Application

A. Geetha · SV. S. Bharathi (✉) · A. R. Bernard · R. Y. Teja · K. Pradeep
Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Tamil Nadu, Chennai, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_30

program on such research are acquiring quality in the area of affection figuring that helps to understand the states of the person psyche.

The following are some method component to the field of neurobiology and computing research as a result of the task:

1. Body of data are previously treated and united into a large dataset.
2. Process mode is easily available.
3. Resolve of the most sorting algorithm to be utilized for, respectively, sentiment.
4. Precision of each classified pattern in EEG data.
5. The time division method of data produces stronger properties for classifiers.
6. Examination and visualizations of data record "emotion analyzing using EEG, physiologic and video signals" (DEAP) dataset.

This paper is divided into five sections: Sect. 1 describes Introduction, Sect. 2 explains Literature work, Sect. 3 proposes System and Methodology, Sect. 4 explores the experimental study, and Sect. 5 explains Conclusion.

2 Literature Work

A great deal of research has gone into identifying sentiment. This segment goes into a few of them.

This paper suggested a method for detecting tension using EEG signals. These signals used the *K*-means clustering approach and methodology for data classification application [1].

Gaurav and Kumar [2] have depicted a method for distinguishing mental stress levels based on physiological parameters. They classified stress into two and three categories using SVM based on the binary classifier, individually. Out of 41 volunteers, thirty subjects were correctly identified in the binary and correctly categorized in the ternary sorter were done by the 26 subjects.

In this proposed work, the mental distress be detected using phonocardiography signals. The PCG interval S1-S1 was used to classify pre-competitive (or test-related) mental stresses, also known as the interval of inter beat signals (IBI). They carried out another experiment in order to characterize and isolate various traits, using the statistical test for Kruskal–Wallis. The extracted characteristics are then fed into an SVM classification, which is based on the smallest squares [3].

This work examined the impact on human stress of English and Urdu music tracks. Levels of the subjects' EEG signals were captured when they listened to either English or Urdu songs. In their algorithm, the authors used four classifiers: minimal sequential optimization, decent stochastic curve, logistic regression (LR), and multilayer perceptron. Stress has been graded in two levels: 2 and 3. Women react better to music than males, according to research [4].

A comprehensive framework has been suggested in [5] to initially identify mental stress by breakdown in electroencephalograms (EEG) and electrocardiograms (ECG). The estimation of therapy effectiveness was stressed as opposed to accuracy.

In the paper, the paradigm demonstrated the tension in four groups. The findings indicate that the stresses and control conditions were significantly different.

In order to deal with driving tension, a machine-based learning solution was suggested. The classification of EEG signals obtained from the subjects had been carried out by three categories, namely neural networks, SVM, and random forest. The study concluded by saying that SVM was higher than others at rest and stress [6].

The investigators used a low-cost portable EEG headset to measure different cognitive stress states in a single experiment. A stop color word intervention test was used to recall moderate stress responses when capturing the EEG in the scalping region of 18 points. The computational function was then derived from each stimulation presentation's EEG signals [7]. The execution of feature extraction was then fed into through classification, quadratic discriminant analysis, logistic regression, and the k-nearest neighbor classifier were used. The accuracy was better here. But the study did not look at cognitive stress in real time [8, 9].

DEAP Dataset

For emotion classification, DEAP dataset is the open-source data which is available online. A group of research people from Queen Mary University of London gathered the results. It comprises a variety of bioelectrical cues that can be used to assess emotions.

To record the EEG signals, 40 pre-selected music and videos, with a time period of 60 s each is shown, where then signals were downgraded to 128 Hz and noise cancelation using Butterworth filters, which are termed band-pass and low-pass frequency range filters. DEAP dataset can be divided using Russell' circumflex model structure. Figs. 1 and 2.

- Here, with real numbers 1–9, it uses self-assessment manikins (SAMs) in the D dataset. Each member was required to indicate any number from the list between 0 and 10. Given, to draw SAM an uninterrupted scale.

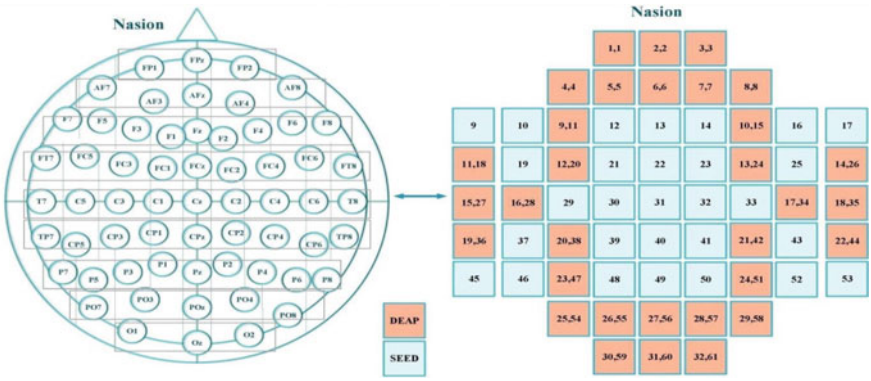


Fig. 1 Diagram—Electrode to channel mapping

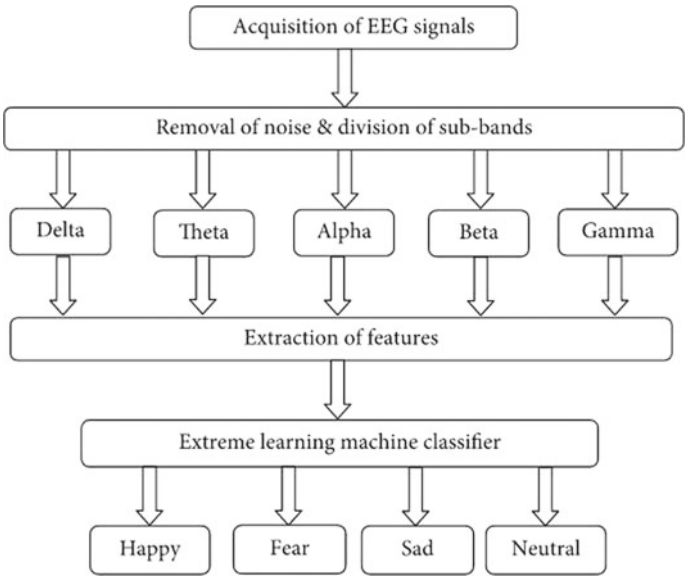


Fig. 2 Proposed methodology graph

- The exfoliation based on self-assessment rating for each subject was collected between 1–5 and 5–9.
- The labeled data was set to high if its ranking was greater than or equal to 5, and it is set to low if it is less than 5.
- High arousal low valence (HALV), low arousal high valence (LAHV), high arousal high valence (HAHV), and low arousal low valence (LALV) to produce a total of four labeled datasets. Where the emotional states are explained by the given labels Fig. 2.
- There are only two types of DEAP dataset labels are available valance and arousal each within a scale of 1–9.
- Labels 1–4 which maps—valance, and 5–9 represent the arousal scale

Electrode to Channel Mapping

To acquire EEG signals, 10–20 International System requires two type of dataset which is DEAP and SEED data, where the 10–20 foreign framework identifies the site of electrodes mounted on the human scalp for EEG signal detection.

- The chosen precise rate of opposing electrodes with 10–20% of the space between head and rear ends, or right and left sides, of the skull. Figure 3 stands for the connected spot of electrodes to the channel numbers are describes by both the datasets.
- DEAP datasets are used by first 32 channels; the other SEED dataset is utilized for the acquiring of EEG signals via 62 channels.

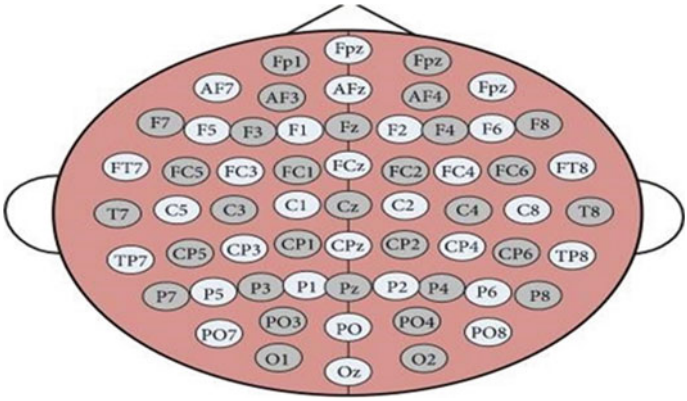


Fig. 3 The 32-electrode position mapping

- The first number maps the DEAP dataset channel number and the second for the SEED dataset as seen in Fig. 3.

The frontal part of the head is mapped by the nation in figure, and channels are maintained using the 10–20 international nation systems Fig. 3.

3 Proposed Methodology

The proposed method includes three key step algorithms, which are as follows:

1. Preprocessing,
2. Feature extraction,
3. Feature selection besides the data acquiring stage. Figure 4 depicts a block diagram of the proposed technique.

Data Acquisition:

- To volunteer the study of people with an average age of 26 years where 7 women and 7 men with normal vision are taken.
- They sit in a dimly lit room, where a computer screen placed at a distance of 120 cm with block of 100 images. Where they are handed with a touch- sensitive button and the EEG signals are recorded through EEG cap with the help of 32 electrodes placed on it.
- Here, the electrodes are separated into two categories forehead electrodes and basilar electrodes and placed in 10–20 arrangement.
- And the picture is shown to people which are divided into targets and non-targets where targets are the snapshot of creature, fish, birds, reptiles, arthropods, and non-targets have indoor and outdoor places, natural landscapes, trees, flowers, and plants.

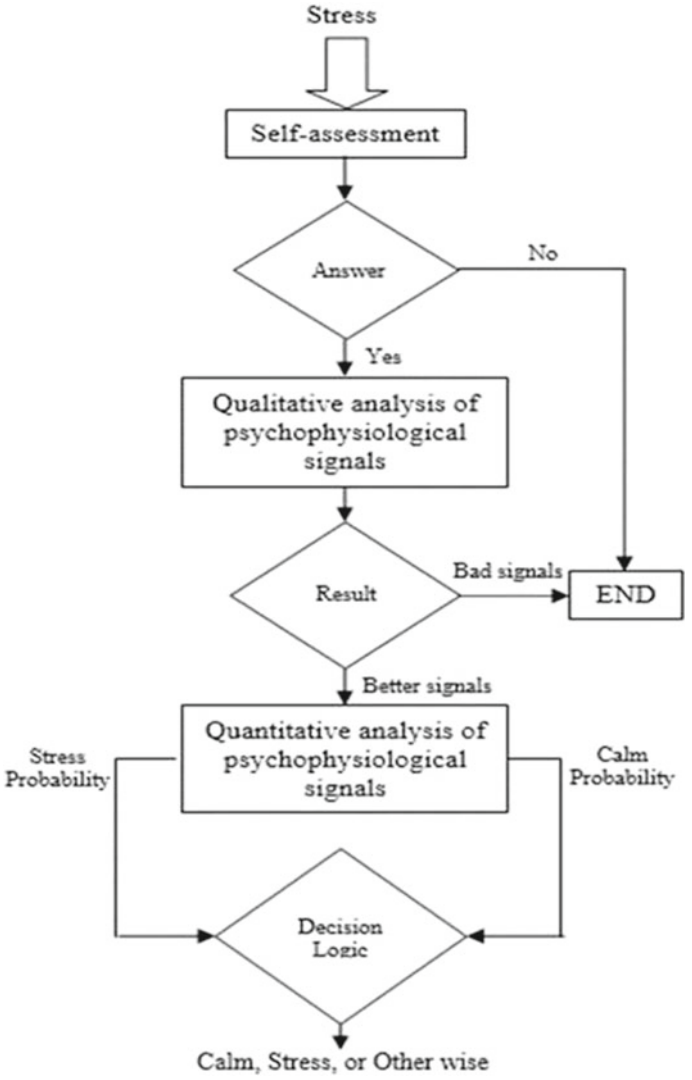


Fig. 4 Flow control diagram

- .The categorization consists of 1000 pictures with 50% distracters and 50% objectives, which are shown only once. In every block, hundred pictures both target and non-target had an equal opportunity during the task. Photograph which is assigned to block viewed with targets image fifty times and fifty different non-target images.
- Every one of the 14 was tested by fifteen targets (with total 210) and 750 non-targets and while training the target image which were flashed continuously.

Participants were suggested to memorize images accurately and immediately with specific sequence of images Fig. 3.

In Fig. 3, the 32-electrode position mapping was utilized in the data acquiring step as per the 10–20 international system.

A Set by Process

1. The input of egg signals is received through the electrodes placed on the scalp of the subject. These electrodes then pick up EEG signals which are generated by the brain due to the neural activity within the neurons of the brain. The COM port is used for serial communication to send the EEG signals which are generated to MATLAB.
2. The EEG signals which are received are split into their four major wavelengths as explained before. These are alpha, beta, gamma, and theta. Wavelength decomposition algorithms are used to extract detailed co efficient data to level *N*.
3. Signal preprocessing, in this step, we remove any and all unwanted signals from the EEG wavelength data received this includes any noise or interference which can cause unwanted signals in the final dataset. This is achieved by using the Butterworth filter, which filters out most of the noise present in the data.
4. After the noise and any unwanted data is removed, the data goes through FFT algorithm to identify each wavelength frequency level fast Fourier transforms are applied to analyze all of the wave frequency ranges.
5. From the FFT algorithm, we can identify and classify all of the wavelength frequency ranges individually and can easily identify any abnormalities.

Acquisition Protocol

EEG signals are measured by the stimulus of photos. These photos are taken from a subset of a system called the international affective picture system (IAPS). IAPS is a database containing a large volume of verified emotionally stimulating pictures. This subset contains 956 different emotionally suggestive picture, which an evaluated by several member. In the study, the picture were used to stimulate the different emotions in the participants. The use of the IAPS help us to standardize and better control the setup for the experiments used. The stimuli pictures used were chose according to rule 1. In which the arousal and valance scales are measured on a scale of 1–9.

LABELING Method of EEG Signals

EEG signals consist of three phase:

- Self-diagnostic phase
- The qualitative study phase
- The quantitative examination of psychological signals phase.

Figure 4 shows the various phases of the operations take place. Later, the trial run there was also a self-diagnostic phase, which is the emotion recognition procedure. The feature vector was equivalent sent into all support vector machine (SVMs), and

the outcome from each SVM classifier was evaluated in the decision line of argument algorithm to select the best emotion. This expansion interpreter can be used as a calculation for the interrelationship scale. The interrelationship scale is granted by the side of the log C for the correlation dimension (r) versus. The curve of record is on a chosen r-series is valued by the slant of this curve at the scaling position35hz filter. This frequency band is used because the frequency bands of interest in the EEG signals are the delta 1–4 Hz, theta 4–8 Hz, alpha 8–13 Hz, and beat 13–30 Hz gamma 30–57 Hz bands. By using this filter, we can remove any high frequencies and also most artifacts.

Feature Extraction:

Extracting some hidden features and useful information from the raw data is knows as feature extraction. Features are classified as the signals that are able to clearly distinguish and differentiate the various emotional states. A common set of values for the EEG signals are used. The features are to be extracted from each of the electrodes of the EEG signal sensors. Two nonlinear feature sets are used; they are multi-dimensional and correlation dimension signals, unwanted noise and refine the signals. By doing this, we remove any environmental noise and drifts.

4 Experimental Study

Analysis of EEG Signals

Preprocessing:

During neuronal activity, the brain generates very small quantities of electrical activity, on the order of millivolts. Electrodes are used to track these signals on the scalp. Since these signals are very weak there is lot of noise. In this paper, that is the reason feature set generation a new random particle is used. Using the function, the features on renewal is calculated where the new particle in MBPO technique is represented by the speed value. The feature is selected when the greater than a randomly formed condition value is between (0 and 1) if not the attribute is not chosen. Hence, using the MBPSO technique optimal feature set is extracted.

Feature Selection:

Modified binary particle swarm optimization (MBPSO) algorithm is used to select the optimal feature set for the identification of stress level issue. Particle swarm optimization algorithm is varied by means of binary value in a proposed method to modified binary particle swarm optimization (MBPSO). To select a new optimal feature set generation a new random particle is used. Using the function, the features on renewal is calculated where the new particle in MBPO technique is represented by the speed value. The feature is selected when the larger than a randomly formed disorder value is between (0 and 1) if not the feature is not selected. Hence, using the MBPSO technique optimal feature set is extracted.

The experimental setup when replicated produces a maximum accuracy, specificity, and sensitivity with an F1 score of about 96.36% accuracy. From the experiment, we notice that adding the adaptation function in the WOA algorithm can increase the effect of the model. The raw dataset used for the training of our algorithm is refined through many per processing by the means of NLMS method, different artifacts such as eye blinking and jerking have been removed. In the feature extraction phase, DCT and MBPSO techniques have been implemented. For the creation of the model, we use an SVM model and to increase its overall effectiveness and accuracy we use the WOA algorithm.

5 Conclusion

The outcome of the test shows the validity of the emotion detection algorithm by computing out the precise assessment with the maximum precision rate. Although the main target of the paper is to accurately identify the emotional state of a person using EEG signals only, we notice a significant increase in efficacy and accuracy if a few biological features are used alongside the EEG signals. Hence, this proposed paper successfully created a model which can detect the emotional state of a person using WOA and SVC models.

References

1. Bernard Patil MN, Patil MS, Patil P (2017) A method for detection and reduction of stress using EEG. IRJET 4(1):1528–1604
2. Gaurav SA, Kumar V (2014) EEG-metric based mental stress detection. Netw Biol 8(1):25–34
3. Cheema A, Mohan Singh (2019) Psychological stress detection using phonocardiography signal: an empirical mode decomposition approach. Biomed Signal Process Control 49:493–505
4. Azri Asif SM, Anwar (2019) Human stress classification using EEG signals in response to music tracks. Comput Biol Med 107:182–196
5. Xia L, Subhani A (2018) A physiological signal-based method for early mental-stress detection. Biomed Signal Process Control 46:18–32
6. Hwang J, Lee S (2018) EEG-based workers' stress recognition at construction sites. Autom Construct 93:315–324
7. Halim Z, Rehan (2020) On identification of driving-induced stress using electroencephalogram signals: a framework based on wearable safety-critical scheme and machine learning. Inf Fusion 53:66–79
8. Sreeja R, Sahay R, Samanta D (2018) Removal of eye blink artifacts from EEG signals using sparsity. IEEE J Biomed Health Inform 22(5):1362–1372
9. Qi1 W (2019) Algorithms benchmarking for removing artifacts in brain computer interface. Cluster Comput 22(S4):10119–10132

Effective Way of Selecting the Industrial Project Team Based on Artificial Intelligence Methods

SV. Shri Bharathi, Angelina Geetha, G. Ranjith Kumar, M. Devi Keerthi Reddy, and Anurag Pandey

Abstract A company makes thousands of new clients every day and receives hundreds of projects. The project which is being delivered must be of high quality, should be delivered on time, and should be completed within the given budget. To deliver a high-quality project, company needs a team of employees who are highly skilled on specific technologies which the project demands. Many companies first take an employee into the team and ask them to learn the technology the project demands parallelly while they work on the project. This method of team selection is quite expensive and takes more time and at some point, quality of the project will be compromised to attain the deadline. The proposed system "Artificial Intelligence Based Team Selection" will be very helpful to choose employees as per the expertise in technologies required by the project. The system will keep account of the performance of the employee based on the certification he has completed, hackathon participations, projects undertaken, technologies known, area of interest, etc. The project manager just has to enter the technical details about the project after which the system will use the power of artificial intelligence to recommend a team of potential employees who will be the best for taking up that particular project.

Keywords Employee expertise · Artificial intelligence system · Recommender systems · Industrial project · Employee team selection

1 Introduction

Although there are lakhs of people who work in a company, the selection of employees to undertake a particular project is a very tricky process. Because at times, projects will be happening simultaneously and many employees will work on them. The management team has to keep account of each and every employee, whether they are in the project or on the bench. After that from the people on the bench, they have to choose the employee whose expertise matches with the project requirements.

SV. S. Bharathi (✉) · A. Geetha · G. R. Kumar · M. D. K. Reddy · A. Pandey
Department of Computer Science and Engineering, School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_31

After this, team formation has to be done, and the project development will start. By following the conventional method of team selection, company loses lots of time and there is a chance the project may face delay.

In some industries, software applications are used to keep account and select employee for the particular task, but those systems work on a normal database search. Even, machine learning-based search is done but the question arises on its performance and accuracy. The proposed system will use artificial intelligence methodology for the selecting potential employees for particular project. The ultimate power of supervised and unsupervised learning will be used to make accurate employee selection. The proposed system will not only consider employees who are in the bench but also consider employees who are currently working on projects, as they may be allocated to some project which may be very relevant to his skills.

The proposed approach uses K-nearest neighbor algorithm, *K*-means clustering algorithm and, Sigmoid Kernel algorithm.

This paper is dividing into 6 sections: Sect. 1 describes Introduction, Sect. 2 explains Literature work, and Sect. 3 proposes System and Methodology, Sect. 4 explains the module description, Sect. 5 explores the experimental study, and Sect. 6 explains Conclusion.

2 Literature Work

AI-Based Design Platform (AID)

The objective of this research paper is to assign or place the deserving student in the right company according to skills they possess and skills requirement by the company. To do so, the author has evaluated and compared the accuracy and efficiency three statistical fitness functions as least square, Pearson correlation, and Manhattan distance. A preference vector table is created for both companies and students; it was based on nine subjects taught in degree [1].

This research paper does the matching of recruiters with job seekers in two different ways. First is based on logs (record) of recruitment agency like curriculum vitae, job statement, and application clicks. It shows that the particulars deduced from their interactions is different from the information contained in the curriculum vitae or resume and job announcement. The second handout of this paper is the hybrid system called 'MAJORE' which stands for matching of jobs and resumes. Here, a deep neural network is formed and trained to contest the collaborative filtering presentation properties [2].

This paper discuses about how artificial intelligence is used in selection of employee. It also discusses about techniques used by companies in recruitment process. This is just a research study based on conceptual papers, journals, books, and websites [3].

This project was developed to keep track of employee particulars. This model is developed for tracking the project in which employee is assigned and daily progress

report of employee is checked. Also, it keeps track of learning of employee. And is capable of performing all manipulation functions including insert, delete, edit, etc. [4].

This paper *K*-means algorithm is used to cluster customers for business application. *K*-mean clustering algorithm of MATLAB is used to develop this system. And *Z*-score normalization is used for two features dataset, i.e., amount of product purchased by each customer and number of visits in a month [5].

To select a potential candidate for job position, there are numbers of tools and softwares that are available in market and many are under development. But there is always doubt on accuracy. In this paper, the author is trying to select best feature or attributes from resume/profile which best describes whether the candidate is qualified enough for particular position. This is being done using profile matching model [6].

In this paper, author has studied a hybrid recommendation system which is more accurate than other general recommendation algorithms available like demographic, collaborative, content-based, knowledge-based and many more filtering techniques. Also used hybrid approaches to address problems such as data sparsity, accuracy, and cold-start [7].

In this paper, different clustering techniques are being compared against a sample data to identify the performance and efficiency of clustering algorithm. Clustering algorithms like artificial bee colony algorithm, firefly algorithm, parallel clustering algorithm, fuzzy *C*-mean model, dynamic clustering, *K*-mean clustering, etc. are studied [8].

In this paper, author has used data mining technique to recommend a job to candidate based on candidate's profile, area of interest, job history, and candidate's behavior. In this, they first get the general preference and then they do content-based matching. This technique provided significant level of accuracy over other basic job recommendation methods [9].

This paper discusses about how the process of sorting of resume can be made more efficient. The author has used the concept of neural network and random forest for parsing of resume which best suits company requirement [10].

3 Proposed System

In the proposed system, concepts of artificial intelligence and machine learning have been used at multiple levels to select team of employees who are the best fit to take up particular project. The system will first select a team with employees at various designations as needed in the project; for example, the project may need developer, software tester, system and programmer analyst, project leader, manager, etc., and then, it will do the designation matching for the selected employees. The way in which user interact with system is shown in Fig. 2 use case diagram.

The powerful combination of unsupervised learning (clustering) and hybrid recommendation methodology will be used to do powerful team selection as shown in Fig. 1 architecture diagram. The proposed system uses the '*K*-means clustering

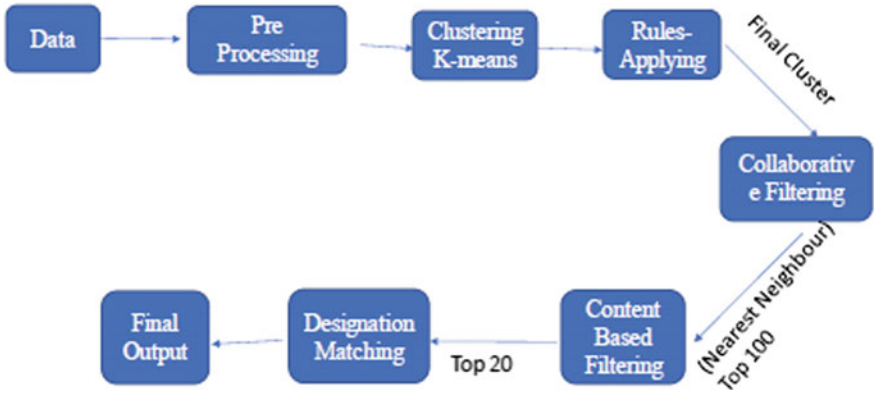


Fig. 1 Architecture diagram

algorithm' for clustering the employees, and then will add rules to clusters to filter out unfit employees and will select the cluster which best fits the rules. The rules will be project-specific; for example, experienced employees and more projects on particular domain. Then will apply Nearest Neighbors based recommendation approach on the selected employee's skills data and will get top 100 candidates who will be the best fit for the given project. After that top 100 candidates apply Sigmoid-similarity based recommendation system to get top 20 best fit candidates. Finally all the employees will be assigned to different levels of designation according to their designation history.

Figures 1 and 2 has shown the architecture and use case diagram.

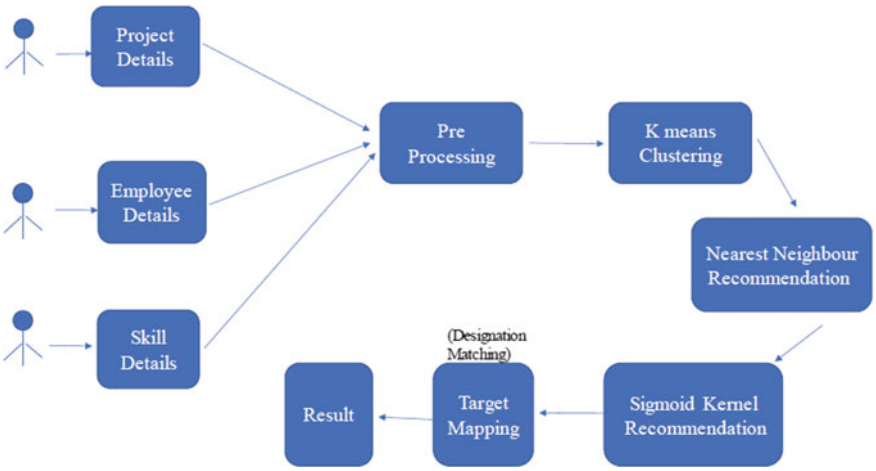


Fig. 2 Use case diagram

Algorithm and Library Description:

This module gives the descriptions about the algorithms used. The algorithms and libraries used to implement system are *K*-means clustering algorithm, nearest neighbor algorithm, content-based recommendation, Collaborative filtering, Scikit learn, and Tkinter for User interface.

K-means Clustering Algorithm:

Clustering refers to segregating the data points in multiple homogeneous groups. Data point in each cluster will be of similar nature. Clustering is an unsupervised learning approach, since there is no label available to compare clustering output. *K*-mean algorithm is a widely used clustering algorithm because of its simplicity and performance.

Nearest Neighbor Algorithm:

Nearest neighbor algorithm can be used for both supervised and unsupervised learning method. "sklearn.neighbors" module provides the functionality for both learning methods. The principle of nearest neighbor algorithm is to discover predefined number of training samples nearest to the new point and forecast the label from there. The most common distance measure technique is Euclidean distance.

Unsupervised nearest neighbors learning is implemented by the class "NearestNeighbour." This library provides interface to three NN algorithms namely Brute_force, BallTree, and KDTree.

Recommendation System:

The objective of recommendation system is to get smaller subset of candidate from a large pool of items.

Procedure of recommendation mostly has three components: candidate generations, scoring system, and re-ranking System. It has been done by scoring each item in the subsets, and then, ranking is done based on score and additional constraints to get final recommended contents.

There are two types of recommendation system:

i. Content-based filtering System
ii. Collaborative filtering system.

In our project, we are using nearest neighbor approach to find out users who are having skills that an ideal employee for that project should have.

4 Module Description

There are seven major modules in this proposed system starting from dataset creation and preprocessing till final team selection.

Module 1: Dataset creation and preprocessing:

A dataset which includes skill details of 1000 employees is created using python programming.

There are two datasets, one with 37 attributes shown in Fig. 3 which are describing about number of project done and area of interest and other with 39 attributes shown in Fig. 4, which describe skills detail.

A total of twenty most used technologies are mentioned in the second table. Various preprocessing steps are applied on the datasets. Some of them are:

```
recommender.data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 37 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Experience                 1000 non-null   int64
 1   Total_projects             1000 non-null   int64
 2   Rating                     1000 non-null   float64
 3   AI_project_count           1000 non-null   int64
 4   ML_project_count           1000 non-null   int64
 5   JS_project_count           1000 non-null   int64
 6   Java_project_count         1000 non-null   int64
 7   DotNet_project_count       1000 non-null   int64
 8   Mobile_project_count       1000 non-null   int64
 9   Area_of_Interest_1_DotNet  1000 non-null   int64
 10  Area_of_Interest_1_JS      1000 non-null   int64
 11  Area_of_Interest_1_Java    1000 non-null   int64
 12  Area_of_Interest_1_ML      1000 non-null   int64
 13  Area_of_Interest_1_Mobile  1000 non-null   int64
 14  Area_of_Interest_2_DotNet  1000 non-null   int64
 15  Area_of_Interest_2_JS      1000 non-null   int64
 16  Area_of_Interest_2_Java    1000 non-null   int64
 17  Area_of_Interest_2_ML      1000 non-null   int64
 18  Area_of_Interest_2_Mobile  1000 non-null   int64
 19  Area_of_Interest_3_DotNet  1000 non-null   int64
 20  Area_of_Interest_3_JS      1000 non-null   int64
 21  Area_of_Interest_3_Java    1000 non-null   int64
 22  Area_of_Interest_3_ML      1000 non-null   int64
 23  Area_of_Interest_3_Mobile  1000 non-null   int64
 24  Language2_Hindi            1000 non-null   int64
 25  Language2_Japenese         1000 non-null   int64
 26  Language2_Kannada          1000 non-null   int64
 27  Language2_Malayalam        1000 non-null   int64
 28  Language2_Tamil            1000 non-null   int64
 29  Language2_Telugu           1000 non-null   int64
 30  Language3_Hindi            1000 non-null   int64
 31  Language3_Japenese         1000 non-null   int64
 32  Language3_Kannada          1000 non-null   int64
 33  Language3_Malayalam        1000 non-null   int64
 34  Language3_Tamil            1000 non-null   int64
 35  Language3_Telugu           1000 non-null   int64
 36  Cluster                    1000 non-null   int32
dtypes: float64(1), int32(1), int64(35)
memory usage: 285.3 KB
```

Fig. 3 Attributes describing project

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 40 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Eid                   1000 non-null   int64
 1   Python                1000 non-null   float64
 2   Machine Learning      1000 non-null   float64
 3   Deep Learning         1000 non-null   float64
 4   Data Analysis         1000 non-null   float64
 5   Asp.Net               1000 non-null   float64
 6   Ado.Net               1000 non-null   float64
 7   VB.Net                1000 non-null   float64
 8   C#                    1000 non-null   float64
 9   Java                  1000 non-null   float64
 10  Spring Boot           1000 non-null   float64
 11  Hibernate             1000 non-null   float64
 12  NLP                   1000 non-null   float64
 13  CV                    1000 non-null   float64
 14  JS                    1000 non-null   float64
 15  React                 1000 non-null   float64
 16  Node                  1000 non-null   float64
 17  Angular               1000 non-null   float64
 18  Dart                  1000 non-null   float64
 19  Flutter               1000 non-null   float64
 20  Vb.Net                1000 non-null   float64
 21  Python_ID             1000 non-null   object
 22  Machine Learning_ID   1000 non-null   object
 23  Deep Learning_ID      1000 non-null   object
 24  Data Analysis_ID      1000 non-null   object
 25  Asp.Net_ID            1000 non-null   object
 26  Ado.Net_ID            1000 non-null   object
 27  VB.Net_ID             1000 non-null   object
 28  C#_ID                 1000 non-null   object
 29  Java_ID               1000 non-null   object
 30  Spring Boot_ID        1000 non-null   object
 31  Hibernate_ID          1000 non-null   object
 32  NLP_ID                1000 non-null   object
 33  CV_ID                 1000 non-null   object
 34  JS_ID                 1000 non-null   object
 35  React_ID              1000 non-null   object
 36  Node_ID               1000 non-null   object
 37  Angular_ID            1000 non-null   object
 38  Dart_ID               1000 non-null   object
 39  Vb.Net_ID             1000 non-null   object
dtypes: float64(20), int64(1), object(19)
memory usage: 312.6+ KB
```

Fig. 4 Attributes describing technologies known

- Data cleaning and outlier detection
- Handling missing values
- Encoding categorical variables
- Standardization and sampling.

Module 2: *K*-means clustering with applied rules:

K-mean clustering is an iterative algorithm that groups the dataset into *k* number of predefined distinct and non-overlapping group known as cluster, where each points of the dataset belongs to only one group. Square distance of each data point from all cluster centroid is calculated and then it is assigned to cluster which is at minimum distance.

The employee skill dataset is being clustered, project-specific rules are applied and the best fit cluster is selected for further processing. This process is known as rules applying. The rules will be project-specific and will be decided by the project manager or clients.

Module 3: Nearest Neighbor-based recommendation:

Nearest neighbor is an unsupervised version of the *K*-nearest neighbor algorithm, where we will be recommending employees with respect skills or technologies the project demands. We will be using cosine similarity for finding the nearest neighbors. This recommendation is being done on the skills dataset of employees. We will have an ideal employee, who will have knowledge on all the skills, the project demands, and we will select employees who are very much similar to this ideal employee. We will take the Top-100 recommendation from this module for further processing. Figure 5 describes distance of the first five employee from the ideal employee after implementing nearest neighbors' algorithm. As each project needs employees to have a particular amount of.

Knowledge in specific skills, this recommendation will be done based on the skills information.

Module 4: Content-based recommendation:

Content-based recommendation is done to produce the final employee's be used in this module. Specific employee attributes will be considered for this recommendation. The selected attributes will be assigned with some weights and the sigmoid kernel will be applied on them to provide final recommendation. This module will help us to find the best fit employees for the given project. We will use the Top 20 recommendations from this process for further processing. This module will select the best out of the best; therefore, this module will make our recommendation more accurate recommendation. We will be using sigmoid-based similarity for providing the final recommendation Fig. 6. The employee-specific data will.

Module 5: Designation matching:

This module will be used to fill the different positions in the team. The above steps will be followed for each designation, i.e., to get the best 'system engineer,' whole program will run once and give the top recommendations, again for team leader system will get details of all team leaders and will recommend most potential employee who can lead a team in this project. The company will make the final

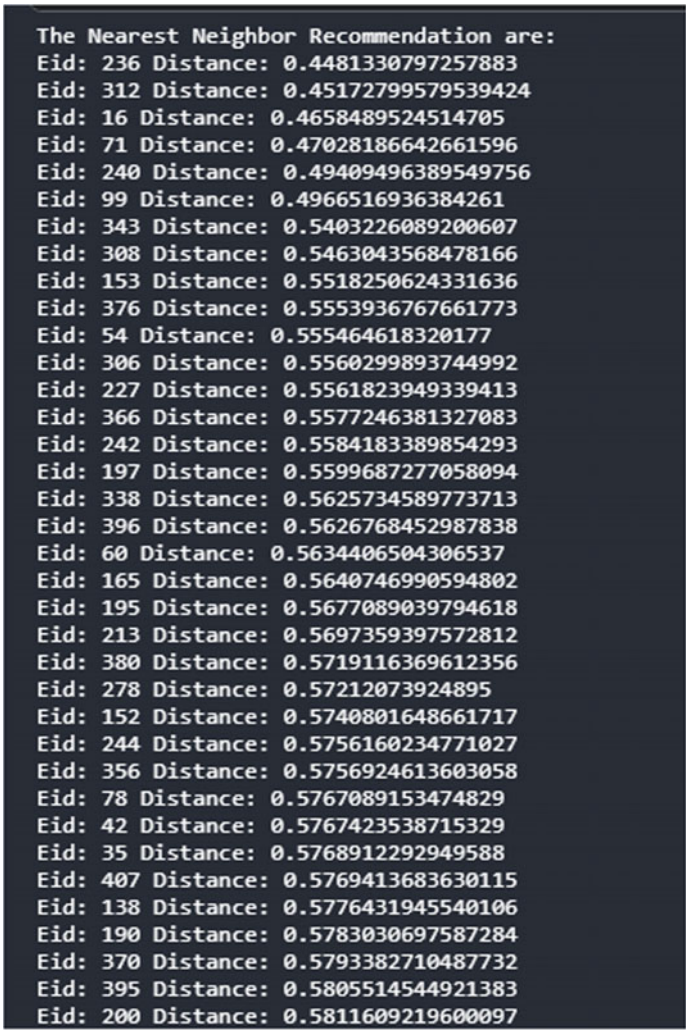


Fig. 5 Top recommendation of nearest neighbor Algorithm

decision, i.e., whether to include that person or not. Designation matching will be done using complex traditional computing and also gets help from some machine learning algorithms. Figure 7 shows the designations of the employees.

Module 6: Displaying selected team:

This is the final module where the top 20 recommended employees will be displayed along with the designation mapped to them.

Since both collaborative and content-based filtering is used for recommendation, it can be said that it is a hybrid recommendation engine.

```
Final Recommendations are:
34          Elsie Holman
35        Tony Hefferon
41        Felicia Lampe
53    Esperanza Galletti
59         Candace Odum
70       Robert Saddler
74        Carol Defabio
77          Alan Duncan
98           Paul Ortiz
106       Ila Emberling
108      Francis Pounds
110          Iris Meier
137    Jennifer Collins
151        Cecilia Dale
152          John Ervin
164        Justin Scott
165           Wade Gass
189           Amy Meyer
190         Sam Sanchez
191   Melissa Hernandez
Name: Ename, dtype: object
```

Fig. 6 Result of sigmoid kernel-based recommendation

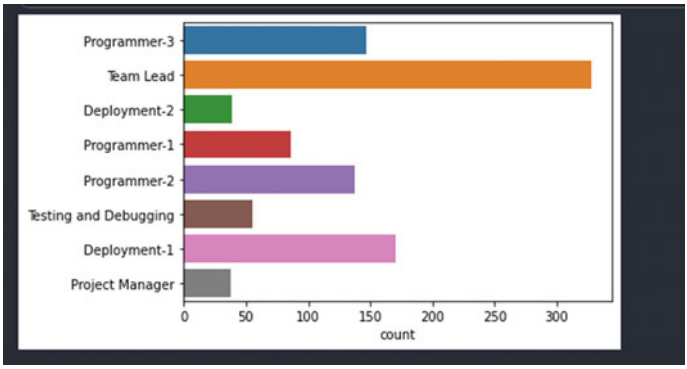


Fig. 7 Designations of the employees

5 Experimental Results

The system successfully selected the team of potential employees who are the best fit for taking up the project.

In module 1, 1000 employees' dataset is created and preprocessed. In module 2 *K*-means clustering algorithm is applied which is forming 3 clusters, Fig. 8 and unwanted clusters are rejected as they do not meet the threshold value after rules applying. And finally, we are selecting the best cluster based on feature importance scoring.

Module 3 gives the name of the top 100 employees who best to be chosen and this is based on the recommendation given by nearest neighbor algorithm (collaborative filtering).

In module 4, top 100 employee details will be provided and based on specific attributes weightage, top 20 employees will be selected. This selection is based on content-based recommendation.

Module 6 will perform designation matching and result display.

Figure 9 shows the clustering result. Also Fig. 9 shows the final recommendation of the work. The overall summary is shown in Figs. 10 and 11.

Case Study

We have tried experimenting the model on recommending people for taking up the project which requires the candidates to know back-end technologies. At first, the data is segregated into five clusters, and we have selected cluster No. 3 because it has the high mean value for the features Java Script Project Count, Dot net Project count, and rating. Now for doing the collaborative filtering, we have considered the

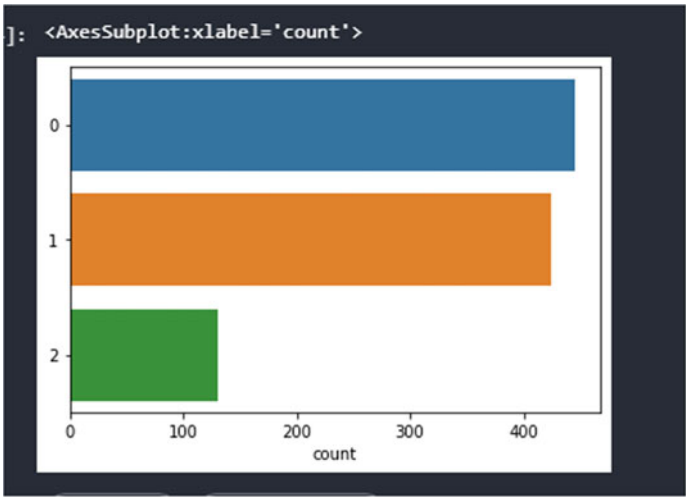


Fig. 8 Clustering result

```
Priority--1
Elsie Holman Programmer-3
Tony Hefferon Programmer-2
Felicia Lampe Testing and Debugging
Esperanza Galletti Team Lead
Robert Saddler Deployment-2
Ila Emberling Deployment-1
Jennifer Collins Project Manager
Justin Scott Programmer-1
----------------------------------------------------------------
Priority--2
Candace Odum Team Lead
Alan Duncan Programmer-3
Iris Meier Deployment-1
Wade Gass Programmer-2
----------------------------------------------------------------
Priority--3
Carol Defabio Team Lead
Cecilia Dale Deployment-1
----------------------------------------------------------------
Priority--4
Paul Ortiz Team Lead
----------------------------------------------------------------
Priority--5
Francis Pounds Team Lead
----------------------------------------------------------------
Priority--6
John Ervin Team Lead
----------------------------------------------------------------
Priority--7
Amy Meyer Team Lead
----------------------------------------------------------------
Priority--8
Sam Sanchez Team Lead
----------------------------------------------------------------
Priority--9
Melissa Hernandez Team Lead
----------------------------------------------------------------
```

Fig. 9 Final recommendation

Fig. 10 Summary-1

PHASES	TASK	APPROACH
1st	Data Pre-processing	Python Script
2nd	Cluster + Rules Applying	Unsupervised Learning
3rd	Collaborative Filtering	Nearest Neighbour
4th	Content Based Recommendation	Sigmoid Kernel
5th	Designation Matching	Python Script

Fig. 11 Summary-2

PHASES	TASK	RESULT
1st	Data Pre-processing	Pre-processed Data
2nd	Cluster + Rules Applying	Selected Cluster of Employee
3rd	Collaborative Filtering	Top 100 Results
4th	Content Based Recommendation	Top 20 Results
5th	Designation Matching	Final Recommendation List

skills such as c#, Asp.net, Ado.net, Java script, Angular, and node as ideal skills that the candidate should know. Now for doing the content-based recommendation, we considered the features area of interest 1-Java Script, area of interest 2-Dotnet, and Java Script project count. So, after the getting the final recommendation, we had done some experimentation on it. The results of the case study are shown in Figs. 12 and 13.

6 Conclusion

In this proposed paper, we demonstrated how unsupervised learning and hybrid recommendation can be used to select the best fit employees from the given set of employees data for the particular project. The objective of the project is to select the team members for the project with particular technical domain and is implemented using *K*-mean clustering, nearest neighbor algorithm for getting best 100 candidates and then sigmoid kernel function is used to get top 20 most potential candidates to work in particular project domain. The system has worked well and its performance is up to the mark. The user interface of project is being developed in Tkinter graphical user interface library which is Python-based, and machine learning is implemented using Scikit learn library.

In future work, we will search for other clustering algorithms whose performance is better than *K*-mean clustering, and we will develop a fully fledged employee selection framework, which can be easily used by the company people to get team recommendation for their project.

	count	mean	std	min	25%	50%	75%	max
Eid	20.0	183.850000	97.341653	2.000000	126.250000	187.500000	280.750000	316.000000
Python	20.0	2.404836	1.100560	1.137795	1.353468	2.493552	3.420180	4.193898
Machine Learning	20.0	2.376111	1.185843	0.458067	1.402867	2.185082	3.527085	4.287146
Deep Learning	20.0	2.350978	1.141811	0.182807	1.387137	2.386284	3.294748	4.174219
Data Analysis	20.0	2.327633	1.210119	0.128248	1.291877	2.295446	3.421686	4.272872
Asp.Net	20.0	2.024812	1.266519	0.524395	1.163149	1.359262	3.190169	4.354950
Ado.Net	20.0	2.097825	1.316792	0.411915	1.250808	1.555168	3.406892	4.471691
VB.Net	20.0	0.267468	0.141475	0.105765	0.176837	0.216672	0.378817	0.541449
C#	20.0	2.163817	1.237155	1.123155	1.315912	1.438428	3.298911	4.594382
Java	20.0	2.010354	1.100487	1.103394	1.203715	1.459353	2.741723	4.259264
Spring Boot	20.0	1.806183	1.253347	0.112460	1.157037	1.440172	2.581426	4.317077
Hibernate	20.0	1.814912	1.314794	0.142468	1.114770	1.471383	2.615187	4.563773
NLP	20.0	1.569525	1.182052	0.233389	0.559065	1.418155	1.543704	4.485827
CV	20.0	1.549126	1.181619	0.210028	0.579915	1.411904	1.545170	4.239863
JS	20.0	2.561624	1.458124	1.138021	1.335368	1.890748	4.238228	4.578862
React	20.0	2.448721	1.516821	0.112276	1.336695	1.878538	4.246520	4.593565
Node	20.0	2.487622	1.589519	0.396021	1.257058	1.865242	4.431683	4.588641
Angular	20.0	2.443780	1.504253	0.116485	1.356304	1.806611	4.182240	4.541952
Dart	20.0	1.685276	0.630505	0.339758	1.418712	1.507806	1.719948	3.281975
Flutter	20.0	1.635369	0.669087	0.110011	1.362585	1.432892	1.759133	3.312867
Vb.Net	20.0	2.090779	1.293041	0.478869	1.309304	1.497376	3.150453	4.505275

Fig. 12 Skills statistical description of selected employees

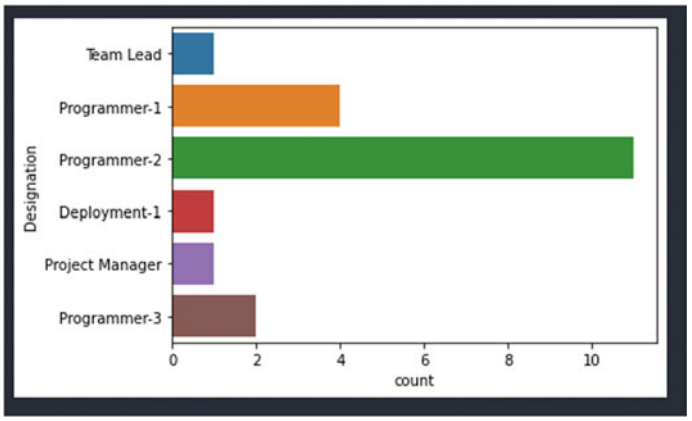


Fig. 13 Designation details of finally selected employees

References

1. Lee D (2018) Artificial intelligence based career matching. JIFS J
2. Schmitt T, Caillou P, Michele (2016) Matching job and resumes: a deep collaborative filtering task, EPiC series in computing. In: 2nd global conference on AI, vol 41, pp 124–137
3. Geetha, Sri Reddy B (2018) Recruitment through AI: a conceptual study. IJMET 9:63–70
4. AkshatLaddha (2019) Employee management system. IRJET, vol 06
5. Chinedu Pascal Ezenkvu (2015) Application of K-means algorithm for efficient customer segmentation. IJARAI. https://doi.org/10.14569/IJARAI.2015.041077
6. Rodriguez LG, Chavez EP (2019) Feature selection for job matching application using profile matching model. In: 2019 IEEE 4th ICCCS conference. https://doi.org/10.1109/sssCCOMS.2019.8821682
7. Morisio M, Cano E (2017) Hybrid recommender systems: a systematic literature review. IOS Press (ResearchGate). https://doi.org/10.3233/IDA-163209
8. Bano S, Khan N, Title P (2018) A survey of data clustering methods. IJAST. https://doi.org/10.14257/ijast.2018.133.14
9. Gupta A, Garg D (2014) Applying data mining technique in job recommender system for considering candidate job preferences. ICACCI
10. Ayishathahira CH, Raseek C, Sreejith C (2018) Combination of neural network and conditional random fiels for efficient resume parsing In: 2018 international CET conference (IC4), IEEE. https://doi.org/10.1109/CETIC4.2018.8530883

A Comparative Analysis on Web Personalization Techniques

Nipun Bansal, Shipra Choudhary, and Utkarsh Patiyal

Abstract The Internet has a large amount of data and is expanding continuously; therefore, it is very important to organize this information in order to provide the users with the best recommendation based on their interests. This is where personalization comes into play. In terms of web pages, personalization refers to the customization of web pages with information that is relevant to a specific user which is determined by tracking the user's surfing history. This paper aims at discussing the various personalization approaches used in web page personalization and concludes the paper with our findings in this domain.

Keywords Recommendation · Personalization · Web page · Domains

1 Introduction

The Internet is full of every type of information, that is why, it is not possible to define a particular type of information to any user. The web itself is up to date with new technology. Since the Internet uses a free style that accepts a formal, informal, ordered, unadorned style to give information on the Internet, not only finding useful data but also organizing it on the basis of customer's interest is also the major challenge nowadays and is called web personalization. The term personalization refers to the act of changing a product to meet someone's individual requirement. When any user puts any query in a search engine, their ways can be divided into two broad sets: searching and browsing. The concept of personalization is purely commercial, and with the growing number of online companies, organizations are becoming personalization oriented to expand customer interaction.

N. Bansal · S. Choudhary · U. Patiyal (✉)
Department of Computer Science Engineering, Delhi Technological University (Formerly Known As DCE), New Delhi, India

N. Bansal
e-mail: nipunbansal@dtu.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_32

2 Related Work

A. ***Effective web personalization system based on time and semantic relatedness*** [1].

This approach aims to employ attributes like semantic information and timing information in addition to navigational patterns for personalization is better at analysing user's interest and activities on the web as compared to only using navigational patterns. The experiments conducted on the model prove that the user profiles generated are better than the traditional approach. Profile accuracy is also improved. Thus, personalized systems' performance can be improved by including semantic and timing attributes.

B. ***Improving web page access prediction using web usage mining and web content mining*** [2].

The authors explore three-session discarding strategies to replace old sessions by new ones. In the first strategy, called "first come first leave", the new sessions are given priority over the older ones, so the oldest sessions are first to be discarded. In the second approach, called "least frequent ones leave strategy", the frequency of each session is taken into account, and the one with the least frequency is discarded first. The last one, "older than some time frame leave strategy", a certain amount of days are used as the time frame. In the event of stored sessions exceeding the capacity, those sessions are discarded which are older than the decided time frame. Both the URLs and the web page content are used.

C. ***Web search personalization using machine learning techniques*** [3].

This study presents a concept-based approach for personalization. A framework is proposed which uses results from a search engine to rank them on the basis of user's interest. The approach uses user's click-through data and past queries to identify user interests. These are then used to predict documents for a user. The obtained data is used for preference mining and for making a profile database. Preferences are stored in user profiles as relations with various concepts. This method gives re-ranked results which are adapted to various users.

D. ***Extraction of behavioural patterns from preprocessed web usage data for web personalization*** [4].

This paper processes user's behavioural patterns for personalization. Web logs have been used for identifying user's interests. This research employs various mining algorithms of frequent sequential patterns (FSP), namely SPADE, WAP-Mine, PrefixSpan for FSP removal from a site. The quality of FSPs is analysed, and then they are used for personalization. At first, a session database is generated, which comprises a pair of tuples.

E. ***Ontology-based web usage mining model*** [5].

In this, the authors aimed to incorporate web usage mining with domain ontology. For this, sequential pattern mining algorithm (CloSpan) has been used. It works on web log file's ontological space to find frequent sequential

patterns. Web pages are recommended to the user by feeding the session data and sequential rules to the recommendation engine. Content-based recommendation fails to match the engine presented by this research. The model makes predictions matching the interest of the user.

F. ***Cluster optimization for enhanced web usage mining using fuzzy logic*** [6].

In this, cluster optimizing method proposed and done by using fuzzy C-means (FCM). Web usage mining consists of the following steps: Data collection followed by log file preprocessing of log file, fuzzy clustering-based pattern discovery, fuzzy cluster-chase algorithm is used for optimizing the cluster, followed by pattern analysis. Web log is used to obtain the input for the mining process. During a user session, the web server records the activity on the site in this log file. This data collected from the navigation activity and then goes through data preprocessing, and using fuzzy C-means algorithm, clusters of similar data are formed. Then pattern analysis is done to create user profiles as pairs of URLs from the clusters secured and find the finest web page used by a good number of the customers.

G. ***An effective technique for personalization recommendation based on access sequential patterns*** [7].

Certain primitive techniques based on filtering, for example kNN, require comparison in real time between the current user session and the patterns obtained from user's usage data history. In order to counter this, the research gives out a more scalable approach for personalization drawn on the basis of certain rules of association. The method devised presents a tree-like structure to store frequent access sequences efficiently and can be merged with an algorithm of recommendation which helps to come up with recommendations without producing each and every association rules first. Since cross beaming of some sub-paths mentions an active window, without contrasting the present customer sessions with each located patterns, this algorithm of recommendation is quicker than several traditional approaches.

H. ***Web page personalization and link prediction using generalized inverted index and FLAME clustering*** [8].

The proposed system consists of some pre-decided parameters which form the basis of personalization. Feedback from new users and an event link access approach is used by the side with domain ontology. The method proposes an inverted index for better estimation of user's needs for which the parameters used are event monitoring and click behaviour along with FLAME clustering.

I. ***Time-driven adaptive web personalization system for dynamic users*** [9].

In this, customer's surfing history is classified and based on that predicts the customer's area of interest. This is done by using an effective personalization method in which the dynamic interests of the user is captured. The proposed model is broadly classified into two parts: online and offline. The offline stage is the static phase, and it processes web logs and represents them in a structured form. A prediction model is devised in the online phase which classifies user request into clusters based on the offline phase. These clusters are then used to provide recommendations using the SVM classifier model.

J. ***Collaborative methodologies for pattern evaluation for web personalization using semantic web mining*** [10].

This paper focuses on the problem of cold start and lack of information in personalization which is overcome by using a hybrid method based on combining information rating system and the content feature selection by using a sample movie dataset which is also applicable to all the other fields.

K. ***Efficient periodic web content recommendation based on web usage mining*** [11].

In this, prolonged algorithm for personalization which considers customer-specific tasks as well as also considers certain other elements connected to the web site uploaded by the visitors to give a better result. This proposed model discovers the customer's surfing action to build a base for an idea along with pre-existing information which are used to forecast the specific customer content.

L. ***A conceptual model for personalized information systems*** [12].

This method covers features such as the levels of personalization, implementation levels, personalization variants and dimensions and costs for integration. Information from different origins was prepared and presented in a consolidated way. The outcome was more of a conceptual framework that would help managers of ICT in defining the best dimension plan to help in the personalization of information systems.

M. ***Web personalization model using modified S3VM algorithm for developing recommendation process*** [13].

It pointed at improving the recommendation services by making use of the data fusion level (S3VM) semi-supervised support vector machine. The MapReduce prototype formed a basis for this approach. The proposed model shows considerable improvement in the performance and precision for extraction of web usage data. The MapReducer model based on data fusion uses S3VM to improve the data classification memory, to reduce memory classification and inaccuracy rate and to boost the implementation performance and lessen the consumption of time.

N. ***An improvised filtering-based intelligent recommendation technique for web personalization*** [14].

It consists of a different technique based on collaborative filtering of boosted item and association rule mining. There are two predictors in boosted item collaborative filtering. These objects give out as an input for the association rule mining. The rule manager assists to choose the right rule from the rule base which results in a personalization which is better than collaborative filtering based on items and filtering based on content as well.

3 Summary of Numerous Approaches

See Table 1.

Table 1 Summary of different methods for personalization

Authors	Focus	Advantages
Vinupriya and Gomathi [11]	Recommendation based on web usage mining	Recommends the best URL present globally
Vinupriya and Gomathi [8]	User behaviour	Better user behaviour analysis
Web personalization using web session for web usage mining	User web sessions	Better recommendation and organization of data
Das and Sajeev [9]	Generating navigational patterns	Effective recommendations for dynamic users
Vivek Arvind et al. [14]	Recommend personalized items	Overcomes the problems of data sparsity and shilling effect
Bharti and Raval [2]	Web content, URL	Reduced prediction time with improved predictions
Doddegowda et al. [4]	Extract behavioural patterns	Recommendations based on user's behaviour and pattern
Ramesh et al. [5]	Ontology with web usage mining	Improvement in the discovered usage patterns
Varghese and John [6]	Similarity of usage patterns	Less prediction time and memory requirement
Tan et al. [7]	Sequential pattern matching	Improves pattern matching speed and supports real-time recommendation

4 Challenges and Issues

The personalization has moved through abounding progressive phases. At first, personalization started as a medium to attract customers by giving them the opportunity to get additional information out of any website and continue their interest. Now, personalization means speeding up the delivery of useful data to the customer in order to achieve the perfect personalization including his or her interests [15]. The five crucial issues faced by personalization are mentioned below:

A. ***Privacy and security***

Customer's safety as well as confidentiality of data is a fundamental problem for personalization because it is formed from customers' personal data. Here, security is an ability of customers and websites to shield their information from illicit access to use and change the data while privacy is a quality of being confidential in the eyes of others.

B. ***Fair dealing and integrity***

Fair dealing and integrity are major repercussions in handling customers on e-commerce sites. Every customer wants to be served fairly when it comes to web site services but when personalization comes in frame, customers act

differently in many ways like interest, time, loyalty and money invested by clients.

C. ***Sense of worth and self-esteem***

Personalization gives a unique opportunity to its customers to maintain and better their sense of worth and self-esteem by providing a customer-friendly policy where customers feel valued, controlling changes and comfortable while having enough variety.

D. ***Usability***

Personalization should be deliberated to be helpful and usable. This is done by providing overall disclosure and other understandable options, so that users can easily select, which best suits their interests at the moment, within a suitable data space.

E. ***Cost***

Actually, personalizing online offers is considered less expensive than converting tangible assets because of the digital type of information sources. Therefore, there should always be an equilibrium between the money spent and the potential gain that can be generated by a decent economy and address the basic trading strategies.

5 Comparative Analysis

See Table 2.

Table 2 Comparative analysis of the various approaches for personalization

Authors	Technique used	Relevance
Khatri and Gupta [11]	Knowledge base model	High
Vinupriya and Gomathi [8]	FLAME clustering	Medium
Web personalization using web session for web usage mining	Data mining technique	Medium
Das and Sajeev [9]	K-means approach	High
Vivek Arvind et al. [14]	Collaborative filtering, association rule mining	Medium
Bharti and Raval [2]	Session discarding algorithm	High
Doddegowda et al. [4]	Sequential pattern mining algorithm-based approach	Medium
Ramesh et al. [5]	CloSpan algorithm	Medium
Varghese and John [6]	Fuzzy C-means algorithm	High
Tan et al. [7]	Tree-based approach	High

6 Conclusion

We have talked about some methods to personalize the web based on distinct areas. Each paper has some edges and flaws, but the purpose of all papers is to make precise recommendations and to provide relevant data and services to each customer at distinct times through these personalization platforms. This is a comparative survey drawn on some specific parameters, and short contrast is given amid all the touched on methods. Henceforth, there should be changes in personalized systems to develop it into a toolkit, which is compatible with a web browser. Therefore, affected domains and service recommendations can be provided to the customer when surfing online.

References

1. Sajeev GP, Ramya PT (2016) Effective web personalization system based on time and semantic relatedness. In: 2016 international conference on advances in computing, communications and informatics (ICACCI), Jaipur, pp 1390–1396. https://doi.org/10.1109/ICACCI.2016.7732242
2. Bharti PM, Raval TJ (2019) Improving web page access prediction using web usage mining and web content mining. In: 2019 3rd international conference on electronics, communication and aerospace technology (ICECA), Coimbatore, India, pp 1268–1273.https://doi.org/10.1109/ICECA.2019.8821950
3. Bibi T, Dixit P, Ghule R, Jadhav R (2014) Web search personalization using machine learning techniques. In: 2014 IEEE international advance computing conference (IACC), Gurgaon, pp 1296–1299.https://doi.org/10.1109/IAdCC.2014.6779514
4. Doddegowda BJ, Raju GT, Manvi SKS (2016) Extraction of behavioral patterns from pre-processed web usage data for web personalization. In: 2016 IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), Bangalore, pp 494–498. https://doi.org/10.1109/RTEICT.2016.7807870
5. Ramesh C, Rao KVC, Govardhan A (2017) Ontology based web usage mining model. In: 2017 international conference on inventive communication and computational technologies (ICICCT), Coimbatore, pp 356–362. https://doi.org/10.1109/ICICCT.2017.7975219
6. Varghese NM, John J (2012) Cluster optimization for enhanced web usage mining using fuzzy logic. In: 2012 world congress on information and communication technologies, Trivandrum, pp 948–952.https://doi.org/10.1109/WICT.2012.6409211
7. Tan X, Yao M, Xu M (2006) An effective technique for personalization recommendation based on access sequential patterns. In: 2006 IEEE Asia-Pacific conference on services computing (APSCC'06), Guangzhou, Guangdong, pp 42–46. https://doi.org/10.1109/APSCC.2006.27
8. Vinupriya A, Gomathi S (2016) Web page personalization and link prediction using generalized inverted index and flame clustering. In: 2016 international conference on computer communication and informatics (ICCCI), Coimbatore, pp 1–8. https://doi.org/10.1109/ICCCI.2016.7479983
9. Das P, Sajeev GP (2017) Time-driven adaptive web personalization system for dynamic users. In: 2017 IEEE international conference on computational intelligence and computing research (ICCIC), Coimbatore, pp 1–6. https://doi.org/10.1109/ICCIC.2017.8524198
10. Bhargava R, Kumar A, Gupta S (2019) Collaborative methodologies for pattern evaluation for web personalization using semantic web mining. In: 2019 international conference on smart systems and inventive technology (ICSSIT), Tirunelveli, India, pp 852–855. https://doi.org/10.1109/ICSSIT46314.2019.8987821

11. Khatri R, Gupta D (2015) An efficient periodic web content recommendation based on web usage mining. In: 2015 IEEE 2nd international conference on recent trends in information systems (ReTIS), Kolkata, pp 132–137. https://doi.org/10.1109/ReTIS.2015.7232866
12. Reis JL, Álvaro Carvalho J (2014) Personalization dimensions: a conceptual model for personalized information systems. In: 2014 9th Iberian conference on information systems and technologies (CISTI), Barcelona, pp 1–5. https://doi.org/10.1109/CISTI.2014.6876933
13. Rajawat AS, Upadhyay AR (2020) Web personalization model using modified S3VM algorithm for developing recommendation process. In: 2nd international conference on data, engineering and applications (IDEA), Bhopal, India, pp 1–6. https://doi.org/10.1109/IDEA49133.2020.9170701
14. Vivek Arvind B, Swaminathan J, Viswanathan KR (2012) An improvised filtering based intelligent recommendation technique for web personalization. In: 2012 annual IEEE India conference (INDICON), Kochi, pp 1194–1199.https://doi.org/10.1109/INDCON.2012.6420799
15. Rajput M, Agrawal R, Kumar Sharma D (2013) A study and comparative analysis of web personalization techniques. In: 2013 IEEE conference on information and communication technologies. https://doi.org/10.1109/cict.2013.6558301

A Machine Learning Technique to Predict the Breast Cancer at Early Stage to Improve the Survivability

K. Sai Vijaya Lakshmi, K. Little Flower, and G. G. Krishna Kanth

Abstract In medical research, identification and diagnosing the disease at an early stage is a challenging task. If we identify the disease at an early stage, we can increase the probability of survival of human life. Nowadays, most of the women are suffering from breast cancer. To predict this breast cancer, we have proposed a supervised machine learning algorithm, i.e., logistic regression. This algorithm categorizes the given data into either cancerous (malignant) or non-cancerous (benign) data. In this paper, the logistic regression results are compared with the results of this logistic regression with random forest and decision tree—the results showed that logistic regression performed better in terms of accuracy.

Keywords Machine learning · Logistic regression · Benign · Malignant · Random forest · Decision tree

1 Introduction

The functionality of the human body depends on the functionality of cells. If any of the cells are malfunctioning, then it creates a problem. In recent trends, women are seriously affected by breast cancer. As per WHO records, breast cancer is the leading cancer, especially in middle-aged women worldwide. 2,68,600 cases were registered with breast cancer in the year 2019. The main challenge is to identify and diagnose breast cancer at an early stage with minimum side effects. Headache, pain, neurological problems are the side effects of this breast cancer.

Predicting this breast cancer is a challenging task. Generally, by using breast ultrasound, diagnostic mammogram, MRI, we identify breast cancer. Early-stage identification of breast cancer is very crucial [1]. If we predict breast cancer at an early stage, then there is more chance of surveillance.

K. S. V. Lakshmi · K. Little Flower (✉)
Computer Science and Engineering, GITAM (deemed to be University), Hyderabad 502329, India

G. G. K. Kanth (✉)
Clians Labs Pvt. Ltd., Hyderabad, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_33

One in eight women diagnosed with breast cancer, but only 80% can expect to receive the specialist's proper treatment and diagnosis. Machine learning is perfect for predicting whether the breast cancer tumor is benign or malignant just by looking at the measurements. Here, we have proposed logistic regression and compared three ML algorithms, i.e., random forest, decision tree, and logistic regression [3].

A random forest classifier is the one that can able to predict the breast cancer, but it is suitable for high-dimensional data. The decision classifier is slightly different; this can output a categorical prediction based on different branches that have been constructed. It is good to determine what class numerical prediction, such as decision categories, by recursively evaluating how well it is able to split the data that comes into its separate category. To predict the breast cancer, we are proposing logistic regression on categorical data [4].

Logistic regression is a type of model used to predict the probability of occurrence of a particular event. This type of regression relies on predicting parameter values based on the prediction equation constructed using training data. However, linear regression predicts the value of the continuous dependent variable.

The novelty of this paper is:

1. Proposed model takes two different class input datasets: +ve class and −ve class.
2. Logistic regression techniques are used to detect cancerous (malignant) data.
3. Our model is used to detect breast cancer at low cost and at high accurate rate.

2 Literature Survey

Most of the breast cancers are identified by the patients by a lump in their breasts. There are different ways to identify their lumps or any abnormalities (Fig. 1).

Typical methods to examine are as follows:

(i) Self-observation method: Examining the breast for any abnormal tumors or lumps by themselves [5].
(ii) Mammography: Here, the examination of breasts can be done through low-level X-rays. It is the initial stage of the investigation to identify breast cancer [6].
(iii) MRI: This examination is done on people who are more toward malignant tumors [7].

Another way to examine breast cancer is by using machine learning techniques. Machine learning techniques are used to predict abnormal tumors, and these tumors are categorized into two types.

(i) Benign tumor: These are not cancer tumors, and they will not do any harm, and these tumors are removable [8].
(ii) Malignant tumors: It is a cancerous tumor that can spread rapidly and harm the human body [9].

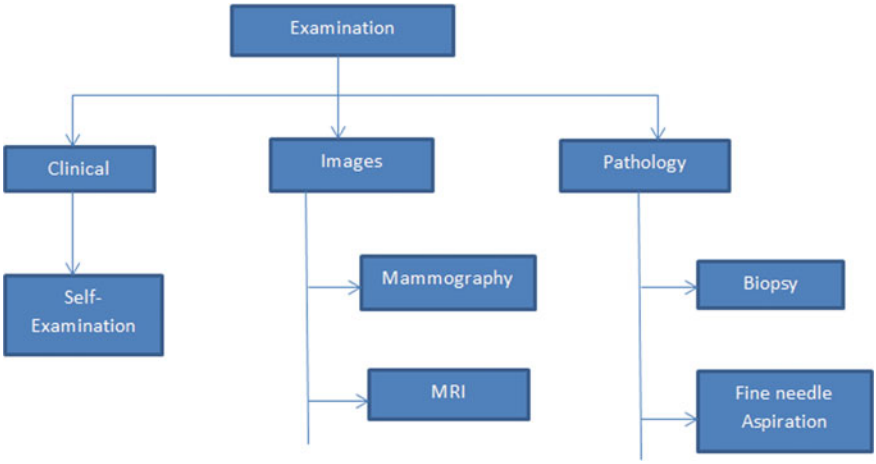


Fig. 1 Typical methods to examine breast cancer

Earlier researchers used random forest classification and decision trees to predict breast cancer. Random forest is a supervised classification algorithm [10] to handle the missing values by itself. Random forest algorithm is suitable for high-dimensional data because accuracy is high for massive data. The decision tree is the most popular machine learning technique to predict the cancerous tumors in the breast. The decision tree can be used after preprocessing the data. It is used to find the data is malignant or benign. This algorithm is used for hierarchical data. Here, we are proposing a logistic regression algorithm to predict the tumor is benign or malignant because logistic regression is best suitable for categorical data.

3 Proposed System

The main motto of this paper is to predict the tumor is benign or malignant. Machine learning techniques have the potential to deal with complex, vast, and high-dimensional data. For this, first, collect the dataset and then apply data preprocessing. After preprocessing the data, divide the given data into two parts, i.e., 70% of data for training and 30% of data is for testing. After division, apply the feature selection method then go for a suitable prediction algorithm [11].

Step 1: Data preprocessing:

Data preprocessing is a data mining technique where we can transform the raw data into an understandable format. Because, in real time, the data are incomplete, inconsistent, and with anomalies. So, the data preprocessing step is applied to resolve these issues.

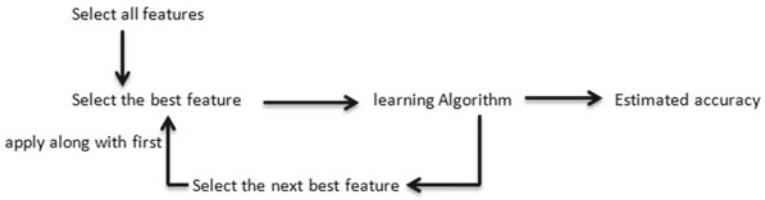


Fig. 2 Filter method workflow

Step 2: Feature selection:

Feature selection means choosing the features that provide accurate results with less data, identifying and removing anomalies, and using the feature selection method.

Step 3: Filter method:

The filter method uses univariate statistics instead of cross-validation methods to measure the features. It is computationally faster and less expensive when dealing with high-dimensional data. Here, we start with the variable whose performance is best against the target and then select another variable whose performance is best against the previous value. We repeat the process until we achieve the present criterion (Fig. 2).

Step 4: Prediction model:

The prediction model is the main phase to identify breast cancer signs and symptoms early to save human lives. To predict this, we are going to use supervised machine learning technique such as logistic regression model to identify the lump is benign or malignant [12].

Logistic regression comes under the classification of the techniques of regression analysis. But, logistic regression can be applied on categorical data, which means it can accept only binary values, i.e., either 0 or 1. So, 0 means it is benign, and 1 means it is malignant. Regression analysis indicates prediction and estimates the relationship between the target and predictor variable [13].

$$D = 1/(1 + e^{-i})$$

The sigmoid function is used to convert the independent variable into an expression of probability with respect to dependent variable. Always the probability is between 0 and 1 (Fig. 3).

We pass the input to this logistic regression function, and then, it will convert the given input into a binary classification, i.e., in between 0 and 1. Here, will fix a threshold value which is in the range of 0 and 1. Based on this threshold value, will identify the given input as cancerous (malignant) or non-cancerous (benign) data [14] (Fig. 4).

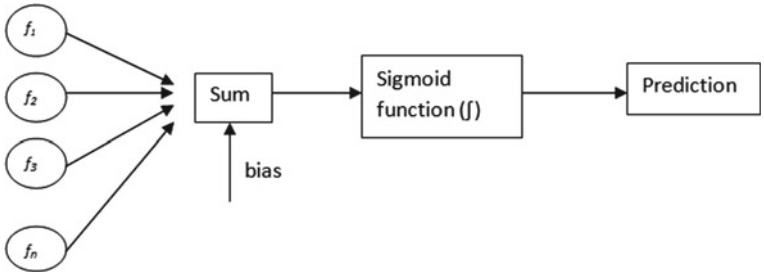


Fig. 3 Sigmoid function

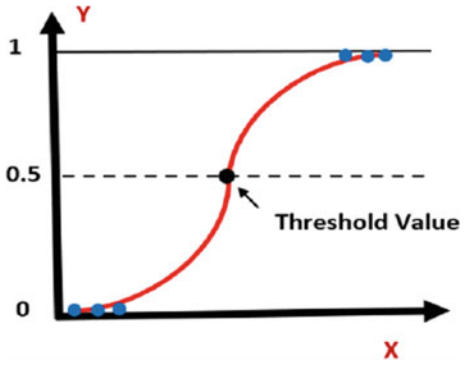


Fig. 4 General logistic regression function

Flowchart for logistic regression:

See Fig. 5.

Pseudo-code for logistic regression:

Input:

Logistic regression model.
Sample input data *i*.
Labels are from *S* [1..*n*].

Function:

for $i = 1$ to n do

Consider a binary vector array Si for each label Si = 1 if it belongs to malignant and Si = 0 if it belongs to benign.

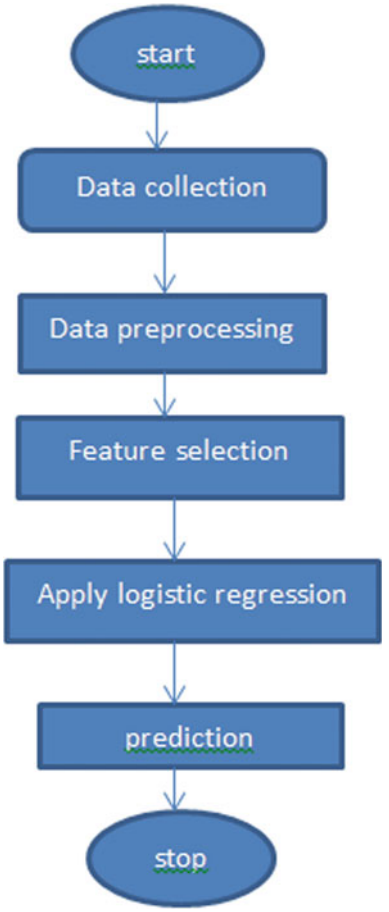


Fig. 5 Flowchart for predicting breast cancer

4 Experimental Results and Performance Analysis

Performance can be measured by using the following parameters. Here, we represent data given categorical classes, i.e., either +ve class or −ve class. In these classes, the possible parameters are as follows:

a. True +ve (tp): Here, the input is +ve, and output is +ve, i.e., correct classification.
b. True −ve (tn): Here, the input is −ve, and output is −ve, i.e., correct classification.
c. False +ve (fp): Here, the input is −ve, and output is +ve, i.e., wrong classification.
d. False −ve (fn): Here, the input is +ve, and output is −ve, i.e., correct classification.

Fig. 6 Measuring the accuracy

Parameters	Decision Tree	Random forest	Logistic Regression
tp	0.91	0.94	0.96
fp	0.02	0.04	0.02
tn	0.97	0.95	0.97
fn	0.08	0.05	0.03
tpr	0.91	0.94	0.96
fpr	0.02	0.04	0.02
precision	0.97	0.95	0.97
recall	0.91	0.94	0.96
accuracy	0.94	0.95	0.97

e. True +ve Rate (tpr): This is measured based on true +ve and all +ves, i.e., $\frac{tp}{tp+fn}$
f. False + ve Rate (fpr): This is measured based on true −ve and all −ves, i.e., $\frac{fp}{fp+tn}$.
g. Accuracy: $\frac{tp+tn}{tp+fp+tn+fn}$.
h. Sensitivity: tpr
i. Specificity: 1 − fpr
j. Precision = $\frac{\text{No. of tp}}{\text{No. of tp+fp}}$
k. Recall = sensitivity.

We have calculated all these parameters for three algorithms, i.e., decision tree, random forest, and logistic regression, respectively, to find the accuracy of predicting the breast cancer effectively to improve the survivability.

Performance measure:

The performance can be measured on accuracy. This accuracy can be evaluated based on tpr, fpr, precision, recall, etc. parameters, and it is shown in Fig. 6.

If we observe the above table true positivity, accuracy is high, and false positivity is low in logistic regression compared with random forest and decision tree. So, the above table proven that logistic regression can predict the breast cancer effectively (Fig. 7).

5 Conclusion and Future Scope

This paper mainly focused on the breast cancer prediction models. Here, we proposed a supervised machine learning algorithm, i.e., logistic regression model, to predict the tumor as benign or malignant. We compared the logistic regression results with random forest and decision tree algorithms. The results shown that logistic regression

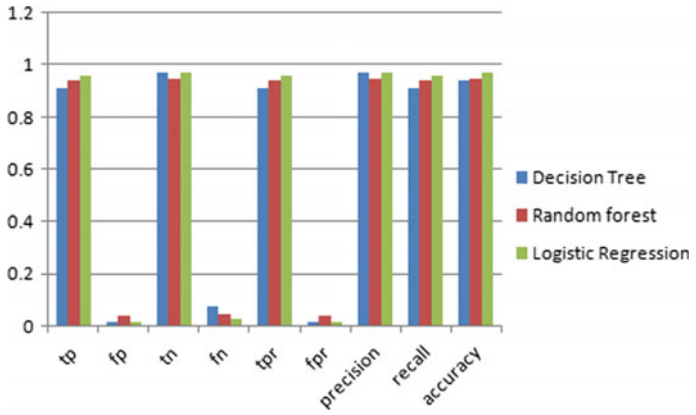


Fig. 7 Graphical representation of accuracy

gives high accuracy (where it depends true and false positivity rate) compared with the decision tree and random forest.

In general, the diagnosis of breast cancer in the medical field is expensive and time-consuming. Using the ML technique, we can predict breast cancer at an early stage and save lives. This prediction acts like a clinical assistant to diagnose breast cancer. In the future, we may apply unsupervised machine algorithms to predict breast cancer.

References

1. Sasikala S, Appavu alias Balamurugan S, Geetha S (2015) A novel feature selection technique for improved survivability diagnosis of breast cancer. Proc Comput Sci 50:16–23
2. Ray S (2019) A quick review of machine learning algorithms. In: 2019 international conference on machine learning, big data, cloud, and parallel computing (COMITCon). IEEE
3. Islam MM et al (2020) Breast cancer prediction: a comparative study using machine learning techniques. SN Comput Sci 1(5):1–14
4. Amrane M et al (2018) Breast cancer classification using machine learning. In: 2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT). IEEE
5. Gowri T, Geetha S (2020) Breast cancer prediction using supervised machine learning algorithms
6. Meenalochini G, Ramkumar S (2020) Survey of machine learning algorithms for breast cancer detection using mammogram images. Mater Today Proc
7. Das K, Behera RN (2017) A survey on machine learning: concept, algorithms, and applications. Int J Innov Res Comput Commun Eng 5(2):1301–1309
8. Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 9(1)
9. Ahmad LGh et al (2013) Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform 4 (124)
10. Cruz JA, Wishart DS (2006) Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2:117693510600200030

11. Asri H et al (2016) Using machine learning algorithms for breast cancer risk prediction and diagnosis. Proc Comput Sci 83:1064–1069
12. Yasmin M, Sharif M, Mohsin S (2013) Survey paper on diagnosis of breast cancer using image processing techniques. Res J Recent Sci. ISSN 2277:2502
13. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260
14. Zhang X-D (2020) Machine learning. In: A matrix algebra approach to artificial intelligence. Springer, Singapore, pp 223–440

Convex Hull: Applications and Dynamic Convex Hull

Vinod Kumar, Mahima, Sahil Verma, and Nishant Nijhawan

Abstract Computational geometry is, basically, the study of algorithms for problems related to geometry. The classical study of geometric objects and geometry is not sufficient for optimal algorithms techniques. Therefore, it becomes important to figure out properties or concepts that provide efficient computation for these geometric problems. The main focus of this paper will be on the convex hull problem, which is one of these geometric problems. We gathered information about convex hulls and its applications along with different algorithms to implement them. The purpose of this paper was to explore the existing applications and come up with new applications of the use of convex hulls. In our research work, the major objective has been to study dynamic convex hull, and propose an algorithm to implement the same. The results of our proposed approach were able to outperform existing implementations of dynamic convex hull.

Keywords Convex · Concave · Polygon · Hull · Convex hull · Geometry · Dynamic convex hull · Wireless sensor networks · Fingertips detection

1 Introduction

The convex hull is an extensively researched structure in the field of computational geometry, having a wide variety of applications like engineering sciences, wireless sensor networks, collision avoidance, and many others. Computation of the convex hull has been widely studied. Several implementations exist for computing a convex hull such as the Graham scan algorithm, Jarvis March algorithm, and Kirkpatrick–Seidel algorithm [1].

Dynamic convex hulls find applications in various areas where points need to be inserted or deleted from the set. The most efficient existing algorithm is the OvL81 with a worst case time complexity of $O(\log^2 n)$ for every insertion or deletion. Unlike

V. Kumar (✉) · Mahima · S. Verma · N. Nijhawan
Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India
e-mail: vinod_k@dtu.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_34

some other algorithms, the Preparata algorithm only deals with insertions, and can insert a point into the convex hull in a time complexity of $O(\log n)$ [2].

Definition of "Convex"

A set $P \subseteq$ Rd is convex if $pq \subseteq P$ for any $p, q \in P$.

The unique property that makes a polygon convex is that it is possible to walk between any two of the vertices straight without leaving the polygon. It refers not only to vertices but to any two points within the polygon. In short, any polygon that has no reflex angles, i.e., angles greater than 180°, is convex. A set having this property is generally referred to as "convex" [3].

A. ***Convex Hull***

Suppose there is a set of points $S = \{a, b, c, d, e, f, g, h, i, j\}$ in the Euclidean plane as shown in Fig. 1. The highlighted polygon abcdef is a polygon such that a segment that connects two points in S lies completely inside the polygon abcdef. The set $S' = \{p | p$ lies in the region bounded by abcdef$\}$ represents a convex set and also represents a convex hull of S [4].

The convex hull for a given set of points can be described as the smallest achievable convex polygon which completely encases all the points belonging to a set. Convex implies that the structure does not have a vertex which is inclined inwards [5].

One helpful way to understand the convex hull is using the rubber band analogy. Assume that the points given in the set are nails and are protruding from a flat surface. Now, think what might occur if we take an elastic rubber band and stretch it around these nails. The elastic rubber band will try to contract itself back to its original length and while this happens, the elastic rubber band will encase the nails, contacting those nails that stick out the farthest from the middle [6].

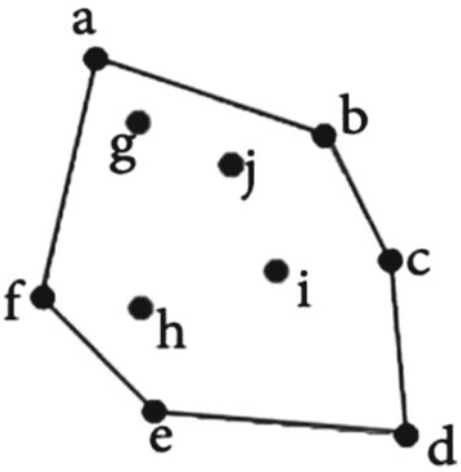


Fig. 1 Convex hull of a set of given points

B. *Dynamic Convex Hull*

The dynamic convex hull problem is a part of dynamic problems in the subject of computational geometry. This problem basically comprises maintenance, i.e., tracking the convex hull for a set of points going through a sequence of changes [7]. These changes comprise points being inserted, deleted, or modified in the set of points belonging to the convex hull.

2 Related Work

A. *Graham Scan Algorithm*

Just by looking at a set of points, we can instinctively make out those points that are possibly going to make contact with the convex hull, and those points which will be enclosed within the hull, and hence be closer to the center [8]. The challenge here is to write code that can allow the machine to do the same, without human intuition of course.

To enable the machines to solve the convex hull problem, we often use the Graham scan algorithm, which is a prime algorithm of computational geometry. This approach gives us the edge-points of the convex hull in the order of where they lie on the graph.

In the beginning of the algorithm, such a point is selected which is sure to lie on the edge of the convex hull that would be generated. Let's call this the starting point. As an example, the point with the smallest ordinate would always be a safe choice. In case multiple points have the same ordinate, the point having the largest abscissa shall be selected. In general, this idea can be used to find other corner points too, not specifically the one with the smallest ordinate or largest abscissa [9].

The next step is to sort the remaining points by the angles that they create at the starting point. If multiple points create an identical angle, the point which is closest to our starting point would be placed sooner in the sorting cycle. This sorting is shown in Fig. 2.

After this, we traverse the sorted array and determine for every point whether or not it happens to be an edge-point on the convex hull. Furthermore, for each set of three points selected, it is to be decided whether they make a concave or a convex corner. In case it is a concave corner, the middle point of the three cannot be an edge-point of the convex hull that would be generated.

Here, "concave" and "convex" corners are described in relation to the whole polygon. These terms alone would have no meaning because there wouldn't be any reference direction.

The edge-points are then stored in a stack so that we can add points into the stack as we encounter them while traversing through the sorted array of points, and remove them whenever we find a concave corner. Technically, the top two positions of the stack need to be accessible so in place of a stack, we make use of a vector, but it can be visualized as a stack because we merely need to focus on the top two elements [5].

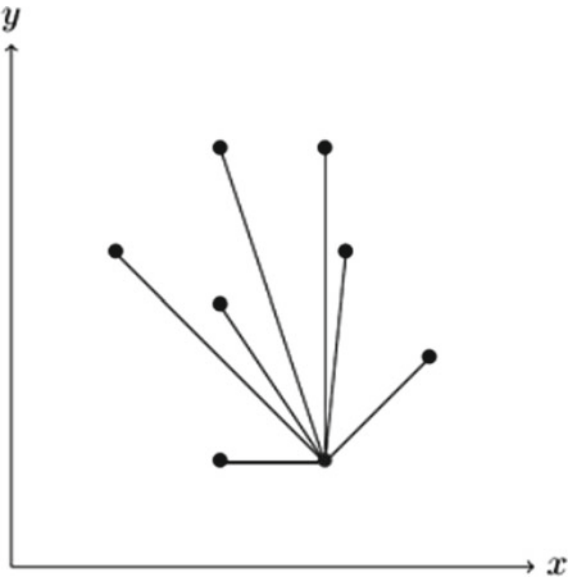


Fig. 2 Graham scan algorithm: sorting the points by angle [10]

Traversing through the points in the sorted array, we add points into the stack. As we keep finding out that certain points aren't edge-points, we pop them from the stack. We can check if a corner is concave or convex by calculating the cross product and checking its sign (negative or positive). As discussed earlier, if a corner is concave, it is to be popped from the stack. If it is convex, it doesn't have to be popped and can remain in the stack for the time being [11].

Like this, we move to the next point and repeat the process: check if a corner is concave or not and make a decision on removing the point accordingly, and so on.

The time complexity of the above implementation is $O(n \log n)$.

B. *Jarvis March Algorithm*

In the Jarvis March algorithm, we pick the point with the smallest abscissa as the starting point, and we go on to enclose points in the anticlockwise direction. Now, if we are at a point p, we use orientation () to find the next point. To find the next point, we set p as the starting point for a cycle. For the next point (say, q), we need the triplet (p, q, r) to be anticlockwise, where r is any other point. Now, we can store q as the point next to p and repeat the cycle by setting q as the starting point of the cycle until we reach the original starting point [12].

C. *The Kirkpatrick–Seidel algorithm*

The Kirkpatrick–Seidel algorithm works on the self-dubbed idea of marriage-before-conquest, which follows the reverse of the steps of divide-and-conquer. In a divide-and-conquer algorithm, the input is first divided into parts, and then, operations are performed on each part after which the parts are merged. The marriage-before-conquest algorithm works in the opposite fashion. The input is first split into parts,

and then, the order of merging is decided, and afterward the subsets are recurred upon. This approach finds the upper portion of the convex hull. The corresponding lower portion can be computed by rotation of the points. The two portions are then merged to find the final convex hull [13].

3 Applications of Convex Hull

Convex hulls have wide applications in numerous fields. Inside arithmetic, convex hulls are utilized to examine polynomials, the eigenvalues in a matrix, and unitary elements, and a few hypotheses in discrete calculation include convex hulls. Convex hulls of pointer vectors of answers for combinatorial issues are vital to combinatorial advancement and polyhedral combinatorics [14]. In financial matters, convex hulls can be utilized to apply techniques for convexity in financial aspects to non-convex areas. In geometry, convex hulls are important for the estimation of boat structures. Certain other applications of convex hull include:

A. ***Finding Evacuation perimeter in case of Chemical or Nuclear Leak***
   Consider a city having sensors placed in a uniform manner throughout. In case of a chemical or nuclear leak, we can find the perimeter of the city that requires instant evacuation by constructing the convex hulls of the areas with radiation levels more than a certain threshold value [15].

B. ***Building blocks for some problems***
   Another application of the convex hull is in calculating the diameter of a set of given points which means, the distance between the two most distant or farthest points. To find, the convex hull is the preprocessing step here. The next step includes finding the pair of points which is the farthest.

C. ***Wireless Sensor Networks***
   The most important utility of wireless sensor networks (WSN) is in event detection. Numerous sensor nodes are to be used in an area to transmit data regarding anomalies in the physical situations to the sink node. Figure 3, these data transmissions can be quite sporadic should a gas leak or wildfire occur. The navigation of moving targets, route reconstruction, and evaluation of the event are all affected by the uneven shape of the event boundary. A convex hull of the occasion can be made in case of navigation. In the instance of a wildfire, people in the woods can be guided out of the area safely just by crossing the convex hull polygon [16]. Obliterated routes in a WSN can be registered effectively by using the convex hull of the event region by updating the lost routes directly by parts of the convex hull. This money on the expense for communication to remake routes and furthermore recuperates the network traffic rapidly.

D. ***Building Containment Zones for COVID-19 prevention***
   This is one of the most captivating applications, given the state of the world, and the impact COVID has had on us, keeping track of the spatial reach out of a corona-infection episode should be possible utilizing the curved structure.

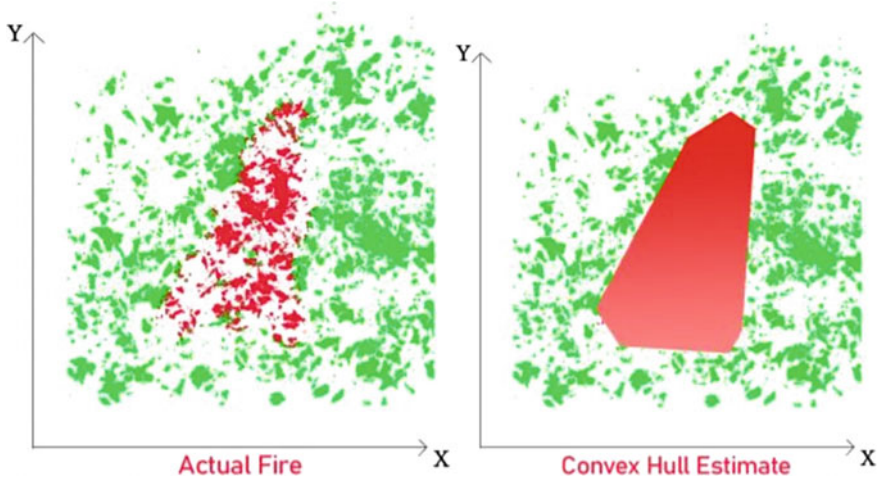


Fig. 3 Convex hull in WSN: convex hull estimate of forest fire

E. ***Engineering Sciences: Life-Cycle Modeling of Remotely Sensed Defects***

One of the methods used for structural condition assessment is remote sensing. Huge amounts of data are being collected by engineers throughout the life cycle of the system. This helps them to calculate the structure's remaining life by interpreting the data to find the defects in the system. The problem with remote sensing is that the defects found are not very useful in calculating the life-cycle dynamics of the system. Using convex hulls for the same can help us overcome this limitation. This reduces the complexity and expressing a defect using its parameterized convex hull allows regular monitoring for the purpose of estimation [17].

F. ***Collision Avoidance***

To check whether a car avoids collision with the obstacles or other cars, we can simply check if the convex hull of that very car avoids collision with those obstacles [18]. Also, Fig. 4, with a convex car, the calculation of paths that avoid collision is way easier.

4 Proposed Approach

We propose to implement dynamic convex hull implementation with insertion operations in a time complexity of $O(\log k)$ and returning the convex hull which consists of k points in $O(k)$ time. We could have done the insertions in $O(\log n)$ and returned the hull in time O(n), but here the key observation is that in most cases $n \ll k$. Therefore: $O(\log k) < O(\log n)$, which optimizes our solution significantly.

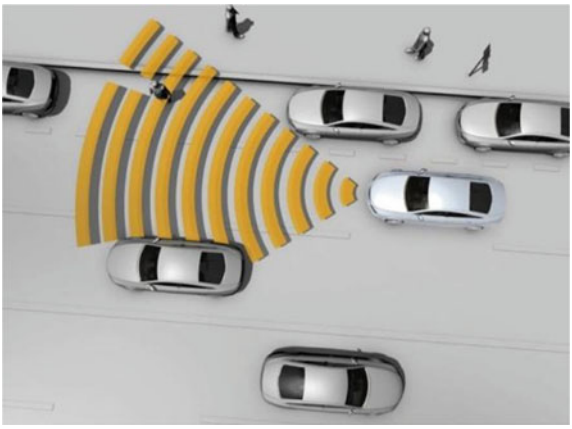


Fig. 4 Applications of convex hull: convex hull in collision avoidance [7]

The following is an elaboration on how to handle insertion of a new point to the convex hull:

If the point to be inserted lies completely inside the existing convex hull, then no change in the convex hull will be required. The following approach is proposed for the insertion of a point that lies outside the convex hull. The structure of the convex hull is proposed as a circular linked list, i.e., the subsequent nodes are connected to each other using pointers, and the last node is connected to the first node. The approach is to find the left and the right tangent from P to the existing convex hull and eliminate all the points which were present in the initial convex hull and will not be present in the new one. In Fig. 5, the points marked with a cross are the points which shall be removed from the hull on the insertion of this new point P. The new

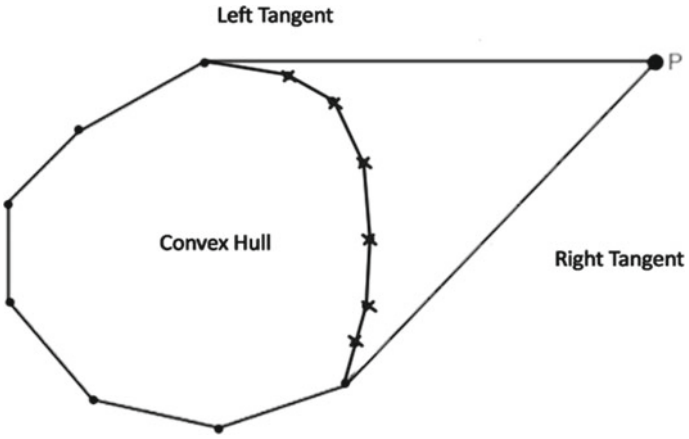


Fig. 5 Implementing dynamic convex hull: finding the left and the right tangents

convex hull will contain *P,* and the cross-marked points will be removed. The naming convention of left and right is relative to the bottommost point of the convex hull.

5 Implementation

A. *The Data Structure*

The structure of the convex hull is proposed as a circular linked list, i.e., the subsequent nodes are connected to each other using pointers, and the last node is connected to the first node.

B. *Finding the Tangents*

First, we will deal with the left tangent and mark the intersecting point. We select the points with the angle greater than the inserted point P with the X-axis. We apply binary search to find the point with the smallest angle with the X-axis for which P lies in the clockwise direction if we consider the hull as anticlockwise moving. It is done using cross product of two vectors made by two adjacent points on the convex hull, and the vector made by joining the bottommost point with the point P. Since the hull is stored in a circular fashion data structure, we can binary search from the first point to the last point and find the desired point with the minimum angle. That will be the point through which the left tangent will be drawn. The time complexity for this job would be $O(\log k)$. Likewise, the right tangent can be computed too.

C. *Deleting the unnecessary points*

After finding both the points, we can just delete the points occurring in between these 2 points and add the new point in our data structure. In the given Fig. 6, points {4, 5, 6, 7, 8, 9} would be deleted by de-linking from both the end-points (i.e., intersection of tangent points). The time complexity for this job would be $O(1)$.

D. *Returning the Resultant Hull*

After we delete the points, we just need to link our new point with both the end-points of our tangent intersections to get our final convex hull.

6 Results and Discussion

Time Complexity Analysis: Let n be the total number of points in the set, and k be the number of points included in the convex hull,

Step-1: Applying binary search for the left tangent would take $O(\log k)$ time.
Similarly, for finding the right tangent. So, the total time required for both tangents,

$$T1 = 2 * O(\log\ k) \tag{1}$$

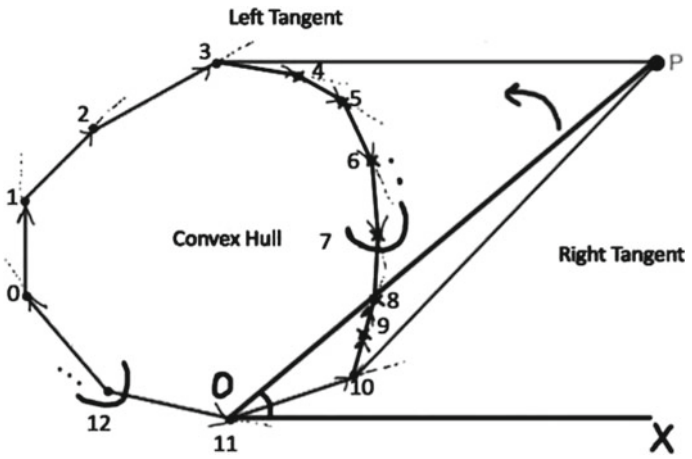


Fig. 6 Implementing dynamic convex hull: deletion of points from left tangent

Step-2: Time required for unlinking the unnecessary points from the hull,

$$T2 = O(1) \tag{2}$$

Step-3: Time required for adding the new point to the convex hull,

$$T3 = O(1) \tag{3}$$

Therefore, the total time taken to insert a new point in the convex hull,

$$\begin{aligned} T &= T1 + T2 + T3 \\ &= 2 * (\log\ k) + O(1) + O(1) \\ &\simeq O(\log\ k) \end{aligned}$$

Also, returning the complete convex hull will cost us $O(k)$ since we have k nodes in our hull data structure, and it is in a form of a circular linked list implementation with set like functionality for making binary searches possible.

We have proposed to implement dynamic convex hull with insertion operations in a time complexity of $O(\log k)$ and return the convex hull which consists of k points in $O(k)$ time. The existing approaches perform the insertions in $O(\log n)$ and return the hull in time $O(n)$, but here the key observation is that in most cases $n \ll k$. Therefore, $O(\log k) < O(\log n)$, which optimizes our solution significantly.

7 Our Contribution

We have studied the existing implementations of convex hull and dynamic convex hull. We have worked on optimizing the time complexity of insertion of points in a dynamic convex hull. We have identified some areas of application where our approach can be utilized.

8 Conclusion

In the field of computational geometry, convex hulls are very important, and they find various uses in the fields of image processing, pattern recognition, and several others. When we need to add a new point to a preexisting convex hull, the traditional methods to compute a convex hull treat it like a new set of points instead of adding one point to the existing convex hull, resulting in a high computation time. We have proposed an optimized implementation of dynamic convex hull which reduces the time taken to insert a new point to an already existing convex hull. We have researched some existing applications of the convex hull, and we have discovered some new applications of the same.

References

1. Convex Hull|Set-2 (Graham Scan), Geeksforgeeks. https://www.geeksforgeeks.org/convex-hull-set-2-graham-scan/
2. https://link.springer.com/article/10.1007/s11786-012-0111-z (2012)
3. https://www.geeksforgeeks.org/convex-hull-set-1-jarviss-algorithm-or-wrapping/
4. https://link.springer.com/article/10.1007/s11590-021-01729-w (2021)
5. Convex Hull Representations (2017) https://link.springer.com/chapter/10.1007/978-3-319-64246-8_2
6. Pascal Sommer, A gentle introduction to the convex hull. https://medium.com/@pascal.sommer.ch/a-gentle-introduction-to-the-convex-hull-problem-62dfcabee90c
7. Kirkpatrick-Seidel algorithm (ultimate planar convex hull algorithm). https://iq.opengenus.org/kirkpatrick-seidel-algorithm-convex-hull/
8. Convex Hull|Set 1 (Jarvis's algorithm or wrapping), Geeksforgeeks
9. The Convex Hull of a Disjunctive Set (2018) https://link.springer.com/chapter/10.1007/978-3-030-00148-3_2
10. Prabhat P (2015) Research proposal: convex hull algorithms. https://doi.org/10.13140/RG.2.1.1934.7689
11. Dynamic Convex Hull, Wikipedia. https://en.wikipedia.org/wiki/Dynamic_convex_hull
12. https://medium.com/@harshitsikchi/convex-hulls-explained-baab662c4e94
13. Science Direct, convex hull, Robert Laurini, in geographic knowledge infrastructure (2017) https://www.sciencedirect.com/topics/earth-and-planetary-sciences/convex-hull
14. Harshit Sikchi, convex hulls: explained, convex hull computation
15. Applications of the convex hull (2016) https://luminousvillage.wordpress.com/2016/05/12/applications-of-the-convex-hull/

16. Convex Hull (2021) Brilliant.org. Retrieved 13:41, Mar 10, 2021, from https://brilliant.org/wiki/convex-hull/
17. Partitioning of the Convex Hull (2017) https://link.springer.com/chapter/10.1007/978-3-319-64246-8_6
18. Convex Hull, Wikipedia https://en.wikipedia.org/wiki/Convex_hull

Typical MANET Design for 5G Communication Network

R. Bharathy, T. Manikandan, P. Keerthivasan, P. Mohamed Fayaz, A. Mohammed Shahid Zafer, and K. Mohana Krishnan

Abstract Versatile ad-hoc networks have being broke down for a more extended time, and these are the associations outlined only from compact user equipment (UE) which work together for interchange of data into a framework-less climate. It is mainly suited for certain applications including the essential edge exercises, lamentable zones, and in the obstructed conditions like grounds and fields, where various clients can trade data clearly with each other or by using different gadgets like switches or switches. Fifth era (5G) of remote interchanges is proposed to give a lot higher information rate. The antenna is designed to operate under sub-6 GHz band. The parameters of this antenna are used as an input in MATLAB 5G toolbox to simulate the modulation/demodulation, waveform generation, uplink/downlink waveforms, and prevention of noise interference.

Keywords 5G · MANET Antenna · MIMO · HFSS · MATLAB

1 Introduction

Wireless fifth-generation technology is proposed for delivering higher data speeds, super-less delay, greater dependability, huge organization limit, and a more uniform client experience to more customers. In communications media, 5G technology is the fifth-age advancement standard for broadband cell organizations, which the cell network affiliations started passing on worldwide in 2019 and is the masterminded substitution to the 4G associations which offer accessibility to most current phones. All the fifth-generation applications in a device are associated with the Web service and the mobile network with the help of radio waves through a nearby receiving antenna in the device. The main benefit of the upcoming organizations is that it

R. Bharathy (✉) · T. Manikandan · P. Keerthivasan · P. M. Fayaz · A. M. S. Zafer · K. M. Krishnan
Department of Electronics and Communication, Rajalakshmi Engineering College, Tamil Nadu, Chennai 602105, India
e-mail: bharathy.r@rajalakshmi.edu.in

T. Manikandan
e-mail: manikandan.t@rajalakshmi.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_35

will possess incredible rates of data transmission, providing increased amount of download speeds, maximum of 10 Gb each second (Gbit/s) [1]. Due to this, the all-encompassing data transmission, it is evident that the affiliations won't simply serve cells like existing cell affiliations, yet similarly, it can be utilized as wide range of Web access providers for workstations and PCs.

2 Existing Methods

The existing methods deal with mobile communication for 5G networks using the millimeter-wave concept. The information transmission need experienced by distant correspondence has been energized for the utilization of the under-used millimeter-wave (mm-wave) range for future 5G broadband consumable versatile correspondence networks. This paper looks at about the attributes of the mm-wave inciting channel, and features the guideline troubles, plans and advantages related with the utilization of mm-waves. The current receiving wire configuration presents the plan of a twofold band PIFA radio wire for 5G applications on a simplicity substrate with littlest structure factor and most stretched out data transfer capacity in the two groups (28 and 39 GHz). The proposed twofold band PIFA radio wire comprises of a shorted fix and a changed U-molded space in the fix. The radio wire shows great coordinating at and around both focus frequencies. A CPW line and a taking care of through are utilized on the base side of the substrate to energize the PIFA radio wire designed on the top side of the substrate which additionally work with the estimations of the receiving wire at mm-wave frequencies. The receiving wire was planned on minimal expense Isola FR406 substrate [2]. Another design deals with the design of a patch array antenna design using HFSS. Because of expansion in the advanced cell clients, the limit request additionally expands, the fifth era (5G) portable innovation would have the option to incredibly build correspondence limit by utilizing the enormous measure of range in the millimeter-wave (mm-wave) groups. This design shows the inspiration for future receiving wire for 5G portable specialized, approach, and offers an assortment of recreation results that show 28.5 GHz recurrence that can be utilized while utilizing steerable directional radio wire. The proposed thought is carried out by utilizing HFSS software tool. The multicarrier transmission strategies have been the most dumbfounding one for the progression of far-off correspondence systems like 4G, long-term improvement and now the substitution 5G [3]. Existing MANET configuration manages the utilization of MANETs for 5G correspondence [4]. Versatile ad-hoc networks (MANETs) are getting noticed for a long time, and there are the organizations framed exclusively from versatile user equipment (EU) that are coordinating in trade information to an infrastructure-less climate; MANET can be utilized for some applications which incorporate the strategic edge activities, heart-breaking regions, and in the blocked conditions like grounds and areas where numerous clients will trade data straightforwardly with one another or utilizing different gadgets like switches.

3 MANET Antenna

A MANET contains different mobile phones that get together to shape an association dependent upon the situation, with no assistance from any current Web system or base stations. A MANET can be described as a self-overseeing plan of center points or MSs (moreover filling in as switches) related by far-off associations, the relationship of which outlines a correspondence network exhibited as an abstract correspondence diagram. This is instead of the remarkable single bob cell network model that maintains the necessities of distant correspondence between two portable hubs depends on the spine-leaf architecture and stationary base stations. In the MANETs, there is no framework for this exists, and organization geography might increase progressively in an eccentric way because hubs permitted to be mobile and every other hub restricts sending power, confining admittance to the hub just in the adjoining limits. So, MANETs are essentially shared, multi-bounce remote organizations in which data bundles are communicated to the intermediate station starting from the source node to a discretionary objective.

Some particular uses of specially appointed organizations incorporate modern and business applications including helpful versatile information trade. There are so many extant and upcoming military systems administration prerequisites for hearty, IP-consistent information administrations inside portable remote correspondence organizations, with a considerable lot of these organizations comprise of profoundly unique self-governing geography fragments. Progressed highlights of mobile impromptu organizations, incorporating information rates viable with media applications worldwide wandering ability, and cooperation with more organization structures are empowering new applications [5].

4 Simulation Results

Figure 1 shows the antenna as a 3D model with *x*, *y,* and *z* axes. The outer box is called the radiation box. 3D view of proposed rectangular antenna is shown [6]. This shape of the MANET antenna has been chosen for easy integration into MIC's or MMIC's. As the patch antennas are taken care of along centerline to balance, it limits excitation of other undesired modes.

The simulation of return loss between 1 to 10 GHz is shown in Fig. 2. The graph's *x*-axis represents the frequency in gigahertz, and the *y*-axis represents the return loss in decibel. The simulation plot shows three frequencies crossing the −10 dB point. The frequency points are 3.55, 5.55, and 9.1 GHz. This shows the MIMO nature of the antenna.

Figure 3 shows the f-OFDM spectrum of the proposed MANET antenna. The bandwidth of this spectrum is 100 MHz, which is the ideal bandwidth for 5G MANET antenna. The millimeter wave is to be propagated through this 100 MHz spectrum [7].

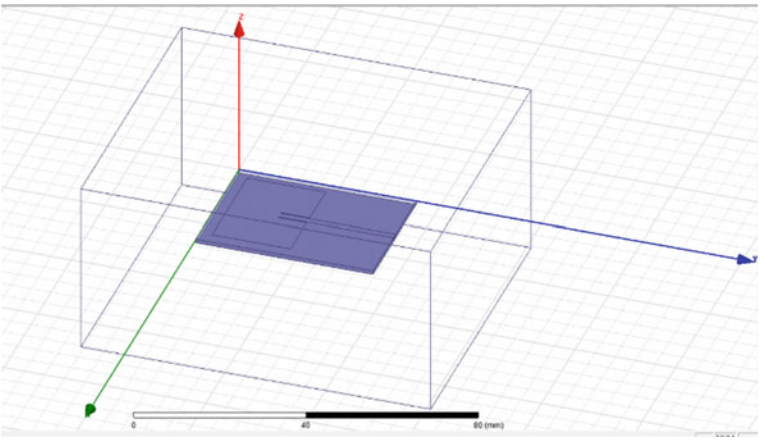


Fig. 1 Antenna as a 3D model

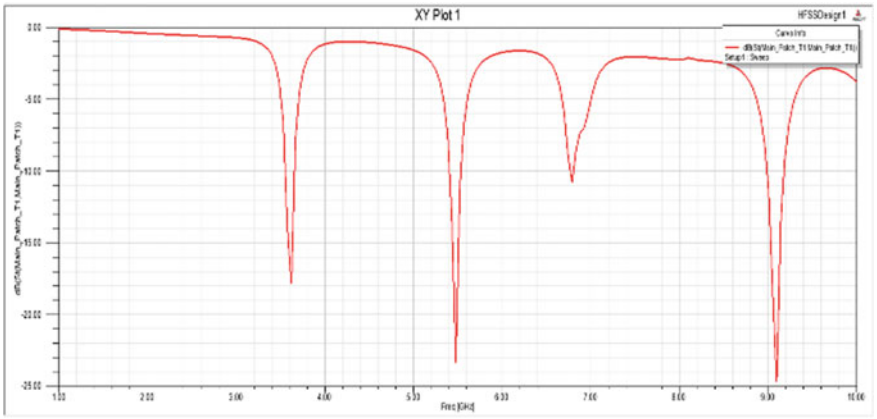


Fig. 2 Simulation of return loss

Shows the polar plot of the MANET gain as shown in Fig. 4. Versatile ad-hoc network gain was found to be 2.626 dB by using polar plot. This gain shows the low power low gain antenna can be used for 5G communication.

The pattern of directivity in 3D is shown in Fig. 5. The maximum value of directivity for this MANET antenna is found to be 14.8 dB. The central red region in this 3D plot shows high directivity followed by yellow region which has a moderate directivity. The blue region shows the less directivity region. The minimum value is −42.4 dB. With the higher directivity, the higher is the power gain in the antenna.

Figure 6 gives the input RF signal. It shows the 3.5 GHz RF signal. This radio frequency signal is proposed to be the input for the MANET antenna. 5G frequency bands are in the range of 3–3.6 GHz; 5–6 GHz, and 8–9.2 GHz. The first two

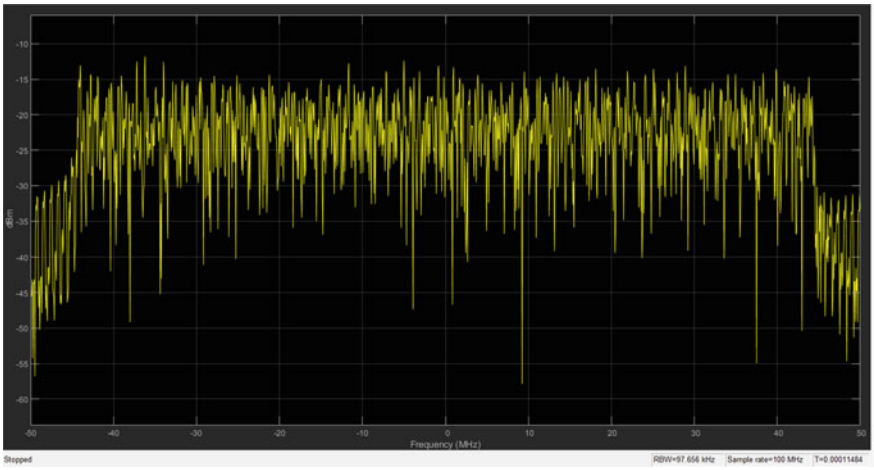


Fig. 3 f-OFDM spectrum of the proposed MANET antenna

Fig. 4 3D polar plot of MANET antenna gain

frequency bands are officially licensed by TRAI [8]. The third band, however, is reserved for experiment purposes. With the upcoming technologies, the 3rd band from 8 to 9.2 GHz can be utilized to achieve higher data rates with a little improvement in the proposed antenna design.

Figure 7 is the transmitted data from the MANET antenna. This data are discrete in nature as this is a digital data transmission. The analog signal is converted into digital signals for transmission purpose.

Figure 8 shows the filtered OFDM (f-OFDM) modulation technique. The transmitted signal is converted into analog signal for propagation into the channel. f-OFDM modulation is best suited for 5G communications in India as the frequency

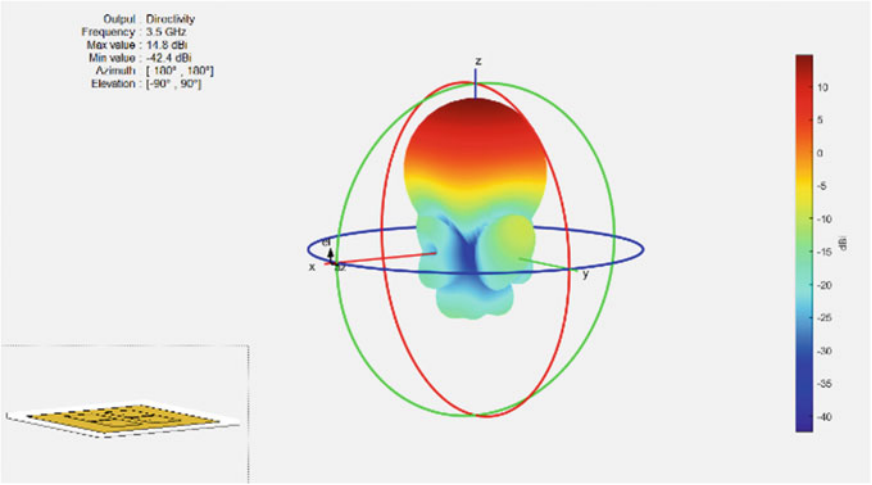


Fig. 5 Pattern of directivity in 3D

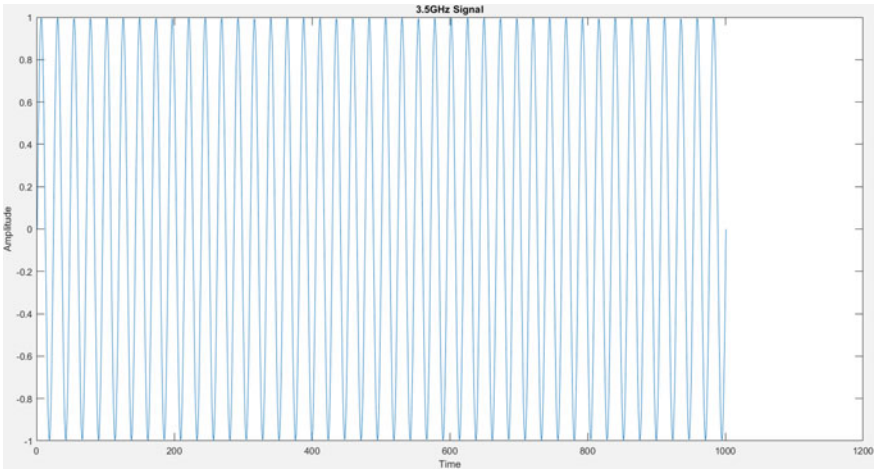


Fig. 6 Input RF signal

bands are under sub-6 GHz band. Conventional OFDM is most suited for 4G, 4G+, and LTE

The received data from the other antenna are shown in Fig. 9. After the signal exits from the channel, the MANET antenna receives the signal in digital form. This signal is then converted to analog signal and used for other purposes.

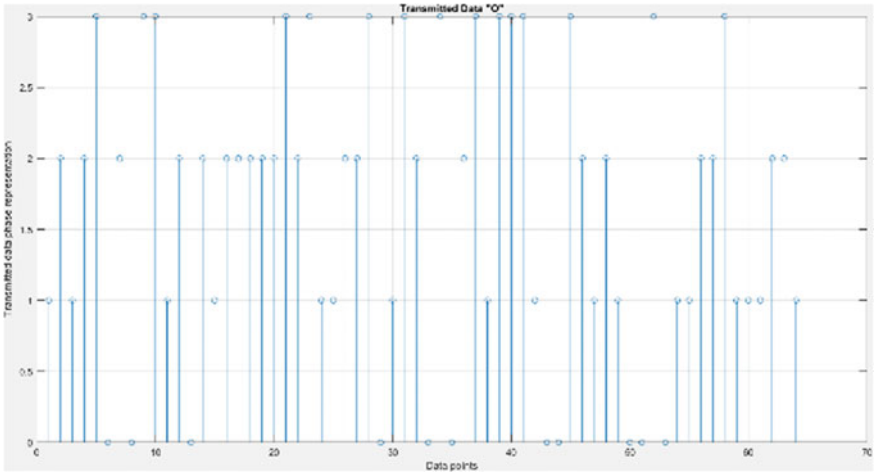


Fig. 7 Transmitted data from the MANET antenna

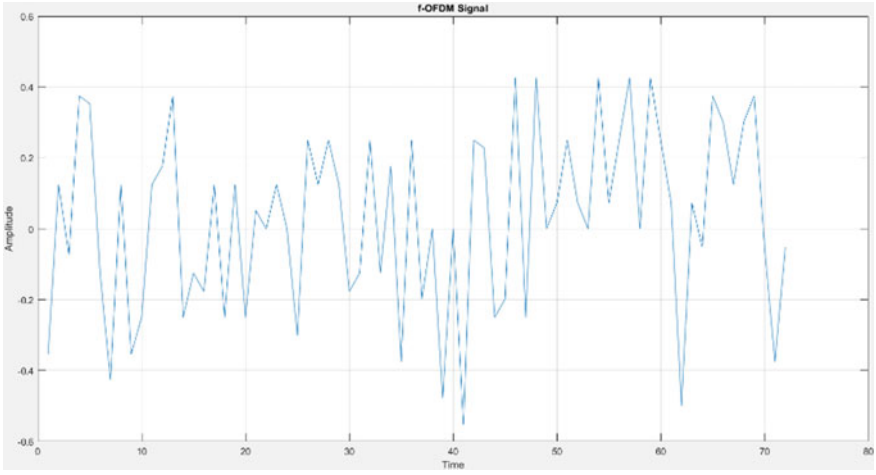


Fig. 8 Filtered OFDM (f-OFDM) modulation technique

5 Conclusion

The proposed strategy examines about the possibility of some notable directing shows for convenient of the cuff associations with millimeter-wave repeat gatherings and advises the most ideal approach to utilize mm-wave repeat to assemble the association efficiency. A couple of limits of the associations are changed to improve execution of this design compared to its previous version wireless fidelity (Wi-Fi). This design of versatile MANET in 5G possess a radio frequency (RF) structure

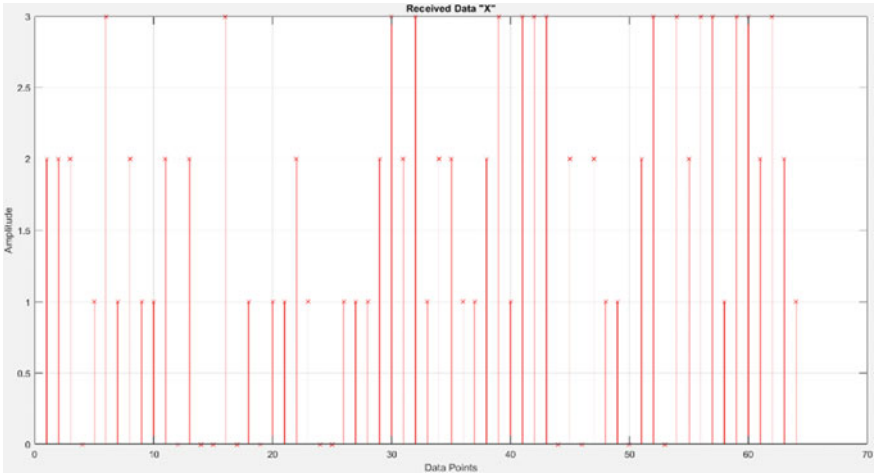


Fig. 9 The received data

zeroed in on extraordinarily enormous data rate, less idleness, reduced energy, and cost. To achieve this, controlling shows in the MANET ought to be versatile, energy-capable, and incredibly execution practical. The antenna is capable of operating under three different bands which enables the perspective of achieving higher data rates with the improvement in design shape.

References

1. Rappaport TS et al (2013) Millimetre wave mobile communications for 5G cellular: it will work! IEEE Access 1:335–349
2. Hong W, Baek K, Lee Y, Kim YG (2014) Design and analysis of a low profile 28 GHz beam steering antenna solution for future 5G cellular applications. In: IEEE international microwave symposium, 1–6 June 2014, Tampa Bay, Florida
3. Agarwal M, Roy A, Saxena N (2015) Next generation 5G wireless networks: a comprehensive survey. IEEE, pp 1553–877X (c)
4. Ojaroudiparchin N, Shen M, Pedersen GF (2015) A 28 GHz FR-4 compatible phased array antenna for 5G mobile phone applications. In: IEEE international symposium on antennas and propagation (ISAP2015), Tasmania
5. Rouy P, Vishwakarma RK, Jain A, Singh R (2016) Multiband millimeter wave antenna array for 5G communications. In: 2016 international conference on emerging trends in electrical electronics & sustainable energy systems (ICETEESES), pp 102–105
6. Ojaroudiparchi N, Shen M, Zhang S, Pedersen GF, A switchable 3D-coverage phased array antenna package for 5G mobile terminals. https://doi.org/10.1109/LAWP.2016.2532607. IEEE
7. Zhang W, Weng Z, Wang L (2018) Design of dual-band MIMO antenna for 5G smartphone application. In: IEEE 2018 international workshop on antenna technology (iWAT). https://doi.org/10.1109/IWAT.2018.8379211
8. Bharathy R, Vannilavan P, Rajasekar K, Vignesh R, Vasantha Kumar S (2018) Network coding as a performance booster for multi-hop transfer of data. Int J Sci Innov Eng Technol 3(3). ISBN No. 978-81-923607-3-7

Classification Rule Diagram to Diagnosis the Tuberculosis Based on Symptoms

T. Baskar and M. Kannan

Abstract Data mining is a vast area in research and applications where it is applicable for many real-time problems hence it proves to be beneficial for large data handling. Especially, when it comes to medical diagnosis, using data mining algorithms always proves to be exemplary where it helps to quickly diagnose diseases using available data based on classification algorithms. For example, diagnosing tuberculosis can be done using knowledge discovery patterns on uncovering various symptoms and associated patterns of behavior which is cost-efficient instead of diagnosing using various medical tests. In this paper, we are employing a classification diagram which is a key element of a much more elaborative process called knowledge discovery, to provide some indicators that are mainly responsible for the occurrence of special disease—tuberculosis and symptoms that are highly correlated with the positive examination test based on various rule-based conditions. Classification rules can be used to infer the defining characteristics based on information relevance of different test components and hence discover hidden knowledge, unexpected patterns, and new rules from the database by using the classification rule diagram to diagnose tuberculosis.

Keywords Data mining · Classification rule diagram · Knowledge discovery · Medical diagnosis

T. Baskar (✉)
Research Scholar, Department of Computer Science and Applications, SCSVMV University, Kanchipuram, India

M. Kannan
Department of Computer Science and Applications, SCSVMV University, Kanchipuram, India
e-mail: saikannan1999@kanchiuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_36

1 Introduction

Data mining has involved an excessive deal of consideration in the info about health care and in society as a whole in recent years, due to the extensive accessibility of vast amounts of data and the forthcoming essential for spinning such data into useful information and knowledge. The data you collect about your patients is one of the best assets. Hidden inside the data are all sorts of precious information that could make an important dissimilarity to the way you diagnose and interrelate with your patients [1]. According to the RNTCP annual report 2019, the estimated TB incidence in India is about 27 lakhs. Also, RNTCP was capable to accomplish an announcement of 21.5 lakhs in 2018. The uniqueness of the affected citizens essentially balances related to the bulk of the precious persons are all in the similar age of 15–69 years and two third of persons are males. TB diseases to stop by 2025, it possibly will be planning of Indian Government. This article is an extension of our previous work [2], the current research article applies classification methods to improve the diagnosis of TB diseases. The classification rule diagrammatic methodology allows easy to classify TB diseases based on symptoms.

2 Objectives

Knowledge discovery and data mining plays a vital role in healthcare industry for clinical diagnostics of medical reports like X-ray where it is based on a predefined pattern of trained model, This can be done with traditional statistical approaches and algorithms like classification algorithms and random forest algorithms but they are prone to time-consuming when processing on large set of data and costly in preparing rules and models for training the mining conditions. In order to simply the process, objectives are set and presented below.

Main objectives of the research articles are:

1. To classify the TB symptoms.
2. To easily diagnosis the TB disease based on the symptoms.
3. To find a piece of hidden information from clinical research using the classification rule diagram.
4. To properly find the clinical patients based on the classification rule diagram to forward for TB testing.
5. To diagnosis using PTB test results, all the patients will be conformed in order to continue with an anti-TB treatment.

3 Literature Review

Alfadhli et al. [3] identify TB diseases through the special method called SURF. The SVM classifier achieved an AUC score of 89% of 64 pixels. Finally results in a small size of detection are minor than at the superior scale. In X-ray, at an upper grid size, the SURF detected the high-level quality.

Imianvan and Obi [4] revised through FCM based on TB symptoms. The significant results of FCM methods, classification, and matching of cluster groups to set of TB symptoms to diagnosing the TB diseases achieved through this method.

Jaegar et al. [5]. Regular transmission for tuberculosis diseases detection in chest X-rays. TB screening is a challenging task and new approaches using computer-aided exposure. Survey of TB infection detection in chest X-rays using various data mining techniques to solve the research problem. Hooda et al. [6]. Conducted the survey of TB detection used in dissimilar stages of CAD system.

Lopes and Valiati [7] presented the investigation of pretrained CNN as extracted in tuberculosis exposure. Results, the three CNN architectures are collectively to form a particular international descriptor with the intention of used in SVM. Through deep learning, medical image classification is achieved.

Shukla and Agarwal [8] analyzed a new approach in favor of tuberculosis record scategorization using OCSB clustering methods. Hence with the aim of exact arrangement possibly will be performed. The results show improvement with the existing algorithm.

Skoura et al. [9] examined imaging of tuberculosis. Diseases diagnosis through CT, PET, FDG, and MRI. Pulmonary tuberculosis, Extra-PTB, and HIV into co-infected TB diseases were diagnosed less time and cost-effective results are achieved.

Vajda, et al. [10] authors analyze the automatic TB screening in chest radiographs for a better diagnosis method to TB diseases compared with other methods like culture and sputum.

Melendez et al. [11] diagnosed TB diseases through automatic recognition on chest X-ray in high-risk transmission. The total data was collected of 38,961 patient's chest X-ray from 2005 to 2010, out of 87 patients were diagnosed with TB. Result, specificity was 55% and negative predictive value was 99%, at a sensitivity of 95%.

Santosh and Antani [12] reveal that chest X-ray selection technique to identify a tuberculosis disease. Accuracy analyzed in dissimilar classifiers: 1 BN, 2 MLP, and 3 random forest algorithms. The data collected from U.S. National Library and National Institute of Tuberculosis in India. Results show that 91% of the ROC curve.

Rakhmetulayeva et al. [13] formative the success of cure about TB based on test drug with the help of classification algorithm using support vector machine. The initial data was 1261 lines of data, collected from the health department. Medicine test is significant used for arithmetical process to forecast an achievable enhancement for the patient health.

Asha et al. [14], three significant algorithms were analyzed such as CBA, CMAR, and CPAR. CPAR be improved better results compared to other algorithms.

Iqbal et al. [15], diagnostic for electronic health records (EHRs) to collect the data for tuberculosis along with insurance claim for the person. Final results, TB affected patient's codes, and insurance claims codes are unmatched. Hence, to improve the coding system is required.

4 Methodology

This research work combines the traditional classification algorithm with modern data mining techniques like parallel clustering to accelerate the speed in processing large set of data.

In this work, we have classified TB suspects based on different symptoms. Such as primary and secondary and it is depicted in Fig. 1.

The diagram Fig. 2 shows the classification rule diagram to suspects TB diseases based on pulmonary and extra-pulmonary symptoms. Through this classification rule diagram, the patients have more than two or three symptoms with dry cough, they can give an advice to take a TB testing.

Figure 3 shows the secondary symptoms of TB diseases. The clusters such as bad habits, contemporary diseases, BCG vaccine, nature of work, nature of spreading and attributes are involved in classifying the secondary TB symptoms.

Figure 4 diagram demonstrations the testing methods of pulmonary TB. This is the input analytical tool used for case finding in RNTCP. The organize testing process is effortless to execute at the peripheral laboratories and not as expensive.

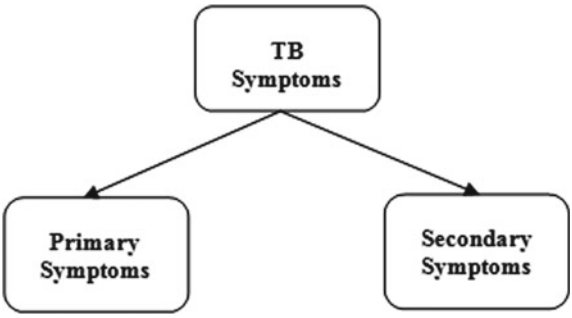


Fig. 1 Classification of TB symptoms

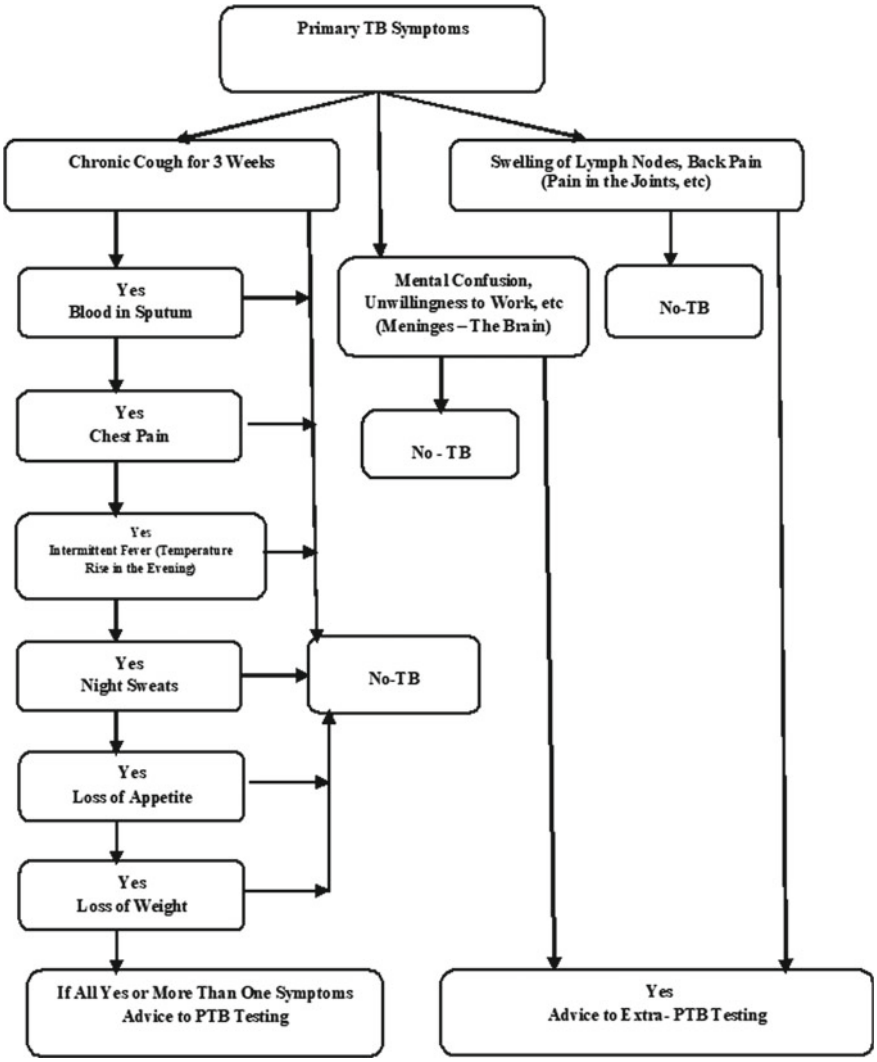


Fig. 2 Classification rule diagram of primary TB symptoms

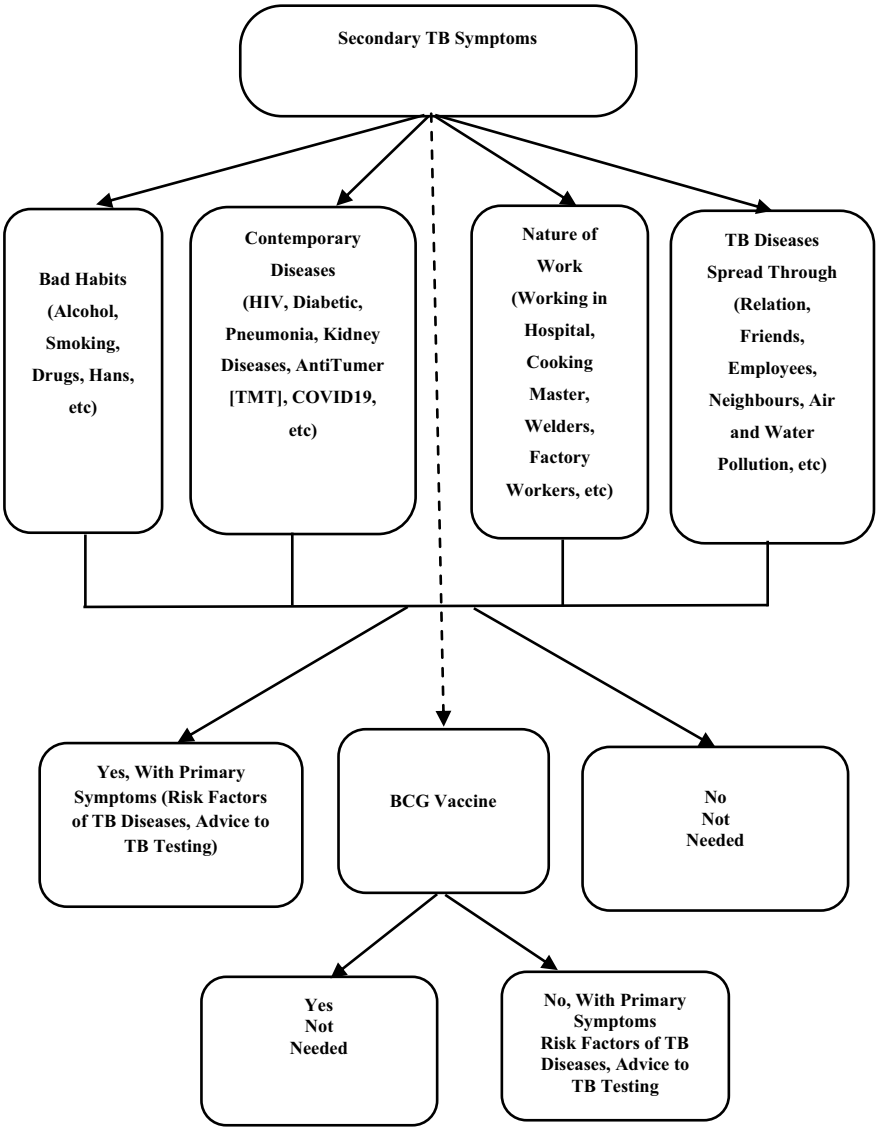


Fig. 3 Classification rule diagram for secondary TB symptoms

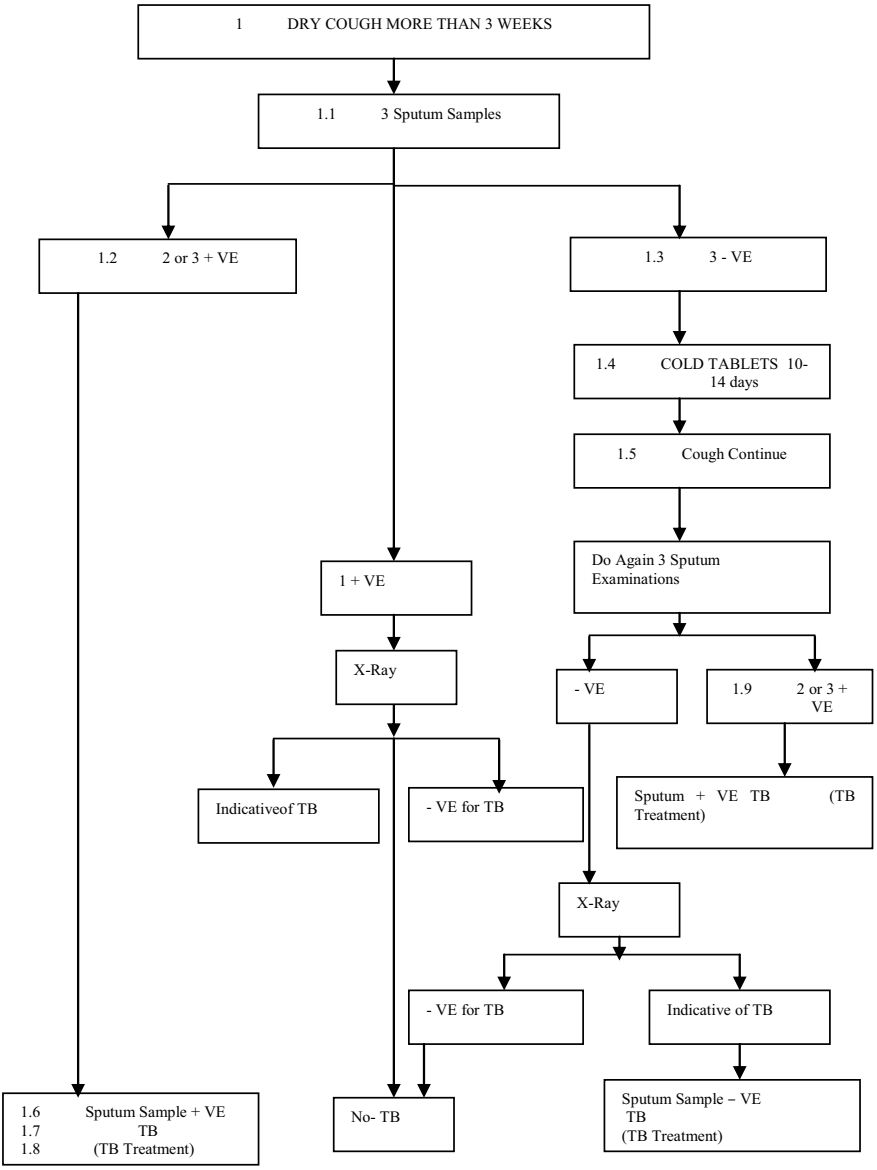


Fig. 4 Classification algorithm for pulmonary TB testing

4.1 Classification Rule Algorithm for PTB Testing

```
If Sputum >=2 Positives Then
Treatment =A
Else if Sputum == 3 Negatives Then
        If X-ray == Positives Then
                Treatment=B
Else
                Treatment=S
Else if Sputum==1 Positives Then
        If X-ray==Positives Then
                Treatment =A
        Else
                Treatment=S
```

Treatment

A = Sputum Sample +^{VE} TB (TB Treatment).
B = Sputum Sample - ^{VE} TB (TB Treatment).
S = Safe(No - TB)

5 Results

The above classification rule diagram, Figs. 1, 2, and 3 shows the symptoms based on the classification of tuberculosis diseases. The classification rule diagram, concentrating on patient symptoms. This is categorized as primary and secondary symptoms. In the primary, diagnosis the pulmonary and extra-pulmonary TB suspects based on symptoms. Next one secondary, the main risk factors for people suspected of developing TB are poor habits, nature of work, existing diseases, such as HIV, BCG vaccine, and spread by others. The benefits of this outcome are that, it will direct patients to make the right decision to think whether TB testing is needed or not. Also, it helps the concerned department employees to find out who needs TB testing. Based on the classification diagram, we can easily filter patient's dataset and give the direction for the patients to take anti-TB treatments.

In this work, researchers have proposed classification rule diagram to diagnose tuberculosis symptoms, through this work, strong association attributes can be filtered, extracted and interpreted.

6 Conclusion

Throughout this paper, researchers have focused on how data mining concepts help to classify TB symptoms based on the classification diagram. The Crux of this result is that anybody can easily identify TB suspects, interpret, and classify the TB symptoms and give the direction for the patients who need the immediate attention and treatment. The limitation of this work is that it is a diagrammatic representation to classify the TB symptoms, scope for further work is that will convert this work into an algorithm with proper dataset.

References

1. India Tuberculosis analysis Report 2019—RNTCP Annual Report, Central TB Research Division, Ministry of Family welfare and health, New Delhi, 2019, pp 1–222
2. Baskar T, Kannan M (2021) Survey on data mining and inferences in tuberculosis medical data. (IJCRT) Int J Creative Res Thoughts 9(1):3064–3070. ISSN: 2320-2882
3. Alfadhli FHO, Mand AA, Sayeed MS, Sim KS, Al-Shabi M (2017) Classification of TB with SURF spatial and pyramid features. In: 2017 International conference on robotics, automation and sciences (ICORAS). IEEE, New York
4. Imianvan AA, Obi JC (2011) Fuzzy cluster means for expert system in the diagnosis of TB. Global J Comput Sci Technol
5. Jaeger S, Karargyris A, Candemir S, Siegelman J, Folio L, Antani S, Thoma G (2013) Automatic screening for the tuberculosis disease in chest radiographs: a research survey. Quant Imaging Med Surgery 3(2):89
6. Hooda R, Mittal A, Sofat S (2019) A survey of CAD methods for tuberculosis detection in chest radiographs. Soft computing: theories and applications. Springer, Singapore, pp 273–282
7. Lopes UK, Valiati JF (2017) Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput Biol Med 89:135–143
8. Shukla M, Agarwal S (2014) Hybrid approach for tuberculosis data classification using optimal centroid selection-based clustering. In: 2014 Students conference on engineering and systems. IEEE, New York
9. Skoura E, Zumla A, Bomanji J (2015) Imaging in tuberculosis. Int J Infect Dis 32:87–93
10. Vajda S, Jaeger S, Karargyris A, Santosh KC, Candemir S, Xue Z, Antani S, Thoma G (2018) Feature selection for the automatic TB screening in frontal chest radiographs. J Med Syst 42(8):1–11
11. Melendez J, Sánchez CI, Hogeweg L, Philipsen RHRW, Hayward HM, Aldridge AC, Abubakar I, van Ginneken B, Story A (2018) Accuracy of an automated system for the diagnosis of TB detection on chest radiographs in high-risk screening. Int J TB Lung Dis 22(5):567–571
12. Santosh KC, Antani S (2017) Automated chest X-ray screening: can the lung region has symmetry to help and detect pulmonary abnormalities? IEEE Trans Med Imaging 37(5):1168–1177
13. Rakhmetulayevaa SB, Duisebekovaab KS, Mamyrbekova AM, Kozhamzharovaa DK, Astaubayevac GN, Stamkulovac K (2018) Application of a classification algorithm which based on a SVM classifier in determining the effectiveness of treatment of TB. Proc Comput Sci 130:231–238
14. Asha T, Murthy KNB, Natarajan S (2012) Data mining techniques for the diagnosis of TB. Understand Tuberculosis—Global Exp Innovative Approach Diagnosis 16:333–353
15. Iqbal SA, Isenhour CJ, Mazurek G, Truman BI (2020) Diagnostic code and agreement for an electronic health records and the claims data for TB. Int J Tuberc Lung Dis 24(7):706–711

Ensuring the Security and Load Balancing in the Cloud Environment by BPA-SAB Method

Abin T. Abraham and E. J. Thomson Fredrik

Abstract Cloud storage is the latest standard, with the benefits of pay-per-use and on-demand applications in the future. The load balance in a distributed setting for the cloud is difficult due to factors such as stability, portability, security, and interoperability. To achieve this the proposed research aims for intrusion detection with a higher data protection quality and load balancing performance by developing a novel hybrid-based Binomial Protection-based Authentication with a Stochastic Agent-based Load Balancing (BPA-SAB) technique. The proposed BPA-SAB technique only permits users with proper licenses. The intruders have been discovered through the user behavior. The load balancing mechanism is handled by the SAB technique. The performance metrics of intrusion detection, computational time are better when compared with the existing work Secure and Energy-Aware Load Balancing.

Keywords Cloud storage · Load balancing · Security · Binomial

1 Introduction

Cloud technology has been an evolving mechanism for quite decades [1]. Cloud storage offers cost savings by boosting volume or competitiveness with a small number of users, and ensuring quick access to knowledge at minimum upfront costs [2]. Simplifying procedures results in projects involving fewer representatives being completed and the cost of resources is being reduced [3]. Cloud computing often reduces the level of emerging software licenses to a minimum, such as stretching and developing [4].

A. T. Abraham
Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

E. J. Thomson Fredrik (✉)
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_37

2 Related Works

In Manasrah and Ali [5], suggested a new technique named Genetic algorithm (GA) with Particle Swarm Optimization (PSO) to reduce expense and time to fulfillment. In Jena and Mohanty [6], use similar technology such as computing, encoding, replication, convergence, and replication methods to improve the schedule of the tasks and executing mostly with aim of reducing time for executing the task. In Phan and Park [7], suggested a strategy for dealing with DDoS threats in a Software Defined Network (SDN). In Nguyen et al. [8], used AES, DES, and some other security mechanisms for encrypting the file header information in SFC messages. In Kumar Mishra et al. [9], analyzed the performance of heuristic-based algorithms for load balancing and carried out the simulation using CloudSim simulator. Comparing with existing works of load balancing algorithms, our proposed BPA-SAB technique performs load balancing efficiently in Cloud Computing.

3 Proposed Model

In the proposed BPA-SAB technique, the intruder attack detection process is carried out with the help of BPA. In the proposed BPA-SAB technique, the BPA is introduced for enhancing data privacy preservation by performing intruder attack detection in an effective manner. In BPA, the Regression Coefficient (RC) is utilized for detecting anomalies on the cloud environment. The RC is the method based on a prediction that determines the occurrence probabilities by the action based on the features it has been observed. Since the users who are unauthorized will not permit for data accessing in cloud storage. Thereby, the confidentiality level of data is improved.

Algorithm 1: BPA for Privacy Preservation

Input: User Requests$UR_i = UR_1, UR_2, \ldots, UR_n$
Outcome: Improvised rate in detecting intrusion with a higher level of privacy
Stage-i: Start
Stage-ii: For: In cloud considering requests from every user
Stage-iii: According to Eq. (2) and Eq. (3) the RC needs to be computed
Stage-iv: If condition ($Y == 1$) then
Stage-v: The particular user was considered an authorized one
Stage-vi: Else
Stage-vii: The particular user was considered an unauthorized one probable to create a threat
Stage-viii: End-If
Stage-ix: Permitting users who are authorized to receiving a cloud data access
Stage-x: End-For
Stage-xi: Stop

As shown in Algorithm 1, the BPA is deployed in the proposed BPA-SAB technique to identify the intrusion attack and to enhance the privacy preservation of data in the cloud environment.

3.1 Stochastic Agent Balancing (SAB) for Effective Load Balancing

In the proposed BPA-SAB technique, the SAB model is introduced for load balancing among a number of servers dynamically in an effective manner. In the proposed BPA-SAB technique, the time taken for cloud data accessing is also significantly reduced through achieving effective load balancing through the SAB Model.

Algorithm 2: SAB for Balancing the Loads Dynamically

Process Input: Total UserRequest$UR_i = UR_1, UR_2, \ldots UR_n$that are Authorized
Number of Cloud Server $CS_i = CS_1, CS_2, \ldots CS_n$
Process Outcome: Enhancing the efficiency for balancing the loads
Stage-i: Start
Stage-ii: For: Considering requests from every authorized user
Stage-ii: Calculate server's weight in the cloud depending on the number of tasks completed by Eq. (6)
Stage-iv: SLS was accomplished by the Centralized Agent
Stage-v: Based on the objectives feature, choose the best server in the cloud
Stage-vi: Allocate the specified tasks for the servers in the cloud which can deliver the required resources
Stage-vii: End-For
Stage-viii: Stop
The SAB-based searching is performed for providing higher load balancing efficiency. The main aim of SAB is to find out the optimal cloud server for providing needed services to the users. Through the Stochastic Agent, the weight of cloud servers is computed for each authorized user request according to the executed tasks by using Eq. (6).

4 Performance Comparison of Models

An experimental evaluation for the proposed BPA-SAB hybrid technique is implemented in Java Environment utilizing Cloudsim simulator. For experiment setup, the number of cloud users and the number of cloud user data is considered to range from 100 to 500 and 200 to 1000 within 20 iterations.

Table 1 and Fig. 1 illustrate that the experimental results of LBE concerning the number of cloud users. For simulation purposes, the number of cloud users is considered the range from 100 to 500 which is taken as input IDR is measured in

Table 1 Numerical comparison of LBE

Cloud users	SEA-LB	BPA-SAB
100	71	88
200	74	90
300	77	92
400	81	94
500	86	96

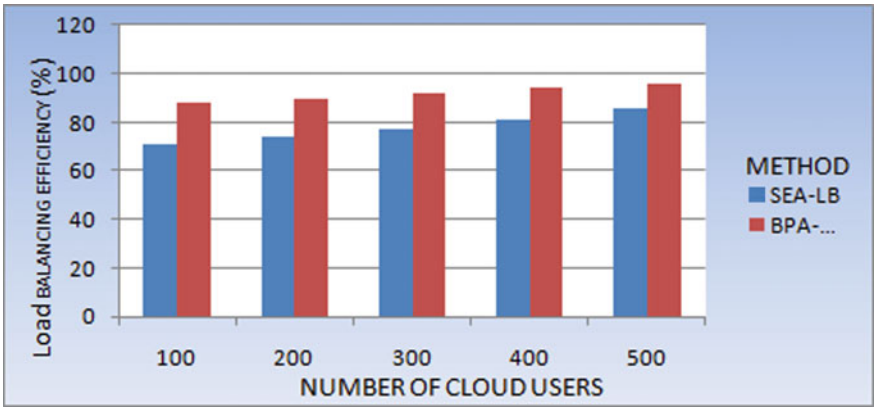


Fig. 1 Graphical comparison of LBE

terms of percentage (%). When the performance of IDR is high then the technique provides improved results for secured data communication in the cloud.

Table 2 and Fig. 2 illustrate that the experimental results of IDR concerning the number of cloud users. For simulation purposes, the number of cloud users is considered the range from 100 to 500 which is taken as input. The performance analysis of IDR of the proposed BPA-SAB technique is compared with the existing method SEA-LB. The IDR is gradually improved in both methods based on the different number of cloud users. As a result, the BPA-SAB technique significantly improves the LBE when compared to SEA-LB. The DAT is determined as the total taken time in the cloud for data accessing.

Table 2 Numerical comparison of IDR

Cloud users	SEA-LB	BPA-SAB
100	63	85
200	69	88
300	75	91
400	80	93
500	86	96

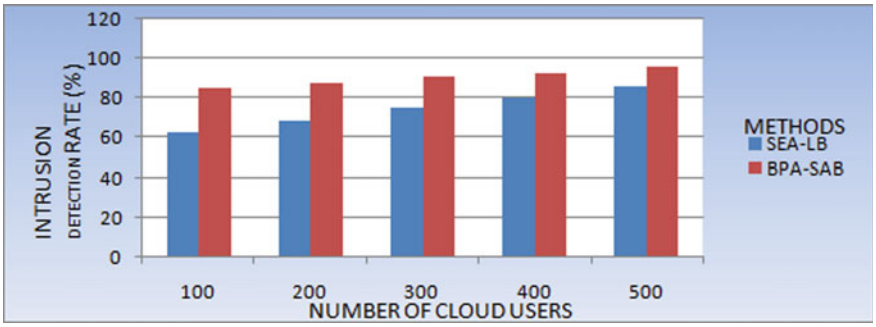


Fig. 2 Graphical comparison of IDR

Table 3 Numerical comparison of DAT

User data	SEA-LB	BPA-SAB
200	25	20
400	50	35
600	75	50
800	100	65
1000	125	80

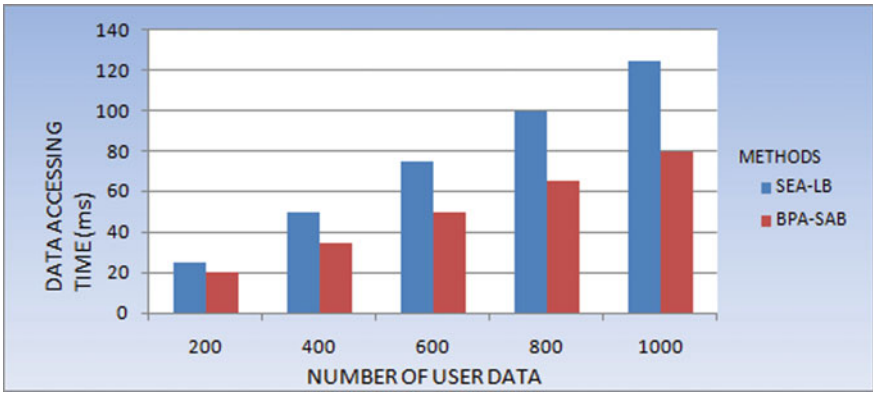


Fig. 3 Graphical comparison of DAT

Table 3 and Fig. 3 illustrate the experimental results of DAT concerning the number of user data. For simulation purposes, the number of user data is considered as the range from 200 to 1000 which is taken as input. The performance analysis of DAT of the proposed BPA-SAB technique is compared with the existing method SEA-LB. The DAT is gradually reduced in both methods based on the different numbers of user data. The BPA-SAB technique significantly reduces the data accessing time when compared to SEA-LB. It is proved from the above results that the performance

metrics of Intrusion Detection, Computational Time are better when compared with the existing work Secure and Energy-Aware Load Balancing (SEA-LB).

5 Conclusion

In the cloud environment, the BPA-SAB technique is introduced for achieving secured data communication by attaining better improvement on privacy level to the data and load balancing efficiency. At first, the user transmits a request to the cloud server. The BPA-SAB technique determines the RE coefficient for each user with the help of BLog it Authentication. The BPA-SAB technique effectively balances the load among the number of servers in the cloud through the deployment of SLS during the data accessing in the cloud. The BPA-SAB technique enhances the LBE also minimizes DAT in a significant manner.

References

1. Ahmed B, Malik AW, Hafeez T (2019) Services and simulation frameworks for vehicular cloud computing: a contemporary Survey. EURASIP J Wireless Commun Network 4
2. Zeng CH, Chen KC (2018) Social network analysis facilitates cognition in large wireless networks. IEEE Trans Cogn Commun Network 4(3):618–634
3. Zhang SM, Sangaiah AK (2018) Reliable design for virtual network requests with location constraints in edge-of-things computing. EURASIP J Wireless Commun Network 2018(1)
4. Wu KC, Liu WY, Wu SY (2018) Dynamic deployment and cost-sensitive provisioning for elastic mobile cloud services. IEEE Trans Mob Comput 17(6)
5. Manasrah M, Ali HB (2018) Workflow scheduling using hybrid GA-PSO algorithm in cloud computing. J Wireless Commun Mob Comput 3:2018
6. Jena T, Mohanty JR (2018) GA-based customer-conscious resource allocation. Arab J Sci Eng 43(8)
7. Phan TV, Park M (2019) Efficient distributed denial-of-service attack defense in SDN-based cloud. IEEE Access J 7
8. Nguyen VC, Vu AV, Sun K, Kim Y (2017) An experimental study of security for service function chaining. In: Proceedings of 9th International conference on ubiquitous future network, July 2017
9. Kumar Mishra S, Sahoo B, Parida PP (2020) Load balancing in cloud computing: a big picture. J King Saud Univ—Comput Inf Sci 32(2):149–158

Further Analysis on Internet of Things (IOT) Applications in Emerging Markets and Vietnam

Nguyen Thi Hang, Dinh Tran Ngoc Huy, Trung-Hieu Le, Sylwia Gwoździewicz, Nguyen Thi Phuong Thanh, Nguyen Thuy Dung, and Dinh Tran Ngoc Hien

Abstract In many industries such as manufacturing, construction and hospitals in recent years, there are many uses and applications of Internet of things (IoT) in many developing countries. Authors mainly use qualitative analysis, including analytical, synthesis and explanatory methods regarding to Internet of things issues and applications and design. Hence, the purpose of this study is to explore issues of Internet of things (IoTs) applications and give out suggestions for IoTs design principles in emerging markets and Vietnam. For example, authors presented current uses such as: applications for smart hospitals and medical equipment with IoTs, as well as suggested further principles to set up IoT design.

Keywords Smart applications · IoT uses · Hospital solutions · Vietnam

N. T. Hang (✉) · N. T. P. Thanh · N. T. Dung
Thai Nguyen University, University of Information and Communications Technology, Thai Nguyen, Vietnam
e-mail: nthang@ictu.edu.vn

N. T. P. Thanh
e-mail: ntpthanh@ictu.edu.vn

N. T. Dung
e-mail: ntdung@ictu.edu.vn

D. T. N. Huy
Banking University HCMC, Ho Chi Minh City Vietnam—International University of Japan, Niigata, Japan

T.-H. Le
Dai Nam University, Hanoi, Vietnam
e-mail: hieult@dainam.edu.vn

S. Gwoździewicz
Jacob of Paradies University, Gorzow Wielkopolski, Poland

D. T. N. Hien
Ho Chi Minh University of Technology, Ho Chi Minh, Vietnam

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_38

1 Introduction

IoTs, Internet of things which is technology helping to cut costs and with better qualified resources and improving planning, has various uses in our work and life, at home and in office or warehouse or manufacturing place, enhanced with protocols such as Bluetooth, Wi-Fi and others, even in healthcare services or hospitals, healthcare workers can change the way they detect the patient's disease and innovate treatment as well.

In our smart houses, IoTs system with sensors can help us to easily recognize strangers to enter our houses, hot water system can be automatically controlled, or smart cameras system, etc.

Internet of things (IoTs) can be considered as a interrelated and wireless system, with digital devices connected which can enable to receive, store and send information via a network without requiring interaction of human-to-human or human-to-computer. He and Li [1] conducted research in internet of things in industries and also by [2–5]. Faquir et al. [6] said there is cybersecurity solutions attached also confirmed by Jaccar and Nepal [7]. Nga and Huy [8] said internet supported banking sector. From wikipedia, internet has been useful in industry (source: [9])

This study will presents some principles of IoTs design system and organized with introduction, research questions, literature review, main results, discussion and conclusion.

Among the paper's contribution is the solutions presented for preventing security risk, as well as stated IoTs has wide range of applications in our industries and life.

2 Literature Review

We summarize previous studies in Table 1.

Besides, He and Li [1] state that it is an emerging technology, the Internet of things (IoT) is expected to offer promising solutions to transform the operation and role of many existing industrial systems such as transportation systems and manufacturing systems.

And Gnoni et al [10] found out a we can design a system with IOT applications in order to manage risks or hazards such as safety of plants as well as workers.

Last but not least, Malik et al. [11] specified that IOTs can help to link any device to Internet, and help decreasing cost and develop more advanced system—Internet in many industries including such as health care, construction, automotive industries, embedded devices, environment monitoring, agriculture, etc.

Then, Thanh et al. [12] stated that not only IoTs but AI applications such as infrared applications are also very useful in industries.

Compared to previous researches, our study will show also security risk prevention solutions in conclusion.

Table 1 Summary of relating studies

Authors	Year	Content, result
Mary and Malaisamy	2019	IoTs allow us to share data and information without much human intervention. Also, less manpower involving in using sensors by automation with various features
Uslu et al.	2020	IoTs allows us to develop so-called smart hospitals by setting tech infrastructure and environment with combination of cloud tech, sensors, protocols, databases, etc. We need to consider components such as challenges, optimized factors, technologies and opportunities
Khanna and Kaur	2020	IoTs has been considered in a few years ago and it is considered under many aspects including challengers, applications and tech
Ghost et al.	2020	Concentrates on applications and uses of IoTs for construction and building (smart objects)
Kelly et al.	2020	IoTs can help a lot for health care services and improve public health, in which it enables to improve coordinated system and accessibility of preventative public health services
Ahmad et al.	2021	As IoT is in fast-growing stage and demand of smart devices also increasing so the manufactures oversight the security aspects and delivering the vulnerable devices in the market attackers easily targeting the devices using these vulnerabilities and performing a large number of DDoS and other types of Attacks to steal user personal information and data from IoT devices
Trung-Hieu Le, Nguyen Thuy Dung, Dinh Tran Ngoc Huy, Nguyen ThiPhuong THanh, Dinh Tran Ngoc Hien, Nguyen Thi Hang	2021	IoTs have wide applications in industrial uses, esp. In Vietnam and emerging markets

3 Methodology

This paper mainly use qualitative analysis with technical issues relating to IoTs and security and network solutions.

We also propose to build a security system design principles in order to contribute to reduce IoTs and their consequences.

4 Main Results

4.1 Further Suggestion on Principles of IOT Application System Design

First, we propose IoTs application system based on the following principles:

- Monitoring from remote locations: Applying IoTs, combined with cloud technology and AI can help to improve better management information system and help managers of manufacturing firms, factories and hospitals can manage data and process patients data from remote areas.
- Network security: we need to ensure data privacy protection and protect clients data, for example, bank clients and credit card data seriously to avoid hackers to steal information and big money.
- Smart applications: Applying IoTs with sensors and AI can help much for smart house, smart building and construction and smart hospitals, etc.

Next, we recognize groups of IoTs applications needed.

4.2 Hand Gestures Recognition and Other IOTs Solution System Design

Hand Gesture Recognition

Figure 1 shows us that data collection process as follows.

Describe data collection step: researchers has uses a data sample and observe gesture of hands involving twelve prior define gestures with thirteen labels. So researchers also recognize difference between mean and variations of gestures.

Moreover, time length of each gesture is changing and different depending on person.

We describe the set of data, with the support of ELAN Multimedia Annotator Tool.

There are hand gestures, total 12 pre-defined gestures, and annotate 13 class labels which is total the gestures are not the same as the twelve gestures.

There are set of data consisting of total 13 class labels, and 21,420 instance. Our figure describes noted data that accelerate and gyroscope in the sense of magnitude accordingly.

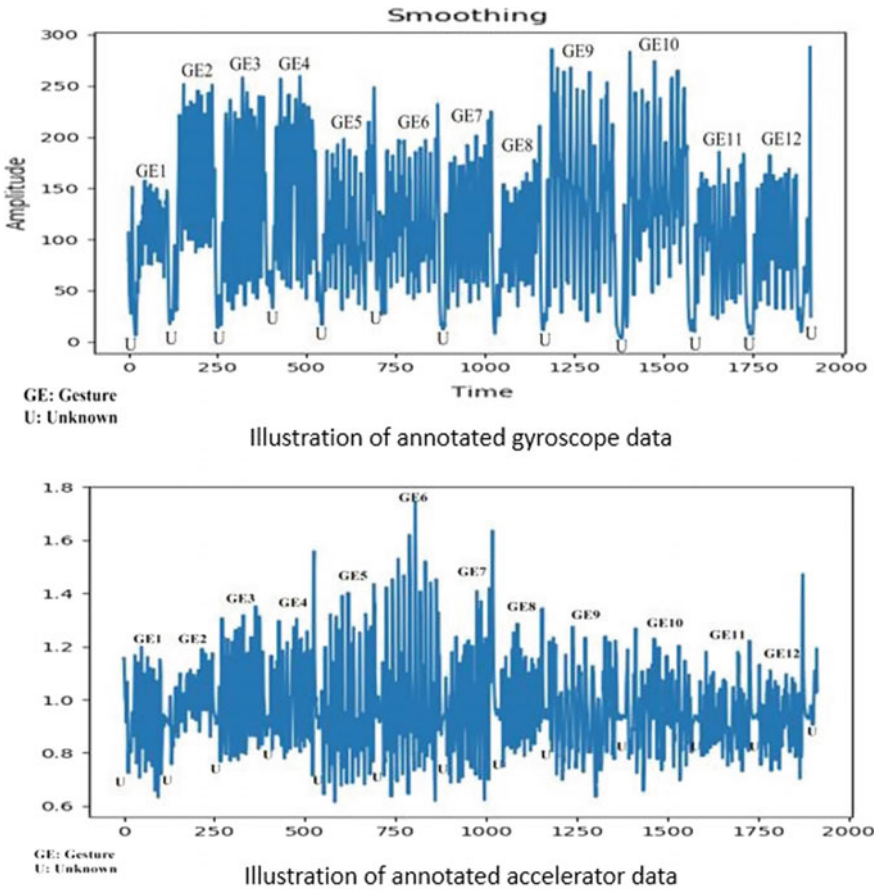


Fig. 1 Collecting data and results

Other Applications of Internet of Things (IOT)

Smart House Application System

Nowadays, with IoTs and sensors, our smart home can be controlled with automatic light system and automatic recognition when strangers enter our house.

Multi-purpose intelligent sensors: gas sensor, smoke sensor, water leak sensor, motion sensor, light sensor, door sensor help control every element of the house from light, temperature, humidity, … From there, adjust and ensure the health and safety of family members. IoT technology will connect all devices such as curtains, lighting, air conditioners, etc. through the central controller. Thanks to that, just a smartphone/tablet with a network connection (3G, 4G/Wi-Fi), you can quickly turn on and off the lights, air conditioning or open the curtains with just one touch. This is the benefit of smart home technology that not all solutions have.

When we mention "house" we consider Internet of things (IoTs) application for the house as a physical structure in which we can take advantage of IoTs applications. Roles of internet with AI and technology showed many applications [13, 14].

Hospital Solution System

This is the solution in which we can apply IoT into monitoring health care of people and patients, or we even can combine IoT with cloud technology and Big Data to process data of patient effectively.

In medical or healthcare sector, there is Breakthrough in Smart Medicine with IoTs and technology that has considerably changed the healthcare services and innovating the way devices and applications in which users interact or connect with.

Nowadays we can recognize the values and uses of Blockchain tech and cloud computing which function as a future development that can bring unprecedented breakthroughs in medical services.

It will benefits both from patient side: he or she could control and proactively monitor his or her health condition, and form doctors side: they can better monitor patient health data with technology help.

Recently, a number of studies have demonstrated that it is possible to remotely monitor the health of patients using advanced biotechnological technologies combined with an embedded IoT device. Non-invasive biosensors promise real-time patient treatment, ensuring timeliness in care, enhancing adherence to treatment, and improving health outcomes.

Researchers also can develop medical equipment to support to recognize or check patients blood clots, which can support patients for their staying in a range of treatment and also help to reduce the stroke risk or bleeding.

Kelly et al. mentioned hospitals at time being can be supported with IoTs benefits that enhance health care delivery, as well as to support predicting, health matters, diagnosing and treating and monitoring patients inside and outside the hospital.

Other industry applications with IoTs

Production and factories

IoTs, together with cloud technology and sensors and AI has played major roles in contributing to connect data for production optimization and better inventory management.

Specifically, in the factory, the digitization of production/intelligent production helps the equipment—equipment system communicate easily and continuously without the supervision and permission of children people in the process. The device-device communication enables sensor systems, actuators and control systems to instantly interact with each other. The devices will be smarter, the data from the device will be brought to the center more. The application of AI technology, Bigdata allows the center to use that large amount of data to make reasonable maintenance decisions to help the system operate in a more durable and reasonable way, avoiding the loss caused by the system shutdown production system to repair.

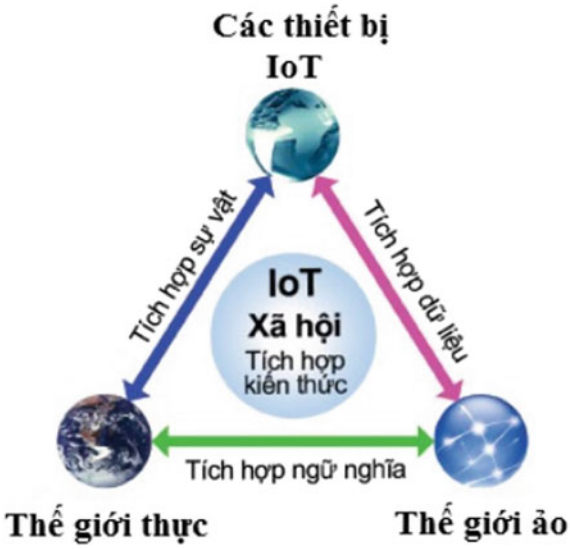


Fig. 2 IoT and real and unreal world linkages. *Source* wikipedia.com

5 Discussion

As we can see in Fig. 2.

We can realize there are some popular applications of IoTs such as: nowadays, there are range of applications and uses of IoTs in hospitals.

For instance, with real-time status tracking using a smart medical device connected to a smartphone application, medical data and other health requests can be collected and data connection available of the smartphone to transmit the collected information to the doctor.

Data of health are collected and transmitted with The IoT device including but not limit to: blood pressure, oxygen and blood sugar, weight and electrocardiogram (ECG).

Then doctors, patient's insurers, can connect to data and also, participating medical companies or consultants, to allow them to view the data. Data is collected regardless of time, location or device.

6 Conclusion

In summary, we analyzed that Internet of things (IoTs) have many applications and uses in monitoring data from remote locations, in smart hospitals, in banking, etc. In health care sector, doctors and nurses can take advantage of internet of things and internet data in order to store big data of clients, whereas in banking sector analysts can analyze risk data from internet technology. Also in any industries, people start

to apply and use robot or robotic, high-end robot and new materials in the industrial revolution. Ghosh et al. [15] said there are future trends of IoTs in construction industry. While Hang et al. [16] stated it is good for labor force training in industry 4.0.

However, we also have to consider solutions to prevent risks from IoTs:

For instance, **Virus, warms and Spywares**: Attackers are trying to send some malicious files as an email attachment when the recipient receives the email and download the attachment or download other files from Internet so it will affect the system. Then, **Phishing attack**: This type of attack usually uses to strip the user's important information such as credit card details, email passwords, etc. in this type of attack the emails or website is used.

Therefore, we need to make sure our security and administration applications are always up to date and this is a necessary action to always do first because older software often has a lot of vulnerabilities that hackers are so familiar with and will definitely use them to attack thoroughly and that is a concern that any administrator should be concerned about.

Cyber security risk prevention:

With the development of Industry 4.0, cyber attacks on innovative ICT technologies can be a fundamental problem, which is why networks and information systems of the combined economy, society or critical infrastructure in the era of digitization require ever stronger and efficient security.

Determining the basic safety recommendations for Industry 4.0 components, services and processes based on risk analysis is the first step to resolve difficult technical constraints in this field. Therefore, several basic recommendations should be made regarding the improvement of cyber security risk management standards in the development of new ICT technologies. These recommendations are only the basis for undertaking further analyzes of the subject taken.

Implementing solutions related to the security interoperability of Industry 4.0

ENISA (The European Union Agency for Cyber security) in the "Industry 4.0 Cyber security …" 2019 report, justifies that with the introduction and integration of Industry 4.0 devices with existing systems, there is a question of interoperability. In industrial environments, ensuring interconnection between different devices is often difficult, especially for devices that are no longer supported. That is why it is important to promote secure solutions ensuring smooth integration of Industry 4.0 devices with older systems and among themselves. When devices and platforms from different suppliers are used, interoperability may not always be possible. Ensuring interoperability between devices/platforms is not only smooth operation but also security. Therefore, it is necessary to ensure and support the interoperability of devices, Industry 4.0 platforms, as well as security practices. The concept of interoperability refers not only to communication protocols and various application frameworks. The concept of security interoperability appears in the complex supply chains of Industry 4.0, which means that it is very difficult to ensure a common basic level of security across platforms, devices, protocols and frameworks. The

weakest link in the chain can have harmful effects on the entire chain, which is why providing a single common layer of cybersecurity in all these elements is a very difficult problem.

Limitation of Research

Authors need to make deeper analysis on IoTs applications in areas such as Big Data, cloud tech, AI combined, etc.

Also, we can expand to analyze further researches for other emerging markets.

Acknowledgements Thank you editors, friends and brothers to assist this publication.

References

1. He W, Li S (2014) Internet of things in industries: a survey. IEEE Trans Ind Inform 10(4):2233–2243. https://doi.org/10.1109/TII.2014.2300753
2. Huy DT, Dat PM, Anh PT (2020) Building and econometric model of selected factors' impact on stock price: a case study. J Secur Sustain 9:77–93. https://doi.org/10.9770/jssi.2020.9.M
3. Huy DTN, Nhan VK, Bich NTN, Hong NTP, Chung NT, Huy PQ (2021) Impacts of internal and external macroeconomic factors on firm stock price in an expansion econometric model—a case in Vietnam real estate industry. In: Data science for financial econometrics-studies in computational intelligence, vol 898. Springer. https://doi.org/10.1007/978-3-030-48853-6_14
4. Le T-H, Dung NT, Huy DTN, Thanh NTP, Hien DTN, Hang NT (2021) Internet of Things (IOT) uses and applications—solutions in emerging markets and Vietnam. Turkish J Comput Math Educ 12(11)
5. Nam VQ, Tinh DT, Le TH, Huong LTT, Huy DTN (2021) Internet Of Things (IoT), Artificial Intelligence (AI) Applications for Various sectors Inemerging Markets—and risk management information system (RMIS) Issues. Des Eng (Toronto) 6:609–618
6. Faquir D, Chouliaras N, Sofia V, Olga K, Maglaras L (2021) Cybersecurity in smart grids, challenges and solutions. AIMS Electron Electr Eng 5(1):24–37. https://doi.org/10.3934/electreng.2021002
7. Jaccar JJ, Nepal S (2014) A survey of emerging threats in cybersecurity. J Comput Syst Sci 80(5):973–993. https://doi.org/10.1016/j.jcss.2014.02.005
8. Nga LP, Huy NQ (2020) Impacts of the Internet on banking services. Industry and Trade Magazine. https://tapchicongthuong.vn/bai-viet/impacts-of-the-internet-on-banking-services-76596.htm
9. https://vi.wikipedia.org/wiki/Internet_V%E1%BA%A1n_V%E1%BA%ADt
10. Gnoni MG, Bragatto PA, Milazzo MF, Setola R (2020) Integrating IoT technologies for an "intelligent" safety management in the process industry. Proc Manuf 42. https://doi.org/10.1016/j.promfg.2020.02.040
11. Malik PM, Singh R, Gehlot A (2020) Industrial internet of things and its applications in Industry 4.0: state of the art. Comput Commun 166. https://doi.org/10.1016/j.comcom.2020.11.016
12. Thanh NTP, Dung NT, Huy DTN, Hien DTN, Hang NT, Thuc HV (2021) Deepening researches on infrared rays principles, applications and products. Turkish J Comput Math Educ 12(11)
13. Van Thuc H, Thao DT, Thach NN, Huy DT, Thanh NT (2020) Designing data transmission system with infrared rays. Psychol Educ 58(2):3406–3411
14. Huy DTN (2012) Estimating beta of Viet Nam listed construction companies groups during the crisis. J Integr Dev 15(1)

15. Ghosh A, Edwards DJ, Hosseini MR (2020) Patterns and trends in Internet of Things (IoT) research: future applications in the construction industry. Eng Constr Archit Manag 28(2):457–481. https://doi.org/10.1108/ECAM-04-2020-0271
16. Hang NT, Tinh DT, Huy DTN, Nhung PTH (2021) Educating and training labor force Under Covid 19; impacts to meet market demand in Vietnam during globalization and integration era. J Educators Teachers Trainers 12(1)

A Novel Hybrid Approach Based on Filters to Ensure Cloud Storage Data Security

Libin M. Joseph and E. J. Thomson Fredrik

Abstract The next-generation architecture of IT organization has been planned for cloud storage. We have proposed a new Hybrid-Secure Cloud Storage (HSCS) approach to ensure data protection during auditing in the cloud. The hybrid approach consists of integrating the Schnorr Elliptical Curve (SEC) framework with the Bloom Filter (BF). Finally, the BF is added to delete redundant keys in table entries to avoid duplication. The performance metrics for SCS and HSCS compared, in this HSCS would provide greater security than SCS, including Auditing-Time, Encryption-Time, Decryption Time, and Storage-Overhead.

Keywords Cloud storage · Trusted party auditor · Schnorr Elliptical Curve · Bloom filter

1 Introduction

As cloud technology develops rapidly, the benefits of cloud such as connectivity, cost-effective including on spot service fulfillment are universally embraced by consumers and organizations [1]. In an attempt to lessen their computing load, users might subordinate complicated calculations mostly to cloud data center [2]. Consumers should still consider outsourcing to the cloud for their massive data to relieve their entire computing burden [3]. The cloud will check whether the file that the user uploads is repeated or not using the file indexes [4].

L. M. Joseph
Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

E. J. Thomson Fredrik (✉)
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_39

2 Related Works

In Shen et al. [5], kept track of its issue of sharing data with contextual knowledge and developed a specific cloud services auditing system that is identity-based. In Li et al. [6], suggested a cloud services control method for authenticating protocol deduplication. In Ding et al. [7] created a compact information security cloud-controlled information security audit system. In Joseph and Thomson Fredrik [8], surveyed various security protocols for secure data storage in the cloud.

3 Methodologies

This SCS-related protocol uses vector variables to only produce the data tags, based on the Ring-Learning with Error-Problem (RLWE). The module for entity initialization is the preliminary step in the proposed HSCS algorithm. There are three modules in the proposed model. In the entity module, there are three cloud security models as "DO" for Data Owners, "CS" for Cloud Server, and "TPA" for Third-Party Auditor. In the Security Module, the initialization of the entity module is based on the fact that perhaps the TPA is real or obscure.

3.1 HSCS Operating Principles

The auditing process is enhanced by using the proposed HSCS with the integration of the SEC framework with the BF. The pseudo-code to perform the HSCS is described as follows:

Step 1: Initialize the domain parameters for the SEC algorithm
Step 2: Select the point on the elliptic curve
Step 3: Consider the pseudorandom number as the secret key
Step 4: Construct the public key using the secret key and base point
Step 5: Create hash value for the message
Step 6: Create the signature set and the block tag for each hash value
Step 7: Create the Index Hash Table (IHT) by using the server-oriented parameters
Step 8: Derive the unique hash value from the entries in the hash table
Step 9: Compute the variations among the fresh entrance and the stored entrants in the IHT
Step 10: While the difference is zero, then the existing hash entry is replaced with the new unique hash entry. Otherwise do, derive the unique hash entry.
Step 11: Compute the new base point and signature for the unique hash entry through proof generation
Step 12: Verify the proof through the octet series and the string to an integer.
Step 13: If the proof is true, then the BF is initiated to remove the duplicate entries.

The process to predict the coefficients of the filters as follows:

Parameters as Input: "$H(m_i)$" as Function of hashing, "n" as Integer, "F" as Field for Message

Expected Outcome: Response from the TPA

(i) Assign bits of "n" by "0"
(ii) Within "M" for every "(m_i)"
(iii) "h_i" value of hashing needs to be calculated
(iv) Repeat this for every "h_i"
(v) "$f(h_i(m_i))$" need to be computed
(vi) "IF $f h_i(m_i)) \neq 1$"
(vii) False will be the TPA outcome, "ELSE"
(viii) True will be the TPA outcome.

4 Results and Discussion

These protocols have been deployed using the Open-Stack cloud environment and it has measured the time required to encrypt, decrypt, and audit the user authentication data access service, storing data service for secure object storage. The system was implemented through the eclipse android simulator which is operating on the Intel-core I5 processor in Windows 7 on desktop at 260 GHz and a RAM of 8 GB.

(i) **Audit-Time**

The audit-time denotes the time required to audit the file for verifying the data integrity. The delay for computation concerning the number of auditing requests is lesser for the proposed system compared to the existing system.

Table 1 shows the auditing-time analysis concerning the number of auditing requests. The proposed method achieves a lower computational delay of about 30% than the existing system.

Figure 1 shows the time to audit concerning a total request for audit. The X-axis of the graph represents the total request for audit and Y-axis represents the computational delay. From the graph, it is observed that the proposed system achieves a lower computational delay than the existing system.

Table 1 Audit-time performance

Auditing request	SCS	IISCS
10	600	450
30	700	500
50	850	550
70	950	650
90	1100	750

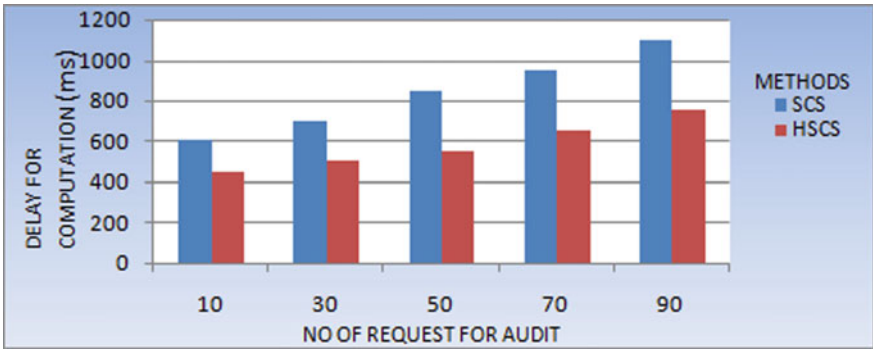


Fig. 1 Audit-time performance graph

Table 2 Encryption-time performance

No. of attributes	SCS	HSCS
10	1	0.7
30	1.9	1.1
50	2.9	1.6
70	3.8	2.1
90	4.8	2.7

(ii) ***Encryption-Time***

The time for encryption is termed as the time took for the method to encrypt the file utilizing a secret key. The Input/Output (IO) file time is not included.

Table 2 presents the comparative analysis of the encryption-time on data owner for the proposed scheme and existing scheme. The variation in the encryption-time is analyzed concerning total attributes. The attributes total varies from 10 to 90 (Fig. 2).

(iii) ***Decryption Time***

The time for decryption is termed as the time took for the method to decrypt the file utilizing a secret key. There is a linear increase in the encryption and decryption time with the increase in the length of the primes. Lower encryption and decryption time shows the efficiency of the proposed security schemes.

Table 3 presents the comparative analysis of the decryption time on data owner for the proposed scheme and existing scheme. Attributes total varies from 10 to 90. Figure 3 shows the decryption time on the user for the proposed HSCS scheme and existing SCS scheme. The HSCS scheme uses a single key for decrypting multiple files.

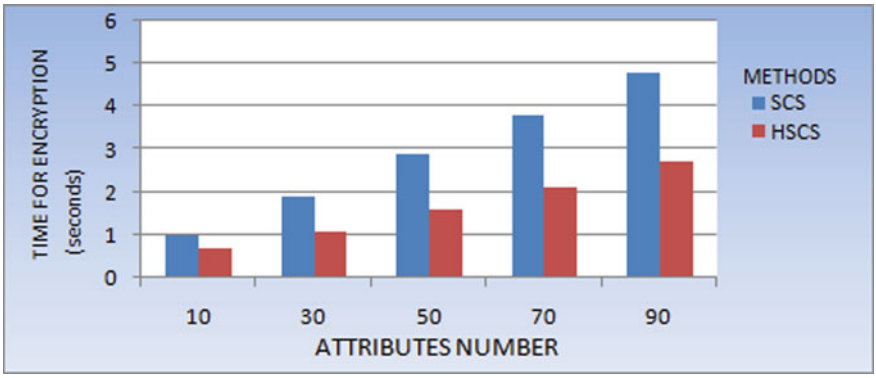


Fig. 2 Encryption-Time Performance Graph

Table 3 Decryption time performance

No. of attributes	SCS	HSCS
10	0.8	0.5
30	1.4	0.7
50	2.1	0.9
70	2.8	1.3
90	3.6	1.7

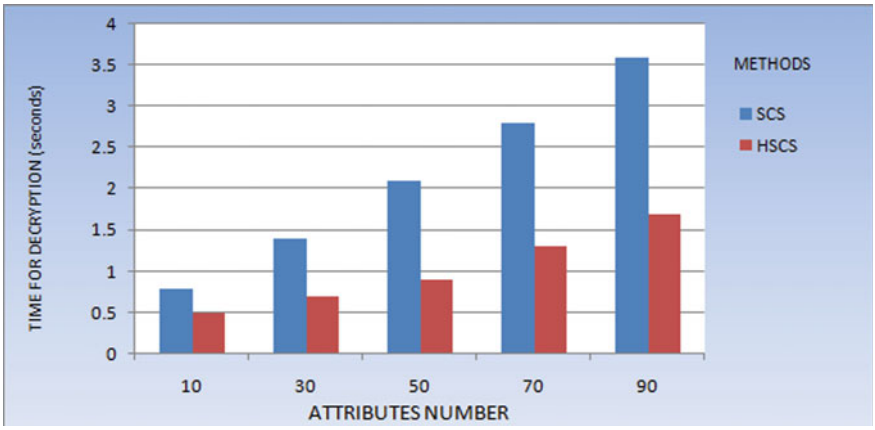


Fig. 3 Encryption-time performance graph

(iv) ***Storage-Overhead***

The storage-overhead is the additional storage space required for storing the necessary information other than the file. Reduction in the storage-overhead at the CSP

Table 4 Storage-overhead performance

Server density	SCS	HSCS
10	2100	1600
15	2400	1700
20	2700	1800
25	3000	2100
30	3300	2400

side is a key feature to reduce the service charge paid by the customer the proposed system is highly efficient than the existing techniques with less storage-overhead.

Table 4 illustrates the storage-overhead analysis concerning server total. Server total ranges from 10 to 30. For 10 servers, the storage-overhead for the proposed system is 1 600 KB/sand the existing system is 2100 KB/s.

5 Conclusion

The proposed HSCS aims to resolve issues in cloud-based security system modeling that are accessible. The effective elimination of non-related material and replication increases device performance thus lowering space burden. By using the proposed integrated evolution model, the detection and avoidance of fraudulent behavior are strengthened. The proposed HSCS effectively minimized the period for auditing, encryption, decryption, and storage-overhead by the total amount of servers and requests for audit due to the optimization in the computational phases by SEC signatures and the deduction of duplicate by BF.

References

1. Zhao P, Yu J, Zhang H, Qin Z, Wang C (2020) Howto securely outsource finding the min-cut of undirected edge-weighted graphs. IEEE Trans Inf Forensics Secur 15:315–328
2. Zheng Y, Duan H, Wang C (2019) Towards secure and efficient outsourcing of machine learning classification. In: European symposium on research in computer security. Springer, Berlin, pp 22–40
3. Fan Y, Lin X, Tan G, Zhang Y, Dong W, Lei J (2019) One secure data integrity verification scheme for cloud storage. Futur Gener Comput Syst 96:376–385
4. Hou H, Yu J, Hao R (2019) Cloud storage auditing with deduplication supporting different security levels according to data popularity. J Network Comput Appl 134:26–39
5. Shen W, Qin J, Yu J, Hao R, Hu J (2019) Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage. IEEE Trans Inf Foren Secur 14(2):331–346
6. Li J, Li J, Xie D, Cai Z (2016) Secure auditing and deduplicating data in cloud. IEEE Trans Comput 65(8):2386–2396

7. Ding R, Zhong H, Ma J, Liu X, Ning J (2019) Lightweight privacy-preserving identity-based verifiable IoT-based health storage system. IEEE Internet Things J 6(5):8393–8405
8. Joseph LM, Thomson Fredrik EJ (2020) Analysis of the recent security protocols in cloud storage system. Int J Grid Distrib Comput 13(06):1440–1452

Grading of Red Chilli, Cardamom and Coriander Using Image Processing

P. Renupriya, H. B. Anita, and Smrithy R. Sunil

Abstract Indian cuisine is known for its wide range of spices. Spices are known as the heart and soul of Indian food. Traditionally, categories are identified based on certain chemical technology or with the help of senses gifted to mankind. In this paper, an image processing technique used to extract multiple features is presented to determine the various categories of spices consumed. This proposed work uses different varieties of common Indian spices such as *Capsicum annuum* (dry red chilli), *Elettaria cardamomum* (cardamom) and *Coriandrum Sativum* (coriander). While creating the image dataset, different categories of all spices were taken from southern region of India. Features are extracted from the manually created image dataset, which forms the base for classification. The result obtained using Multilayer Perceptron (MLP), Naive Bayes and Random Forest classifier is found to be optimal.

Keywords Image processing · Chilli · Cardamom · Coriander · Multilayer perceptron · Naïve Bayes · Random forest · Fast Fourier transform · Discrete cosine transform

1 Introduction

Food is an important component of daily life. The type of food intake determines the well-being of human body. It is important to intake right amount of nutrients to stay healthy. Nowadays, the methods followed by farmers to cultivate crops have been posing a threat to the food industry.

P. Renupriya (✉) · H. B. Anita · S. R. Sunil
Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
e-mail: p.renupriya@cs.christuniversity.in

H. B. Anita
e-mail: anita.hb@christuniversity.in

S. R. Sunil
e-mail: smrithy.sunil@cs.christuniversity.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_40

In this era, numerous chemicals or toxins are added to food supplements for preserving and to increase their shelf life. It is necessary to check the quality of raw ingredients that goes into making food. If there is no quality control, then it gets difficult to track down the amount of chemicals present on them.

The grading of spice helps the farmers, industry and consumers to identify the quality of the products being used. In this paper, image processing techniques are used to study and analyse spices and in turn classify them in accordance with their category. This study focuses on three commonly used variety of spices such as red dry chilli, cardamom and coriander. Bulk samples of coriander and cardamom is captured for this purpose, whereas individual samples of red dry chilli are used. The environment for this work is kept constant and the images were recorded, and further ample number of features were extracted to categorize the samples.

The proposed work uses various features obtained from the images of the sample to identify the class each belongs to. Image processing techniques are used in multiple disciplinary such as classification, image retrieval, image compression, restoration and object detection to name a few. It is highly used to classify images into various categories. It is a simple yet efficient approach to make use of image processing technique for analysis of each sample.

2 Literature Survey

From the findings, it is observed that limited work has been performed that concerns classification of Indian spices. Very few researchers have worked on chilli dataset and hardly any with coriander or cardamom. It is recognized that most of the techniques are either automated or chemical analysis of the species. A brief survey is conducted to study the various techniques to increase the scope of the proposed model.

To classify spices such as dry red chilli, a rule-based approach is presented in [1]. This algorithm uses 9 colour features such as RGB, HIS and L * A * B from bulk chilli samples. The accuracy obtained for chilli without stalk was found to be 90% and as for chilli with stalk it was 82.5%. To check the authenticity of black pepper, diffuse reflectance mid-infrared Fourier transform spectroscopy (DRIFTS) combined with chemometrics was used. This study was used to identify pure or impure black pepper samples. The result obtained showed that DRIFTS combined GA-SVM, and PLS-DA approach gave 100% accurate result for pure samples and for fake pepper samples it was 98% accuracy for train data and 96% for predicted set [2]. A curvature-based approach was used to classify anthurium flower images. B-spline algorithm was used to detect the curve of the flower. Several algorithms such as SVM, KNN, Decision Tree and Naïve Bayes was used to classify the flowers based on their cultivar. SVM and Naïve Bayes classifier gave a resultant classification accuracy of more than 98% [3]. Vanilla species before flowering were compared and classified using image analysis of their leaf patterns. An image processing model was built using OpenCV Java alongside SVM classifier was used to obtain the classification result. It was observed that when four features were used as a feature vector the rate

of accuracy ranged from 80 to 90%. But when two features were considered, the accuracy varied from 66 to 80% [4]. A three step approach was used to check the purity and authenticity of saffron. It involved ISO3632, microscopic examination and DNA barcoding. A total of 36 samples of commercially available saffron samples were used for this study. Using this method, an overall accuracy of 66.66% was detected to identify true saffron species. Around 27% of the samples contained certain form of adulterants, whereas 5% of the samples could not be classified [5]. A software programme was designed using image processing techniques to recognize the colour and shape of chilli leaf image, in order to identify the plant disease. The recognition accuracy of this model was about 93.3% [6]. A low-angle X-ray scattering (LAXS) was used for spices such as anise, coriander, cumin, fennel and black cumin, to detect over-irradiation. The full width at half maximum (FWHM) of the peaks obtained from LAXS are investigated for each of the category. The experimental results suggest that the maximum amount of irradiation should be 30 kGy to avoid damaging the molecular structure of these spices [7]. A 2-D local discriminant-based algorithm was used to detect multispectral data of hazelnut kernels and red chilli peppers to detect aflatoxin and fungal contamination. An accuracy of 92.3% and 90% were obtained for aflatoxin contamination in hazelnut kernels and red chilli peppers and 95.6% in detecting fungal contamination [8]. Similarly classification of rice pulse has been offered in [9]. Commercially available south Indian rice seeds were collected, and flatbed scanning technique was used to capture the images and create dataset. Fast Fourier transform was also used to extract feature which represents the inner details of the rice seeds. An accuracy of 95.78% was obtained using Naïve Bayes Tree in the first stage. Whereas SMO classifier resulted with 87.07% accuracy for the second stage classification. A thin layer chromatography combined with image analysis (TLC-IA) and chemometrics technique was used for validating authenticity of saffron and identify the adulterants. This study was carried using PCA, K-means, PLS-DA and LDA techniques [10]. To check the authenticity of black pepper, DRIFTS combined with chemo metrics was used. The result obtained showed that DRIFTS combined GA-SVM, and PLS-DA approach gave 100% accurate result for pure samples and for fake pepper samples it was 98% accuracy for train data and 96% for predicted set [11].

3 Implementation

3.1 Data Collection

The dataset for various categories of dry red chilli, cardamom and coriander were not readily available. Hence this dataset is manually created for the proposed work. The spices were collected from various vendors across South Indian region.

Red dry chilli: It is a commonly used Indian spice which adds pungent taste and colour to the food. They are used in various forms such as whole, coarse or ground.

Table 1 Sample dataset for Chilli

Varieties of red dry chilli	
 Kashmiri chilli	 Byadagi chilli
 Normal chilli	 Hot chilli
 Guntur chilli	

They act as antibiotics and pain killers. It can help provide relief from lung disease in certain cases. For the proposed research work a sample of 300 images of the following varieties of dry red chilli is used. The sample dataset contains5 chilli varieties that belong to types such as Kashmiri, Byadagi, Normal, Guntur and Hot chilli, they are shown in Table 1.

Cardamom: This spice is used in both savoury and sweet dishes. The images were captured of the whole green pod as a bulk sample. To help aid this research a total of 50 images of bulk samples of different variety of cardamom based on quality. They are Bold 6, Bold 7 and Bold 8. The weight of each sample captured is maintained constant throughout, the samples are displayed in Table 2.

Coriander: The coriander seeds are used as a spice and is widely used in many savoury dishes. Coriander has been used in traditional to get relief from constipation. It comes packed with varied amounts of calcium, magnesium, potassium and phosphorus. The dataset of 110 images of bulk samples of different types of coriander that are shown in Table 3.

The images were taken under a standard environment setup. The environment is built such that the following constraints are kept stationery throughout the data collection procedure:

- A normal webcam mounted on a tripod stand to get standard images
- A white background for optimization purpose
- Light intensity should be balanced using a LED camera panel

Table 2 Sample dataset for Cardamom

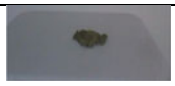
Varieties of cardamom	
 Bold 6	 Bold 7
 Bold 8	

Table 3 Sample dataset for Coriander

Varieties of coriander	
 TYPE 1	 TYPE 2
 TYPE 3	 TYPE 4
 TYPE 5	

- External disturbance is kept minimal
- The base height for placing the item is kept constant

3.2 Feature Extraction

Features are basically used to describe an image and to identify the differentiable factor between non-related sample images. The chosen feature set should be able to enhance the differential factor by a manifold. In this proposed work, boundary descriptor, regional descriptor, statistical and frequency-based features are being used. MATLAB is used for the purpose of feature extraction. The following algorithms elaborate more on this:

Algorithm 1 Input: Colour image of red dry chilli

Output: Feature vector.

Method:

1. The following is extracted for each colour image:
    a. The mean value of red and blue channel (2 features)
    b. Standard deviation of red and blue channel (2 features)
    c. Entropy of green channel
    d. The mean value of saturation and intensity (2 features)
2. Converted the colour image into grayscale and calculated the below mentioned features from the grayscale image:
    a. Mean value
    b. Standard deviation
    c. Total area of the image is calculated

d. Apply Fast Fourier Transform (FFT) to the grayscale image and compute the standard deviation of FFT image.
e. Apply Discrete Cosine Transform (DCT) to the image and compute the standard deviation of the DCT image.
f. Apply DCT to the image and consider the DC component of the DCT image.

3. Binarize the input image to compute the following:

a. Calculated on pixel density
b. The total number of pixels in an edge detected image
c. The mean value

This accumulates to a feature vector of length 16.

Algorithm 2 Input: Colour image of cardamom.

Output: Feature vector.

Method:

1. Taking colour image of each sample, the following features are computed:

a. The mean value for green channel
b. Standard deviation of green channel
c. Entropy of green channel

2. Converting the image into its grayscale, the following features are extracted:

a. Grey scale mean value
b. Grey scale standard deviation
c. Entropy of the image

3. Convert the image into black and white and calculate the mean value.

This totals to a feature vector of size 7.

Algorithm 3 Input: Colour image of coriander.

Output: Feature vector.

Method:

1. Calculate the following for each colour image sample as a feature:

a. The mean value of red and green channel (2 features)
b. Standard deviation for red and green channel (2 features)
c. Entropy of red and green channel (2 features)
d. The mean value of hue, saturation and intensity (3 features)

2. Convert the image into grayscale and compute the following:

a. The mean value
b. Standard deviation

c. Entropy of the image
d. Compactness of the image
e. Apply Fast Fourier Transform (FFT) to the grayscale image and compute the standard deviation of FFT image.
f. Apply Discrete Cosine Transform (DCT) to the image and compute the standard deviation of the DCT image.
g. Apply DCT to the image and extract DC component of the DCT image.

3. Binarize the image and calculate the following:

a. Total number of 0's and 1's (2 features)
b. Mean value

This algorithm gives feature vector of size 19.
Total feature vector of length 42 is obtained by all the above-mentioned algorithms.
The system work of the proposed model is shown in Fig. 1.

Fig. 1 System model

3.3 Classifiers

To classify, WEKA software is used, which has built-in machine learning algorithm which makes this process efficient. The proposed work experimented with Random Forest, MLP and Naïve Bayes classifier. The Random Forest classifier creates several decision trees based on a dataset. The average estimate value of the trees is calculated to choose the best fit model which decides the class of an object. Similarly, MLP is a supervised artificial neural network model. The input and output layer are analysed to obtain the correlation between the pairs. In contemplating the best model, the error needs to be minimized by performing backward and forward propagation. Whereas a Naïve Bayes classifier works on the principle concept of the Bayes theorem. It is a probabilistic algorithm that considers that each feature is completely independent from the other. It is a simple yet powerful algorithm that is highly used for the purpose of classification.

In this work, the input dataset is partitioned into training and testing module. The train dataset is used to train the model while the testing dataset is used to evaluate the trained model.

4 Experimental Results

The algorithms for classification of spices are evaluated on the dataset collected. A total of 460 image samples of spices are collected. This entire dataset is refined to evaluate the proposed system model. The dataset is divided into 6:4 ratio denoting training set and testing set, respectively. Features which distinguish the class of red dry chilli, cardamom and coriander are stored for processing. The respective algorithms are explained in Sect. 3.2. The data is normalized before it is used to train the model. The model is then trained to classify each of the subcategories of each spice variety. This type of classification result helps to classify each variety of spice. The performance of the extracted features is examined using various classifiers.

In terms of the classification results for red dry chilli, it has been identified that MLP gave a better accuracy of 83.33% whereas Naïve Bayes and Random Forest classifier produced an accuracy rate of 75.33% and 80.67%, respectively. This result can be seen in Table 4.

Similarly, for cardamom, Random Forest and MLP produced better result of 95% each when compared to Naïve Bayes with accuracy of 80%.

Table 4 Result for red dry chilli, cardamom and coriander using classifiers mentioned in Sect. 3.3

	Random Forest (%)	MLP (%)	Naïve Bayes (%)
Red Dry Chilli	80.67	83.33	75.33
Cardamom	95	95	80
Coriander	76.36	83.63	69.09

The accuracy for classification of coriander obtained through MLP classifier is 83.63% which is more prominent when compared to the accuracy of 69.09% and 76.36% obtained using Naïve Bayes and Random Forest classifiers.

5 Conclusion

In the proposed work, classification of spices using image processing algorithms are observed. The data collection procedure was conducted manually for the spices belonging to various types such as red dry chilli, cardamom and coriander. As per the findings, it is observed that there is minimal research work related to classification of these variety of Indian spices. In the anticipated work, the data set is created using commercially available red dry chilli, cardamom and coriander from different vendors of south India. A total of 460 image samples were collected which includes all varieties of these spices. MATLAB and WEKA tool was used to extract features and classification. An overall average accuracy of 87.32% was obtained using MLP classifier. In future the dataset size can be increased, and deep learning technique can also be applied. The proposed method can be applied to other variety of spices. Subsequently a digital camera with higher resolution can be used to collect images of better picture quality.

References

1. Sajjan M, Kulkarni L, Anami BS, Gaddagimath NG (2016) A comparative analysis of colour features for classification of bulk chilli. In: 2016 2nd International conference on contemporary computing and informatics (IC3I)
2. Chiu S, Lin M, Ho W, Lin H, Chen P, Huang H (2018) Application of food description to the food classification system: Evidence of risk assessment from Taiwan as Acrylamide of grain products. J Food Drug Anal, 1–8.
3. Pour AS, Chegini G, Zarafshan P, Massah J (2018) Postharvest biology and technology curvature-based pattern recognition for cultivar classification of Anthurium flowers. Postharvest Biol Technol 139:67–74
4. Ambika NK, Supriya P (2018) Detection of Vanilla species by employing image processing approach. Proc Comput Sci 143:474–480
5. Khilare V, Tiknaik A, Prakash B, Ughade B, Korhale G (2019) Multiple tests on saffron find new adulterant materials and reveal that 1st grade saffron is rare in the market 272:635–642
6. Joseph A, Anita HB (2017) Leaf classification using image processing: a review. Int J Adv Res 5(3):28–32
7. Desouky OS, Ashour AH, Abdullah MI, Elshemey WM (2002) Low-angle X-ray scattering from spices 64:267–271
8. Kalkan H, Beriat P, Yardimci Y, Pearson T (2011) Detection of contaminated hazelnuts and ground red chili pepper flakes by multispectral imaging. Comput Electron Agric 77(1):28–34
9. Philip TM, Anita HB (2017) Rice grain classification using Fourier transform and morphological features. Indian J Sci Technol 10(14):1–6
10. Sandasi M, Chen W, Vermaak I, Viljoen A (2018) Phytochemistry letters non-destructive quality assessment of herbal tea blends using hyperspectral imaging. Phytochem Lett 24:94–101
11. Hu L, Yin C, Ma S, Liu Z (2018) Assessing the authenticity of black pepper using diffuse reflectance mid- infrared Fourier transform spectroscopy coupled with chemometrics. Comput Electron Agric 154(September):491–500

Smart Mirror: A Magical Gadget for Diverse IOT Services

R. H. Aswathy, R. Saravanan, B. S. Sudhan, and D. Sasikiran

Abstract The Internet of Things (IoT) is a popular technology in today's world. It influences every human life and extends its path to many applications like education, healthcare, agriculture, logistics, business intelligence, and wearable devices. The smart mirror is an innovative application in home automation that looks like a standard mirror acting as a smart mirror. In the future, "smart mirror" will be used for many purposes because the mirror is a common sight for everyone. The smart mirror is designed in our proposed work, showing the time, temperature, news updates, and weather report to display on the mirror. The IoT-based smart mirror designed using a Raspberry Pi microprocessor with sensors for perceiving the external environment, and a mirror act as an actuator to show the information. This intelligent screen will have the ability to display real-time information and connect to the user's smartphone who uses the Android application. This mirror enables the features like display live news updates, managing events through google calendar, and shows meteorological information and biological information. Besides this, the user can update the additional functionalities and features themselves by specific agreements.

Keywords Raspberry Pi · Python · Internet of Things (IoT) · Mirror · Application programming interface (API) · Display

1 Introduction

In today's world, most people are connected to the Internet. Everything is online with the Internet's help, such as bank transactions, information sharing, and education—the drastic changes in people's lifestyle due to the active computer and embedded system. Various devices and products based on this communication technology have introduced around the world. Internet of Things (IoT) plays a significant role in automation. An "object" in IoT can be anything that can collect, correlate, and

R. H. Aswathy (✉) · R. Saravanan · B. S. Sudhan · D. Sasikiran
Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore 641407, India
e-mail: rhaswathy@kpriet.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_41

transmit data over a network without human intervention. Devices connected technically can be controlled and monitored using the remote monitoring technique. With the great use of automation, the data are exploded through the Internet and make people updated. Our work intends to develop an intelligent mirror similar to various options such as smartphones, tablets, and smart television. This smart mirror looks like an ordinary mirror, but the mirror works as a smartphone that contains a screen display that can show time, weather, and date. Mirrors are part of everyday use because people interact more with mirrors. The combination of mirror and ingenuity with technology proves that access information and communication from smartphones, smartwatches, and even we can allow technology into any of the things we use in everyday life. Smart mirror contains camera and mic which uses for face and voice recognition. It will be wise enough to recognize a person's face and voice and thus see the users. People will be able to retrieve data from the Internet by contacting the screen directly. A user might get the weather conditions like temperature, humidity, and rainfall in the following day to plan the trip or work. The main intention of the proposed work to provide a summary of the important data to plan the work like user location, weather updates, news feeds, and today's program only by knowing the user's face. The focus is to develop an intelligent mirror that identify the user and tailor the experience.

2 Discussion of Related Works

Technology is an integral part of everyone life. "Smart mirror" is an application that brings technology to the traditional home screen and makes them more intelligent. The smart mirror helps the users to organize the daily activities and find the essential information such as meteorological data, news updates, and calendar events. IoT is an interconnection of portable devices with embedded devices connected to exchange data. IoT communicates between people and objects and between objects. Intelligent home automation screen is another domain with IoT applications. IoT applications vary. Instead, this project paper helps to use IoT to make life smooth. The mirror displays the news, time, date, weather conditions, checklist, reminders, and the screen is upgraded to function as browsers. We can gain access to the news or watch YouTube videos. In 2003, Phillip sunk mirror TV developed based on the principles of smart mirrors. The standard TV mounted behind a two-way screen and TV changed as a mirror when turned on. There is an option of making the screen more prominent than the TV performance testing. The author, James Law, developed a cyber-mirror that shows rudimentary information like weather, date, time, and system that integers social communication. So, the user can view their information anywhere. That cyber-secured, the mirror controlled through remote-controlled methods. The mirror can display the patient biological information such as blood pressure, glucose level, weight, height, and heart rate [1]. In an intelligent home [2], the smart mirror shows the daily data and coordinates household items. This application's fundamental concepts are to remind the user's daily activities like a scheduler

and remainder [3]. The face recognition and biometric-based authentication models discussed to customize the mirror according to user needs. The user wants to select the location of the mirror. In the article [4], interaction added with the model recognizes the user command, gesture, and product inculcate all the qualities of personal digital assistance (PDA) such as mobile phones, tablet, and laptop. The author [5] builds Raspberry Pi to build the smart mirror with interconnection with real-time Web sites. In article [3], the Philips home lab acts as the best bed for developing an interactive home environment. The author [6] discussed the LED mirror, which mimics the functionalities of the standard mirror. It shows the personalized information and gives updates of our interest like home automation, sports, refreshment, and hobbies carried out in the article [7]. The mirror works like a widgets work based on command [8–10]. The mirror has designed for updating date and time touch screen information.

3 Methods and Materials

3.1 One-Way Mirror

A one-way mirror gives the mirror users true identity. It is truly a magical mirror as it has a bright spot on one side and is also visible with great power. The screen rests on the front, where the user can view the screen at parallel timing. It gives the monitoring light access to travel through and ensures that the interface is available [UI]. It is the front-end object of the system. The user could physically view the things through the one-way mirror.

3.2 Monitor

The monitor module is integrated with Raspberry Pi via an HDMI module that offering display and speech output. The touch infrared receiver [IR] is a self-monitoring monitoring device which displays the information in the mirror.

3.3 Infrared Frames

Infrared receiver frames provide a touch screen on a magic mirror. Infrared receiver frames have the IR sensors on own place and is connected to the Pi via USB interface. That made the mirror more visible.

3.4 Raspberry Pi 3-Model B

The Pi plays a major in a smart mirror, forming a mirror processing unit. Pi is like a motherboard with all the necessary components that make up a good CPU. Its credit-card size can also work as a complete computer- Raspberry Pi editing with Python programming language. Programs are initially embedded, then upgraded and integrated into the platform like windows, macOS, Linux, and work on Pi. Pi has its built-in integrated development environment (IDE) for editing in programming languages such as Java, Python, C++, C. Installing the OS on the Raspberry Pi is a straightforward process. Raspbian is just a flavor of Debian OS. PyCharm community serving is an excellent IDE to develop our proposed system. Qt Designer is a fantastic tool for customizing the Python UI. It has built up Wi-Fi and Bluetooth to connect and allows four USB devices to connect to it.

3.5 Channel Relay

The eight-channel transmission is connected to the GPIO high power input power source and power supply on the Raspberry Pi. GPIO pins lack to control household appliances. They have a meager output power of almost 5 V, a circuit board that transmits eight-channel that closes the home appliance circuit when given a maximum of 5 V.

3.6 Camera

The critical part of the smart mirror is the camera; this device is used for face recognition. A method called virtual-based products or service implemented with the help of an image processing technique. For example, people walk around the mall looking for different clothing's. We cannot wear the clothes we like. Virtual technology solves this complication by adding or uploading a picture of those dresses collected in the memory of the mirror. The user has to stand in front of the mirror and selects the garments; the mirror display shows the users image with an opted garment.

3.7 Microphone

The microphone is installed to get the user voice input on the screen. It converts sound into an electrical signal. It can sense voice input and makes the mechanism (system) more accurate with a more robust operation. This microphone observes

voice input from the user and perform the correct action. It must be noiseless and identify the user voice even if interruption of other sound occurs.

4 Design Description of Proposed Work

The smart mirror designed as a real-world display device that features on model view controller (MVC) pattern. The plugin used for displaying the data incorporated in the model pattern. This plugin retrieves information from the cloud database and acts as an application programming interface (API). The view pattern is the display screen showing real-time information to the user, and the controller pattern controls each plugin which drives the other two patterns. A proposed screen intended to display the user image as and provide customized source information. Using a one-way mirror wall, this proposed model can display relevant user details like updated and present weather conditions, local time and date, news, user health, and personal circumstances appointment updates from google calendar. The proposed architecture diagram of the smart mirror is shown in Fig. 1.

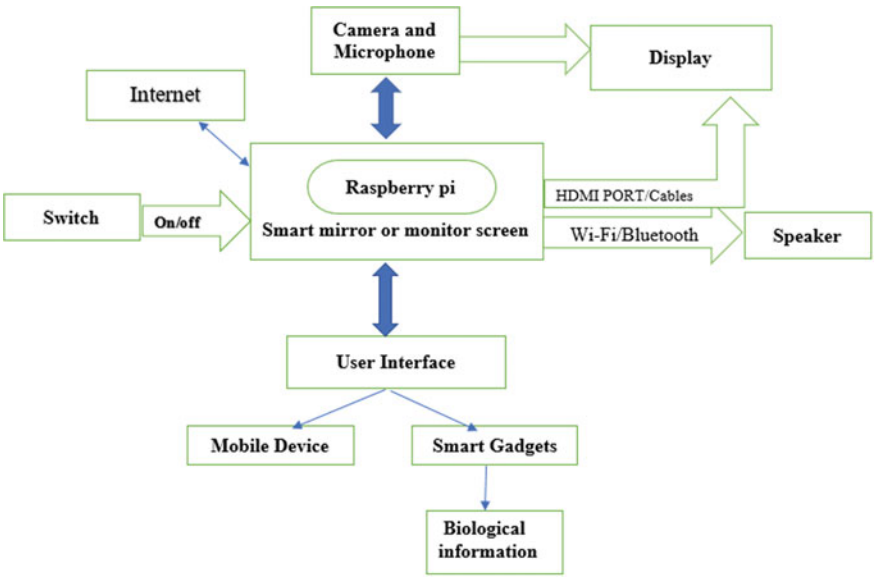


Fig. 1 Architecture diagram of smart mirror

4.1 Control Unit

The control unit works based on two taxonomies. The first one is the detection of human presence with authentication, and the second is the mirror triggered by on/off units. The system detects the human presence in front of the mirror through the PIR, a motion detection sensor. The PIR motion detection sensor is attached to Raspberry Pi, and the resisters are placed to receive the static voltages. The PIR sensor is connected to the Raspberry Pi by getting the voltage during motion detection.

4.2 Network Connectivity

The network connectivity collects information like news updates, meteorological conditions from the Internet, and smart mirror shows. The Raspberry Pi microcontroller is correlated with Wi-Fi modules and an Ethernet port, which retrieves the information to the LCD mirror screen. The Raspberry Pi has many in build module microprocessor, and it has a flexible way to parse the information.

4.3 User Interface

This user interface module permits the user to perform the updates in the mirror settings. The domestic Ethernet connection/Wi-Fi displays their geographical location. The Raspberry Pi microprocessor configured to run as an access server to display the data and edit the mirror settings. The microprocessor acts as an access point with many inbuilt functionalities like Wi-Fi, card reader, and Ethernet port. The live user location and meteorological information displayed by fetching the zip code, which forwarded by the network system. During the login, the UI system checks the authenticated login credentials. The display mirror is stationary, and information is displayed.

4.4 Display Unit

The display unit collects information from the Raspberry Pi through the network unit and displays the textual form's output. This information is passed to the LCD through the SPI module, and the data display depends on the last input to the user's interface unit. The network system continuously updates and exchange real-time data to update the information in the smart mirror. The overall display unit is shown in Fig. 2.

Fig. 2 Flow diagram of display unit

There is five display screen such as date and time, methodological information, news update, google calendar update, and patient biological information. The Adafruit LCD is finalized for flexibility and bright display. D1, D2, D3, D4, and D5 are the display, and Display 1 uses the float value to display the current time from the network system in 12 or 24 h format depending on the user choice. Display 2 is the meteorological information data which each numerical represents an updated weather condition. The meteorological condition suc as raining, cloudy, overcast, storming, and snowing is displayed in the left corner of the screen. The display units are shown in Fig. 3.

Display 3 shows the Google calendar as text information to give remainder about the upcoming task. User can easily manage their time and schedule the work accordingly. Display 4 shows the news data in the form of text-the news updates received from the Internet subsystem displayed to the smart mirror. The left lower corner in the smart mirror shows the news updates. Display 5 shows the patient biological information. The biological data such as weight, heart rate are display in the mirror.

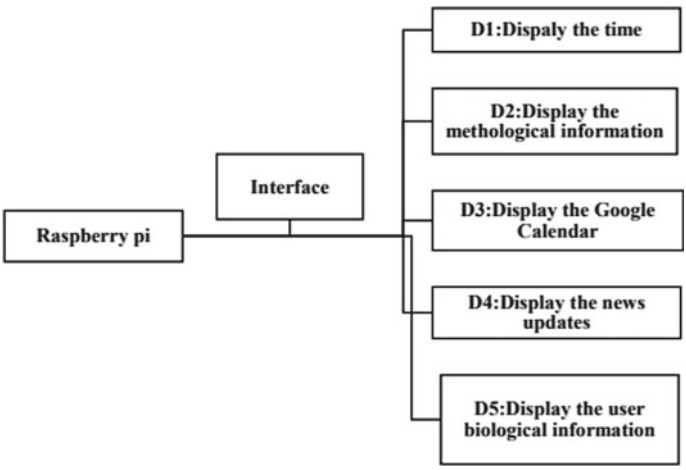


Fig. 3 Display units (D1, D2, D3, D4, D5) of proposed work

The user wears the health band. The health band's data updated to Raspberry Pi microprocessor shows the data through the smart mirror.

4.5 Speaker

The speaker system allows the user to trigger the music button and play music via Bluetooth connection. Once the Bluetooth device is triggered, it seamlessly receives data, and music will play according to the user's wish.

5 Flow Process of Proposed Work

User weight and personal health status are viewed live, whenever a user stands in an integrated stock exchange.

This information performed to lift and encourage the user to lose weight if necessary. Replacement tools can also control the blinds of window and light conditions. The status is forwarded to the screen app which includes a communication module. The excellent quality mirror works with reflexing mirrors. The lower quality mirror makes the GUI a clear one. The mirror acts as a screen, and LCD monitors lighting up, and black pixels start to glow. The magic screen is the Raspberry Pi, the display based on the Internet, which takes over again, indicates the required information. Three models used in the proposed work with 512 MB RAM, Linux base with 700 mA. The monitor acts as a mirror for our work. The user is present in front of the mirror. It authenticates the user, and the information is displayed, which indicates this helpful information using the standard GUI Python library module. The flow diagram is shown in Fig. 4 described. At the initial phase, the system identifies the user's image. If the user is authorized, it process to the next step or else terminated. If the user is confirmed, it asks for options such as date and time, daily updates, biological information, meteorological information, and google calendar. The user can choose one or multiple choices. It helps the user to make his day very productive. The system will shut down if there is no movement.

6 Results and Discussions

The magic screen is the Raspberry Pi, the display based on the Internet, which takes over again, indicates the required information before the user. Table 1 shows the various testing phase of smart mirror. The control units are ON/OFF, room temperature, music, meteorological data, Google calendar, and biological information. The steps describe the input given to the smart mirror and tested with outputs. The system works exactly in all given inputs.

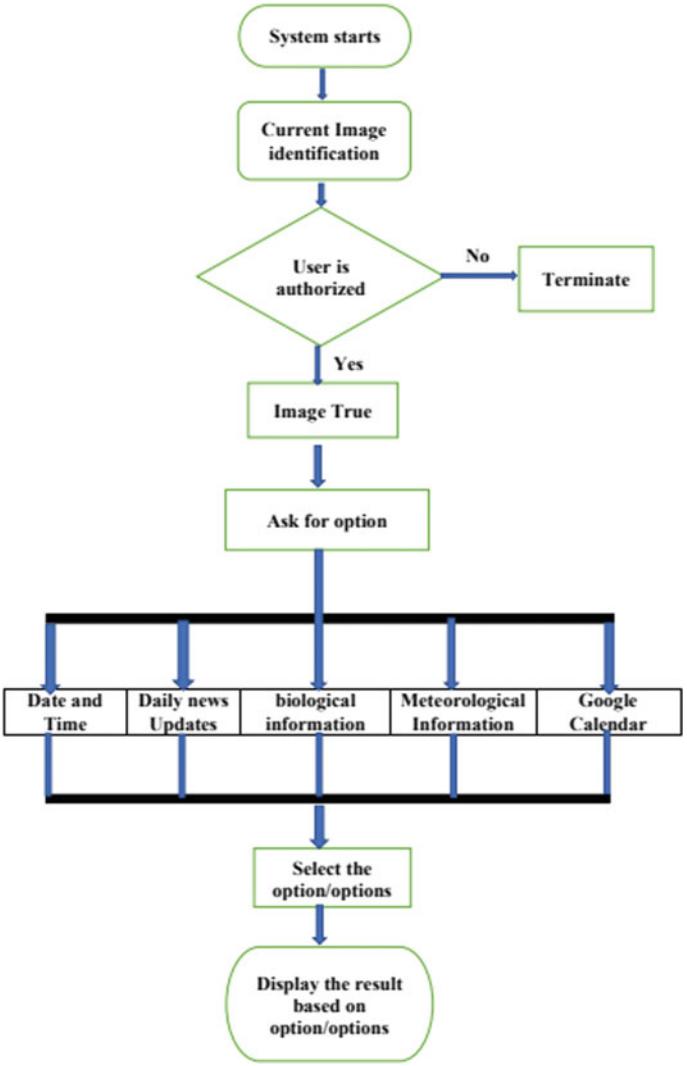


Fig. 4 Flow diagram of proposed work

Figure 5 shows the authentication during first phase. Figure 6 shows the output of the smart mirror; left upper corner is meteorological information. The left lower corner shows the news updates; the right upper corner is time, date, and google calendar, and the right lower corner shows the patient biological information.

Table 1 Testing with input

Control unit	Steps	Output
On/Off	The user moves in from of the smart gadget (mirror)	The mirror keeps active for 5 min if there is no movement
Room temperature	The temperature sensor senses the room temperature and forward data to the microprocessor	The minor displays the room temperature continuously based on the current temperature until the mirror off
Music	The user can operate the minor to play the music in two ways (1) Through mobile device (2) Through voice commands	The user triggers the music; it plays severing until the user stop instructs to stop
Meteorological information	The meteorological information such as weather, humidity raining, and fog is displayed depending on the real-time observations	The real-time meteorological information is sensed continuously and display in the smart gadget
Google calendar	The remainder and schedulers are displayed when the gadget is turn-on	The remainder and schedulers in the google calendar will alert the user at the particular time given by the user
Biological information	The patient biological information like heart rate, weight, and walk rate is displayed	The biological information is displayed whenever the user basic information through voice commands

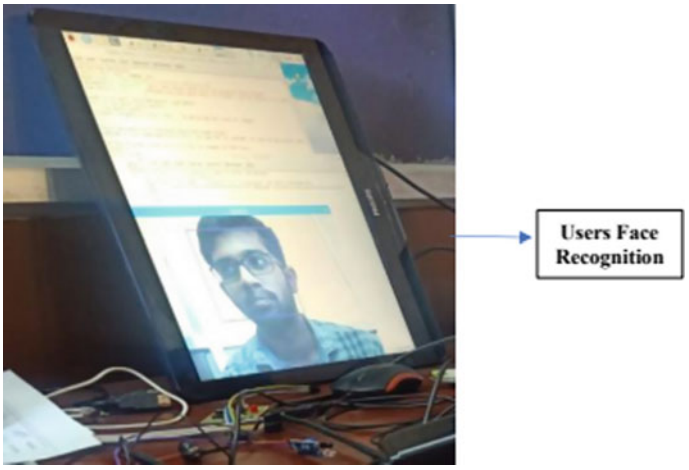


Fig. 5 User authentication during initial phase

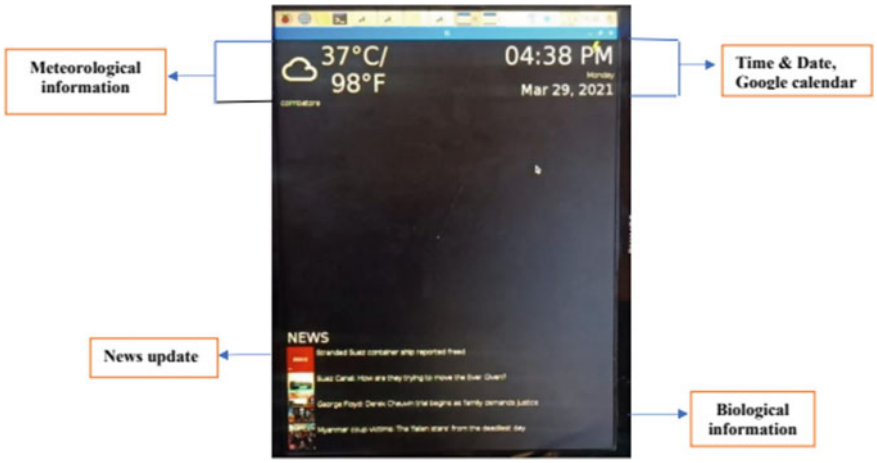


Fig. 6 Output screen of smart mirror

7 Conclusion

This smart mirror is an innovative technology for intelligent life. This reduces job management and increased productivity with brand new technologies. This technology provides instant access to information flexibly and conveniently. It has the significant potential to improve the user experience for viewing the social data, managing daily routine, and monitoring biological parameters. The project makes our everyday life easier. The user has control over the smart mirror to choose their needed information. The user need not bother about to switch off the system because it will detect the motion and work accordingly. The future enhancement of this work is to control the smart mirror with hand gestures.

Acknowledgements We acknowledge all referees who have given the useful suggestion to enhance this work.

References

1. Ab Latif MS, Ismail AA, Zariman A (2019) Smart mirror for home automation. Int J Recent Technol Appl Sci 1(1):1–11
2. Ceccaroni L, Verdaguer X (2004) Magical mirror: multimedia, interactive services in home automation. In: Proceedings of the workshop on environments for personalized information access, pp 10–21
3. Hossain MA, Atrey PK, El Saddik A (2007) Smart mirror for ambient home environment, pp 589–596
4. Ding JR, Huang CL, Lin JK, Yang JF, Wu CH (2008) Interactive multimedia mirror system design. IEEE Trans Consum Electron 54(3):972–980

5. Raspberry Pi (2019) Magic mirror—Raspberry Pi. Available at: Micheal Teaw's official website
6. Dasgupta R, Roy P. Human interacting smart mirror. http://www.cc.gatech.edu/fcele/cl/projects/ejaVu/mnmjindex~hml
7. Kashimoto Y, Fujiwara M, Fujimoto M, Suwa H, Arakawa Y, Yasumoto K (2007) ALPAS: Analog-PIR-sensor-based activity recognition system in smarthome. In: 2017 IEEE 31st International conference on advanced information networking and applications (AINA), pp 880–885
8. Athira S, Francis F, Raphel R, Sachin NS, Porinchu S, Francis S (2016) Smart mirror: a novel framework for interactive display. In: 2016 International conference on circuit, power and computing technologies (ICCPCT), pp 1–6
9. Yusri MM, Kasim S, Hassan R, Abdullah Z, Ruslai H, Jahidin K, Arshad MS (2017) Smart mirror for smart life. In: 2017 6th ICT international student project conference (ICT-ISPC). IEEE, New York, pp 1–5
10. Arun Kumar K, Uday K, Veeranjaneyulu K (2020) Smart mirror design using Raspberry Pi. CVR J Sci Technol, pp 53–57. https://doi.org/10.32377/cvrjst 1908

Smart Sustainable Agriculture Using Machine Learning and AI: A Review

A. Menaga and S. Vasantha

Abstract Artificial intelligence and machine learning are all about using data for efficient inferences and predicting the future and decisions. These decisions are made human-like, by machines; machine learning and big data are having a greater impact on the way we live. Scholars and scientists are looking at machine learning as a pioneer opportunity to create a positive impact in our day-to-day life, especially in the field of agriculture domains. The research reviews and project popular machine learning models used in the field of agriculture such as (a) crop management (crop yielding, fruit picking weed, and diseases detection), (b) soil management, (c) water management; the paper aims to introduce different types of machine learning methods and algorithms used in machine learning, and how machine learning reaches the agriculture, by implementing the machine learning into farmers by a remote sensor. The farmers will get benefited in decision-making such as risk reduction, quality seed selection, and easy monitoring with software. At the same time, the present study also focuses on the way artificial intelligence and IoT, if introduced in the agricultural sector, can boost the productivity of the sector sustainably.

Keywords Artificial intelligence · Internet of Things (IoT) · Big data · Machine learning · Smart sustainable agriculture

1 Introduction

Agriculture plays a vital role mainly in developing countries for economy wellbeing; this is the main source of income, employment, and also, it plays a significant role in international trade because the most developed country is not engaging much in the agricultural activities; it provides the large source of revenue to the government by transportation (majority of trains and trucks carries the agriculture products; finally, a very main point is food security.

A. Menaga · S. Vasantha (✉)
School of Management Studies, Vels Institute of Science & Technology & Advanced Studies (VISTAS), Chennai, India
e-mail: vasantha.sms@velsuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_42

Machine learning technology is an area of AI computer and is automatic data learning without human intervention; it has a greater impact on the global economy, especially machine learning made a huge difference in world trade mainly in communication barrier by making the availability of three hundred languages, banking and FinTech, (by making automatic communication with client and reduces the risk of abuse) and agriculture sector (crop management, soil management, livestock management, etc.).

The research aims to project and review different machine learning models used in the field of agriculture; the paper also introduces different terminologies and abbreviations used in machine learning such as BM Bayesian models, DL deep learning. The paper will be helpful to get an outline of how machine learning can be used in agriculture, and what methods are available in machine learning, especially for the same field.

The research explains the concept of machine learning and reviewed various machine learning techniques used in the agriculture field, to gain knowledge of AI techniques, and also, the paper aims to unfold the usage of IoT in agriculture to make a sustainable agriculture model.

1.1 Objective

1. To review various machine learning technology used in the agriculture field
2. To examine the application of the AI model for sustainable agriculture.

2 Literature Review

The article is divided into two levels one—the explanation of algorithm limited to the field of agriculture, level two is segregated into three general categories; namely, crop management, water management, and soil management. Crop management is categorized into crop yield, crop quality, and weed and diseases detection; the research is searched in Scopus, Web of science also in pub med, article considered the period from 2000 to the current year.

The article is structured as follows: part one explains the literature review of various machine learning models followed by reviews of crop management, water management, and soil management; part two applications of artificial intelligence and IoT in Sustainable Agricultural Practices are explained, part three discussion and conclusion.

3 Machine Learning

In simple term, machine learning is a data which teach the computer human-like, and these data are known as training data, which makes human work accurate and easy and improves the experiences; data in machine learning will be in four categories, numerical data, categorical data, time-series data, text [1]; to measure the performances of ML, various statistical tools are used to predict the result (Fig. 1).

Machine learning is widely classified into supervised and unsupervised data, supervised data is manually trained data whereas unsupervised data is the machines interpretation. Refer Table 1.

3.1 Machine Learning Algorithm Models

The machine learning algorithm is a code that runs on data and produces the decision-making output. The output or a result in a prediction made using the past experiences [2], e.g., in agriculture, weather prediction is made when the temperature is less than 17 degrees soil will get moist, and hence, it is not a good time to crop or yield, the research limit to unfold the explanation which is relevant to agriculture.

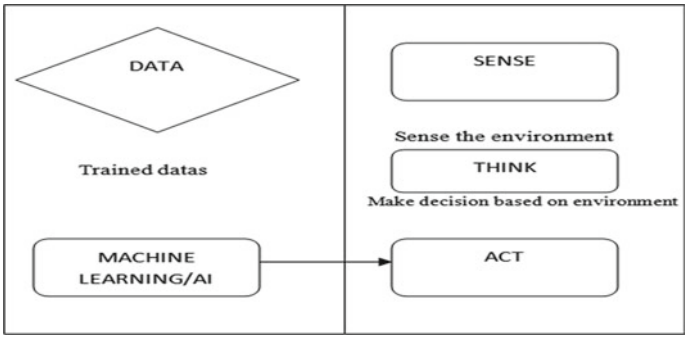


Fig. 1 Machine learning approach

Table 1 Authors model

Types of machine learning method		
Task learning	Explanation	Types
Supervised	Data are labeled and trained	(1) Regression-True/false
Unsupervised	Data are not labeled machine will make inferences of data with close similarity and association	(1) Association-rule-based approach (2) Cluster-similar object

3.2 Supervised Learning Models—Regression

Regression help to find out the relationship between two variable, and it is mainly used to predict the causal effect; regression is classified into typo types 1-single regression 2-multiple regression under which (i) linear regression (ii) non-linear regression is subcategorized in single the features, or variable will be one in case multiple variables; there will be more than one feature;, in machine learning regression has linear regression, polynomial, support vector, decision tree, random.

3.3 Classification-Bayesian Models (BM)

In Bayesian models (BM), probability is used for the output result; the Bayesian model can be used both in classification and regression; some of the Bayesian model algorithms are Naive Bayes, Gaussian Naïve Bayes, the mixture of Gaussian, Bayesian network [3] (Pearl, Duda & Hart).

3.4 Artificial Neural Networks (ANNS)

ANN is adopted and design similar to human brain neurons networks, which has an interconnection analyzing system or processing unit; the number of layers are arranged into (i) input layers (ii) hidden layers (iii) output layer [4].

A deep artificial neural network is also termed deep learning [5]. DL has multiple layers of processing data refer to Fig. 2; DL is a new algorithm in ML where data extraction is done by itself; this type of data is processed both in supervised, and unsupervised comprehensive deep learning is found in learning [6] research literature.

4 Machine Learning and Agriculture

4.1 Crop Management

The method of training the machine and producing the decision is widely used for predicting the future; machine learning is simplified human thinking with break down the complexity of the problem with easy future decision-making result; in that way, agriculture plays a very important role in the global economy; global crop yielding is an important part to address for global food security, predict and reduce the climate changes [7].

Crop yielding forecast is a significant agricultural hitch. The yield depends on weather settings (rainfall, high temperature, etc.), pesticide. Precise information

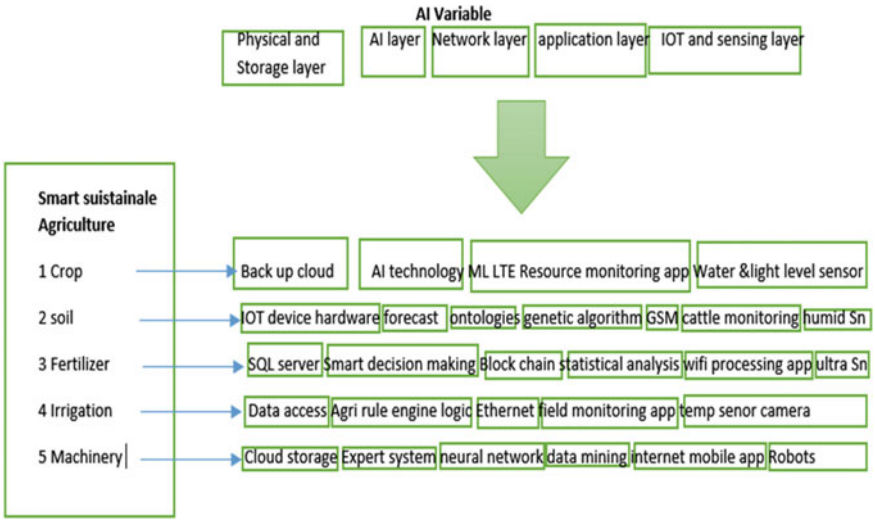


Fig. 2 Applications of artificial intelligence and IoT in sustainable agricultural practices. Authors model

regarding the past crop yield is significantly in decisions-making related to managing agricultural risk Ali et al. [8]. According to [7] he classified the crop management into (i) Fruit picking and forecasting (ii) Automatic fruit picking model green citrus detecting the ingrown citrus; in this circumstances, machine learning will provided information for growth and help the farmer for maximum yield [9].

4.2 Disease Detection

The common treatment for disease plant is spraying chemical in the farm; the major advantage of ML is that in traditional method chemical pesticide is sprayed in the enter farm land in order to prevent the diseases, but in ML, the trackers are attached with camera and computer will detect the diseased crop and spray only in that area instead of spraying chemicals exactly. Another technique under disease detection is detection of parasites in strawberry and spray chemicals in the field. The method of screening the bakanae diseases in crops the is to detect the disease and optimize the maximum yield, in wheat crop, wheat is prominent source used worldwide found the health wheat canopies based on genetically super nature imaginary data; another study. Lastly [10] found the disease detection method using the algorithm with image sensory. Found the method in CNN-based algorithm with image sensors.

4.3 Weed Detection

Weed is unwanted leaves grown, and it will take all the nutrients from the crops and plant; removing the weed is a major problem in agriculture; the first study of weed deduction in machine learning is found by Pantazi et al. [11] based on counter-based image sensor captured with drones (UAS) identify the *Carduus marianus,* and those were removed another study by Pantazi et al. [12] he categorized weed and main crop perfectly to be accurate in weed detection; lastly, the research reviews [13] weed detection by SVN in the plain crop.

4.4 Water Management

Water plays a vital role in agriculture management; it is difficult by a traditional method due to accuracy in prediction in climatic changes also balance in agronomical; the paper explores in evaporate transpiration (it is a period between the absorption of water by land to atmosphere and by evaporating it other surface plants) of the daily and monthly forecast; this system will help farmers to manage and store the water and plant the field accordingly, Mehdizadeh et al. [14] in his research he found that dried and semi dried land evaporates faster and causes a climactic changes ,to read the data of those climactic change he set up the weather forecast station and collected the data.

4.5 Soil Management

The soil is an important source for the continued existence on earth; the layer of the soil is used for planting, and other layers are used in soil fertility, microbes, etc., to predict climate change; soil temperature alone will play a major role in prediction. Firstly [15], the aim of the study is to find the soil dryness with data and evapotranspiration; the purpose of this study is to provide information for remote decision-making; another study done by Morellos et al. [16], he developed a self-evolution method which is called as SAE_ELM where the soil is tested with six different depth from 5 to 100 cm; the aim of the study is accuracy in soil management.

In agriculture AI, there are different model developed for managing farms (i) crops, (ii) water, (iii) soil management. Tables 2, 3, and 4 explain various models of AI in agriculture.

Table 2 Machine learning method for crop management

References	Crops	Type	Functionality	Method used	Algorithm
Crop management					
Amatya et al. [7]	Fruit picking and yield management	coffee	Categorized coffee fruit into ready for harvest, not fit for harvest lastly fruits which are matured by overseen for harvesting	Colored digital imagine	SVM
Sengupta and Lee [9]		Tomatoes	Segregated between green red and yellow tomatoes and picked which is ready for harvest	RGB image	Clustering/EM
Senthilnath		citrus fruit	Identify the number of ingrown in outside condition and grown in green gas	Digital image	ENG
[7]		cherry	identify fruit with foliage	RGB image	SVM
Pantazi	Crop	Wheat	yield predication	Soil parameter and satellite imagine sensor	ANN
zDiseases detection					
Pantazi	Crops	Wheat	Detecting yellow rust and nitrogen stressed	hyper Spectral reflectance and imaginary data	ANN
Moshou		Rice	Detecting fungus diseases like bakanae in rice seedling	Morphing and colure traits	SVM
Moshou et al.		Wheat	Detecting yellow rust and water stressed	Spectral reflectance	SVM/LS

(continued)

Table 2 (continued)

References	Crops	Type	Functionality	Method used	Algorithm
Ebrahimi et l. [17]	Fruit	Strawberry	Detecting patricides and thrips	Reign index	SVM
Weed detection					
Pantazi et al. [11]	weed	Weed	Detection of *Silybum marianum*	Spectral reflectance	ANN/CP
Pantazi et al. [12]		Weed	Diffracting weed 56% accurate and maize 100% accuracy rate	Hyper-spectral reflectance and imaginary data	ANN/SOG/MOG
Binch and Fox [13]		Weed	Grass versus weed	imaginary data	SVN

Table 3 Machine learning method for water management

References	Type	Functionality	Algorithm
Mehdizadeh et al. [14]	Evapotranspiration	Done with monthly mean with the arid and semi-arid region.	MARS
Feng et al. [18]	Evapotranspiration	Done daily, weekly, and monthly interconnected to weather forecast station (6) and trained and tested with trained data	ANN/GRN
Patil and Dika	Evapotranspiration	The same test was conducted with connecting with two weather forecasting stations, and data were tested	ANN/ELM

Table 4 Explains the machine learning method for soil management

References	Type	Functionality	Algorithm
Coopersmith et al. [15]	Soil drying	The study is to find the soil dryness with data and evapotranspiration; the purpose of this study is to provide information for remote decision	IBM/KNN
Morellos et al. [16]	Soil conditioning	Testing soil organic carbon, nitrogen, and moisture content	SVM/LS-SVM
Nahvi et al. [19]	Temperature	Where soil is tested with six different depth from 5 to 100 cm	ANN/SAE

4.6 Applicability of IoT and AI/ML in SSA practice

Smart sustainable agriculture [SSA] is becoming a need of the hour. The present-day pollution due to the chemical-based production of crops is spoiling the environment. Not only the environment but also the health of human beings and the formal community is at risk. If this scenario continues in the existing period, the chances of the slowdown of the economy will upsurge. This is because to satisfy the need of the ever-increasing population; the government needs to import the food product from other countries of the world.

Fortunately, the solution to this intense problem is known to the scientist. The applicability of modern technology to the field of agriculture can reduce the issue of pollution due to cultivation practice and will also lead to the availability of abundant food crops for the consumption of people. Artificial intelligence [AI] and machine learning [ML] along with the Internet of Things [IOT] can prove beneficial in reducing the adverse agriculture impact on the environment and people. Figure 2 explains the model for sustainable agriculture by using AI, adopting the AI IOT model and implementing in the field of agriculture will increase the yield; Usage of IoT in agricultural field will reduce the labor work and labor cost requirement in the field.

The above presentation of systematic layers is the essential component of SSA. The layers are described as follows:

1. Physical hardware and storage layer: This layer compromises of strong hardware required to store the virtual technology
2. AI and data management layer [DM]: This layer deals with various processes including in the segment of collection, from the SSA clouds.
3. Network layer: Networking between various components serves to fulfill the actual demand of the architecture. The networking panel uses the recent technologies to establish the connection of the flow of data from one aspect of the SSA model to other include Wi-Fi, Internet, and GSM/CDMA.
4. Security layers: AL algorithms are hindered in many cases by the security of the data inputs. The presentation layer will enable its functional capabilities in managing the dangers arising from malware, viruses, data theft related to all the SSA layers. In case of the absence of this layer, the entire process can give either defective conclusions or may altogether be unable to generate one.
5. Application layer: This comprises full or partial monitoring of the data flow and the implication of all the layers to the ultimate SSA layer. The data flow here is monitored using the integrated mechanism in agriculture applications including monitoring of cropping culture using a drone, disaster management, geographical area spread, potential acquaintance of area under cropping, and many more.
6. Internet of Things [IOT] and sensing layers: This is the most crucial layer applicable to the collection of data related to agricultural practices. It is using the sensor required to track the element directly or indirectly controlling the

agriculture parameters such as sensor sensing humidity, moisture, temperature, and biological elements.

7. Domain SSA layer: The base layer around which all the layers are rotating. Various agriculture domains are included in this layer which has to be modified with the help of other layers.

The growth of agriculture products in a sustainable manner is the need of the hour. Today all the possible experiments are carried out to come up with smart sustainable agricultural practices so that with putting pressure on existing natural resources and degrading the environment the cultivation period can be reduced along with the improvement in quantity, quality, volume, and variety of crops available for mass consumption. This can be achieved only with an integrated approach where IOT and AL/ML and be combined with SSA practices.

5 Discussion

General discussion—total number of the article involved in the study is 40, the search thread covered the partial crop management and water management and soil management in the popular method used between period 2015–2021; all the methods are briefly discussed in tabulation format; it is slightly unlike to cover the entire paper; the review research is value to threat and validity [20].

Specific discussion—search strategy is thin downed with the relevant scope of the research; a manual search is done with using the word-like machine learning and yield prediction with Google scholar engine, Scopus, and Web of sciences exclusion category was made like publication in another language, article which is not available open-source, publication before the period 2014; all those which comes before 2014 used for common algorithm terms for machine learning.

The study discussed the sustainable agriculture model by introducing IoT (Internet of Things) which makes the study novel in nature, besides the study, reviews machine learning models until 2021 which fill the review gap of past literature.

6 Conclusion

There are many algorithm models are used; the paper summarizes eight models in which crop management has SVM, EM, least square regression, cluster, SOG/MOG, SVN for soil management models like IBM/KNN, SVM/LS-SVM, and ANN/SAE are used; finally, for water management, MARS, ANN/GRN ANN/ELM is reviewed.

Machine learning is used in agriculture sector mostly in crop management prediction in which majority study on crop yield prediction and diseases detections. Machine learning has shown a vast development in the field of agriculture sector like

fruit picking, crop analysis from drones, and satellites with ninety-five percent accuracy weed eradication with digital imaginary camera and spraying the pesticides in the weed alone, real-time weather forecasting like seeing temperature, rainfall, and humidity with solar radiation; the prominent work ML can do to agriculture is autonomous tractors intelligent machines, decision-making with algorithms remote monitoring; the study shows that there is frequent usage of ANN, SVM for crop management as well as soil and water management, using machine learning in agriculture can develop the country as a whole for future study; the government scholar and experts should spread the usage of machine learning to real-time farmers and normalizing the machine learning like traditional study will save the farmers as well the impact of lands.

References

1. Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, ... and Lander ES (2017) Systematic dissection of genomic features determining transcription factor binding and enhancer function. Proc Nat Acad Sci 114(7):E1291–E1300
2. Dippé MA, Wold EH (1985, July) Antialiasing through stochastic sampling. In: Proceedings of the 12th annual conference on computer graphics and interactive techniques (pp 69–78)
3. Russell SJ, Norvig P (1995) Artificial intelligence: a modern approach, vol 9. Prentice Hall, Upper Saddle River. ISBN 9780131038059
4. Hecht-Nielsen R (1987) Counter propagation networks. Appl Opt 26:4979–4983
5. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
6. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge, MA, pp 216–261
7. Amatya S, Karkee M, Gongal A, Zhang Q, Whiting MD (2015) Detection of cherry tree branches with fullfoliage in planar architecture for automated sweet-cherry harvesting. Biosyst Eng 146:3–15
8. Ali I, Cawkwell F, Dwyer E, Green S (2016) Modeling managed grassland biomass estimation by using multi-temporal remote sensing data—a machine learning approach. IEEE J Sel Top Appl Earth Obs Remote Sens 10:3254–3264
9. Sengupta S, Lee WS (2014) Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. Biosyst Eng 117:51–61
10. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Comput Electron Agric 145:311–318
11. Pantazi XE, Moshou D, Oberti R, West J, Mouazen AM, Bochtis D (2017) Detection of biotic and abiotic stresses in crops by using hierarchical self organizing classifiers. Precision Agric 18(3):383–393
12. Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM (2016) Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric 1(121):57–65
13. Binch A, Fox CW (2017) Controlled comparison of machine vision algorithms for Rumex and Urtica detection ingrassland. Comput Electron Agric 140:123–138
14. Mehdizadeh S, Behmanesh J, Khalili K (2017) Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. Comput Electron Agric 139:103–114
15. Coopersmith EJ, Minsker BS, Wenzel CE, Gilmore BJ (2014) Machine learning assessments of soil drying for agricultural planning. Comput Electron Agric 104:93–104

16. Morellos A, Pantazi X-E, Moshou D, Alexandridis T, Whetton R, Tziotzios G, Wiebensohn J, Bill R, Mouazen AM (2016) Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst Eng 152:104–116
17. Ebrahimi MA, Khoshtaghaza MH, Minaei S, Jamshidi B (2017) Vision-based pest detection based on SVM classification method. Comput Electron Agric
18. Feng Y, Peng Y, Cui N, Gong D, Zhang K (2017) Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. Comput Electron Agric
19. Nahvi B, Habibi J, Mohammadi K, Shamshirband S, Al Razgan OS (2016) Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. Comput Electron Agric 124:150–160
20. Šmite D, Wohlin C, Gorschek T, Feldt R (2010) Empirical evidence in global software engineering: a systematic review. Empirical Softw Eng 15(1):91–118
21. Pantazi XE, Moshou D, Bravo C (2016) Active learning system for weed species recognition based on hyperspectral sensing. Biosyst Eng 146:193–202
22. Robertson AG, Shih J, Yau C, Gibb EA, Oba J, Mungall KL, ... and Thiessen N (2017) Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. Cancer Cell 32(2):204–220

Analyzing Data in Decision Making for Educational Universities Using Machine Learning

G. Mohana Priya, M. Latha, and K. S. Archana

Abstract In an educational field, the application of machine learning algorithms is currently the main focus for researchers and scientists. This document focuses on describing the usability and advancement of machine learning for increasing the accuracy of decision making for educational universities; popular machine learning methods are analyzed. The concert of each machine learning algorithms is assessed in requisites of prediction of time, accuracy, and the same has been documented. The various machine learning algorithms used in educational universities have been developed and tested in many regions of developed countries. Such tools are an immense boon for developing countries which can be leveraged even across high school level since assimilation of data is difficult due to the availability of limited resources. The rationale of this study is to establish the prediction of the rate of graduation of undergraduate programs in the university. The management of educational universities plays a vital role in making most pronouncements that have collision on the strategic level and the student, professors, graduates, and the whole institution's academic society. In the process of decision making, the directors depend on educative tools to support their various missions and many barriers including rigid government structure changes in rules, technological conditions. This may hinder the support required by university directors and managers. Therefore, this document seeks an outline of the popular usability of the machine learning algorithms in the educational university with the focus on techniques used for graduation rate prediction.

Keywords Supervised algorithms · Strategic decision · Machine learning · Decision making

G. Mohana Priya (✉) · M. Latha · K. S. Archana
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

M. Latha
e-mail: latha.se@velsuniv.ac.in

K. S. Archana
e-mail: archana.se@velsuniv.ac.inssss

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_43

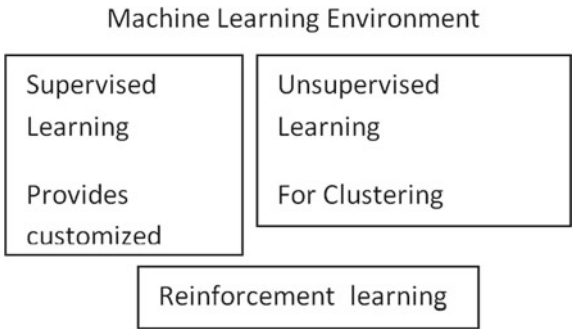


Fig. 1 Machine learning environment

1 Introduction

The decision-making process in education universities long has been based on traditional ways. Manually stored data does not suffice decision making for managers. The usage of a machine learning algorithm is crucial to improvise this development. The goals of machine learning algorithms are to program machines to analyze given information or experience to resolve a precise problem. Machine learning techniques are used in some areas such as natural language processing, pattern detection, educational field, compute visualization, bioinformatics, etc. Daily education is changing; the application of machine learning is worn to make the process easy.

The specific use by machine learning in the educational university is to support management by predicting student learning performance, student dropout, test student, course allocation, faculty allocation for a student, and choice of major student outcomes, etc. The learning process of students in education software is required to be very precise. Educational software is necessary to suffer a testing procedure previous to implementation. The program is validated with attributes of the system and to meet the goals. Metrics and some of the specifications of the program organize flow graphs, execution data graphs, failure reports, and elements of learning which are crucial requirements of the software. The intelligent learning environment (Fig. 1) uses collectively supervised learning and unsupervised learning.

This paper describes how the most suitable learning algorithm shall be chosen with an objective and investigation of past data.

2 Literature Review

This paper [1] exploits academic decisions for the rate of graduation from five undergraduate engineering programs, using two algorithms fit for academic decision making, viz. SVM and ANN through the confusion matrix.

Categorization detain has been proposed for envisage student's performance in blended learning based on machine learning. The SVM and RF classification algorithms through tenfold cross-validation are designed for categorization purposes, and LMS modules are used to observe student's performance and classify their status as pass or fail [2].

This paper [3] refers to the prediction of the rate of graduation to facilitate that will help decision essentials at the strategic level by use of Machine Learning algorithms. The student's academic performance is calculated with the help of analysis of numeral of graduated students and not graduated students. Learning methods are worn to procedure the examination such as Logistic Regression, Random Forests, and Decision Trees.

Ankita Kanojiyaand Dr. Viral Nagori: This paper [4] explains the literature survey helps in result the benefits of Decision Support Systems (DSS) in different features of edification. The comparative learning of selective four papers taken has consideration for the following research work to address.

Huma Samin and Tayyaba Azim: This paper [5] provides university through suitable course allocation and supervisor assignment. The two popular probabilistic topic models used in this paper are Latent Dirichlet Allocation and Author Topic Model. The perplexity (standard evaluation metric) is used to measure the generative performance of LDA and ATM.

Wen Cao et al. [6] suggests a CJDE algorithm hazard toward solve (JDE) harms. These dynamic problems use the RCJDE algorithm. The MJDE is a kind of dynamic problem and a modified recursive CJDE risk.

J. Stimpson et al. [7] discussed the usage of Logistic Regression and Linear Regression algorithms for small course datasets; this had an analogous performance with intricate algorithms (SVM, ANN). Regression and classification algorithms are useful to the records set to envisage the final course score.

Kiel F. Harell: In this paper [8] seeks the study of Foxfire (professional development retreat) for classes for teachers educational for researchers and practitioners dedicated to speculative democracy in teacher edification. The decision-making process is to support democratic learning and administering participants' awareness of problems. The Foxfire courses for teachers with chronicles of numerous iterations of the course.

Ioanna [9] predicted a quitter in e-learning programs. The accepted machine learning methods are feed-forward neural networks (FFNN), probabilistic ensemble simplified fuzzy ARTMAP (PESFAM), and SVM using three diverse decision schemes (overall accuracy, sensitivity, and precision).

Asil Oztekin [10] discussed an assessment method for e-learning systems using the machine learning method. Machine learning methods are SVM, Decision Trees, and neural networks along with various linear regressions that are used for assessment.

Jae Young Chung and Sunbok Lee: In this paper [11], the Random Forest model is used to envisage students' dropouts by Decision Trees. Machine learning is applied to patterns extracted from data as part of predictive modeling importance

on the conclusion of a model. The Fandom Forest model is used to predict students' dropout (Table 1).

Table 1 Literature survey of distinct papers

Paper	Method used	Result	Advantage	Future work
[1]	The two algorithms fit for the academic decision making are SVM and ANN using the confusion matrix	The evaluation among the algorithms is detained via the confusion matrix and the ROC curve	Graduation rates	Development of relevance software that makes easy the search, correlation, repossession of data, and visualization essential to craft educational decisions for university higher officials
[2]	RFs, SVMs used	The projected system uses RF and SVM via tenfold cross-validation	Monitor students learning recital	Estimate the period by following the time stamp of the customer and logout data for every occasion to calculate the event period in all the Moodle events
[3]	Decision Trees, Logistic Regressions and Random Forests wherever RF achieves the best outcomes	The 70% training and 30% validation and to process the test machine learning algorithms were used	Graduation rates	Various characteristics used and its record that are normalized can be added
[4]	None	None	Literature survey of DSS	The scope of learning can be measured as an assessment of study work that has been tackled in different areas of the edification domain such as admission, class arrangement, choosing, suitable pedagogy, administration, and others

(continued)

Table 1 (continued)

Paper	Method used	Result	Advantage	Future work
[5]	The two popular probabilistic topic models worn in this paper are Latent Dirichlet allocation and Author Topic Model	By evaluating the confusion of LDA and ATM with a sample data. The innovative control of LDA is greatly enhanced than ATM	Course and faculty allocation for a student in the specific department	The paper aims to extrapolate this system for the staff of other programs based on the current concept registered for computer science staff
[6]	RCJDE algorithm	RCJDE does better than the two-level approaches toward requisites of JPM and becomes an extremely easier calculation than RJDE	To solve JDE problems	Dynamic regularities, an enhanced algorithm should envisage this prospect strength enhanced in terms of various distance dimensions
[7]	Linear regression algorithm and the Logistic Regression algorithm are used in support of the diminutive classes database had an analogous performance with more intricate algorithms (SVMs and ANNs)	Linear regression used in support of regression scrutiny and Logistic Regression used in support of class recommended results	Prediction of final course score	Further research of this could investigate the mentioned details of transforming prediction and features support toward bigger about additional complicated datasets effects under specific attack progresses in the understanding of domain specialists should be additional exploration
[8]	The chronicle of multiple iterations of the teacher education classes	In general democratic teacher education, the scrutiny of three case revises (3 weeks) of Foxfire course for teachers is too inadequate to achieve saturation level	The decision-making process is to support democratic learning and directing member's notice to problems	Chronicle on supplementary recursive of the classes would give to a superior perceptive of the Foxfire course for teachers

(continued)

Table 1 (continued)

Paper	Method used	Result	Advantage	Future work
[9]	FFNN, PESFAM, and SVM	Achieved 75% to 85% in general student categorization rate from the initial segment of the two programs	Student's dropout in e-learning courses	The planned dropout prognostication technique on additional kinds of studies further distance learning, e-learning, compounded learning, and classical education is all investigated. Performing more reliable outcomes utilizing various learners' attributes, also methods for dropout prediction, and to investigate retention strategies of educational institutions
[10]	Three machine learning techniques SVM, all along with various linear regressions are worn	The tentative consequences showed that the overall usability based on the metric severity index	Assessment for e-learning systems	The application of partial least squares or structural equation modeling techniques can be incorporated to endow with a learner cause-and-effect association among load and return variables
[11]	The Random Forest (RF) model used to predict Students' dropout. Dataset of 3–10-fold cross-validation and confusion matrix	Dataset of high school (165, 715) is separated keen on 80% training (data = 132572) And 20% testing (data = 33143) dataset to evolve the dropout. The receiver operating characteristic curve is used for trained RF	To predict high school student's dropout	By using machine learning it continues to develop an early dropout warning system. The predictive model aids classroom teachers to take away the risks through suitable interventions and held to enhance the protecting factors around the students

3 Results and Discussions

The prevalent algorithms used in the educational areas from the facts percentage are SVMs 21%, ANNs 26%, DT (Decision Tree) 16%, Logistic Regression 16%, RF (Random Forest) 16% and linear regression 5% on related works. Also, it is experiential that current studies are not paying attention to higher officials as the majority of the chronicle was inadequate to the learner's career and teacher's profession. The administrator from universities which contain scrupulous vision outcomes requirements and comprise no done addressed earlier than. Important pronouncements are sustained when the correct data is specified to the higher authorities of the universities would be exposed.

4 Conclusion

The effectiveness to analyze data at educational universities using machine learning is still materializing. The machine learning algorithm of various sets of data and various sets of prediction is used. In academic decision making, different algorithms SVMs, RFs, Decision Trees, regressions, neural networks are used. In this research, these techniques are extremely helpful to predict the rate of graduation, student learning performance, faculty allocation, student dropouts, and choice of a major subject in educational universities.

References

1. Y Nieto V García-Díaz C Montenegro RG Crespo 2019 Supporting academic decision making at higher educational institutions using machine learning-based algorithms Soft Comput 23 12 4145 4153
2. Nespereira CG, Elhariri E, El-Bendary N, Vilas AF, Redondo R P D (2016) Machine learning based classification approach for predicting students performance in blended learning. In: The 1st international conference on advanced intelligent system and informatics (AISI2015), 28–30 Nov 2015, Beni Suef, Egypt. Springer, Cham, pp 47–56
3. Nieto Y, Gacía-Díaz V, Montenegro C, González CC, Crespo RG (2019) Usage of machine learning for strategic decision making at higher educational institutions. IEEE Access
4. Kanojiya A, Nagori V (2016). Analysis of Decision Support Systems implemented for Suggesting a course for higher education. In: Proceedings of the second international conference on information and communication technology for competitive strategies. ACM, p 107
5. H Samin T Azim 2019 Knowledge based recommender system for academia using machine learning: a case study on higher education landscape of Pakistan IEEE Access 7 67081 67093
6. W Cao J Lan XR Li 2015 Conditional joint decision and estimation with application to joint tracking and classification IEEE Trans Syst Man Cybern Syst 46 4 459 471
7. AJ Stimpson ML Cummings 2014 Assessing intervention timing in computer-based education using machine learning algorithms IEEE Access 2 78 87
8. KF Harell 2019 Deliberative decision-making in teacher education Teach Teach Educ 77 299 308

9. I Lykourentzou I Giannoukos V Nikolopoulos G Mpardis V Loumos 2009 Dropout prediction in e-learning courses through the combination of machine learning techniques Comput Educ 53 3 950 965
10. A Oztekin D Delen A Turkyilmaz S Zaim 2013 A machine learning-based usability evaluation method for eLearning systems Decis Support Syst 56 63 73
11. JY Chung S Lee 2019 Dropout early warning systems for high school students using machine learning Child Youth Serv Rev 96 346 353

Mediating Effect of Impulsive Buying Tendency Between Shopping Enjoyment and Online Impulsive Buying Behavior

M. Ruby Evangelin and S. Vasantha

Abstract Online impulsive buying has drawn greater attention among consumers during this electronic era. The current study aims to showcase the shopping enjoyment that would drive the consumers' stimulus in impulsive buying behavior, and this study proves that the mediating factor (impulsive buying tendency) influences the consumer in online impulsive buying. The theory of social capital theory was taken for the conceptual model. The tenacity of this research is to deliver comprehensive knowledge of online impulsive buying. The research content was examined from various research articles and used for writing this research article, to find the various factors influencing online impulsive buying. This research paper will be of great help to the nurturing marketing professionals and researchers toward the vast dimensions of purchaser's impulsive buying behavior.

Keywords Buyer's behavior · Buying tendency · Online impulsive buying · Shopping enjoyment

1 Introduction

The marketers have found a new way to attract the customer. They have tried to understand the customer taste and perceptions of varied products, brands, colors, etc.… and to what extent do these factors influence them to buy the products. Nowadays, the Internet plays a vital role among the customers who does online shopping. The reason for the exploration is to clarify the relationship among the shopping enjoyment which encourages the clients to make impulsive buying. Online impulsive buying is challenging for the shopping websites which stirs up the emotions and moods of the customer to make an immediate purchase [1]. Due to the surge, the retailers attempt

M. Ruby Evangelin · S. Vasantha (✉)
School of Management Studies, Vels Institute of Science Technology and Advanced Studies, Chennai, India
e-mail: vasantha.sms@velsuniv.ac.in

M. Ruby Evangelin
e-mail: rubyevangelin.sms@velsuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_44

to increase such purchases to all the customers. Therefore, the shopping enjoyment of the customers is imperative for this study.

Verhagen and Dolen [2] exposed that 60% of the customer does online shopping, whereas 40% of the customers with their enjoyment for shopping are intended to do online shopping. Due to this change in the scenario, online shopping has given more importance to impulsive buying behavior. This shopping enjoyment acts as a trigger for the customer to go for a repeated impulsive buying [3].

The impact of the study is to apply the concept of flow and social capital theory for the study on online impulse buying, including the craving to buy, shopping enjoyment, and impulsive buying. According to Li-Ting Huang (2015), the author explored about the flow and social capital theory uses the urge of buying as an independent variable, impulsive buying tendency as the control variable. Among these factors that urge buying are shopping enjoyment and impulsive buying tendency as the mediating factor.

The concept of flow and social capital theory gives the insight of social interactions among the consumers which stimulates them to make a purchase using the social media. The concept has helped the other researchers to analyze the buying behavior of the consumers and enable them to make a unplanned purchase depending upon their positive mood; in return, the consumers are emotionally stimulated to make an unplanned purchase (Song 2015). The research discussed here attempts to fill the gap in this area of impulsive buying; there are enormous literature of offline impulsive buying, but however there are only few facts on online impulsive buying which are remained unexplored.

To abridge, our study explores the effects of shopping enjoyment that leads to online impulsive buying and also finds the contribution of impulsive buying tendency as the mediating factor for the study.

2 Research Framework and Hypothesis Development

See Fig. 1.

The study's main focus is to probe into the cognition reaction that leads to impulsive buying and the mediating factors (shopping enjoyment) which enhances the relationship between the cognitive reaction and online impulsive buying.

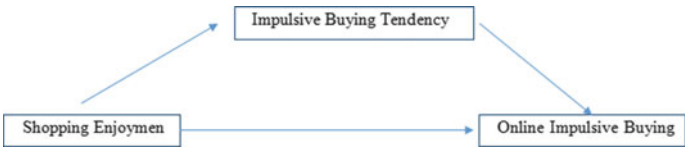


Fig. 1 Conceptual framework

Shopping enjoyment is the most vital concept in the research toward customer perception and marketing research. Shopping enjoyment speaks the person's character, quality, connecting the shopping behavior with the incredible delight and charming viewpoints. Shopping motivation which is associated with shopping enjoyment has a vital role in consumer behavior for the past decades (Zanna and Rempel 1988).

3 Shopping Enjoyment and Online Impulsive Buying

Shopping enjoyment kindles the tendency of impulsive buying tendency which in turn will influence the customer to adapt to online impulsive buying. Shopping enjoyment also influences the customer's attitude, beliefs, and behavioral motives for online shopping. Moreover, shopping enjoyment also acts as shopping motives.

As indicated by Goyal (2007), the concept of shopping enjoyment toward online shopping is been considered as a buyer's willingness to buy the product that addresses the inclination to discover shopping's more awesome to experience more unmistakable shopping delight than others.

Shopping enjoyment is the factor instigating the particular differences, where customers believe procuring to be a creator of fun, where the customers do not follow a particular routine in their own purchasing list, and likewise, it will also instigate the customers buying behavior to make a second purchase (Sharma and Sivakumaran 2004).

One full of feeling reaction attached to store ascribes is shopping enjoyment. Shopping enjoyment is the degree to which a shopping involvement in an online portal is seen to give protection in its own right (Childers et al. 2001).

Shopping enjoyment has gained importance in retail studies and is an important motivation in explaining several consumer attitudes, such as cognitive attitude and channel usage intentions (Martos-Partal and Gonzàlez-Benito 2013; Frasquet et al. 2015).

During online shopping, customers would scroll for the wants and needs, and the customer would greatly experience shopping enjoyment during the search for the desired product due to the motivation which would raise the customer's enjoyment factor to dominate and end up in doing an online impulsive buying (Holbrook and Corfman 1985).

The previous study emphasizes that shopping enjoyment influences the impact of the customer's behavior (Nicholls et al. 2000). Due to this factor, the customer will be induced to make a repurchase or repatronage intention. This shopping enjoyment will increase the customer's urge to frequently visit the online shopping portals and create a positive impact on the online shopping portals and lead to online impulsive buying. (Saad and Metawie 2015).

Shopping enjoyment plays a vital role in consumer behavior, so it is vital to find antecedent factors for shopping enjoyment. The antecedents of shopping behavior are numerous. Due to a lack of researchers toward the shopping enjoyments, further

study can be made over the factors that induce the customers to enable shopping enjoyment with themselves.

It has been proved from the previous research work that the person who spends more time shopping using the online shopping portals will tend to spend more money on doing online impulsive buying [4].

The shopping enjoyment's characteristics are related to the engagement of the online portals through the customer's behavior such as the exposure of the advertisement, which would lead the customers to have shopping enjoyment which indeed will lead to online impulsive buying.

Shopping enjoyment is a factor which would kindle to start the motivation conduct of customer's buying behavior. The joy of shopping on the web is a kind of delight for the clients who just do not cling to an equivalent routine shopping list, however impact the client to continue to the online indiscreet purchasing (Sharma and Sivakumaran 2004).

Shopping happiness is characterized as the delight one gets in the shopping process. The customers who get engaged in shopping enjoyment get the tendency to shop unplanned products and get great excitement psychological rewards for their shopping behavior process [5].

Notwithstanding the effect of character characteristics and culture, the connection between shopping satisfaction and online imprudent purchasing merits due thought. Goyal and Mittal [6] checked shopping experience impacts the joy as the buyer's benchmark practice to discover shopping seriously satisfying and to encounter more striking shopping fulfillment than others.

Therefore, the characterization features to some extent that shopping enjoyment might have alluded as a particular inward suffering propensity of the person to appreciate the shopping cycle. The path that for specific individuals enjoyment begins from shopping estimates characteristically inside themselves because of their shopping behavior (Jung and Lim 2006; Bong Soeseno 2010); it would be beneficial to think of it as a significant factor.

Buyers also companion shopping with enjoyment and enthusiasm which they would experience as a pleasurable action to them nevertheless of online or offline shopping. Shopping enjoyment is portrayed as how much web shopping is in a little while flawless and entertaining [7].

According Broekhuizen and Huizingh [8], the online shopping causes enjoyment in shopping which would be fun-filled and lively in its place of shopping task fulfillment; consumers prefer buying their merchandise through online shopping portals as perceived enjoyment, and the outward diversion will be viewed as enjoyment to them.

Thong et al. (2006) suggested that shopping enjoyment has a significant consequence of online shopping by using online shopping portals. If considered offline shopping, online shopping creates an entertaining moment for the customers to shop through online shopping portals; these intentions will influence the customer to purchase their desired products by their unplanned buying behavior. The main purpose of the study is to understand the concept of shopping enjoyment and its relationship on the online impulsive buying. The shopping enjoyment results in the

positive mood to make an online purchase, which brings a greater happiness and satisfaction among the consumers.

4 Impulsive Buying Tendency Toward Online Impulsive Buying

The occurrences of impulsive buying tendencies were considered as unreasonable behavior among the consumers in 1940. Online impulsive buying behavior can be picturized as the behavior of the customer who tends to have the impulsive buying tendency nature to buy the products regardless of differences of culture, region, and other factors.

On the whole, online impulsive buying happens from the two key perspectives: state of mind by the shopping enjoyment or the individual behavior that influences the consumer to adapt to the online impulsive buying.

Chich et al. (2012) examined that impulsive buying tendency is entirely different among the individual's buying perception toward online shopping, and the researcher concluded that impulsiveness can be used as the substitute for the impulsive buying tendency that is influenced by the shopping enjoyment leading toward the online impulsive buying.

Impulsiveness is the attitude of the customer who is probably going to make unintended, spontaneous, and unreflective buys (Kim and Lennon 2000). The previous study illustrates that the impulsive buying tendency is consistent across the different product categories. Psychologists believe that individuals vary in their tendency to act impulsively while shopping using online portals (Beatley and Ferell 1998).

Impulsive buying tendency will impact the purchaser's rash conduct to be generally steady and liable for the easygoing impact on the hasty purchasing utilizing the online portals [9].

Online impulsive buying and purchasing incautiously are used alternately. It is misleading for a couple of researchers to use the extent of indiscreet purchasing propensity as a substitute for imprudent purchasing conduct. Rash purchasing propensity is not equivalent to hasty purchasing conduct, since this indiscreet purchasing inclination impacts purchaser conduct which will encourage the customer to do the imprudent shopping toward the end (Zhang et al. 2007).

As indicated by Gottwald and Weinberg [10], impulsive buying tendency happens with the passionate clash between planned and unplanned purchases. According to Rook [11] referenced that there are different emotional inspirations for impulse buying that individuals may insight into shopping. A visual experience with an item or a special stimulus can persuade the unexpected liking to purchase. In any case, impulse buying conduct does not generally come from direct visual support. Once in a while, with no reason or stimulation, individuals are abruptly propelled to shop.

Impulse buying tendency has been conceptualized as an individual attribute (Dholakia 2000) which is been highly influenced by environmental factors. On

the other hand, Stern (1962) then again recognized categories of impulse buying tendency: unadulterated impulsive buying tendency thoroughly breakdowns the typical purchasing behavior. It occurs right when the customer has no purchase objective aside from the thing moves sentiments that in the end lead to the exhibition of buying; update rash purchasing inclination happens when the purchaser sees a thing and keeps in mind that the stock at home is depleted, or audits a business or other information about the thing and a previous wish to get it; recommended indiscreet purchasing propensity happens when the buyer sees a thing out of the blue and perceives a need that it can satisfy; and arranged hasty purchasing inclination happens when the buyer enters the store with the assumption to get some specific things yet furthermore desires to make diverse purchases dependent upon the unprecedented offers and progressions that the individual being referred to finds at the store.

Greenfield (1999) and LaRose et al. [12] contend that the Internet shopping gateways can ask the buyer to make more drive purchasing than the offline shopping experience: the more prominent item range, the simplicity of making buys every minute of every day from any area, and the utilization of advanced showcasing methods, for example, personalization of the items as indicated by the craving of the clients, these components can urge the client to make impulsive buying through the web-based shopping entrances. However, not many factors, for example, late conveyance or the transportation costs would dishearten the clients to pull out themselves from doing web-based shopping. This study of impulsive buying tendency affects the behavior of the consumer and helps the consumer to make an unplanned purchase. The literatures states that the impulsive buying tendency is high compared during the online purchase than the offline purchase.

5 Conclusions

The research paper elucidates the result that the authors have expanded the collection of information as for seeing how and why shoppers with fluctuating levels of impulsive buying tendency toward the online unplanned purchase. In conclusion, unplanned buying behavior plays a vital role to increase the profit of the sellers and gives shopping enjoyment while shopping through the online portals. Psychological factors, for example, character attributes, full of feeling states, interest, item contribution, observation, standardizing assessment, and other intellectual components, assume pivotal and direct functions in influencing indiscreet purchasing conduct. The impulsive buying tendency of individuals is unreflective so they buy cannot be made without deciding on the item. Individuals have a propensity for zeroing in on moment fulfillment as opposed to settling previous inconveniences which prompt moment buys.

References

1. Beatty J (2003) On the product nature on impulse buying tendency. J Bus Res 56(7):505–511
2. Verhagen T, van Dolen W (2011) The influence of online store beliefs on consumer online impulse buying: a model and empirical application. Inf Manage 48(8):320–327
3. Xiao S, Nicholson M (2013) A multidisciplinary cognitive behavioural framework of impulse buying: a systematic review of the literature. Int J Manage Rev
4. Verhoef P, Lemon K, Parasuraman AP, Roggeveen A, Tsiros M, Schlesinger L (2009) Customer experience creation: determinants, dynamics and management strategies. J Retail 85:31–41
5. Bellenger DN, Korgaonkar PK (1980) Profile the recreational shopper. J Ret 56(3):77–92
6. Goyal BB, Mittal A (2007) Gender influence on shopping enjoyment: an empirical study. Ind Manage Stud J 11(2):103–116
7. Chiu CM, Chang CC, Cheng HL, Fang YH (2009) Determinants of customer repurchase intention in online shopping. Online Inf Rev 33(4):761–784
8. Broekhuizen T, Huizingh E (2009) Online purchase determinants: is their effect moderated by direct experience? Manage Res News 32:440–457
9. Zhang Y, Rossow W, Stackhouse P (2006) Comparison of different global information sources used in surface radiative flux calculation: radiative properties of the surface. J Geophysi Res
10. Weinberg P, Gottwald W (1982) Impulsive consumer buying as a result of emotions. J Bus Res 10:43–57
11. Rook DW (1987) The buying impulse. J Consumer Res 14(2):189–199. https://EconPapers.repec.org/RePEc:oup:jconrs:v:14:y:1987:i:2:p:189-99
12. Larose R, Mastro D, Eastin M (2001) Understanding internet usage: a social-cognitive approach to uses and gratifications. Soc Sci Comput Rev 19
13. Bong S (2016) The influence of impulse buying toward consumer store loyalty at Hypermarket In Jakarta. Bus Entrepreneur Rev 10:25
14. Huan (2015) Flow and social capital theory on online impulse buying. J Bus Res
15. Kim J, Lennon S (2013) Effects of reputation and website quality on online consumers' emotion, perceived risk, and purchase intention: Based on the stimulus-organism-response model. J Res Interact Market 7
16. Morris B, Holbrook (1999) Introduction, in consumer value: a framework for analysis and research. In Morris B (ed) Holbrook. Routledge, New York, pp 1–28
17. Greenfield PM (2018) Tudying social change, culture, and human development: a theoretical framework and methodological guidelines. Develop Rev
18. Sharma B (2003) Impulse buying and variety seeking: two faces of the same coin? Or, maybe not!. North American Association for Consumer Research (ACR) Conference at Toronto, vol 31
19. Thong H (2011) The effects of post-adoption beliefs for the expectation-confirmation model for information technology continuance. Int J Hum Comput Stud 64(9):799–810
20. Wang C-J (2006) "The relationship between service quality and customer satisfaction" the example of CJCU library. J Inf Opt Sci
21. Shannon R, Mandhachitara R (2008) Causal path modeling of grocery shopping in hypermarkets. J Prod Brand Manage
22. Xiao S, Nicholson M (2013) A multidisciplinary cognitive behavioural framework of impulse buying: a systematic review of the literature. Int J Manage Rev 15. https://doi.org/10.1111/j.1468-2370.2012.00345.x

COVID-19 Disease Forecasting Using Machine Learning Approach

Sangram Sanjayrao Dandge and Pon Harshavardhanan

Abstract In terms of peri-operational performance, the estimation systems based on AI (ML) have shown their centrality for improving dynamics in the future operation. For some time, ML models have been used in several fields of implementation involving identifiable evidence and priority for antagonistic objects. Some methods of anticipation are well used for working with prediction problems. This study indicates that ML models are capable of conjecturing the quantity of imminent COVID-19 patients that is by and by seen as potential threat to humanity. As COVID-19 has now had a tremendous effect on the planet, and in excess of 8 million individuals in excess of 100 nations are tainted. To contain its spread, various nations distributed control measures. Nonetheless, it is not known when the plague will end in worldwide and different nations. Anticipating the pattern of COVID-19 is a critical test. By and large, there are distinctive gaging models accessible, yet specific, four standard estimating models, for example, concentrated in speculation of undermining factors of the COVID-19 are the most un-total shrinkage and selection operator (LASSO), support vector machine (SVM), exponential sweep (ES), and linear regression (LR). In the paper report, the use of these methods to cope with the COVID-19 pandemic flow situation is a promising aspect, just as the prevention of COVID-19 infection model is proposed.

Keywords ML · COVID-19 · Modeling · Forecasting

1 Introduction

In the last decade, AI (ML) has proved itself as an unequivocal area of analysis by overcoming several highly difficult and complicated certifiable problems. Virtually, all of these present rooms, e.g., medical care; an AV, business applications; daily

S. S. Dandge (✉) · P. Harshavardhanan (✉)
School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India
e-mail: sangram.dandge2019@vitbhopal.ac.in

P. Harshavardhanan
e-mail: pon.harshavardhanan@vitbhopal.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_45

language practice (NLP), clever robots; sports, show of the atmosphere; speech and photo-handlings, were included in the program territories. The learning of ML calculations is typically based on very inverse experimental methods in standard calculations, which conform to programming directions depending on choices, such as if [1]. Perhaps, most of the ML regions are estimated [2]; multiple standard ML estimates have been used to handle the potential work needed within a wide variety of fields including climatic expectancy, outbreak determination, currency exchange, and visualization of disease. In later expectation of patients' situation with a specific disease, various relapse and neural organization models are very important [3]. A number of experiments have been conducted to foresee different AI diseases, such as coronary tract sickness [4], estimation of cardiovascular infections [5], and predicted development of the bosom malignant [6].

The investigation [7] focuses in particular on live figures of the confirmed COVID-19 cases, and study [8] also focuses on the flare-up and early reaction statistic of COVID-19. This statistical mechanism can be helpful in order for early intercession to be effective in coping with these circumstances. Coronaviruses constitute an immense community of pathogens that can cause illness in creatures and humans. Some people know that coronaviruses cause respiratory diseases from the regular virus to more extreme diseases.

The coronavirus 2 respiratory syndrome (SARS-CoV-2). The case that was previously reported was followed up on November 17, 2019. Starting from June 18, 2020, 188 countries and territories had over 8.36 million cases, resulting in more than 449,000 cases. There has been regeneration of over 4.09 million residents. Basic incidents cause fever, hacking, fatigue, windiness, and scent and taste loss. While several cases cause mild incidents, some progress to intense respiratory disorders (ARDS) can be fostered by cytokine storm, [9] multi-organ deception, sepsis, and clumps of blood. The time span from transparency to the start of indication is typically roughly five days, but may be between two and 14 days [10].

The infection spreads primarily by tiny beads that are produced by hacking, whirling, and chatting, between people during close contact. The beads most of the time go to the ground or to objects rather than into the air. Worse people may get tangled by touching a defiled surface and contacting their face later. It is usually contagious within the three initial days after the start of the signs, but dissemination may be conceived before any side effects arise. The standard research technology is the reverse transcription chain reaction of a nasopharyngeal swab in real time (rRT-PCR). CT imagery can be also helpful to determine in people who have high concerns that they are contaminated by side effects and risk factors; however, guidelines do not advocate using CT imagery for regular screening.

1.1 Mode of Spreading

Groups of people will get the contamination by approaching a person with hack and wheezing that is caused by the infections. Coronavirus was mainly transmitted

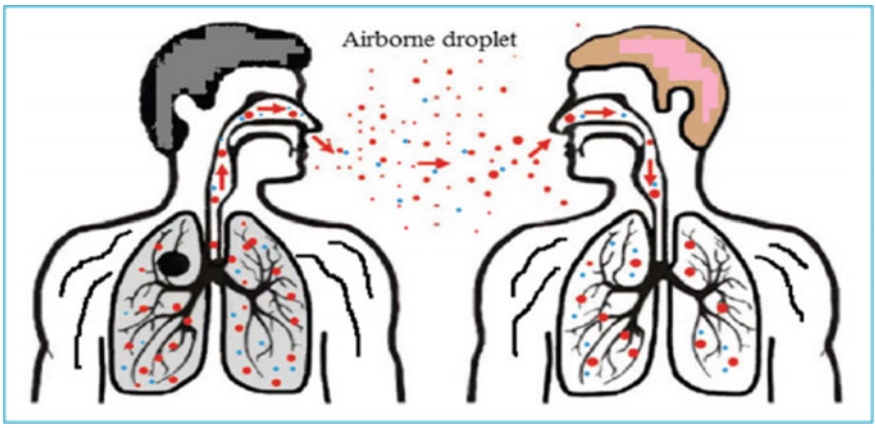


Fig. 1 Transmission of coronavirus via airborne droplets

by airborne zoonotic beads. The ciliated epithelium that induced cell damage and disease was regularly contaminated at the site of the contamination. A giotensin that changes over catalyst 2, a film exopeptidase used by coronavirus in the receptor for human cells, is recorded by a study distributed in 2019 (Fig. 1).

1.2 Types Coronaviruses

Coronaviruses are in the Coronaviridae family with the subfamily. The extreme and the nature of the resulting disease are different among various forms of human coronaviruses. Experts are actually experiencing seven kinds of coronavirus, which can taint individuals.

1.3 Signs and Symptoms

Fever is the most-known side effect of COVID-19, but it is an especially serious and introductory consideration for certain seasonal, disabled, or fundamentally ill people who do not have fever. When attending an emergency room, only 44% of people had fever, while 89% had fever sooner or later after their hospitalization. A lack of fever does not confirm that anyone is without illness.

Hack, lack of appetite, fatigue, windiness, and development of sputum as well as muscle and joint torment are all fundamental consequences. For example, signs of queasiness, heaviness, and loose bowels in changing rates have been found. More irregular symptoms include wheat, flushing the nose, sore throat, and damage to the scalp. In a few cases, chest snugness and palpitations were initially issued in China.

A reduced scent or disturbing taste influences may occur. In 30% of affirmed cases in South Korea, lack of scent was an introductive side effect [11].

1.4 Protection from Covid-19

Keep track of the new COVID-19 statistics available on the WHO Web site as well as on the general public and local welfare department. In most countries around the world, COVID-19 is occurring, and many are witnessing outbreaks. Experts are abundant in China and other countries to alleviate their episodes. In any event, the conditions are capricious, so review the current news regularly.

By avoiding potential harm, you could decrease the chance of becoming infected or spreading COVID-19:

- Clean your hands frequently and totally by rubbing or cleaning them with water. What is the reason? Washing your hands with cleanser and water by using a liquor-based rubber hand that performs infections on your hands.
- Hold the distance between you and others at any rate 1 m. What is the reason? They spill tiny fluid beads from their mouths or nose that can hold bacteria as someone pulls them, wheezes, and speaks. If you are so close, you can take the beads, including the COVID-19 virus, if you have the disease.
- Stop swarming in areas. What is the reason? When people gather in groups, you are forced to come into touch with someone with COVID-19, and it is tougher to keep 1 m away.
- Face, nose, and mouth should not be contacted. What is the reason? Many surfaces are contacted, and infections can occur. Hands will transfer your eyes, nose, or mouth from defiled infection. The infection will invade your body from that point on and twist you.
- Ensure that you and those around you obey great cleanliness in your breath. This means that as you cough or wheeze, the mouth and nose are filled with the twisted elbow or tissue. Dispose of your tissue and wash your hands quickly at that point. What is the reason? Beads infection spread. You protect those around you from illness, such as chill, influenza, and COVID-19 by practicing high breathing cleanliness.
- Stay home and confine yourself except to small symptoms like hacking, brain pain, moderate fever, before you heal. Get someone to deliver equipment to you. Carry a mask and ensure not to contaminate anyone if you find yourself off of your building. What is the reason? Communication with others can defend them from conceivable infections such as COVID-19 and other [12].
- If you suffer from fever and coughing and respiratory problems, look for health treatment, but dial early if possible and meet the nearest well-being authority leading. What is the reason? Public and close experts in your vicinity have the latest knowledge on the circumstances. Bringing your doctor in advance would

encourage your physician to lead you easily to the appropriate welfare office. It will also help you avoid pathogens and multiple diseases from spreading [10].

- Stay up to date on the new trustworthy info, such as the WHO and the local and public wellness experts. What is the reason? Nearby and public experts are ideally qualified to urge residents in their general region to protect themselves.

1.5 Use of Machine Learning in COVID-19

Every ounce of mechanical growth and inventiveness, as the planet wrestles with COVID-19, takes humanity one step closer to defeating this pandemic. Essential activities for improved communication and emergency response in COVID-19 are man-made conscience (AI) and IC. AI creativity allows PCs to copy human intelligence to easily recognize examples and interactions by gathering vast quantities of information.

The war against COVID-19 was waged rapidly by alliances in a few aspects with their AI capabilities: scaling the exchange of customers, looking at how COVID-19 is expanding, discovery, and therapy accelerating.

Any organization seeks new ways to function adequately and to tackle the problems with its clients and leaders, regardless of whether small or enormous or public or private, as social isolation and isolated steps remain. AI engineering plays on an essential role in empowering the machines by providing them with far-reaching connectivity, empowering telemedicine, and providing food security [11].

The chatbots are funded by AI for contactless screening of COVID-19 side effects and to address questions from population as a whole in medical care and government agencies. One model is Clevy.io, a starting French and AWS client, who sent a chat to ease the exploration of official government interchanges over COVID-19 by individuals.

2 Problem Definition

The COVID-19 spreading all over the planet endangered mankind. This kind of disease spreads mostly through intimate contact, usually by minute droplets caused by coughing and sneezing. The drops are frequently slumped to the floor and on objects instead of traveling via air. Unusually, individuals could be sick but instead contact the faces by contacting a contaminated object. The very first three days following the beginning of symptoms that is most infectious, however, dissemination prior to symptoms and from persons who exhibit no indications is conceivable. Most people who have had COVID-19 virus get an autonomous and recovered infection. Researchers understand, though, that even a fraction continues to suffer from more serious illness with an intensive care unit in 10% of instances. Unfortunately, certain

victims are going to die. The management and medical authorities are stressing a lot to accommodate people with probable COVID-19 symptoms.

Depends on the resources provided by health care, a quick decision system can assist physicians analyze individual data and estimate the risk of developing a certain disease effectively. This calls for a quick and personalized health care system, but mostly people who are quite precise though with a significant cost for time and resources rely on the present pneumonia detection system.

3 Proposed Work

This research is intended to develop the early model from now on, coronavirus is an intense challenge of human life worldwide. By 2019, the infection first emerged in a city named Wuhan, China, where innumerable citizens produced proof of pneumonia [11]. It effects the human physique in a wide range of different ways, including extreme acute respiratory disease and multi-order deceit. A large amount of people globally, including multiple activities the next day, are affected by this pandemic. Many new individuals from nations around the world are consistently optimistic. The infection mainly spreads through close connections, via breathing beads or through contact with defiled surfaces. The most testing aspect is that a person can get an infection for a long time without showing symptoms. Almost all nations have proclamated either halfway or requesting lockdowns through the affected local and urban populations as the reasons for its extension and thought about its risk. Research scientists worldwide are currently obligated to locate an effective antibody and medication for the disorder. Since no drug has been proven to date for the death by infection, the parliamentarians of all nations have taken measures to discourage the spread. There is no medication. Of all promises, the 'be educated' is regarded as crucial in nearly all facets of COVID-19. In addition to this aspect of the data, different scientists analyze the different components of the pandemic and yield findings for the human race [10].

The following schematic view of the designed framework is shown in Fig. 2 and trying to work on automated framework to recognize harmful disease by considering some images dataset to avoid the side effects of those disease. The framework is capable of acquiring histological and current datasets by using the various methods, and those datasets are useful for acquiring and predicating the harmful disease in human being [9]. Also, the framework consists of different phases of traditional analysis, such as data acquisition, data analysis, classification of data, training selected data, and measuring the performance to predict the disease by using ML algorithm. This traditional diseases predication processing framework consists of different tasks, such as:

Framework consists of different tasks, such as:

- Data gathering
- Noise removal

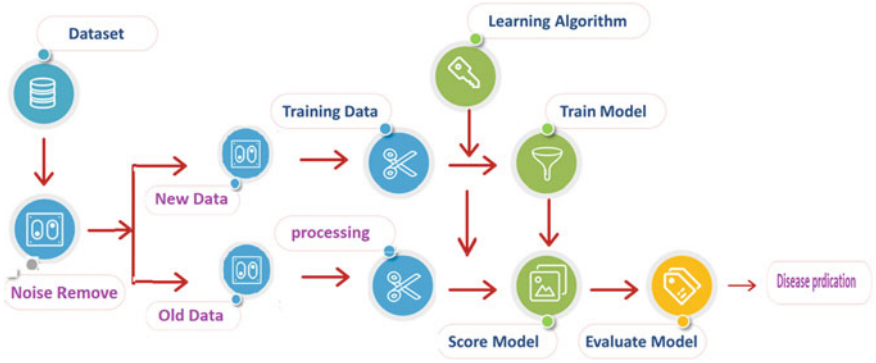


Fig. 2 Proposed diseases predication processing framework using ML

- Data preprocessing
- Data extraction
- Use of learning algorithm (like SVM, KNN, DT with proposed machine learning-based hybrid approach, i.e., combination any two classification approaches like SVM, KNN, DT to improve the accuracy)
- Training the data
- Predication of diseases.

3.1 Four Regression Models Have Been Studied of COVID-19 Future Forecasting

(1) *Linear Regression*

In relapse demonstrating, an objective class is predicated on the autonomous highlights [9]. This strategy can be consequently used to discover the connection among autonomous and subordinate factors and furthermore for estimating. Straight relapse, a sort of relapse displaying, is the most usable factual procedure for prescient examination in AI. Every perception in direct relapse relies upon two qualities, one is the needy variable, and the second is the free factor. Direct relapse decides a straight connection between these needy and free factors. There are two elements (x, y) that are associated with straight relapse investigation. The condition underneath shows how y is identified with x known as relapse.

$$E(y) = \beta_0 + \beta_1 x \tag{1}$$

The mistake term here utilizations to account the changeability between both x and y, β_0 speaks to y-block, β_1 speaks to incline. To place the idea of direct relapse in the AI setting, to prepare the model, x is spoken to as information preparing dataset and y speaks to the class marks present in the info dataset. The objective

of the AI calculation at that point is to locate the best qualities for β_0 (capture) and β_1 (coefficient) to get the best-fit relapse line. To get the best-fit suggests, the distinction between the real qualities and anticipated qualities ought to be least, so this minimization issue can be spoken to as:

$$\begin{aligned} &\text{minimize}\frac{1}{n}\sum_{i=1}^{n}\left(\text{pred}_i - y_i\right)^2 \\ &g = \frac{1}{n}\sum_{i=1}^{n}\left(\text{pred}_i - y_i\right)^2 \end{aligned} \tag{2}$$

Here, g is known as cost work, the average root square are of $y(\text{pred}_i)$ and $y(y_i)$ figures, and n is the entire number of centered details.

(2) *LASSO*

Rope is a model of recurrence with a position with a direct recurrence technique using retrenchment [13]. Shrinkage here relates to the procurement of exceptional data test figures for the focal standard. Consequently, the shrinking cycle increases the stability of LASSO and eliminates the mistake [14]. For multi-country scenarios, rope is known as a more realistic model. Since the model conducts L1 regularization, the penalty included is proportional to the size of the coefficients. LASSO makes the repeat smoother in comparison to the number of highlights it uses. It uses a method of regularization to punish the other highlights. In other words, the effects of the recession cannot be as large as possible and should be assumed to be zero.

A traditional multivariate relapse uses all the highlights and assigns each of them a relapse coefficient. Furthermore, if the current part would not boost its health by adding this highlight, it will not be added as zero. LASSO attempts to add these two in turn, and it will no longer increase in value. Then, the power of regularization is that it will of course assess us by using the penalty word for the extra highlights. So, the models are limited to a few coefficients because the loop eliminates the coefficients, when their qualities are equal to zero.

$$\sum_{i=1}^{n}\left(y_i - \sum_{j} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \tag{3}$$

It sets the coefficient that can be decrypted in a minute, where μ|slope | is penalty word (amount of square residual + μ|slope|).

(3) *Support Vector Machine*

A SVM is the managed ML calculation used both as retrograde and as a solution [15] . This is a kind of managed ML calculation. SVM reciprocity is a non-parametric method based on a variety of computational capabilities. The configuration of the capability named component shifts the contribution of knowledge to the ideal structure. SVM manages relapse problems with straight power such that the info vector

(x) is translated to n-dimensional space, called part space, while handling non-direct relapse issues (z). After this direct relapse, the planning is completed with non-direct planning techniques for space. In ML, put the concept with a multivariate dataset preparation (x_n) with N number of impressions, with y_n as a category of noticed reactions. The straight potential can be seen as:

$$f(x) = x'\beta + b \tag{4}$$

In this way, the aim is to make the evaluation of $f(x)$ with $(\beta'\beta)$ as pointless standard properties as possible as imaginable. The dilemma then fits into reducing work as:

$$J(\beta) = \frac{1}{2}\beta'\beta \tag{5}$$

For a single status of figures not exceeding p, as in the following case, with all the residuals:

$$\forall : \left| y_n - \left(x_n'\beta + b \right) \right| \leq \varepsilon \tag{6}$$

(4) *Exponential Smoothing*

Gaging is performed based on details from the past time in drastic smoothing techniques in the household. The effect of the previous data expectations is dramatically rotating as they become more knowledgeable. The weight of different slack qualities in these lines is declined mathematically. ES is an unbelievable time scheme that specifically estimates univarious intelligence technique [7]. The current time measurement (F_t) in ES is:

$$F_t = \alpha A_{t-1} + (1 - \alpha) F_{t-1} \tag{7}$$

Here, α smoothing cost where $0 \leq \alpha \leq 1$, $A_t - 1$ is the genuine estimation of the past period in time arrangement, and F_{t-1} is the figure estimation of the past gage.

Above study has some key findings like:

- Different ML calculations that appear to perform better in various class expectations above are some essential models identified with gaging.
- Most ML equations require a sufficient measure of knowledge in order to forecast the future, since the scale of the dataset increases, the model indicates.
- ML model-based guessing will help leaders' control COVID-19 pandemics.

4 Methodology

The objectives of this task are twofold.

(1) To investigate advancement of an AI calculation to recognize chest X-ray of people with respiratory disease testing positive for COVID-19 from other X-ray.
(2) To advance revelation of examples in such X-ray through AI interpretability calculations.

A model can been prepared on a dataset made out of X-ray named positive for COVID-19 disease, typical X-ray, and X-ray portraying proof of different pneumonias. As of now, we are utilizing machine learning as the interpretability technique being applied to the model. This work needs more aptitude and more information. The promptness of this work could not possibly be more significant, as any experiences got from this work might be helpful for medical care professionals and analysts as the COVID-19 pandemic keeps on advancing.

5 Conclusion

The current COVID-19 pandemic is obviously an overall general ailment. There have been quick advances in our opinion on the microorganism, how it corrupts cells and causes infection, and clinical characteristics of disease. In view of speedy transmission, countries all throughout the planet ought to grow thought into disorder surveillance structures and scale up country status and response exercises including setting up quick response gatherings and improving the constraint of the public laboratory system.

In this investigation, a ML-based forecast framework has been read for anticipating the danger of COVID-19 episode around the world. Generally speaking, the investigation gages in this way can likewise be of extraordinary assistance for the specialists to make ideal moves and settle on choices to contain the COVID-19 emergency.

This investigation will be improved persistently later on course, and next, we are intending to investigate the expectation approach utilizing the refreshed dataset and utilize the most precise and fitting ML strategies for determining. Ongoing live estimating will be one of the primary centers in our future work.

References

1. Jin YH, Cai L, Cheng ZS, et al (2020) Zhongnan Hospital of Wuhan University novel coronavirus management and research team; Evidence-based medicine chapter of China International Exchange and Promotive Association for Medical and Health Care (CPAM). A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil Med Res 7:4
2. Pan Y, Guan H, Zhou S, et al (2020) Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur Radiol

3. Song F, Shi N, Shan F, et al (2020) Emerging coronavirus 2019-nCoV pneumonia. Radiology
4. Pan F, Ye T, Sun P, et al (2020) Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. Radiology
5. Fang Y, Zhang H, Xu Y, Xie J, Pang P, Ji W (2020) CT manifestations of two cases of 2019 novel corona-virus (2019-nCoV) pneumonia. Radiology
6. Lei J, Li J, Li X, Qi X (2020) CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology
7. Shi H, Han X, Zheng C (2020) Evolution of CT manifestations in a patient recovered from 2019 novel corona-virus (2019-nCoV) pneumonia in Wuhan, China. Radiology
8. Huang, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China Lancet
9. World Health Organization. WHO coronavirus disease (COVID-19) dashboard. https://covid19.who.int
10. Kathole AB, Chaudhari DN (2019) Pros & Cons of machine learning and security methods. http://gujaratresearchsociety.in/index.php/ JGRS, ISSN: 0374-8588, 21(4)
11. Kathole AB, Halgaonkar PS, Nikhade A (2019) Machine learning & its classification techniques. Int J Innov Technol Exploring Eng (IJITEE) ISSN: 2278-3075, 8(9S3)
12. Kathole AB, Chaudhari (2019) Fuel analysis and distance predication using machine learning. Int J Future Revolution Comput Sci Commun Eng 5(6)
13. Apostolopoulos ID, Mpesiana T (2020) Covid-19: Automatic detection from X-Ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43:635–640
14. Rubin GD, Ryerson CJ, Haramati LB, et al (2020) The role of chest imaging in patient management during the COVID- 19 Pandemic: a multinational consensus statement from the Fleischner Society [published online ahead of print, 2020 Apr 7]. Chest 2020, S0012-3692, 30673–30675. https://doi.org/10.1016/j.chest.2020.04.003
15. Ozsahin I, Sekeroglu B, Mok GSP (2019) The use of back propagation neural networks and 18F-Florbetapir PET for Early Detection of Alzheimer's Disease Using Alzheimer's disease neuroimaging initiative database. PLoS One 14:e0226577

Face Mask Monitoring for Educational Institutions and Organizations

Angelina Geetha, S. V. Shri Bharathi, Sai Laasya Vabilisetty, Naga Chakradhar Rao Rayasam, and Yuva Teja Aagollu

Abstract In the year 2020, we had seen a disastrous effect on the society by the outbreak of coronavirus which has demanded the world to mandatorily wear a mask. Therefore, one major protection for people is to wear mask in public areas. Thus, we proposed a face mask monitoring system, which uses convolution neural networks to train the model, Caffe-based DL detector with MobileNetV2 architecture to extract the facial features for mask detection and allowing our monitoring application to be deployed on any embedded system easily. The main objective of this paper is to automate face mask monitoring so that it reduces time and human effort administered in this process and, thus, helps in controlling the spread of virus.

Keywords Face mask · Neural network · Caffe-based DL detector · MobileNetV2

1 Introduction

1.1 Pandemic and Need for Face Mask

The outbreak of coronavirus in 2019 has developed into an uncontrollable situation and lead us to a pandemic in 2020 by affecting millions of lives every single day. All types of businesses irrespective of size and reach had to face a significant amount of loss, and all educational institutions had to be shut for an indefinite period of time. This pandemic situation has transformed the lifestyle of all the individuals. It has also affected all the business sectors irrespective of the size and reach.

This has forced countries all over the world to initiate new rules for wearing a mask. Studies had shown that virus spread is more in busy and populous areas like schools, market, offices. Therefore, wearing a face mask was made mandatory throughout the world. So, face mask has become a day-to-day essential in the present-day world. Face masks, when worn properly by everyone, can reduce viral transmission by 70%. So, wearing a face mask is an easy way of reducing the risk of spreading any infection

A. Geetha · S. V. Shri Bharathi · S. L. Vabilisetty · N. C. R. Rayasam (✉) · Y. T. Aagollu
Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_46

or virus unknowingly. Thus, there is a great need for everyone to wear a face mask in the present-day world.

This becomes the motivation for the main objective of this project which is proposing a model that detects whether a person is wearing a face mask even when the person is in motion. This proposed model uses Caffe-based DL algorithm which works well for static images as well as live video. This paper focuses on deploying this model to any embedded systems like surveillance network. To achieve this, we have used MobileNetV2 architecture which makes this model easier to be deployed onto any embedded systems.

2 Related Works

Facial recognition using Viola Jones algorithm [1–3] and semantic segmentation [4, 5] is performed previously. This model works fine with unmasked faces showing 85–90% accuracy, but it drops to less than 70% accuracy when masked faces come into action. This way, we cannot detect a person whether masked or not accurately.

Qin and Li [6] proposed a method which uses a one-stage detector. This consists of feature pyramid network with multiple feature maps which helps this model to get decent accuracy. This face mask detector succeeded in removing objects from images which have low confidence. However, this model cannot differentiate between mask and other objects which cover face. This leads to misinterpretation as person is wearing a mask.

Another face mask model was proposed by Mohammed, Gunasekaran and Khalifa [7, 8] which consists of two components. First component is ResNet-50 which is used for feature extraction and the second component is YOLO v2 which is used for detecting medical face mask. This face mask detector model detects only medical face masks in static images.

Facial recognition using principal component analysis was performed by Ejaz [9]. The model works fine with unmasked faces showing 85–90% accuracy, but it drops to less than 70% accuracy when masked faces come into action. This way, we cannot detect a person whether masked or not accurately. Also, facial recognition using deep residual learning was proposed by Zhang, Sun and Ren [10]. Face mask detection was also done using OpenCV, Keras and Tensor Flow [11].

To summarize all the above studied works, we can say that all the above studies were successful in detecting masks from faces in static images and frontal faces but lacked the accuracy in detecting faces from video streams, noisy images, and images with non-frontal faces.

3 Detection Methodology

3.1 Facial Detection

Facial identification or detection is the approach of detecting people in an image or a video with the help of their facial features. It can be referred as an advanced computer technology which is used to identify facial structures in static images. With these continuous enhancements in technology, face detection has also been extended to videos which involves people movement.

In comparison with other features, facial features provide good amount of accuracy.

3.2 Need for Face Detection

Facial feature detection or identification plays a crucial role in this digital world. It is mainly used for security and verification purpose. It can also aid forensic investigations by automatically recognizing individuals in security footages or any other videos.

As mentioned above, the need for face detection in the field of security has increased a lot. The application of face detection may vary from an individual to a corporate company. Several corporate companies, nowadays, use face detection for monitoring employee presence, and almost all the individuals, nowadays, use face detection as security for their electronic gadgets like mobiles, laptops.

3.3 Proposed System and Novelty

We propose this paper with an objective of getting an automated face monitoring system which identifies whether a person has mask or not, by using facial detection algorithms to extract the features from an image or video and convolution neural networks to train the model. The face detection algorithm used in this model is Caffe-based DL detector. This model detects even the non-frontal faces and predicts as masked or non-masked. We have used MobileNetV2 architecture which allows our application to get deployed onto any embedded system easily. Our model is capable of taking static images as well as live or recorded video as input to detect faces and predict mask or no mask.

Our model uses Caffe framework for feature extraction during image processing. This framework makes our model unique and novel compared to existing models. Caffe is a brand new deep learning framework which is developed to especially increase speed and modularity and decrease time taken in the process of image processing. So, our model is very fast and accurate when compared to all the existing

models. And, the MobileNetV2 architecture, which we have used, also makes our model novel, as it helps in deploying our model onto embedded systems and mobile devices very easily.

3.4 Caffe-based DL Algorithm

In our project, we use this Caffe-based DL algorithm to extract the features of a face during the process of face mask detection.

Caffe is a deep learning framework made with speed and modularity in mind and used for image processing and feed forward networks. It is suitable for fine-tuning the existing deep neural networks. Caffe supports multiple types of deep learning architectures. These can range from image classification to image segmentation. Caffe-based DL detector gives high accuracy and detects the faces even in motion.

3.5 MobileNetV2

MobileNetV2 [12] is known for its performance on mobile devices. It is based on the inverted residual structure in which the residual attachments are between the layers. As a whole, the architecture of MobileNetV2 consists of the initial complete convolution layer with 32 filters and 19 residual bottleneck layers. The basic structure of MobileNetV2 is shown in Fig. 2.

3.6 Dataset Characteristics

This paper conducted its experiments based on the dataset which is created by the images which are collected from, Kaggle, RMFD [13], and as well as Bing API. This dataset consists of 4095 images. Thus, the dataset we used is divided into two parts:

Masked images—2165.

Unmasked images—1930.

Our model is also capable of detecting faces with and without masks from live or recorded videos. Figure 1 shows the examples of images present in the dataset.

Fig. 1 Examples of images in face mask dataset

3.7 Data Preprocessing

All the raw input images are preprocessed before they are used for training in a five-staged process. First, the raw image is resized (256 × 256). Then, color filtering (RGB) is applied over the channel. As we have used MobileNetV2 architecture, our model supports 2D 3-channel image. Then, the image is cropped with the pixel value of 224 × 224 × 3. Finally, the images are converted into tensors (similar to NumPy arrays).

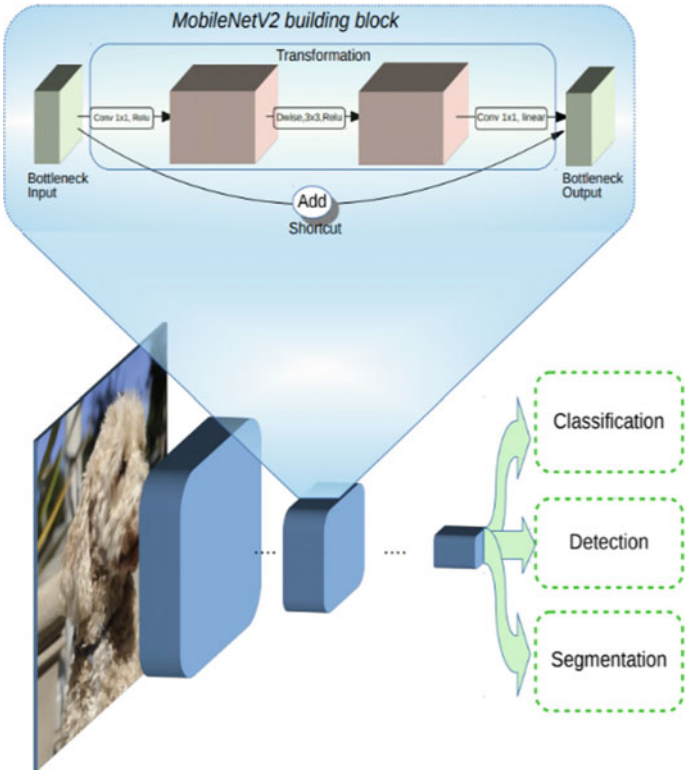


Fig. 2 MobileNetV2 architecture

3.8 Architecture

Figure 3 presents the architecture diagram of the proposed detector model. The introduced model consists of two important modules: training module and the deployment module.

In the training module, dataset is loaded from the disk, and it is trained using the Caffe-based DL algorithm with the help of MobileNetV2 architecture. We split the train and test data into 80% and 20%, respectively, to obtain more accuracy. Once face mask model is trained, it is serialized and stored on the disk.

Now in the deployment module, the model which we have trained previously is loaded from the disk, and the faces from static images or a video stream are loaded as input. These images are preprocessed using OpenCV and PyTorch and made ready to apply the model. Now, the trained model which is loaded from the disk is applied on these images to determine whether the images are masked or unmasked. The result indicating mask or no_mask is displayed finally as result.

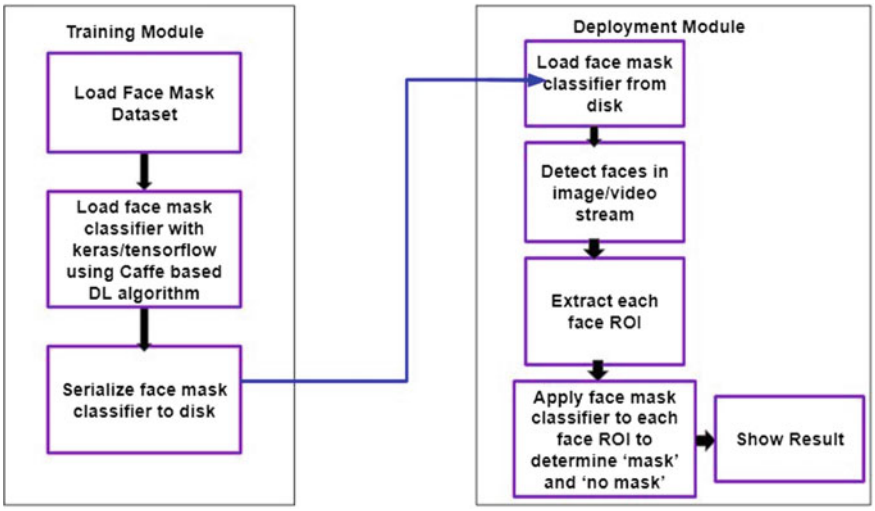


Fig. 3 Architecture diagram

4 Results and Analysis

Figures 4 and 5 depict the output of this paper. As shown in these figures, when we give the raw images or video streams as input, the model will detect all the faces with mask and indicates them with the help of a green box around the face, and it will detect all the faces without mask, and red box is used to indicate them. The model also presents the percentage of accuracy of the detection above these colored boxes.

The dataset is split into 75–25% training images and testing images, respectively. Our model is trained for 30 epochs, and as shown in Table 1, we have achieved an accuracy of 99% for both masked and unmasked images with the precision, recall, and F1-score being 99%. This proves that Caffe-based DL algorithm with MobileNetV2 is computationally efficient in face mask detection.

Fig. 4 Example of output

Fig. 5 Input and output images

Table 1 Accuracy table

	Precision	Recall	F1-score	Support
With_mask	0.99	0.99	0.99	388
Without_mask	0.99	0.99	0.99	386
Accuracy	–	–	0.99	774
Macro-avg.	0.99	0.99	0.99	774
Weighted avg.	0.99	0.99	0.99	774

5 Conclusion

To reduce the transmission of coronavirus, certain measures like wearing a facial mask, washing hands must be ensured. We have modeled a face mask monitoring system using Caffe-based DL algorithm with MobileNetV2 architecture to extract the face features and train the model. Our model can be easily deployed onto any embedded system like CCTV cameras, as we have used MobileNetV2 architecture. We have used a dataset which is created by the images which are collected from Kaggle and RMFD datasets. This model works effectively for both static images and live video streams. This face mask monitoring system can be implemented in all workspaces and educational institutions. The scope of our project can be further expanded to the public areas like airports, bus stations, railway stations, which are usually very crowded.

As our model uses Caffe framework and MobileNetV2 architecture, the accuracy of the model is observed to be 99 percent which is very high in the context of facial recognition models. Thus, this model sets a new benchmarking in the automated monitoring systems when administered in colleges and workspaces. This way, our

model ensures the reduction in time consumption and human efforts in the process of face mask monitoring and also helps in controlling the spread of coronavirus which is very essential for the society at present times.

6 Image Sources

Figure 2—https://analyticsindiamag.com/why-googles-mobilenetv2-is-a-revolutionary-next-gen-on-device-computer-vision-network/?hcb=1

Figure 3—Self-designed architecture diagram of our model

Figures 1, 4, and 5—These images are taken from the RMFD dataset and given as inputs to our model which resulted in the output as shown in the figures.

RMFD dataset—[1] https://paperswithcode.com/dataset/rmfd.

References

1. Jiang M, Fan X, Yan H (2020) Retina face mask: a dace mask detector. Department of Electrical Engineering, Hong Kong
2. Lippert C, Ali R, Azed S, Ahmed A, Shahihar HM, Mojumender M (2020) Face Mask Detector. University of Potsdam
3. https://www.researchgate.net/publication/326667118_Face_Detection_Techniques_A_Review
4. Meenpal T, Ashutosh A (2019) Facial mask detection with semantic segmentation. National Institute of Technology Raipur
5. Meenpal T, Balakrishnan A, Verma A (2019) face mask detection using semantic segmentation. National Institute of Technology, Raipur
6. Qin B, Li D (2020) Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Zhejiang University, Hangzhou 310058, China
7. Mohamed A, Gunasekran, Hamed N, Taha D, Khalifa NEM (2019) Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. University of California, USA
8. Khalifa EM (2020) Fighting against COVID-19: a novel deep learning model based on YOLO-V2 with ResNet50 for medical face mask detection. University of California, Davis, USA c College of Information and Electrical Engineering. Asia University, Taiwan
9. Ejaz MS, Islam MR, Sarker A (2019) Implementation of principal component analysis on masked and unmasked faces. Bangladesh Army University of Engineering and Technology, Bangladesh
10. Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition
11. https://ieeexplore.ieee.org/document/9342585 Face mask detection using tensor flow, Kera and OpenCV
12. https://arxiv.org/abs/1801.04381—A brief paper on MobileNetV2
13. https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset

Deep Learning-based Traffic Identification and Accident Detection in Remote Environment

S. Sathya Priya, Vikram Kumar, Pavan Teja, Samanthapudi Anil, and S. Revathy

Abstract Events like traffic jam, road accident, vehicle fire, robbery, vehicle breakdown are the most common events that occur on the roadside. In cities, traffic control department, fire brigade, and other agencies are available to handle such events. However, their work fully depends on the information, faster the information, faster will be the action, and the intensity of damage will be reduced accordingly. But unfortunately, currently used method for surveillance, like camera, police and highway patrol, traffic signal, manual traffic signal, control rooms is not sufficient to tackle such situation. The demand for an intelligent and a fast system that would be capable to detect the multi-class events immediately and alert the same to the concerned authority arises. The proposed system uses deep learning algorithm, specifically convolutional neural network, to process the real-time video and to detect and alert the authority. The proposed system can be used in the city to regulate the traffic and also to detect accidents or mobs like situation real time. Also, in remote highways and roadways, it is capable of detecting accidents, vehicle breakdown, out of gas, and other suspicious activity.

Keywords Convolution neural network · Deep learning · Internet of Things · Monitoring system

1 Introduction

Deep learning is a sub-genre of AI and is being used in number of domains as per the requirement, few areas where deep learning is used most extensively are automation, robotics, image processing, etc. Concept of deep learning can also be used in transportation system for observing the non-rural road and freeway, for checking amount

S. Sathya Priya (✉) · V. Kumar · P. Teja · S. Anil · S. Revathy
Departmaent of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India
e-mail: sathyap@hindustanuniv.ac.in

S. Revathy
e-mail: srevathy@hindustanuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_47

of traffic jam using videos and images. Surveillance cameras are used everywhere for monitoring congestion. Cameras are used, but it cannot detect congestion automatically [1]. Manual congestion detection needs lot of man power. It is very difficult to find congestion through all the surveillance cameras by watching it continuously [2] all the day. It is almost impossible to cover all the cameras manually all the time. Therefore, prompt detection of congestion is important for large-scale areas. This system will help in detecting dense or sparse vehicle on fire and road accidents. Once the accidents are identified, then it can be notified to the road safety team and then corrective measures can be taken by the road safety.

Hence, this system is capable of providing four types output, dense, sparse congestion, and fire or road accidents. Domain used is deep learning [3, 4], and it is subset of machine learning in AI. Algorithm used is convolution neural network, which is highly capable in identifying visual imagery that makes it more suitable for the proposed work as this work is more focused on identifying the traffic and accidents from the images taken by surveillance cameras [5].

2 Related Works

2.1 An Intelligent Monitoring System of Vehicles on Highway Traffic

The intelligent highway traffic monitoring system [6] is simple, efficient, and economically cheaper than the other monitoring systems available at that moment of time. This system was capable of predicting direction and speed of moving vehicle. The technologies used in this work are digital image processing, pattern recognition, and mathematical calculations. The main component of this system is a high-definition camera which is placed in a traffic pole in highway. The proposed system takes video of vehicles [7] which are moving on highway, then converts it into frames of resolution 480 * 640 pixels, then background elimination operation is being performed to focus on vehicles, and this is done using Gaussian distribution. Then vehicle extraction and noise filtering are done using preceding subtraction method and median filtering. Gaussian mixture model and rich feature hierarchies are used to detect vehicle, and step gradient is used to detect the front of vehicle to find moving direction. Finally, the speed of moving vehicle is calculated using simple mathematical formula $S = V \times T$, where time 'T' value depends upon video frame rate.

2.2 IoT-based Smart Traffic Monitoring System Using Vehicles Count

This system gives an IoT-based solution [8] to solve the problem of traffic jam. This work is suggested keeping in mind the situations, in which there is a huge traffic in one lane, while at the other lane, there may be less or no traffic. But still, each lane gets equal time, i.e., of 60 s. This system was made to give more priority and time to the lane with higher or denser traffic. The major component of this traffic monitoring system is an IR sensor which will be placed on both sides of the road, and these sensors are responsible to keep an account on the number of vehicle currently waiting in queue and also monitors the rate at which vehicles are adding in the lane. Then, these data are being transferred to the centralized server where KNN algorithm is implemented over the recorded data to prioritize and allocate signal open time on each side. Finally, the priority and open time are send to a traffic control system which switch red and green lights accordingly.

2.3 Modern Traffic Control System Using Concept of Fuzzy Logic

This system uses the concept of fuzzy logic [9] to increase and decrease the signal open and close timings. This system keeps an accounts on two factors: first, number of vehicles in the queue and second, the rate of vehicle coming to the traffic signal pole. Acoustic detector, Doppler radar, and microwave sensors are used to detect the sound generated by the vehicles as it passes the sensor and reports to the centralized system wirelessly. The input is being provided into the Mamdani fuzzy inference system. The fuzzy logic toolbox in MATLAB which has a graphical user interface is used to proficiently design and implement the control system.

The existing systems focus on monitoring and predicting the traffic patterns [10]. A system for multi-class detection is proposed in this paper.

3 Proposed System

The proposed system is focused on identifying dense or sparse traffic based on the input image provided by surveillance cameras. This system does multi-class detection, i.e., fire, accident or any accident detection and prediction of dense, and sparse traffic. The proposed work system uses convolution neural network (CNN) algorithm that classifies the input images very efficiently. Images after getting detected are sent to the centralized server for further processing, and an alert is being sent to either the traffic police or the concerned department.

Figure 1 shows the architecture diagram of the proposed system. CNN is used to

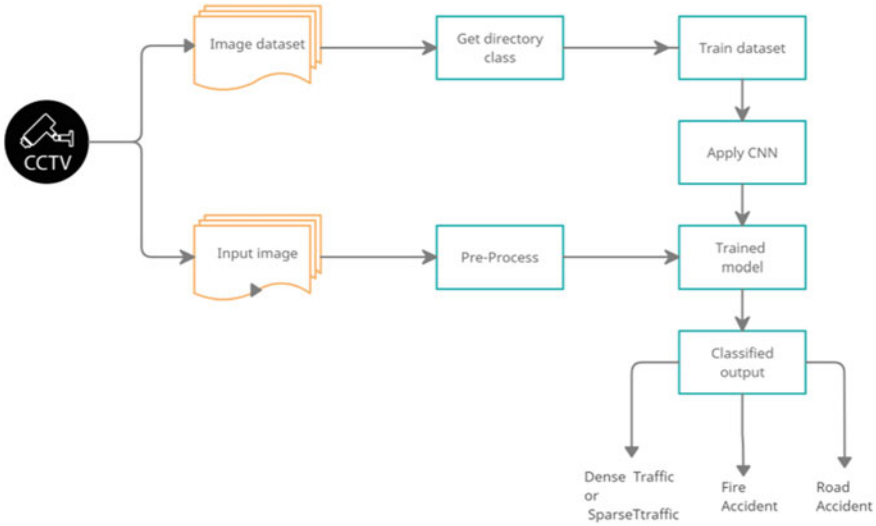


Fig. 1 Architecture diagram

train the dataset to detect accidents and to predict the traffic pattern.

4 Module Description

The proposed system is being designed and developed as the following stages:

- Collection of dataset
- Image preprocessing
- Training using convolution neural network
- Recognition
- SMS alert.

4.1 Collection of Dataset

Different types of input images of traffic are collected from online, and the folders contain input images of dense traffic, sparse traffic, accident, fire. These collected images are used for training and validation process.

4.2 Image Preprocessing

Much of preprocessing for these input datasets is not required before implementation. Training and test datasets are accessed from folders and are given as input to function from Keras 'flow_from_directory,' which provides required preprocessing like reduction of dimensions. Input image for test input is turned to NumPy array and dimension reduction.

4.3 Training Using Convolution Neural Network

Convolution 2D neural network from Keras is used for training and testing models [11]. Conv2D architecture is given below (Fig. 2).

4.3.1 Sequential Model

The model type that is being used in the proposed system is the sequential model. Sequential is an easy way to create a model in Keras. In this, layer-by-layer model is built. 'Add method' is used to add layers to the model. Sequential model allows to simply stack sequential layers and even recurrent layers from network as input to output. First line is used to declare model type as sequential function.

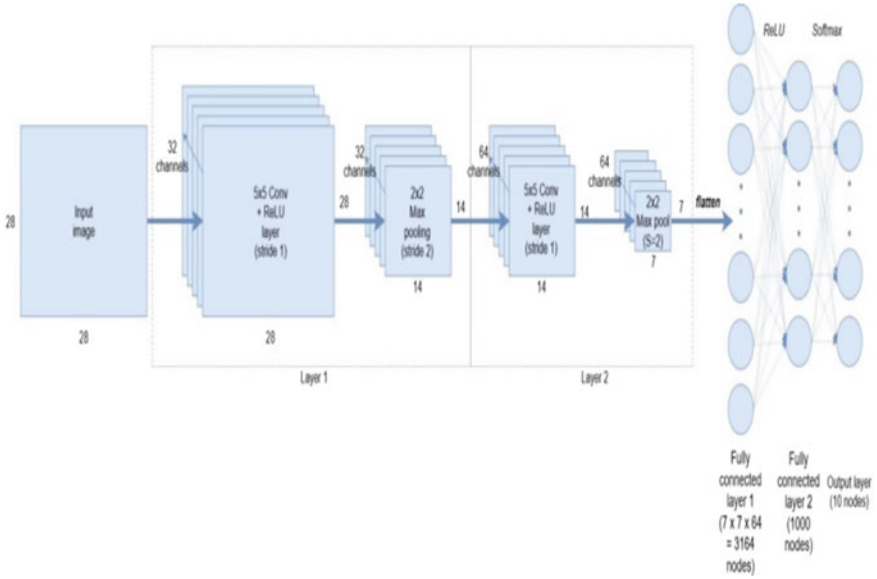


Fig. 2 Training using CNN

4.3.2 Training Neural Network

After declaring sequential model, multiple Conv2D layers [11] are being added with multiple parameters like number of nodes kernal size, activation function, and finally, a dense layer is used for output layer. Now, the model is compiled which will take three parameters, namely optimizer, loss function, and matrices. The optimizer controls the learning rate, and the optimizer used for the proposed system is 'Adam.' The loss function that is being used is 'categorical cross entropy,' and a lower score of loss indicates that model is performing better. Also, 'accuracy' matrix is used to see the accuracy score on the validate set when the model is being trained.

Now, the model is trained by using 'fit()' function on the model, and it includes parameters like training data and label, which are being split in preprocessing stage. For validating, test data and corresponding label are used. The model will cycle number of times as given epoch value; the more the number of epochs, the more efficient will be the model. But, after a certain point of epoch, it needs to be stopped since further epochs will not contribute in model improvement. After training is completed, the model is tested; good accuracy is achieved, and the same is demonstrated in graph using Matplotlib library.

4.4 Recognition

In this module, validated or test data are passed to the fit function, and the input image passed is being converted to an array using NumPy [4] and is compared with the trained model and finally predicts whether the traffic condition in frame or image is dense traffic, sparse traffic, fire, or accident.

4.5 SMS Alert

When heavy traffic is detected by traffic detection system, it sends an alert message to the controller. SMS gateway connection is used to dispatch SMS alert, which gives timely alerts to the system.

5 Results and Discussion

If the traffic recognition encounters massive traffic jam, an alert is sent to the controller. Here, SMS gateway connection is used to dispatch the SMS alert through fastsms.com, which provides message of massive traffic jam with time. Hence, system will be getting alerted on time (Fig. 3).

The following figure shows the model loss in our experimental method (Fig. 4).

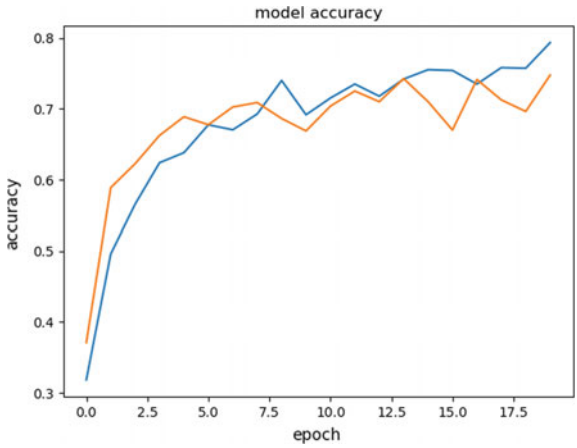


Fig. 3 Model accuracy

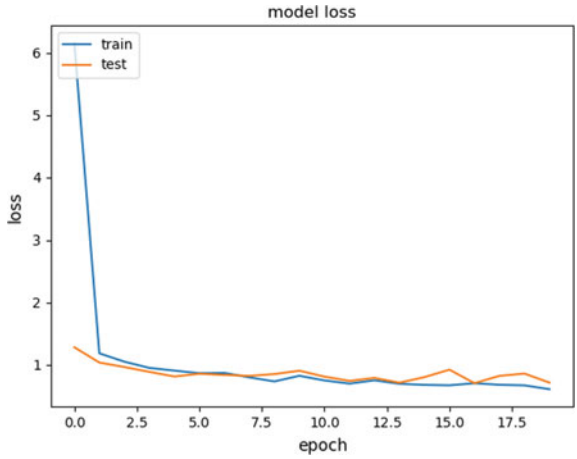


Fig. 4 Model loss

The sample output image showing the condition of heavy traffic is shown in Fig. 5.

6 Conclusion

The system is implemented and tested by giving different input images and videos. The performance of the proposed system is satisfactory, and also the performance and accuracy of the proposed system are better than the other algorithms and other existing systems. The feature of multi-class detection is incorporated in the proposed model that takes care of accident and fire prediction coupled with monitoring and classifying the traffic as sparse and dense traffic. Issues like longer response time are observed when the image of video of large size is passed and accuracy of test data

Fig. 5 output

is bit lower than expected in such conditions. These issues have to be addressed in the future. An effective, efficient, and fully working system with additional features like detecting no vehicles on one lane of the traffic signals, traffic density in signals could be proposed in the future.

Image Source: Free images available online in google images are collected randomly, and the database is created manually. Any other standard image database is not incorporated in the proposed system.

References

1. Zhu X, Wang Y, Dai J, Yuan L, Wei Y (2017) Flow-guided feature aggregation for video object detection. In Proc ICCV 408–417
2. Lei H et al A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning. Pattern Recognit 79:290–302
3. Yu K, Lin L, Alazab M, Tan L, Gu B (2021) Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system. IEEE Trans Intell Transp Syst 22(7):4337–4347. https://doi.org/10.1109/TITS.2020.3042504
4. Zhao Z, Chen W, Wu X, Chen PC, Liu J (2017) LSTM network: a deep learning approach for short-term traffic forecast. IET Intell Transp Syst 11(2):68–75
5. Barth M, Boriboonsomsin K (2009) Environmentally beneficial intelligent transportation systems. IFAC Proc Volumes 42(15):342–345
6. Khan S, Ali H, Ullah Z, Mohammad F (2018) An Intelligence monitoring system of vehicles on highway traffic. 2018 ICOSST. https://doi.org/10.1109/ICOSST.2018.8632192, Published On: Dec 2018
7. Gundeep K, Sharma S (2017) Traffic management using digital image processing. IJCST 8(2)
8. Senthil Kumar J, Veeramanickam MRM, Arun S, Narayanan K, Anandan R, Javed S (2018) IOT based smart traffic signal monitoring system using vehicles counts. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i2.21.12388
9. Koukol M, Zajíčková L, Marek L, Tuček P (2015) Fuzzy logic in traffic engineering: a review on signal control. Math Prob Eng 2015, Article ID 979160:14. https://doi.org/10.1155/2015/979160

10. Maha Vishnu VC, Rajalakshmi M, Nedunchezhian R (2018) Intelligent traffic video surveillance and accident detection system with dynamic traffic signal control. Cluster Comput 21:135–147. https://doi.org/10.1007/s10586-017-0974-5
11. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Proc Adv Neural Inf Pro-cess Syst 1097–1105

Detection of Active Attacks Using Ensemble Machine Learning Approach

S. Revathy, S. Sathya Priya, Shaik Rafi, P. Pranideep, and D. Rajesh

Abstract Cyber attackers entrap online users and companies by stealing their sensitive information without their knowledge. Sensitive information such as login credentials, bank card details, and centralized servers of organizations are hacked by the attackers. Some of the cyber attacks are phishing attack where attackers make the online users to trust their websites as a legitimate website and retrieve their personal information by making them as a prey for cyber attacks. Malware attack is a one where attacker will inject a malicious software into the company server or any online user's device, without their knowledge and steal all the data in their devices and servers. Intrusion is an invasion, where attacker will attack the network and theft all the network resources. But, there are many types of cyber attacks solutions such as visual similarity-based approaches, intrusion detection system, signature-based, heuristic-based, specification-based, anomaly-based methods that are proposed, but they have some disadvantages. Because of few unsecured HTTP websites and lack of cyber knowledge, cyber attacks are increasing day by day. In our proposed system, a unified ensemble approach (UEA) is proposed by combining different machine learning algorithms using ensemble approach that gives better accuracy and detection rate. This model aims to detect the intrusion, phishing attack and prevent the malwares thereby mitigating the cyber attacks encountered by individual and organization.

Keywords Ensemble approach · Feature selection · Intrusion detection · Machine learning · Phishing detection

S. Revathy (✉) · S. Sathya Priya · S. Rafi · P. Pranideep · D. Rajesh
Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India
e-mail: srevathy@hindustanuniv.ac.in

S. Sathya Priya
e-mail: sathyap@hindustanuniv.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_48

1 Introduction

The new technologies are evolving day by day, and most of the people are using digital devices for communication. The users for online services are increasing day by day, and the attackers are taking this as a free-hit and stealing the sensitive data.

1.1 Phishing Attack

Phishing attack is a way to steal sensitive information from the users, which means attacker will create a duplicate, trustworthy, legitimate site, and directing online user into providing personal and sensitive data. The attacker creates a copy of website and sends it to the victim and waits till he receives victim sensitive or personal information.

The reasons for trapping victims using phishing attack are as follows:

1. Lack of computer knowledge.
2. Limited knowledge on cyber security.
3. Not cautious toward warnings and proceeding further undermining existing system tools strength. Phishing attacks can occur in various forms such as websites, email. In email, phishing attackers design fake email in such a way that it looks trustworthy and send the mail to many people assuming few of them will fall for it.

Some Types of Phishing Attacks Are

1.1.1 Website Phishing

In this phishing attack, the attacker will create a copy of the legitimate website and collect the information from victims by advertising this website in social networking sites, and they also manage to give security to the website such as HTTPS connection.

1.1.2 Spear Phishing

In this phishing attack, they target a specific community in an organization and send a [1] mail in such a way that the mails are sent by their colleagues and pretending them to provide sensitive information of the organization.

1.1.3 Whitelist and Blacklist

All the legitimate URLs are present in whitelist. The browser downloads the website after checking whether the URL is present in the whitelist. There is a drawback in the URL that if a legitimate website is not registered in whitelist, the browser will block the website. All the phishing URLs are present in blacklist. The drawback of the blacklist is that if any phishing URL is not registered in blacklist [2], the browser downloads the website thinking it as a legitimate website.

1.2 Malware

This is a type of cyber attack where attacker tries to install a malicious software into the personal system or a company sever to gain access to them and steal the information present in them; usually, this is done for money gain.

The reasons for trapping victims using malware attack are as follows:

1. Clicking on the unwanted links sent to their mails and messages
2. Opening unsecured HTTP websites
3. Allowing unknown persons to access laptop or PCs.

Some Types of Malware attacks are

1.2.1 Worms

Computer worm is a malware program, when injected in a computer system, it will make it as a host, and without the help of host, the worm will scan the computer and replicate itself into many and injects them into other computer systems connected to host system The biggest disadvantage with a worm is that it can replicate many of its type [3], i.e., a single worm can generate thousands of its copies.

1.2.2 Adware

Adware is a malicious program which will generate pop-up advertisements in the system. This advertisement contains a harmful code, which on clicking the advertisement link will download and install the software, and our system will get compromised.

1.3 Intrusion

It attempts to attack the computer network, destroy the security of the network, gain access to the network, and download all the network resources without the knowledge of the user.

1.3.1 Common Gateway Interface Script

CGI is used in the network for communication between the client and server; it also provides easy opening called backtracking through which attacker will enter the network and access the secured files in that network.

2 Existing System

2.1 Heuristic-based Techniques

This approach extracts features from the URL and compares them with phishing attack features for the detection, but the accuracy of the algorithm is very low as some phishing URLs are designed with non-phishing attack features [4]. List-based comparison is not used in this approach. Hence, the system detects zero-day phishing attacks which are not detected by list-based techniques.

2.2 Visual Similarity-based Technique

Here, the attacker makes an exact copy of legitimate site to cheat the online users by making them believe that it is a legitimate site without any doubt. This anti-phishing technique will compare the phishing and legitimate website database. Based on the similarity ratio, this technique will detect whether the given website is phishing or legitimate. In such a way that if there is a less similarity ratio, it will detect that it is a phishing website, and if there is high similarity ratio, it will detect it as a legitimate website.

2.3 Machine Learning Techniques

Irfan [5] In this, features are extracted from the website and then compared with the trained machine learning algorithm. Based on the comparison, it will detect whether

the given website is legitimate or not. It is a combination of heuristic feature and machine learning algorithms.

2.4 Signature-based Detection

A signature database is maintained which is used to detect the malware by comparing the malware; the signature is generated by examining the disassemble code of malware binary. After analyzing the disassemble code, its features are extracted. Signature is generated using those features.

2.5 Anomaly-based Method

It is used to detect the unknown malware attack as new malwares are created day by day. It is based on machine learning. A trustful machine learning model is developed; when any one tries to inject malware into the system, then it will compare them with that system; if it is not found in that model, then it is declared as the suspicious.

The existing heuristic-based approach [6] cannot detect a website when it is designed with new feature that was not included in the training. Signature-based detection can deal only with known attacks. In visual similarity-based approach, false negative classification rate is higher, and their accuracy is very low and users have to use different systems to detect each cyber attack; hence, the UEA model is designed to address the challenges as a combined model.

3 Proposed System

As technology is increasing, people are relying on the newest features for relaxed life. In the same way, cyber attacks are increasing for the sake of money and data. This system is going to detect some of those cyber attacks such as phishing attack, malware, and intrusion with their respective datasets using five supervised classification machine learning algorithms, feature selection, feature scaling, and ensemble methods. After execution of each algorithm, they produce test metric values and accuracy for each cyber attack. Predicted accuracies are extracted and processed in such a way that it predicts the algorithm with best accuracy for each cyber attack dataset.

The uniqueness in the system is it is built using ensemble approach to find the accuracy of the system using different machine learning algorithms.

The system follows the following steps for execution:

Step-1: Select any of the cyber attacks, and the system starts the process.

Step-2: If the selected cyber attack is phishing, then users have to specify how many features to be selected.
Step-3: The system will train and test different types of machine learning algorithms using the dataset for the selected cyber attack and extracts accuracy.
Step-4: The extracted accuracy is processed, and the system gives the name of the algorithm which has the highest accuracy.

4 Module Description

The modules included in our project are as follows:

- Dataset preprocessing
- Phishing detection
- Intrusion and malware detections
- Voting classifier.

4.1 Dataset Preprocessing

There are many features available in the given dataset, and we are going to select only the top most features in the given dataset by using feature selection for phishing dataset. Malware detection and intrusion detection are huge datasets; hence, data processing will be slow. So, feature scaling [7] is applied here to covert the values in dataset in such a way that the values are in between 0 and 1.

4.2 Phishing Detection

After data processing, using feature selection process, the selected top most features are now tested and trained using five different types of algorithms in such a way that 75% of dataset is used for training and 25% is used for testing. Accuracy and the metric values for the algorithms are generated and stored.

4.3 Intrusion and Malware Detections

Intrusion and malware detection datasets are huge, so the dataset is converted in such a way that the values will come in the range of 0–1. This is done by using feature scaling process. Once the values are converted, then testing and training processes will start where 75% is used for training and 25% is used for testing the system using dataset. Accuracy and metric values are generated and stored (Fig. 1).

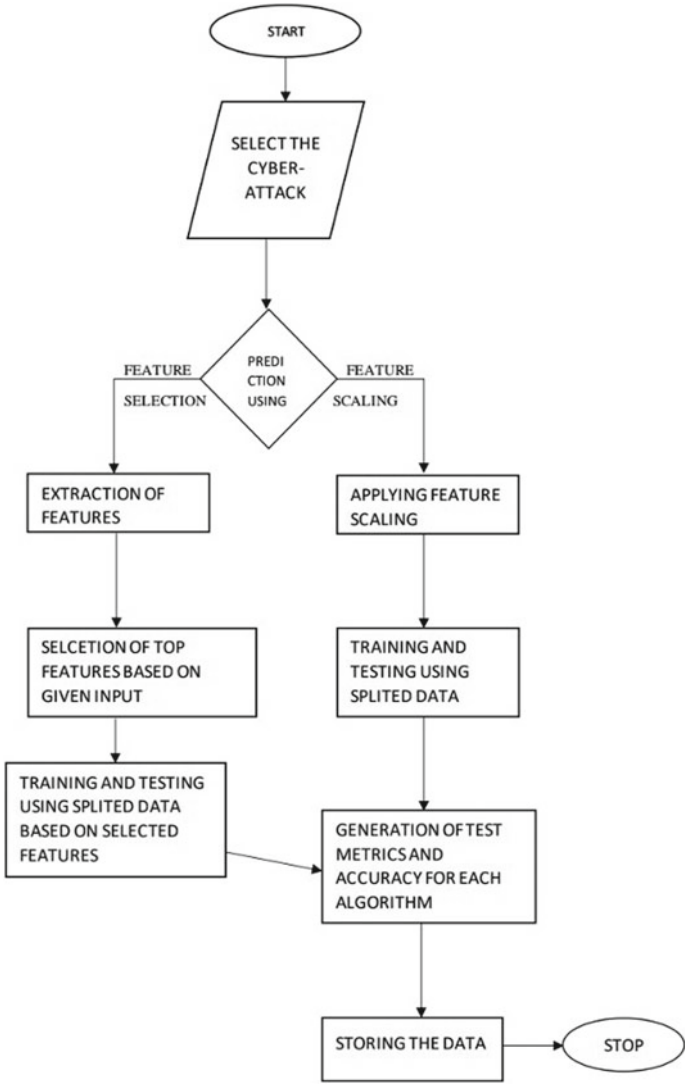


Fig. 1 Flow diagram for phishing, malware, and intrusion detections

4.4 Voting Classifier

In this module, the dataset is trained and then tested using five different models and then compared using the hard and soft voting classifiers in such a way that hard voting works are based on the majority and soft voting works are based on the probability (Fig. 2).

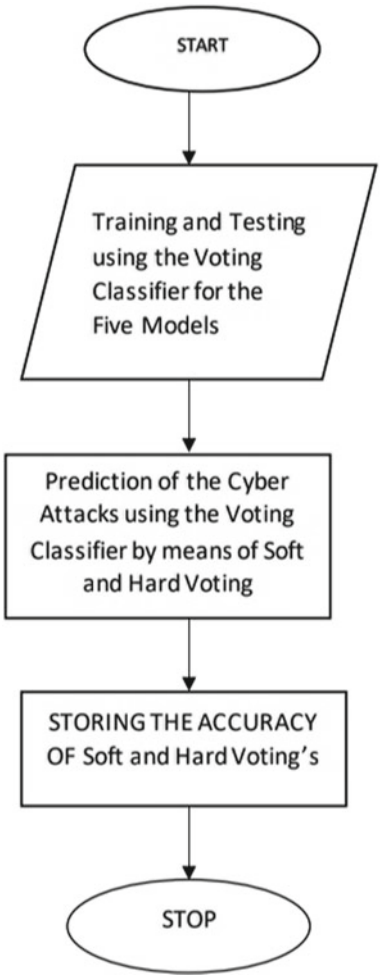


Fig. 2 Voting classifier

5 Workflow

See Fig. 3.

6 Results and Discussion

The proposed UEA is implemented using Tkinter as a front end and Anaconda Jupyter Notebooks. The dataset used for the intrusion detection is KDD CUP 99 dataset, [8] phishing dataset for phishing detection, and malware dataset for malware detection. The accuracy of phishing website detection using machine learning (ML) is more

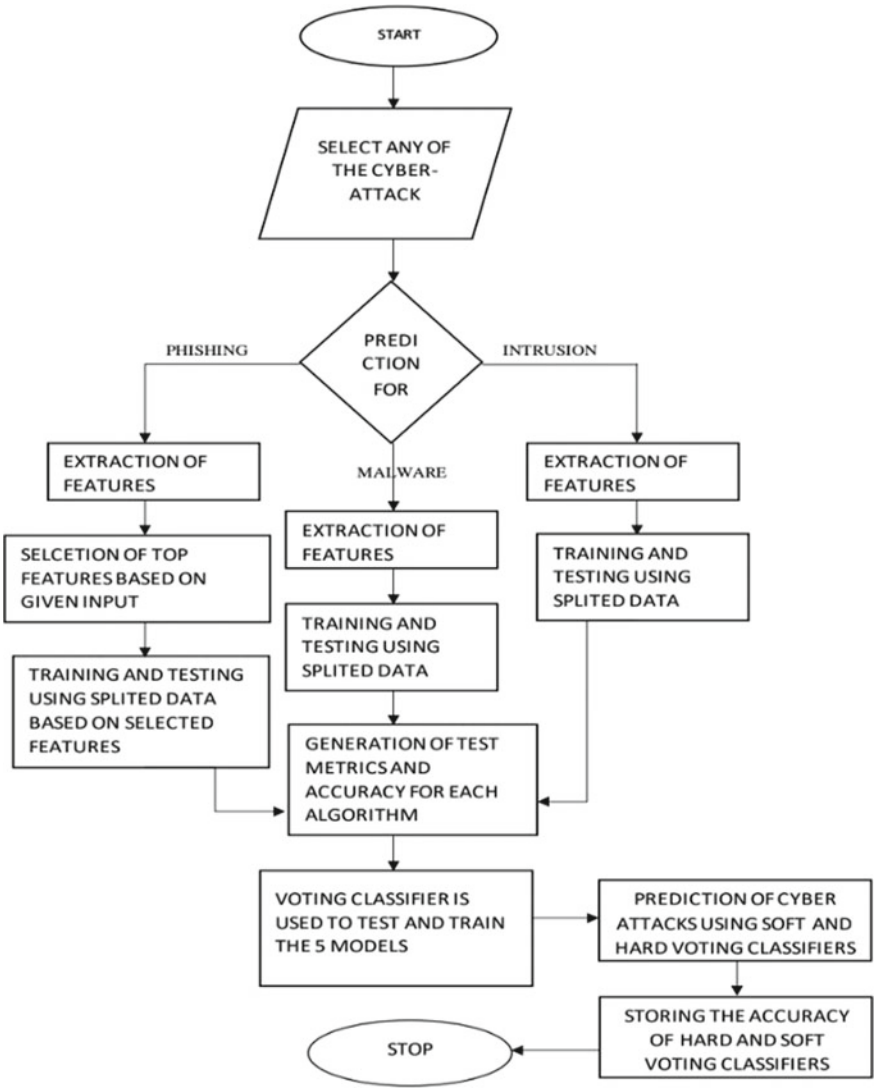


Fig. 3 Workflow

than 90%, and we cannot predict which algorithm shows the highest accuracy because we are using feature selection process, and accuracy changes for every execution; the highest accuracy of malware and intrusion is more than 90% which is given by random forest [9] without using the ensemble approach. The UEA voting classifier is combining all the five algorithms and compares them based on training and testing using hard and soft voting, and it is giving more accuracy than the single learning

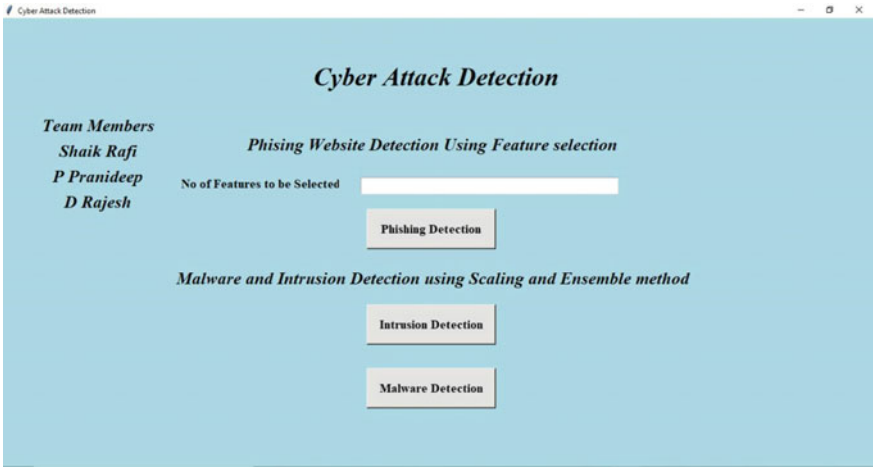
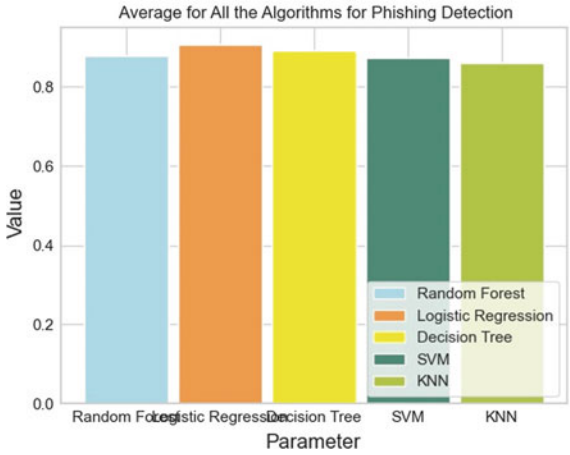


Fig. 4 User interface

Fig. 5 Phishing accuracy values



algorithms. The metric values are generated and visualized for all the three cyber attacks (Figs. 4, 5, 6, 7, 8, 9, and 10).

7 Conclusion

The proposed machine learning model combines phishing, malware, and intrusion detections into a unified model using ensemble approach which is not present in the existing approaches. The individual or the organization who needs to prevent the above cyber attacks can use the proposed unified model. This model detects

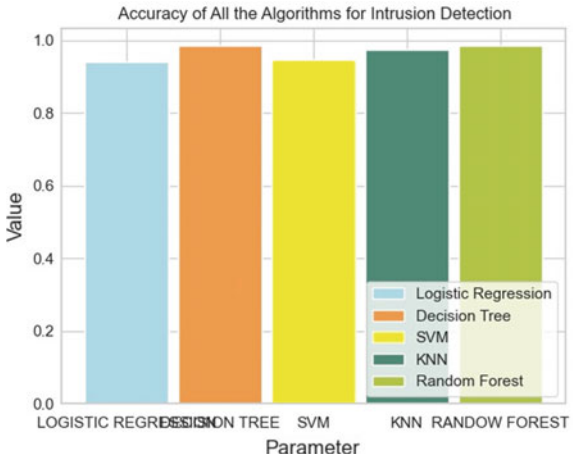


Fig. 6 Intrusion accuracy values

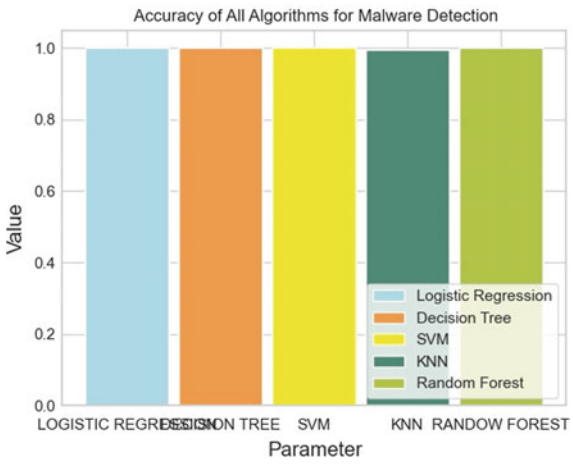


Fig. 7 Malware accuracy value

the phishing, intrusion, and malware attacks which primarily utilize online services to grab individual sensitive data and prevents the user from the cyber attacks. The accuracy of the proposed unified model is comparatively higher than the individual approach which is cost-effective for the users who are searching for a different system to prevent from each cyber attacks.

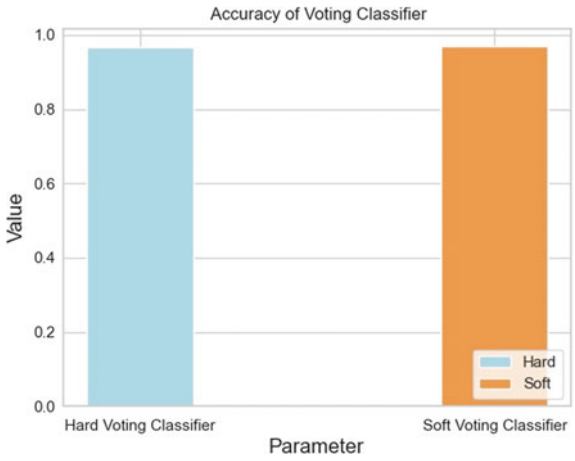


Fig. 8 Phishing voting classifier accuracy

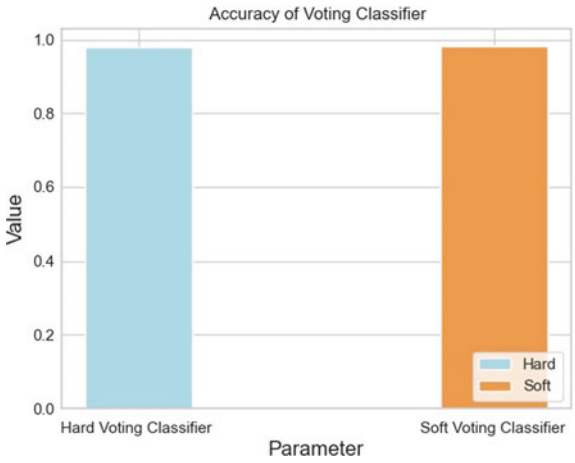


Fig. 9 Intrusion voting classifier accuracy

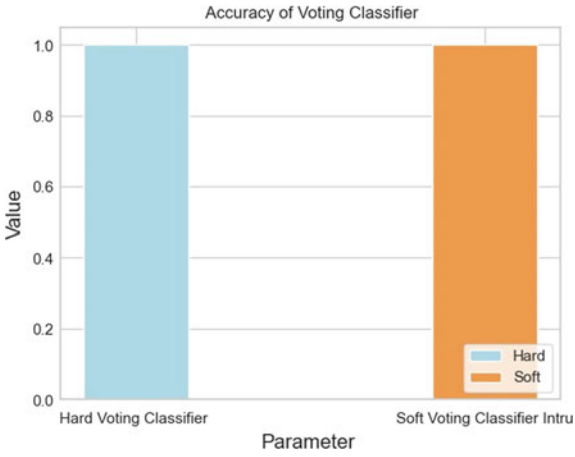


Fig. 10 Malware voting classifier accuracy

References

1. Fette I, Sadeh N, Tomasic A (2007) Learning to detect phishing emails. International conference on WWW., May 2007, pp 649–656
2. Sommer R, Paxson V (2010) Outside the closed world: on using machine learning for network intrusion detection. IEEE
3. Jisan BMY, Taher KA, Rahman MM (2019) Network intrusion detection using supervised machine learning technique by using feature selection. IEEE
4. Rahman A, Chowdhury M, Islam R (2017) Protecting data from the malware threats using machine learning. IEEE
5. Irfan R (2018) Phishing website detection using machine learning algorithms. IJCA 181(23)
6. Tiwari P, Singh R (2015) Int J Eng Res Technol (IJERT) 4(12), ISSN: 2278-0181
7. Breiman L (2001) "Random Forests", "Machine Learning" 45:5–32
8. Machine Learning Repository—UCI (2012) http://archive.ics.uci.edu/ml/
9. Ghosh S, Bandyopadhyay SK, Roy S (2012) A tutorial review on text mining algorithms 1(4)

Cyber Attacks and Security System Design Solutions in Emerging Markets and Vietnam

Nguyen Thi Phuong Thanh, Trung-Hieu Le, Hoang Van Thuc, Dinh Tran Ngoc Huy, Nguyen Thuy Dung, and Dinh Tran Ngoc Hien

Abstract The purpose of this study is to present the concept of cyber attack and cyber security risk, and security system design in Vietnam. The 2nd purpose is to principles of security system design solutions, as well as some solutions for preventing and detecting cyber attacks such as multi-story firewalls, and system solutions for a specific case study of FPT group in Vietnam. There are network and cyber attacks rising in recent years which put threats on security system in many companies and many developing countries. Authors mainly use qualitative analysis regarding to cyber security risks, cyber attacks and security system design for emerging markets including Vietnam. Therefore, this papers aims to analyzes it and proposes some recommendations for security system design solutions in emerging markets and Vietnam. For instance, authors propose multi-story firewall system, IDS/IPS intrusion detection and prevention system, as well as principles to set up security system design.

Keywords Cyber attacks · Multi-story firewall system · Security system solutions · Vietnam

N. T. P. Thanh (✉) · H. Van Thuc · N. T. Dung
Thai Nguyen University of Information and Communication Technology, Thái Nguyên, Vietnam
e-mail: ntpthanh@ictu.edu.vn

H. Van Thuc
e-mail: hvthuc@ictu.edu.vn

N. T. Dung
e-mail: ntdung@ictu.edu.vn

T.-H. Le
Dai Nam University, Hanoi, Vietnam
e-mail: hieult@dainam.edu.vn

D. T. N. Huy
Banking University HCMC, Ho Chi Minh City Vietnam - International University of Japan, Niigata, Japan

D. T. N. Hien
Ho Chi Minh University of Technology, Thành ph´ô H`ô Chí Minh, Vietnam

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_49

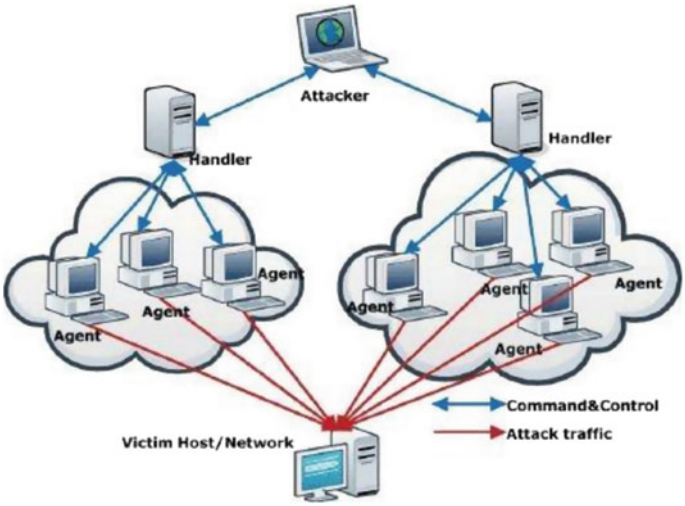


Fig. 1 Network and cyber-attacks. (*source* internet)

1 Introduction

Eian et al. [1] *mentioned in below figure that:*

Denial of Service (DoS) attacks (executed by teardrop or smurf attack): put threats on a system by blocking clients service requests and taking place during COVID-19 with effects of slowing down responses to corona viruses (Fig. 1).

Currently, cyber attacks are prone to growing quickly, both in terms of quantity as well as the sophisticated in each attack. The safety of the systems of enterprises requiring early detection and warning of abnormal changes, hazards are threatening the system. It is therefore necessary to have several solutions that improve cybersecurity security issues.

This study organized with introduction, research questions, literature review, main results, discussion and conclusion. We can conclude that: under Internet of things (IoTs) and network effects, there is threats from cyber attacks as well as hazards for the system in corporations.

2 Literature Review

We summarize previous studies in Table 1.

And Huy et al. [2] said that enhancing IT skills are needed for enhancing quality of laborers for EVFTA in Vietnam. Nam, Duong Thi Tinh, Dinh Tran Ngoc Huy, Trung-Hieu Le, Le Thi Thanh Huong (20210 stated that Internet of things (IoTs) has

Table 1 Summary of relating studies

Authors	Year	Content, result
Singh	2014	Proposed a 5 stage model to reduce effects from cyber risks or crime. This model applied for stages of cloud computing It helps to reduce cyber attacks through phases: cleaning computers, data storage and security during transmission
Nepal and Jaccard	2014	Presented an overview on software, network and hardware vulnerabilities and the need of developing malware defense mechanisms
Abomhara and Koien	2015	While Internet of things (IoT) becoming popular, the number of cyber attacks in IoT devices also increasing Hence, there is a need for cyber defense
Parikh and Patel	2017	Proposed cyber attacks solutions and recommended that we can use technology to reduce effects of cyber attacks
Eian et al.	2020	Describe their main components of cyber attacks including vulnerability, fear of attacks and spectacularity factor
Williams et al.	2020	Projected that expenses for cyber attacks in 2021 can be up to 6 trillion USD, and number of attacks increase 5 times more after COVID-19

various effects on Vietnam industries. This content also very good for enhance quality of IT trainers and teachers in Vietnam (Huy et al. [3]). And many job opportunities will come to laborer during EVFTA if they have IT skills (Hang et al. [4]). Moreover, in electrical and engineering fields, we need solutions preventing for cyber security risk (Huy et al. [3]). Also Thao et al. [5] stated e-government online public services will go with IT risk solutions to ensure quality. This topic is necessary in the context of many schools apply IT in history and geographic teaching (Huong et al. [6]).

We can compare our study to precious related studies that: Our study aims to analyze general issues of cyber attacks and cyber security risk and propose solutions of system design for a specific company in Vietnam, FPT. As we have seen above, not many previous studies has been done on this scope for Vietnam market.

3 Methodology

This paper mainly use qualitative analysis with technical issues relating to cyber attacks and security and network solutions.

We also propose to build a security system design principles in order to contribute to reduce cyber attacks and their consequences.

4 Main Results

4.1 The Principles of Security System Design

Cybersecurity must be set up based on the following guidelines:

- Deepening defense: it would be better to divide into some floors and various layers in the system in which each floor and class will be implemented with different security or containment policies. Moreover, we might note that, there is a precaution to prevent when a layer or a class is compromised, and it cannot be affected by other floors or classes.
- Different technologies can be used: company might recognize they might not rely on only one technology product for network security of the firm. It is more meaningful to apply defense policy with layered as the same products in the network will be easily passed and the stratigraphy. Hence, people might consider to use various technological products of various carriers to limit the down side when conducting stratigraphy. At the same time using a variety of technology and security solutions combined to strengthen the power of the defense system as a coordinated Firewall as the direct prevention tool, proactive defense reactions, antivirus to filter viruses…, etc.
- International standards: Today there are some international standard certifications ISO/IEC 18405:2005 EAL4, etc.

Based on the following criteria and guidelines, there are several groups of solutions needed.

4.2 Group Solution System for Preventing and Detecting Attacks

Planning and Design Solutions Team

SOA- Service—Oriented architecture is standard model applied to system of advanced network which became base for large network design.

- Designing infrastructure according to the SOA model
- Classification design method—Hierarchical
- Service deployment models and user management
- Partitioning virtual LAN (VLAN)

System Solutions for Preventing and Detecting Attacks

Multi-story Firewall System

This is the solution in which we can divide VLAN (i.e., a logical network of devices set up based on functional elements and applications) into various systems that can help to get the highest results, performance, security, management.

The firewall system is an access control system between the Internet and the local network. The firewall has both hardware element and software component. The hardware has a stable performance, does not depend on the operating system, virus, malicious code, blocking good protocols at the network layer in the TCP/IP Reference Model.

IDS/IPS Intrusion Detection and Prevention System

Currently, the attack forms of people with bad intentions are increasingly more and sophisticated. For example, in units that can manually install tools on the working PC or laptop to proceed to eavesdrop or scanning directly on the server, from which can retrieve email accounts, changing work planning, school fees, exercise scores….

IDS/IPS, functioning as an effective way, is an extremely important security system, it can detect attacks based on preset signs or malicious, unusual snippets on network traffic; at the same time, it is possible to remove them before they can harm the system.

We can conclude this section that: there are several solutions including multi-story firewall system or IDS/IPS system, etc. That follow certain principles in designing an effective solution.

5 Discussion

We can realize there are some popular security design models in past years, for instance: in below client server model, there are clients and server components (Fig. 2).

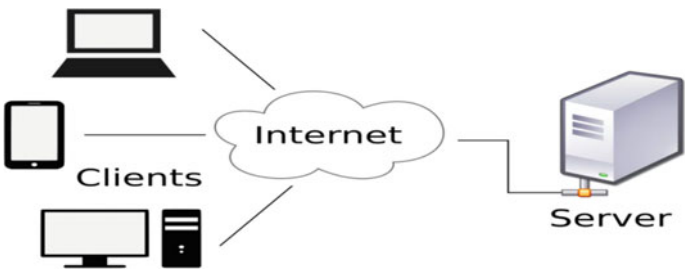


Fig. 2 Client server design model. (*source* internet)

Other Solution Groups

- System building updates, centralized fixes.

The first stage of hackers when conducting an attack is to survey the target system to find out the fault of the operating system, of the services, of the applications when they have not been updated on the vendor's website.

There are evidence hosing us that in many cases of enterprises and agencies, the use of software products is almost less updated with patches, which are also individually on personal computers, which is the opportunity for hackers to use the public to exploit security vulnerabilities. To update patches for all clients in the entire system over the Internet takes time and consumes a lot of bandwidth and consistency.

- Logging, monitoring, system monitoring
- Data encryption and transmission solutions
- User training
- Antivirus System.

We can conclude this section that: there are other groups of solutions that companies might consider to use to prevent cyber security risks and cyber attacks, such as data encryption, password and tokens, antivirus system, etc. This solution aims to create safety for system between clients and server.

6 Conclusion

In above section, we also suggest principles of security system design solutions, as well as some solutions for preventing and detecting cyber attacks such as multi-story firewalls, and IDS/IPS intrusion detection and prevention system.

Pivarnikova et al. [7] said that cyber attacks take place and increase in the world today and they can use Bayesian math to estimate cyber attacks, how they take place and recognize by dividing into stages to discover attacks earlier.

Beside, we also could consider other group solutions, for instance, in an example case of FPT group in Vietnam

Big Data technology that solves security algorithms

- Information security monitoring and FPT Incident Response service. Eagle Eye MDR (short for MDR is Managed Detection and Response) is capable of monitoring the system, analyzing unusual behaviors based on Big Data, collected from major customer episodes on Data Lake. The core technology used is Endpoint Detection and Response (EDR), sensors collects activity on endpoints to respond.
- Namely, FPT. Eagle Eye MDR can record the entire behavior that takes place on every server, the employee's workstation on a report screen; Monitor and control

the degree of compliance of personal equipment; Control the overall level of security across the entire system; Quickly detect the unsafe behavior, quickly perform analysis steps, process remove malicious code completely from the computer.

- FPT services. Eagle Eye supports effective monitoring of information safety and automatic security incident response. This is also a service offered in the first SaaS model in Vietnam, helping businesses quickly detect the network attacks and calculate the costs to be spent.
- Soon after, the experts experienced security and troubleshooting of FPT will make a warning, recommendation and quickly be present to accompany with the 24/7 troubleshooting business. Information will be continuously exchanged between FPT's team of cybersecurity experts and enterprises through the customer portal.

Solution designing or architecture

First, data will be collected regarding to structure of client and diagram of network.

Second, in case of FPT, applications such as antivirus, firewalls, such devices used to install for the agent to collect log. Then, center of FPT (cybersecurity) will receive data through an encrypted transmission channel (which guarantees the confidentiality of the information).

Here, we clean the data and transfer/transport to 2 parts:

- Log Storage: The log storage unit, which serves for several purposes such as Forensic investigation when there is a problem, machine learning…
- SIEM: Conducts data correlation analysis (Correlation) for unusual detection based on predefined laws (Security Rules) and displays warnings on the Dashboard.

Hence, we can go to conclude that: taking advantage of system design principles such as create many layer or multi floors for deepening defense, and other groups of solutions (Eagle Eye, Log storage, etc.) big corporations such as FPT in Vietnam can protect Big Data for clients and for itself. Last but not least, our study has tried to propose some principles and recommendations for security system design solutions in emerging markets and Vietnam.

References

1. Eian IC, Yong LK, Li MYX, Qi YH, Zahra F (2020) Cyber attacks in the Era of Covid-19 and possible solution domains. Preprints. https://doi.org/10.20944/preprints202009.0630.v1
2. Huy DTN, Hanh NTT, Hang NT, Nhung PTH, Thao NTP, Han LT, Sang DT (2021) General solutions for enhancing quality of teachers during globalization in emerging markets including vietnam-and some pedagogy psychological Issues. Psychol Educ 58(4)
3. Huy DTN, Van Ngoc P, Ha NTT (2021) Education and computer skill enhancing for Vietnam laborers under industry 4.0 and evfta agreement. Elementary Educ Online 20(4)
4. Hang NT, Tinh DT, Huy DTN, Nhung PTH (2021) Educating and training labor force Under Covid 19; Impacts to Meet Market Demand in Vietnam during Globalization and Integration

Era. J Educators Teachers Trainers 12(1). Retrieved from https://digibug.ugr.es/bitstream/handle/10481/69215/JETT_Volume%2012_Issue%201_Pages%20179-184.pdf?sequence=1
5. Thao NTP, Van Anh N, Huong LTT, Huy DTN (2021) Measuring citizen's satisfaction when using E-Government online public services: lessons from Vietnam. J Contem Issues Bus Gov 27(1)
6. Huong NM, An TTL, Huy DTN, Thanh NTP, Lan NTP, Huong CTM, Van Nam M (2021) Teaching methods of history and geography and it influence for students in primary schools, Vietnam. J Educators Teachers Trainers 12(1)
7. Pivarnikova M, Sokol P, Bajtos T (2020) Early-stage detection of cyber attacks. Information 11(12):560. https://doi.org/10.3390/info11120560

Alzheimer's Disease Detection Using Deep Learning-CNN

Aditya Singh, Nishad Kharkar, Patel Priyanka, and Suhasani Parvartikar

Abstract Alzheimer's disease is a stage-progressive brain disorder and also an incurable disease that causes the brain cells to shrink and die. One of the most common cause of dementia is Alzheimer's disease. Treatment for this disease has not been found yet. Early detection of Alzheimer's disease may help doctors to conducts trials which may temporarily improve or slow progression of symptoms and prevent brain to tissue damage. The proposed system predicts Alzheimer's disease using a deep convolutional neural network with the help of brain MRI scans. This model can deliver superior performance for early-stage diagnosis thus helping the doctor to diagnose the Alzheimer's disease accurately and lower the number of false negative results in those critical cases.

Keywords Alzheimer's disease · Deep learning · MRI data · Old age

1 Introduction

The human brain is the central organ of the human nervous system. It controls the activities of the body, stores the information into the memory, processes it, integrates, and coordinates the information it receives from the sensory organs and makes decisions.

Alzheimer's diseases also called as senile dementia is a progressive form of dementia or neurological brain disorder that causes the brain to shrink (atrophy) and brain cells to die. For most people, first symptoms appear in their mid 60s, but there are people whose first symptoms appear during their mid 30s called early onset Alzheimer's. It is not known that what exactly causes it and when does it starts, but how the process takes place [1], how it starts spreading, and slowly, how it deteriorates the brain is known. It takes lot of time for this process to happen before showing its first symptoms. As earlier mentioned, this disease is incurable, and it gets worsen over the time. Patients with this disease slowly lose their memory then unable to recognize the language, people, or anything. A patient may suffer from

A. Singh · N. Kharkar · P. Priyanka · S. Parvartikar (✉)
Department of Computer, Saraswati College of Engineerirng, Kharghar, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_50

cognitive impairment such as problem in understanding things and unable to make judgments, difficulty in recognizing language, or speaking or reading. They also suffer from behavioral changes and mood swings such as anger, hallucination, and lose track while walking, and lastly, patients get paralyzed. To diagnose it, patient goes through CT and MRI scans and also some test such as mental status test, blood, and urine test. Doctors also check for patient's family history [2].

Our system proposes the early detection of the Alzheimer's diseases based on the given X-ray images of the brain with the help of deep learning-CNN-ResNet architecture at early.

2 Literature Survey

We are going to analyze and assess different methodologies that have been implemented earlier. This helps us to draw a better picture of the advantages and shortcomings, also other theories, so as to develop a good understanding for our work in this field of deep learning. This does not classify as research, but a rediscovery and acknowledgment of work done previously in this field.

Moradi et al. [3] proposed the paper on how the authors have done their research on recognizing mild cognitive impairment (MCI) as an alteration stage between Alzheimer's and age-dependent cognitive decline. The authors also suggest MRI-based biomaker using ML techniques on Alzheimer's disease neuroimaging initiative ADNI database. Paper claims that their algorithm has achieved the score of 0.9020 under the curve (AUC) in distinguishing between (pMCI) and (sMCI).

Several different techniques of image processing and then feeding those back into machine learning algorithms were the workflow adopted by the authors. Monitoring the MCI patients and their age to see if they turn out to have Alzheimer's was one of the main focus as well.

Chetan Patil et al. [4] put forth an approach for ahead of time detection of Alzheimer's disease with the help of image processing technique on MRI images. The authors have used k-means clustering, watershed algorithm, wavelet transform, and a tailored algorithm for their analysis. In their proposition, the T1-weighted MRIs, boundary detection algorithm with k-means clustering were used for image processing for evaluating hippocampal atrophy, to draw out the region of interest, segmentation, respectively. Authors are confident about their proposition and how it can be useful to the community.

Having delved thoroughly the discoveries of the given research paper, we found out that the said paper in question focused more on the hippocampus of the brain. That being said, it took into account not only the medical history of the patient, but also the hippocampal volume of the patient where a direct link between the volume being less and the patient having AD was found.

Mahmood and Ghimire [5] proposed an automated functioning system relying on mathematical as well as image processing techniques for classification of Alzheimer's disease. They have used principal component analysis to decrease high dimensional

vector space into 150. They have completed the training and testing of the classifier using the OASIS database. According to authors, their proposed approach gathers approximately 90% of accuracy.

Exploring the research paper, we got to know that they have focused on dimensionality reduction of the image using PCA and then applied it to generic feedforward model for classification.

Acknowledging the mentioned authors and their commendable work in the said field, we got to know that hippocampus part of the brain can help us to detect AD; so, the dataset we used are the MRI scans of hippocampal part. Also, applying different technologies on images and then passing it to ML algorithms, we adopted the idea to feed the MRI data to our deep learning algorithm, to have a reliable result to detect AD in early stages. And for dimensionality reduction of images, instead of going for PCA and then a generic feedforward model, we went for ResNet 18 architecture. This simplifies the process since ResNet does it for us in every pooling layer of the architecture [6].

We shall emphasize on the validation losses along with the accuracy. This will lead to eventual lower number of false negative results, which our team focused on greatly. We shall work on achieving this by taking the simplest routes, helping the general populous to understand the whole process without much confusion.

3 CNN

Convolutional Neural Network (CNN) is one of the deep learning algorithms which gives quite good results then other classification algorithm because it doesn't require much preprocessing. Also, coming to complicated images, it gives good accuracy than other methods which have strong pixel dependencies throughout network in fully connected layer which is not the case in CNN making it easier to train [7].

A CovNet architecture is better in lowering the number of parameters used and the reuse of weights; also, using the application of relevant filters, it successfully captures, spatial and temporal dependencies inside an image. Training the network can help understand the image preferably.

There are four layers of CNN:

1. **Convolutional layer**: This layer uses matrix filter and executing convolution operation to reveal patterns within the image by moving the filter all over the image and multiplying input pixel and filter pixel corresponding to each other. Then, add them up and divide it with total no. of pixel.
2. **ReLU layer**: ReLU activation function is added to the convolution layer to gather a correct feature map of an image. It replaces every negative input values with 0.
3. **Pooling layer**: Pooling is a compression operation that decreases the dimensionality of a feature map by moving the window across the image and picking the max value from each window.

4. **A fully connected layer**: The final layer where actual classification of the image is done.

4 Proposed System

CNN has various architectures like VGG Net, ResNet, Dense Net, Inception Net, AlexNET, and many more. For this particular project, we are using ResNet architecture. Figure 1, the proposed system uses CNN to classify the Alzheimer disease [8].

A. **Data Collection**: OASIS dataset consists of 6.4 K brain images of both Alzheimer's and non-Alzheimer's patients considering the three scan angles as individual truth thus expanding the dataset by 3x.

B. **Data Preprocessing**: Data preprocessing of images generally involves data augmentation. Data augmentation of an image is a technique where we enlarge the size of our training dataset synthetically by creating modified or altered versions of images. Training neural network models with more data yield more accurate models, and this also helps in creating diverse images which will refine the ability of the fit models as it got new images to learn and generalize. In our project, we are augmenting dataset also flipping it horizontally and vertically.

C. **Dividing Dataset**: The dataset is separated into training and testing set. Training dataset is applied to train the model. Here, the model we use is ResNet. Testing dataset is applied to test our trained model.

D. **Choosing Model**: CNN model which we are using to classify the Alzheimer's diseases is ResNet 18 architecture (Fig. 2).

(1) Creating a deep network that is filled with numerous hidden layers doesn't necessarily mean that it will help make the network perform better, rather it creates a problem that we refer to as the vanishing or exploding gradient problem. This is where ResNet comes into effect.

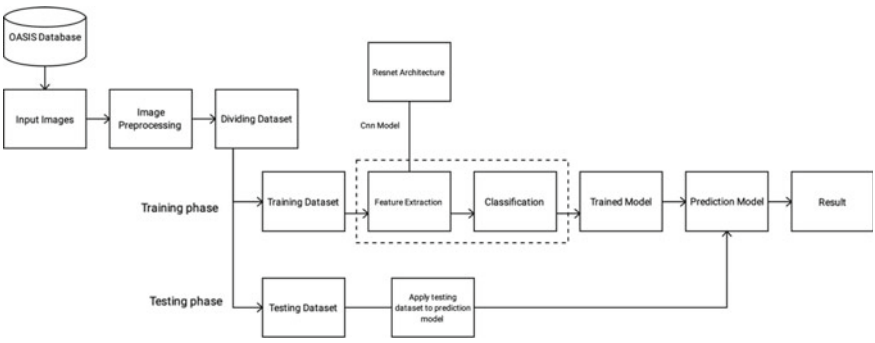


Fig. 1 Flow diagram of proposed methodology

(2) With the support of residual blocks, ResNet enabled us to increase the number of hidden layers that are enough without dwelling upon the vanishing or exploding gradient problem. Residual blocks permit the network to keep a track of things learned before by possessing an identity mapping weight function where the output is equal to the input, irrespective, if the layer got skipped or dropped during propagating the network.

Model Details:
Resnet18 architecture contains 17 convolutional layers and 1 fully connected layer. Figure 3 below shows the architecture of ResNet 18.

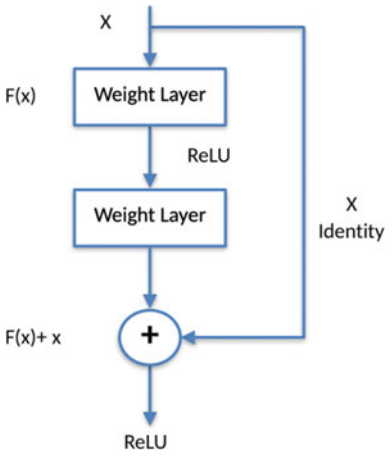


Fig. 2 Residual block

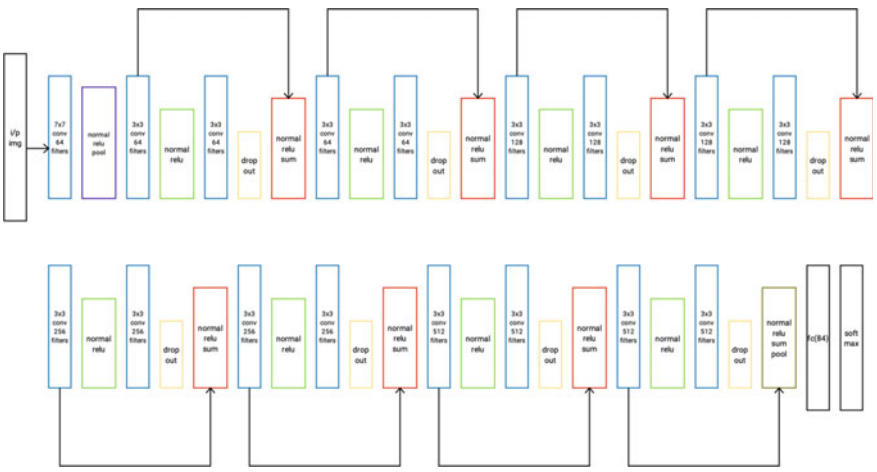


Fig. 3 Resnet18 Architecture

1. First convolutional layer contains 64 filters of 7 × 7 size followed by normal ReLU and pooling layer. Convolution layer basically serves neural network filters also called as feature maps used to capture the features from the input image so to gain some understanding of what features our CNN detects. The purpose of applying the rectifier function (ReLU) is to increase the non-linearity in our images.
2. Remaining convolutional layers are of 3 × 3 size but contain different no. of filters. After first convolutional layer, next four convolutional layers have 64 filters; then, next four have 128 filters; then, next four have 256 filters; then, next four have 512 filters.
3. Between each convolutional layers, a normal ReLU layer and a ReLU summed with the input of the previous convolution layer are placed alternately forming a residual block. And at the end, we have fully connected layer with softmax. Pooling layers help decrease dimensions of the feature maps (Fig. 3).

A. **Evaluation**: After the model is trained, evaluation is done. It enables testing the model with data that is unseen and unused.
B. **Prediction:** We are now ready to use our model inferring results in real-life scenarios.

5 Implementation

The project is designed and tested on Jupyter platform. As the project name suggest, we are using deep learning, and so, we have used PyTorch library to implement CNN. The project is divided into two parts first model building and second predict the outcome using the build model.

Firstly, for building model, OASIS database is used. Initially, dataset was divided into training and testing. We have 6.4 K images; out of which, 2560 images are used for training, and 640 are for validation. PyTorch vision transforms are used for image processing such as resizing, horizontal flip, apply tensors, normalizing and converting images into RGB. For testing data, we don't apply horizontal flip. Next, data loaders are prepared in which we pass datasets and above mention transforms. This way a custom dataset is made.

After this, batches were created as we can't use whole training dataset at every training step. Keeping batch size of six, we get total training batches 844 and number of test batches 214. Shuffling is done, so we don't get same batches at every epochs making the model robust.

We have used pretrained ResNet18 architecture. Training process starts here where training dataset is passed through ResNet model, and training is done. We calculate lose function, do back propagation and then optimizers are to reduce losses. After every 10 steps, evaluation is done and validate on test dataset. Validation loss and prediction accuracy are calculated. Same is done for every epochs. Training stops when we get accuracy between 85 and 90% with minimal validation loss, and model is saved for further use.

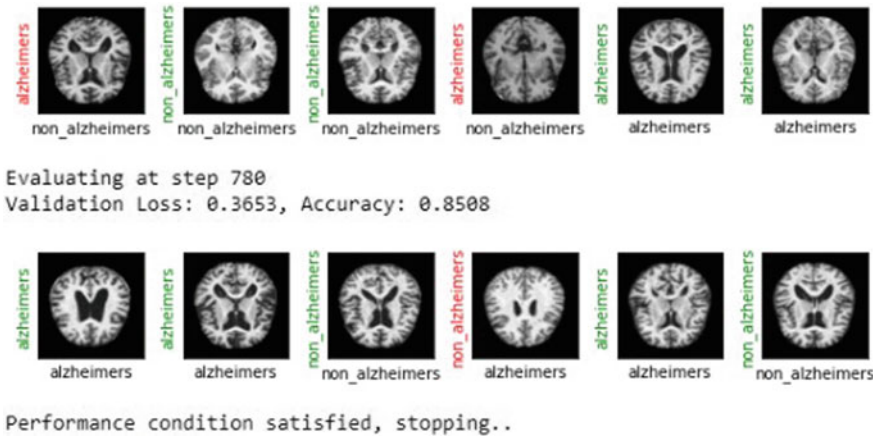


Fig. 4 Training process

For second part of the project, we take the image from user. For creating user, interface Tkinter is used.

After getting image, we apply above mentioned transforms for image processing. ResNet18 model is called, and our saved model is loaded into it (new ResNet model). The preprocessed image is passed through this new ResNet model. Figures 4, 5, 6 prediction is made then which says if patient is suffering from Alzheimer's or not.

6 Result

7 Conclusion

We believe that the use of ReLU activation function was not tried but was necessary to eliminate vanishing gradient problem, also not using too many layers to avoid overfitting and slowing the system as well as the application. Creating this program with this methodology, fine tuning the system, we were successful in a cognizant and time-bound approach.

This program can act as a very solid stepping stone and a framework in all directions of further research into this disease.

In this application, we are trying to emphasis on the premature detection of the Alzheimer's disease as it is the best possible scenario. We aim for our project to aid the doctors in their final verdict and early detection for a particular patient so that he/she has a good, and a reliable source of information regarding the patient's condition, helping the patient to have an early diagnosis and help the doctor to further improve the patient's quality of life.

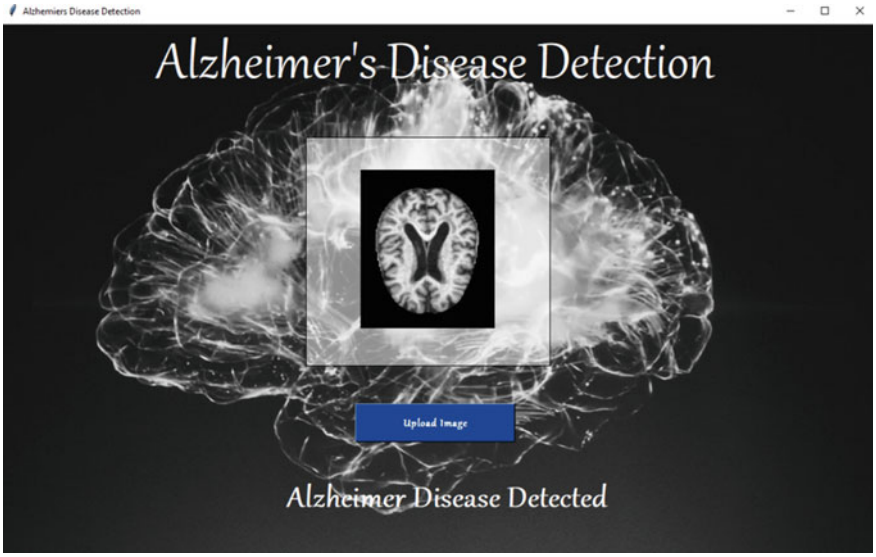


Fig. 5 GUI showing Alzheimer's detected

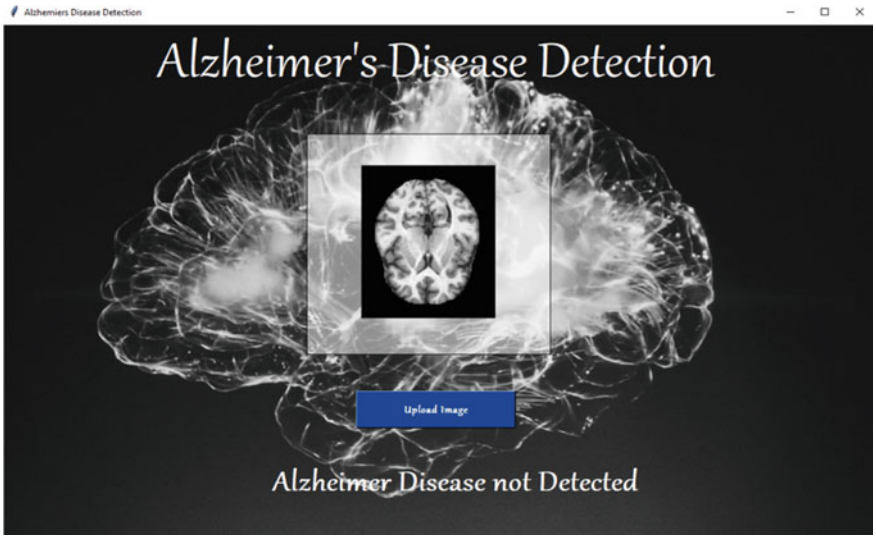


Fig. 6 GUI showing Alzheimer's not detected

Acknowledgments The background images used in GUI for our project is from unsplash; they provide worldwide copyright license to download, copy for free. Also, the dataset used in our project is from OASIS.

References

1. Evanchalin Sweety M, Wiselin Jiji G Detection of Alzheimer disease in brain images using PSO and decision tree approach. 2014 IEEE international conference held on advanced communications, control and computing technologies.https://doi.org/10.1109/ICACCCT.2014.7019310
2. Saraswathi S, Mahanand BS, Kloczkowski A, Suresh S, Sundararajan N Detection of onset of Alzheimer's disease from MRI images Using a GA-ELM-PSO Classifier. 2013 fourth international workshop held on computational intelligence in medical imaging (CIMI). https://doi.org/10.1109/CIMI.2013.6583856
3. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2014) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, Corpus ID: 2105617. https://doi.org/10.1016/j.neuroimage.2014.10.002https://doi.org/10.1016/j.neuroimage.2014.10.002
4. Chetan Patil, M, Madhumitha, SD, Fernandes M, Venugopal A, Unnikrishnan B Early detection of Alzheimer's disease using image processing on MRI Scans. 2015 IEEE international conference held on signal processing, informatics, communication and energy systems (SPICES)https://doi.org/10.1109/SPICES.2015.7091517
5. Mahmood R, Ghimire B Automatic detection and classification of Alzheimer's disease from MRI scans using principal component analysis and artificial neural networks. 2013 20th international conference on systems, signals and image processing (IWSSIP). https://doi.org/10.1109/IWSSIP.2013.6623471
6. Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J, Yuan T-F (2015) Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. https://doi.org/10.3389/fncom.2015.00066
7. Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehéricy S, Benali H (2009) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51(2):73–83. https://doi.org/10.1007/s00234-008-0463-x, Epub 2008 Oct 10.
8. Arpita Raut, V A machine learning based approach for detection of alzheimer's disease using analysis of hippocampus region from MRI scan. 2017 International Conference on Computing Methodologies and Communication (ICCMC). https://doi.org/10.1109/ICCMC.2017.8282683

Automatic Detection of Leukemia from Blood Cells Using Soft Computing Methods

K. Aldrin Karunharan and X. Anitha Mary

Abstract Hematologists have benefited greatly from medical image processing techniques for the diagnosis and investigation of various illnesses such as leukemia. Soft computing algorithms can be used to effectively treat chronic granulocyte leukemia. Proliferation of white blood cells results in leukemia, a lethal disease with an unknown cause that generates fear of death. Leukemia can be detected in white blood cells. However, these diseases can be minimized and controlled by the early diagnosis and treatment. In this paper, we have discussed about clustering and histogram based techniques to detect leukemia blast cells and evaluated their performance based on error rate and computation time.

Keywords Gradient · Immunophenotyping · Etiology · Primitive · Eccentricity · Obscuring · Manifestation and contingent

1 Introduction

Flow cytometry, fluorescent in situ hybridization [FISH], and immunophenotyping were employed in the early days to diagnose leukemia and hematological neoplasms. Diagnosis of chronic lymphoproliferative disorders, ALL [Acute lymphoblastic leukemia] and AML [Acute myeloid leukemia] can be furthered as of the recent developments of these smart algorithms.

Soft computing techniques can be used to diagnose chronic lymphoproliferative diseases, ALL [acute lymphoblastic leukemia], and chronic granulocyte leukemia. Proliferation of white blood cells results in leukemia, a lethal disease with an unknown cause that generates fear of death. Early detection and treatment, on the other hand, can help to limit and control these disorders.

The general steps involved in the detection of leukemia are given in Fig. 1.

K. A. Karunharan (✉) · X. A. Mary
International Maritime College Oman, Karunya University, Coimbatore, India
e-mail: aldrin@imco.edu.om

X. A. Mary
e-mail: anithamary@karunya.edu

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_51

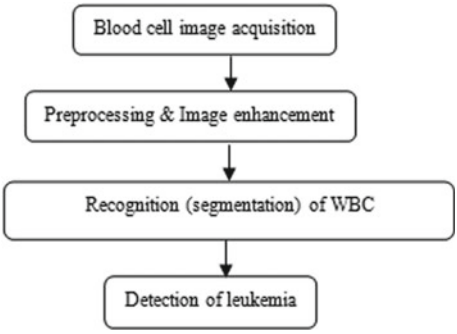


Fig. 1 General steps involved in detection of leukemia

The distinction between chronic and acute leukemia is that the former is caused by an increase in the number of slow-progressing cells, while the latter is caused by an increase in the number of immature blast cells. As indicated in Fig. 2, the standard identification procedure consists of the following five steps.

The membrane must first be recognized in order to successfully separate and accurately identify following neighboring cells. When it came to eliminating noise, the median filter performed admirably. For segmentation, the most crucial step is to transform the image's Red Green Blue model [red, green, and blue] to HIS [Hue saturation and intensity].

The performance can be assessed by comparing the thresholded image to the original gray image. The opening technique also removes undesired pixels and artifacts. Only the finally segmented nucleus and its extracted features can be used to identify lymphoblast cells. The SVM approach outperformed the traditional FISH approach.

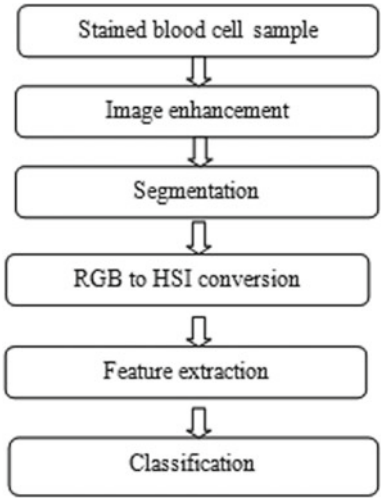


Fig. 2 Steps involved in classification of leukemia

2 Literature Survey

Somasz Mickiewicz and Stanislaw Osowski classified myelogenous leukemia using Support Sector machines for their paper for recognition myelogenous blood cancer using support vector machine algorithm in the year 2005 [1] and have addressed the issues such as unequal distribution of species and closely parallel life stage along with improved species and life stage recognition which was proved by comparing the sensitivity [SE] and positive prediction values [PPV]. It can be concluded that the species recognition technique can be extrapolated to further similar sets of corresponding features. Patil Tejashri G. & V. B. Raskar were very successful in finding the radius and center coordinate form each lymphoid cell affected with blood cancer even if the cells are connected on different points. The ground fact that leukemia is a cruel disease which affects the blood, bone marrow, and the lymph nodes [2]. They achieved this by polynomial approximation and curve segmentation which consisted of three major steps namely circle modeling, circle adapting, and circle merging.

We propose and implement a fully automated system based on image processing to assist in the diagnosis of acute lymphoblastic leukemia in human blood samples by recognizing and counting contaminated white blood cells. For early illness detection, the approach delivers excellent speed, accuracy, and scope. The algorithm is developed in MATLAB, and it achieves an average accuracy of more than 90%.

3 Segmentation of Leukocytes

A gradient primarily based eolotropic diffusion model with an accommodative electrical phenomenon perform is utilized and analyzed for the segmentation purpose. Eolotropic diffusion is Associate in Nursing unvaried procedure [3]. In eolotropic diffusion, the most common strategy is to smooth the consistent parts of the image, while increasing the sides, resulting in a piecewise constant image from which segmentation boundaries can be easily derived. Perona and leader were the ones who first predicted eolotropic diffusion. The Perona-Malik flow is frequently created with u as the initial image. Is the sting stopping perform that permits edges below sure strength to be ironed and stronger edges to be sharpened? The divergence operator is denoted by div, and the gradient operator is denoted by g.

This study employs continuous partial differential equations as geodesic active contours, but instead of developing a single curve, the level set of each image is evolved. In edge-based curve evolution, a geodesic active contour may be a standard methodology. Wherever, the GAC formulation is given, and u is the Gaussian ironed image. This definition of g introduces a nonlinear scaling to b.

This evolution arises from the native decrease in the complete curve length weighted by g, excluding the constant enlargement term. GAC has a problem with concavities inside the boundary 1. Because moving within the concavity might increase the weighted curve length, once the curve hits a small gap within the

boundary, this can be an area minimum. Going into this concavity, on the other hand, results in a better response in the end. The vector fields are only visible in the concavity's immediate vicinity. Smoothing occurs in the tangential direction of the image's amount sets, whereas sharpening occurs in the orthogonal direction.

The median filter's output could be a smudged replica of the first image. Several details in the image have been eliminated, most likely because they were just one component wide and height. There is no distinction made here between edges and flat sections (Figs. 3, 4, 5, 6).

The results of two alternative segmentation methodologies are compared to a typical district developing portioned strategy, with the purpose of identifying new ways for object-based classification and data fusion. The key operation in an automatic differential blood cell counting method is nucleus segmentation. The objective here is to have the capacity to consequently recognize the core of the WBC, having an expert portrayal of what can be distinguished in the picture. Both the fuzzy clustering method and the suggested nonlinear diffusion model are iterative procedures with promising outcomes.

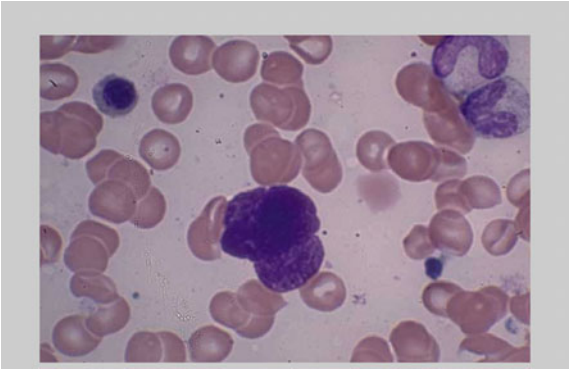


Fig. 3 Leukocyte image

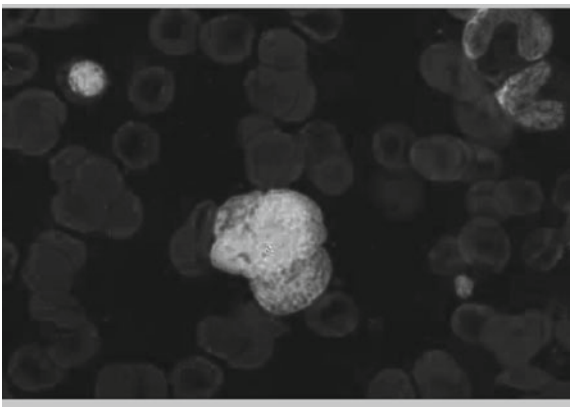


Fig. 4 Segmented image

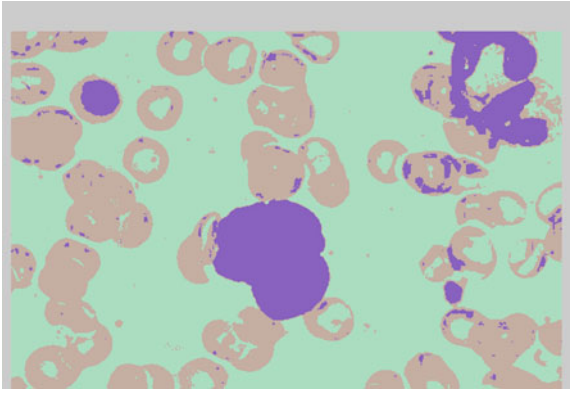


Fig. 5 Cluster labeled segmented image

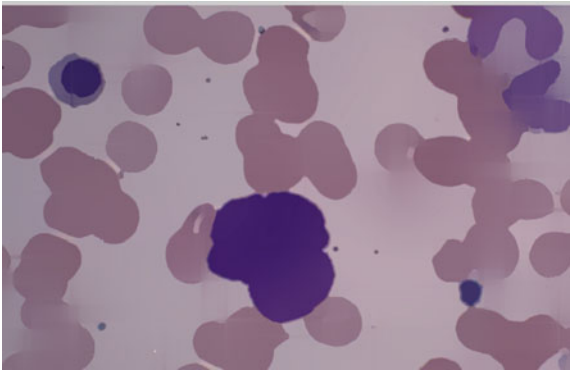


Fig. 6 Final segmented image

Because this extension gives higher differentiation using pattern dimensional space, easier classification is achievable. FLANN allows for a faster pace of training convergence. The following paragraph goes over the step-by-step approach for segmenting lymphocytes in a two-dimensional feature space using this suggested supervised FLANN (Fig. 7).

Due to the uneven shape of each cell, separating clustered cells is a significant challenge. Functional link model can be used to separate the groups. However, applying the FLAAN segmentation algorithm to the gradient image rather than the original image allows for more accurate image segmentation. The "nearest" minimum is the one found at the end of the sharpest decline path. In terms of topography, this occurs if the point is within that minimum's catchment basin. The dissociation of clustered cells is seen in Fig. 6. Our segmentation is observed to follow the nucleus of each cell's border, resulting in a non-circular appearance (Table 1).

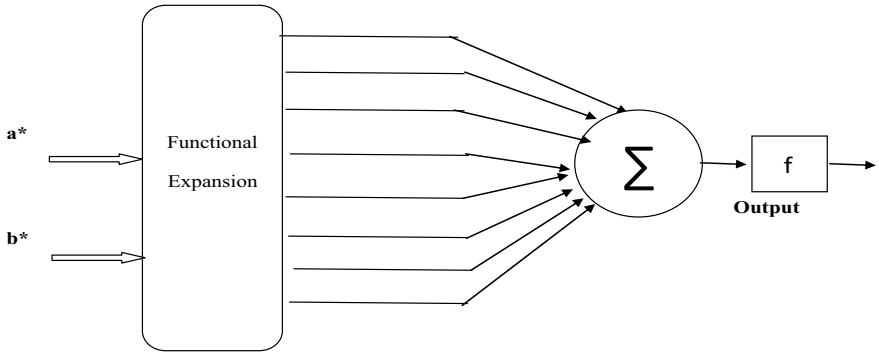


Fig. 7 FLAAN structure for pixel classification

Table 1 Training pattern corresponding to assigned pixel class label

*a**	*b**	Index
0.02609	0.4235	D1
0.2488	0.4356	D1
0.8538	0.2345	D2
0.8514	0.9342	D2
0.1287	0.8567	D3
0.2312	0.9234	D3

4 "Kernel Induced Rough Fuzzy C–means Algorithm [KIRFCM]"

The following steps are given and explored for removing the nucleus and cytoplasm regions from the background and segmentation of each WBC sub image kernel induced rough *C* means algorithm. The conversion of the original color image into RGB format is the first step. The $L * a * b$ image must then be created using the $L * a * b$ color space conversion. After that a feature vector for the input must be created. Afterwards input feature will be transformed onto higher dimension feature space by means of a nonsynchronous charting role.

As a matter of fact, a kernel nonlinear function is then used to do preliminary *C* clustering in this feature space. After that the labeled image is extracted after the huddled outcome. The red–blue–green color image is then segmented and reassembled for each class, which represents a distinct morphological region. Light microscopy is a quantitative microscopy detecting process that is qualitative (Table 2; Fig. 8).

This theory proposes a creative picture preparation and machine knowledge-based framework for quantified depiction of lymphocyte images and recognition of ALL in PBS images. Every lymphocyte image is separated into a developing lymphocyte or a lymphoblast using the predicted method (Table 3).

Table 2 Comparison of algorithms based on error rate

Sample	FLANN	KIRFCM
1	0.3456	0.2345
2	0.4213	0.2465
3	0.3567	0.2678
4	0.2987	0.2652
5	0.3198	0.3100

Fig. 8 Error rate of FLAAN and KIRFCM

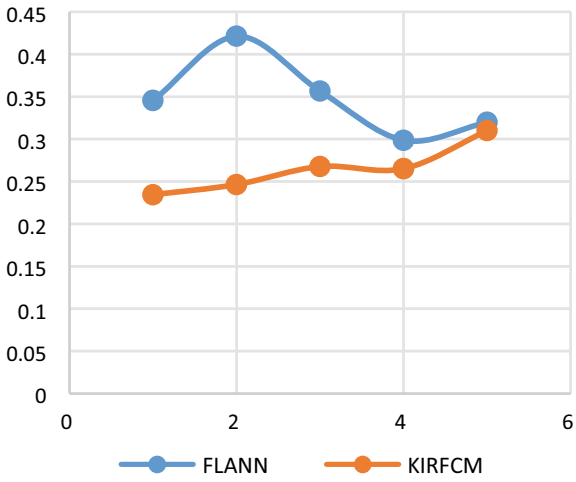


Table 3 Image information for each segmentation scheme

Segmentation scheme	Underlying issue	Appearance data
FLA neural network	Cataloging	Color concentration
KIR fuzzy C means	Bundling	Contextual

Falling division can be used to create PC-supported differentiating proof frameworks, which contain deliberation and order subsystems.

In this endeavor, the morphological, textural, and shading highlights are separated from the lymphocyte division presented in this part.

Table 10.4 summarizes the current visual benchmarks for detecting lymphoblasts in blood tests, which are used by many hematopathologists around the world. The morphology contrasts in this table are easily discernible. Affording to expert reflection, which is universally stared as a few examples, the cell size of large lymphocytes is comparable to that of small-scale bearings. Other morphological confinements, such as nucleus–cytoplasmic (N:C) proportion and core chromatin development, are considered crucial critical variables for the appearing in these examples.

A thorough explanation of each registered highlight's clinical notoriety may be found in Table 4. The following morphological, textural, and shading highlights can

Table 4 Comparison of features of lymphoblast

Feature	Lymphocyte	Lymphoblast
Cytoplasm nucleus ratio	Low	High
Cell size	Small	Large
Nucleus size	Less	Large
Nucleus shape	Oval	Intended
Nucleus perimeter	Suave	Bumpy
Cytoplasm density	Abundant	Light
Nucleus color	elaborate	Red florid
Cytoplasm perimeter	Smooth	Rough

be easily retrieved from the core and cytoplasm image parts of each lymphocyte picture's parallel, dim, and shading image variations, respectively.

FLAAN and KIRFCM use features that are distinct from those used by other classifiers. The dark blue and violet regions were measured for diameter, outline width and height, area, fullness, blood sample image, center coordinates, color standard deviation, color mean, and count of occurrences, as mentioned in the background context. The average area was calculated in cases where more than one area was present.

Using the FLAAN and KIRFCM to automatically detect and count lymphoblasts intends to provide further support to medical practitioners when diagnosing acute lymphoblastic leukemia. Self-learning methods could be added to the algorithm to improve accuracy and enable better separation of clustered lymphocytes.

5 Conclusion

The project's assessments will be utilized to see how good the above classifiers are at detecting leukemia in the PBS image. Presentation measures such as specificity, accuracy, and compassion are determined using the matrix misperception, which illustrates the differences in opinion between the classifier and the hematopathologists. In the binary classification, the positives and negatives are evaluated rejected and recognized accordingly. KIRFCM outperformed FLAAN in the statistical terms and parameters. The system provided 92.76% accuracy on the image data set that was examined. When compared to parallel computing approaches, the computational speed of KIRFCM and FLAAN is significantly faster. With the use of advanced classifiers, the accuracy level is expected to increase to 98% in our further research.

References

1. Mickiewicz S, Osowski S (Aug 2005) Automatic recognition of the blood cells of myelogenous leukemia using SVM. Proceedings of international joint conference on neural networks
2. Patil Tejashri G, Raskar VB (Jan 2015) robust segmentation method for acute lymphoblastic leukemia detection. ISI Int J Res Comput Softw Eng 5(1)
3. Sheikh, Michelle (March 1996) Blood cell identification using neural networks. Proceedings of the IEEE international conference on bioengineering

Survey on Edge Intelligence in IoT-Based Computing Platform

R. Ramya and S. Ramamoorthy

Abstract The enormous development of Internet of things (IoT) and the establishment over Internet services with cloud have moved to another form computing environment, i.e. edge computing; it can process the information towards the end of the organization named as edge nodes on the cloud-based network environment. The issues related to the network response time requirement, transmission capacity, data security and privacy on cloud infrastructure are effectively addressed by the edge computing technique. The proposed research paper listed out various challenges in clouds computing, the definition and layered architecture of edge computing. The model enhances the edge layers for computing operations by adopting different algorithms in machine learning. The introduction of machine learning-based edge computing hybrid model lead into several advantages of edge-based intelligence on IoT platforms. Finally, the paper also listed out various challenges and future research implementation works involved in the edge computing environment. The objective of this survey paper is to open new avenues of research problems by adopting machine learning-based algorithm in the implementation of edge computing.

Keywords Internet of things (IoT) · Cloud computing · Edge computing · Machine learning · AI

1 Introduction

The implementation and applications of the "Internet of things (IoT)" [1] have come to a significant part in our everyday lives. A huge number of interconnected devices produce information every minute of every day. Investigation uncovers new, fascinating examples with limitless applications for buyers and organizations. There

R. Ramya (✉) · S. Ramamoorthy
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India
e-mail: rr7135@srmist.edu.in

S. Ramamoorthy
e-mail: ramamoos@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_52

is a remarkable development in worldwide information traffic, quickened by the expanding prevalence of end user devices, e.g. smart phones. As per "The Intersectional Data Corporation (IDC)", worldwide information shows up that the 70% information can be created by end user devices of IoT by 2025. The IDC expresses there may be higher than 150 billion user devices may associated worldwide by 2025 [2]. For these circumstance, the centralized model of cloud computing is not sufficiently gainful to manage the data made by the edge.

1.1 Challenges of Cloud Computing

The cloud servers are works like a centralized processing model in the cloud computing, and it takes care of the figuring and capacity issues, which empowers the cloud administrations to make monetary advantages. However, with regards to IoT, there are some limitations; they are listed below [3].

(1) **Latency**: With in the conventional cloud computing model, applications transfer information to a server and then acquire a reaction, which expands the framework latency.
(2) **Bandwidth**: An edge device creates a lot of information and sending it to the cloud in an ongoing way will cause extraordinary pressure on framework bandwidth.
(3) **Availability**: The availability of cloud services is one of the challenging issues for an increasing Internet services requirements of day by day lives. Hence, it is a major test for cloud specialist to provide services for 24×7 with guarantee.
(4) **Energy**: Lot of energy consumed by servers to provide cloud services such as calculations and transmissions; it makes difficult in the development of cloud computing services.
(5) **Security and Privacy**: Large number of devices is connected to Internet with more personal and all other information. There must be security and privacy assurance required for all users, but it is very big challenge in every day updating of Internet and cloud services organizations.

To get the solutions for these difficulties, there is a technique of edge computing, which will work with information at the edge of organization or network. With this, there is a possibility to lessen latency, bandwidth and increase of accessibility and to ensure information protection and security.

1.2 Edge Computing

In general, the cloud computing in the IoT works like provider to the end users; it provides data computing and storage virtually through Internet. The new technology

Table 1 Cloud versus edge computing platform

Functionalities	Cloud	Edge
Deployment	centralized	Distributed
components	Virtual resources	Edge nodes
computational	unlimited	Limited
storage	unlimited	Limited
Response time	slow	Fast

edge computing makes these services near to the end users to get better performances in IoT services.

When comparing the parameters of both cloud and edge computing results clear understanding about concepts that is tabulated in Table 1.

Cloud is centralized virtual resource with unlimited storage and computation, with slow response time, but the edge is distributed edge node with limited storage and computation, with fast response time. So the implementation of the edge concept is to get a fast response in the cloud environment of IoT.

1.3 Functional Difference Between Cloud and Edge

The following Table 2 is to compare and differentiate both cloud and edge in terms of various important factors.

2 IoT Integration with Edge Computing

This section includes the details about architecture of integration with both IoT and the edge computing, advantages of this integration process, functional architecture of this technique and finally the challenges of this integration. These will explain over all benefits and the challenges of this technique indication for the future updating requirements.

2.1 Overall View of Integrated Architecture

Edge computing integration will help to the improvements of performance of IoT services. The edge computing devices get placed as a middle layer between top layer and bottom layer of the end user IoT components that shown in Fig. 1 [1]. This will make less latency and better computation services by providing all requirements near to the end devices.

Table 2 Functional difference between cloud and edge computing

Factors	Cloud	Edge
Network delay	Latency in data exchanges between cloud services across the Internet can be much in higher. This delay can result in higher costs to users of multiple cloud services, which may mean that an organization is effectively locked into single provider's solutions	Edge computing decentralizes processing power to ensure real-time processing without latency while reducing bandwidth and storage requirements on the network
Computing power	Cloud computing is more suitable for the organizations which deal with massive data storage	We can use edge computing to save financial resources
QoS parameters	Availability and robustness, flexibility, maintainability and readability	Scalability, usability and accessibility, platform compatibility, load balancing and scheduling
Load balancing	Cloud loading balancing holding the circulation of workload traffic and demands that exist over the Internet	Edge load balancing is achieved by the method of task assignment and estimating the task completion time based on the transmission rate between edge nodes, the computation speed, and the current tasks calculation time
Task scheduling	Task scheduling is the main problem in cloud computing which reduces the system performance. To improve system performance, there is need of an efficient task scheduling algorithms. Existing task scheduling algorithms focus on task-resource requirements, CPU memory, execution time and execution cost	An on-demand computing resource allocation can be achieved by adjusting the task schedule of the edge gateway via the lightweight virtualization technology
Edge node clustering	Cloud technology is moving towards more circulation across multi-clouds and the consideration of different devices, as evident through IoT and network integration with regards to edge cloud and fog computing	Edge is a feature of cluster manager that permits organizations to deploy and centrally manage computing resources in distributed areas as a single clustered framework, from a single interface

The end user's devices of bottom layer can get the advantages of good computation capacity and the required level of storage from the top layers of the above architecture that is from cloud or from edge supporting devices. So these processing can frame the ability to fulfil IoT application requirements of tolerable computation limit, storage and quick reactions, even the middle layer has restricted computational capacity and storage.

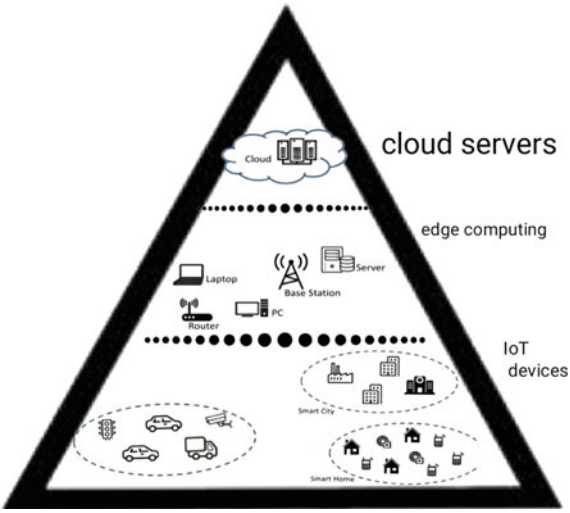


Fig. 1 Layered representation of cloud and edge computing-based IoT

2.2 Benefits of Edge Computing with the Integration of IoT

This segment includes the evaluation in benefits of integration of edge computing with the IoT [4].

1. **Transmission**: Edge computing can provide a significant advantage of quick transmission with quality of services for time-based transactions and bandwidth requirements.
2. **Storage**: Cloud computing has provided the centralized storage services, which is complex in multilayer systems. But the edge computing makes this storage services as distributed by load balancing technology to fulfil the QoS requirements in IoT.
3. **Computation**: Compared to the cloud computing, the edge nodes has less computation power. The computation requirements can be fulfilled in edge-based IoT by using the concept of offloading and task scheduling schemes based on various objectives.

2.3 Functional Architecture of Edge Computing

To achieve the ultimate goal of above all mentioned benefits, the middle layer of edge computing has three stages of processes. They are the data sensing process at the end user of IoT requirements, the data collection in the middle of processing and the real-time data analytics at the cloud computing server side. The data sensing can provide quick reaction for the end user requirements with the help of data collection and real-time data analytics of edge computing. The data collection can get required information's from the cloud servers with the help of real-time data analytics to

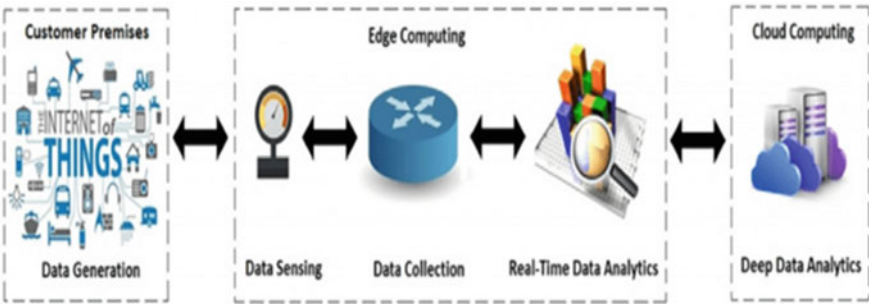


Fig. 2 Edge computing process between cloud computing and customer

provide the short time services for the end user's. The real-time data analytics can do the analysis for the required data and provide support to overall real-time services of IoT.

Figure 2 [3] provides the edge computing process between cloud and customer premises, which can explain the step by step works with in edge computing by using the artificial algorithms mentioned. The given model represents general way to implement any kind of algorithms to get better solutions for IoT environments.

2.4 Challenges in the Integration of Edge Computing with IoT

Here, the detailed explanation is about the difficulties in the edge computing integration with IoT.

1. System integration: As all known the IoT includes different kind of devices and different service requirements. It makes difficult to build a program or tool to manage resources and information to provide services [1].
2. Resource management: There are some difficulties to understand the integration of edge computing with IoT to make the resource management services. Because the resources are different types and they require different services.
3. Security and privacy: In any perspective these are challenging issues to manage. In this integrated edge computing and IoT also difficult. Because it includes different kind cloud Internet storage for large number of users, it makes difficulties to provide security and privacy.
4. Advanced communication: With communication advances in correspondence advances, edge computing will additionally advance as integration of these innovations gets inevitable.
5. Smart system support: The coordination with the sensor devices makes difficulties to manage and control in real-time application.

3 Edge Computing and Artificial Intelligence

This section consists the details of how artificial intelligence (AI) can supports to the implementations of edge computing to provide the better communication and quality of services in the IoT, rather than the cloud alone services. The combination of two technologies artificial intelligence and the edge computing is becoming the new name of edge intelligence in the IoT implementations. These supports as a key to get solutions for above section mentioned challenges of integration of edge computing with the Internet of things.

3.1 Edge Intelligence—Overview

The AI is practically fundamental for all kind of fast analysis of huge amount of information and processing on those information. The edge intelligence can make use of this intelligence for the IoT data-based communications [5]. The intelligence is available in different types according to the type of network processing; some important network processing are "deep neural network (DNN), conventional neural network (CNN), generative adversarial network (GAN) and recurrent neural network (RNN)" [6]. In the last few years, these are very important thing in the network processing. Based on the type of processing, the required technology can be used for implementation work. Here also, these intelligence can support to improve the overall quality of Internet services in the integration of edge with IoT. The overall main subdivisions of AI are explained in the next section.

3.2 Machine Learning Introduction

An artificial intelligence (AI) is one of the more important things in every branch of science and technology. In the artificial intelligence, the machine learning (ML) is a subset with depth algorithms for autolearning of machines. The deep learning (DL) is inner subset of machine learning for neurons level process handling. The relation reference of AI, ML and DL [2] is shown in Fig. 3.

3.3 Types of Machine Learning-Based Algorithms—Overview

In Table 3, machine learning algorithms types list out the main and subtypes of machine learning. It shows there are four types, but the top two types of machine learning are more familiar and important to study and implement. The first and

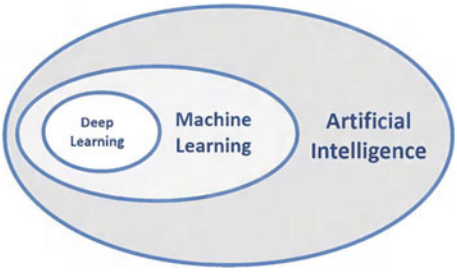


Fig. 3 Relation reference of AI, ML and DL

Table 3 Machine learning algorithms types

Machine learning (including deep learning using conventional neutral networks and recurrent neural networks)	I. Classical learning	1. Supervised learning	i. Classification
			ii. Regression
		2. Unsupervised learning	i. Clustering
			ii. Dimensionality reduction (generalization)
		3. Semi-supervised Learning	iii. Pattern search medical predictions (where tests and expert diagnoses are expensive, and only part of the population receives them)
	II. Reinforcement learning		
	III. Ensemble methods IV. Neural nets and deep learning		

main type of machine learning is named as classical learning which includes the major and main concepts three subtypes; they are "supervised, unsupervised and semi-supervised learning" [7].

Table 4: Supervised learning algorithm types and applications mention all of the specific types of supervised learning algorithms by name and the current applications of those algorithms.

In Table 5, unsupervised and semi-supervised learning algorithm types and applications list out the types of unsupervised and semi-supervised algorithms with those current specific applications. The subtypes of unsupervised learning are clustering, dimensionality reduction and pattern search; those all are having different specific algorithms for different applications as mentioned in Table 5. Similarly, semi-supervised learning algorithm application is also mentioned in the same table.

In Table 6, reinforcement, ensemble and neural nets and deep learning algorithm types and applications list the remaining algorithm types and real-time applications. These all are the new and updated types of artificial intelligence concepts

Table 4 Supervised learning algorithm types and applications

1. Supervised learning	Type	Algorithm	Applications
	i. Classification	K-nearest neighbours (K-NN)	Prediction for classification is taken with the consideration of similarities. This method mainly used for analytics in retail for finding customer buying similar products
		Naïve Bayes	This method used in spam channels, report order, forecast and conclusion investigation ventures
		Support vector machine (SVM)	This method provides more accuracy and efficiency of data mainly for medical field in image processing
		Decision trees	Used in pattern recognition, data exploration, option pricing in finances and identifying disease and risk trends
		Logistic regression	1. To identifying risk factors for diseases and planning preventive measures 2. Classifying words as nouns, pronouns, and verbs 3. Weather forecasting applications for predicting rainfall and weather conditions 4. In voting applications to find out whether voters will vote for a particular candidate or not
	ii. Regression	Linear regression	Deals estimating, chance evaluation examination in medical coverage organizations and requires negligible tuning
		Polynomial regression	Used in many experimental procedures to produce the outcome using equation
		Polynomial regression	Technique for analysing multiple regression data

Table 5 Unsupervised and semi-supervised learning algorithm types and applications

2. Unsupervised learning	Type	Algorithm	Applications
	i. Clustering	K-means	Gathering pictures into various classes, identifying diverse action types in movement sensors and for checking whether followed information focuses change between various gatherings after some time
		Agglomerative	Used at any stage as in indicator of natural or forced cluster formation
		Density-based spatial clustering of applications with noise (DBSCAN)	Used in data mining
		Mean shift	Used in image processing and computer vision
		Fuzzy C-means (FCM)	Used in medical image segmentations
	ii. Dimensionality reduction	t-SNE	Used for visualization in a wide range of applications, including computer security research, music analysis, cancer research, bioinformatics and biomedical signal processing
		Principal component analysis (PCA)	Used when there are multiple independent variables or in combining the features having a correlation in them. It is mainly used to reduce the number of features
		Latent semantic analysis (LSA)	Compare the document in low dimensional space, find relationship between the terms, analysis word association in text corpus
		Singular value decomposition (SVD)	Used to represent large matrix. This is helpful in large computation
		Linear discriminant analysis (LDA)	Providing a projection of a training dataset that best separates the examples by their assigned class

(continued)

Table 5 (continued)

	iii. Pattern search	Euclat	Used in mining frequent item sets vertically
		Apriori	Used to determine association rules which highlight general trends in the database
		FP-Growth	Used in market basket analysis

in machine learning; they are used in many different types of research works for different application updating and implementation [8–13].

Through this, overall analysis study may help to solve any real-time problems in any type of applications by considering and choosing the right algorithm to update and implement. There are many possible algorithms to solve and update the IoT and relevant area problems.

3.4 Future Implementation Directions

This is a review on the writing encompassing edge intelligence. There is examination of some unmistakable challenges in integration of edge computing with IoT devices [4]. For further research work, there are some intelligent implementation types through artificial intelligence and machine learning. Researchers can have various missions, e.g. information assortment, information handling and information investigation, which have distinctive asset utilizations. Future endeavours could zero in on tending to these difficulties.

4 Conclusion

The advancement in the field of cloud services in the IoT makes the invention of edge computing a supporting tool for Internet-based processing. This paper reviews the challenges of cloud computing-based services for IoT end user devices; the edge computing integration with the Internet of Things (IoT) explains the benefits and challenges; the artificial intelligence concepts supported and types of edge computing implementations in the Internet services are explained; and finally, the listed future implementation directions of all concepts are mentioned. There is a hope this paper will bring the attention of the learners as well as researchers community to learn and research in the field of edge computing with machine learning concepts.

Table 6 Reinforcement, ensemble and neural nets and deep learning algorithm types and applications

Type	Algorithm	Applications	
II. Reinforcement learning	Genetic algorithm	Used to optimize a wide range of different fit functions	
	Q-Learning	Used to learn quality of actions telling an agent what action to take under what circumstances	
	Asynchronous advantage Actor critic (A3C)	Used in the implementation of policy gradient	
	State action reward state action (SARSA)	Used for learning a Markov decision process policy	
	Deep Q-Network (DQN)	Used for stock market portfolio management	
III. Ensemble methods	Stacking	Used for expression evaluation and pattern checking	
	Bagging	Random forest	Used in industrial applications such as finding out whether a loan applicant is low risk or high risk, predicting the failure of mechanical parts in automobile engines and predicting social media share scores and performance scores
	Boosting	AdaBoosting	Detect fraudulent activity in banking transactions
		Cat Boost	Used for gradient boosting on decision trees
		Light GBM	Used in cross entropy, multi classification
		XGBoost	Used to solve regression, classification and ranking
IV. Neural nets and deep learning	Conventional neural networks (CNN)	DCNN	Used to classify images
	Recurrent neural networks (RNN)	LSM	Used in time serious analysis
		LSTM	Used in language modelling
		GRU	Used in smaller and less frequent datasets
	Generative adversarial networks (GAN)	Used in photograph editing, blending, inpainting and translating	

(continued)

Table 6 (continued)

Type	Algorithm	Applications	
	Autoencoders	Seq2seq	Used for language processing
	Artificial neural networks (ANN)	Perceptions (MLP)	Used in research to solve problems stochastically

References

1. Yu W, Liang F, He X, Hatcher WG, Lu C, Lin J, Yang X (2018) A survey on the edge computing for the internet of things. IEEE 6
2. Chen M, Challita U, Saad W, Yin C, Debbah M (2017) Machine learning for wireless networks with artificial intelligence: a tutorial on neural networks. ARXIV CoRR, vol. abs/1710.02913
3. Shi W, Pallis G, Xu Z (2019) Edge computing. IEEE 107(8)
4. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenge. IEEE Internet Things J 3(5)
5. Deng S, Zhao H, Fang W, Yin J, Dustdar S, Zomaya AY (2020) Edge intelligence: the confluence of edge computing and artificial intelligence. IEEE 10
6. Zhu G, Liu D, Du Y, Zhang J, Huang K (2018) Towards an intelligent edge: Wireless communication meets machine learning. arxiv, corr.vol.abs/1809.00343
7. Shamir O (2014) Fundamental limits of online and distributed algorithms for statistical learning and estimation. In: Advances in neural information processing system, pp 163–171
8. Sun X, Ansari N (2016) EdgeIoT: mobile edge computing for the Internet of Things. IEEE Commun Mag
9. Sun Y, Peng M, Zhou Y, Huang Y, Mao S (2020) Application of machine learning in wireless networks: key techniques and open issues. ARXIV CoRR, vol. abs/1809.08707
10. Cao K, Liu Y, Meng G, Sun Q (2020) An overview on edge computing research. IEEE 8
11. Psaras I (2018) Decentralised edge-computing and IoT through distributed trust. Munich, Germany Mobisys
12. Introduction of Edge Computing Consortium (2016) [Online]. Available: http://en.ecconsortium.org/
13. Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J (2019) Edge intelligence: paving the last mile of artificial intelligence with edge computing. Proc IEEE 107(8)

Enhanced Data Privacy Preservation Model for Mobile Crowdsensing System Using Blockchain Technology

M. Arulprakash and R. Jebakumar

Abstract With the recent advancement in computing and sensing capabilities of smart terminals, mobile crowdsensing systems built on smart and ubiquitous mobile terminals have become one of the rising applications allowing better sensing and scalable systems. However, the prior mobile crowdsensing system mostly employs centralized architecture, which is subject to low reliability and is vulnerable to malicious attacks. Furthermore, it is difficult to guarantee the privacy with the transparency of MCS systems. These factors have been a major hindrance to the growth of MCS applications. Hence, it is imperative to ensure security and privacy of such systems in order to meet the wide spectrum of its applications. So we propose a decentralized blockchain crowdsensing system to harness the benefits of this sensing paradigm. Our objective is to develop a decentralized crowdsensing system and meet the security and privacy requirements without affecting data sensing quality, achieving a trade-off between data quality and privacy preserving. We plan to integrate blockchain technology into the crowdsensing system. Meanwhile, we also focus on how to protect data privacy in blockchains to develop an accountable MCS system. By leveraging these technologies, we believe that MCS system will have a good performance in privacy protection and security enhancement.

Keywords Mobile crowdsensing (MCS) · Blockchain · Ethereum · Smart contracts · Security

1 Introduction

Data sensing technology has evolved significantly in recent times as prominence of the data has been realized. As a part of that, fixed-location sensing approaches

M. Arulprakash (✉) · R. Jebakumar
Department of CSE, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu 603203, India
e-mail: arulpram@srmist.edu.in

R. Jebakumar
e-mail: jebakumr@srmist.edu.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_53

have become a trend for data collection. Fixed-location sensing approaches include web-based devices which make use of systems like sensors, processors and other hardware devices in order to gather the data from the surroundings analyse the data collected and transmit the processed data. Though these are helpful in reducing and/or eliminating human intervention, these data sensing approaches do not meet the requirements of practical applications as they are constrained due to the less geographical coverage, mobility, high deployment cost and high maintenance cost [1]. Mobile crowdsensing has emerged as a novel data sensing approach to overcome the disadvantages of the above mentioned approach. Mobile terminals are becoming increasingly ubiquitous, and crowdsensing has drawn the attention as the next generation sensing technology attributing to its merits that suits the happening informatization era.

Mobile crowdsensing is a sensing paradigm in which anyone possessing mobile terminals equipped with the robust sensors is capable of sensing the data of common interest [2, 3]. This can also be implied as crowdsourcing of sensory data from mobile devices about anything, at any time and from anywhere. In this technique, intelligent mobile terminals equipped with high sensing and computing capabilities are leveraged to gain knowledge and deliver high quality data. Typically, crowdsensing is comprised of three main parties: requesters, workers and a crowdsensing system. The requesters are sensing task publishers. They issue a sensing task to the system and acquire the sensing data by recruiting the workers. The workers are sensing data contributors.

With the crowdsensing gaining popularity, it has unfolded opportunities for several applications and provided a proper way of assessing the environment (e.g. weather monitoring, noise, air quality), infrastructure (e.g. road condition, traffic analysis) [4], social (e.g. healthcare, travel) and many more. Several technology companies are leveraging this technology to provide the services based on the collected data, some of the famous examples being Google and Uber. The prior mobile crowdsensing systems mostly employs centralized architecture where the system is controlled by a central server while such systems are subject to low reliability and are vulnerable to malicious attacks. With centralized architecture, there is also a need to expend on the central platform. MCS has high concern about the privacy too as the sensitive information of the participant may be leaked to the platform for service processing, and it is not easy to guarantee the privacy of the users [5, 6]. Furthermore, it is imperative to ensure the privacy in order to gratify the participants for their contribution.

Recent advances in sensing, computing and networking have paved the way for the emerging paradigm of mobile crowdsensing (MCS). However, the existing mobile crowdsensing models are mostly based on a centralized platform, which is not fully trusted in reality and results in the existence of fraud and other security problems. The goal of the project is to implement a mobile crowdsensing (MCS) systems which address security threats such as jamming, spoofing and faked sensing attacks during both sensing and information exchange processes in large-scale dynamic and heterogeneous networks. This is what motivated us to go through with the project idea and implement a decentralized architecture to create a model that enhances data privacy in the MCS—blockchain.

2 Related Work

Security of crowdsensing applications has become a major concern of the research association and is highly demanded. While it is crucial to understand and adapt to the new sensing technology, it is equally important to embed trust in a potentially untrusted environment because of the high degree of modularity. Also it is hard to set restrictions on the users and data contributors in the crowdsensing applications. Significant works have been carried out to develop the secure applications based on crowdsensing. We briefly explain the related existing work in MCS in this subsection. In [7], the authors propose a multiblockchain-based smart parking system with mobile crowdsensing technology in which the public blockchain is used by data contributors, and private blockchain is used by the service providers. The bridge node acts as an interface between the two blockchains. The intuition behind the proposed idea is that a huge number of participants can be accommodated while mitigating the risk of personal exposure. But the proposed schemes have not taken into account the potential security threats of the bridge nodes. CrowdBC [8] is a decentralized framework developed based on blockchain for crowdsensing. It is an Ethereum public test network-based tangible framework which eliminates the need of the conventional centralized system. The proposed framework was successful in mitigating the malicious security attacks like single point of failure by extracting the benefits of blockchains and Ethereum smart contracts. Nonetheless, CrowdBC is not capable of handling the complex situations and hence considered to be impractical in the real time spaces. CrowdBLPS [9] is a location privacy preserving crowdsensing system which integrated the concept of blockchain into crowdsensing, understanding the need of decentralization to get rid of security issues. Leveraging the concept of smart contract, the proposed scheme has also achieved its' purpose in preserving the privacy of the location and in controlling the data quality by following a two-stage approach. The authors have planned to enhance the reliability of the system to ensure that the system meet the requirements of the real-world.

In [10], the authors realized the importance of secure, privacy preserving and incentive crowdsensing services. So they proposed a reliable incentive mechanism for motivating and encouraging the sensing participants to participate in the crowdsensing tasks. They also implemented a privacy preserving scheme by leveraging differential privacy technique through adding the noise data assuming the participants as not totally credible. Then they used homomorphic encryption for securing the sensing data. In future they plan to concentrate on the trade-off between accuracy and privacy. In [11], the authors propose mobile crowdsensing by leveraging the algorithms of deep learning to enhance the security. To thwart security attacks of the crowdsensing systems, the authors investigated the authentication and protection of privacy based on deep learning. The security approaches based on DL are found to be propitious for providing the quality of experience-based multimedia services. Nonetheless, it was realized that the authentication based on deep neural networks needs high training times, and the hardware devices used for DL-based computation are not compatible to the MCS devices thus found to be complicated for the practical

implementation of mobile crowdsensing systems. In [12], the authors analysed the challenges and the case studies of mobile crowdsensing systems. They have analysed the fact that the illegitimate task requesters are engaged in keeping the servers busy while needlessly expending energy of servers. The proposed scheme focused on the need to differentiate legitimate and illegitimate tasks and leverages the AI-based machine learning techniques in order to provide a solution to the security vulnerabilities of crowdsensing. However, the proposed work is conducted solely on a location-based and energy-based DoS attacks and plan to consider more practical scenario in future.

SenseCrypt [13] is a secure framework that encrypts the sensitive information of the participants of mobile crowdsensing. The proposed framework leveraged the K-means algorithm for the clustering of data into sensitive and non-sensitive data. In this, the encryption techniques are leveraged in order to secure the sensitive location data of the crowdsensing users. The novelty of the proposed framework is that it has alleviated the cost of computation and communication overhead by leveraging the efficient data compression techniques. However, the paper solely considered the location information as the sensitive data of interest. In [14], the authors proposed a custom-defined location sharing strategy for protecting the privacy of the participant in mobile social networks. They considered the vicinity area as a small bounding rectangle for the proposed scheme where the accuracy of the results is affected by the vicinity area [15]. In [16], the authors focuses on providing the incentives in a privacy preserving manner and they aimed for achieving security and privacy protection in MCS. Ethereum enables extending the functionality of blockchain using smart contracts. Ethereum uses ‘proof of work’ which enables the network of Ethereum to accept on the state of information recorded on the Ethereum.

2.1 Research Findings

- The survey papers aim at the three open issues in mobile crowdsensing: reliability, security and incentive mechanisms.
- These papers majorly integrate the idea of blockchain into crowdsensing, realizing the decentralization of crowdsensing to avoid the security problems.
- It can also be inferred that there exists a trade-off between data quality and privacy preserving.
- Incentive mechanisms motivate people to contribute to MCS efforts while common challenge is providing incentives in a privacy preserving manner; users should be gratified without associating themselves with the data they contribute.

3 Proposed System

We propose a secure blockchain-based crowdsensing framework. The proposed system replaces the conventional triangular architecture by an efficient decentralized framework. The single point of failure attack can be avoided by using the decentralized framework. We leverage the smart contracts deployed on blockchain technology. The task requesters are initially required to deposit certain amount of money through which the enthusiasm of miners and workers can be raised. By this, certain attacks like sybil attack, false-reporting attacks can be prevented efficiently. Thus, the proposed system provides an advantage of alleviating the malicious attacks and mitigating the charge of the central platform. The privacy can also be guaranteed by allowing the participants to participate without true identity.

4 Working Procedure

In a traditional or conventional centralized system, all users or nodes are connected to a central server or network. However, we propose an approach that makes up for the lack of security in the former system. Instead of connecting directly to a server, we will access our application via a web browser, and we will correlate to the client-side app that we will build, a simple app on a web server. This client-side app would not talk to a web backend and database; instead, it will directly speak to a blockchain. We will have code written with Ethereum smart contract that will contain our pre-conditions and logic for our use case on the blockchain. That is how a blockchain works fundamentally and how it is different from a traditional web chain application. A blockchain is a separate peer-to-peer network of nodes that talk to one another, a distributed network and so different computers talking to one another. We can connect to an individual node on the network; our web app is doing it here. All nodes participate in running the network; each node contains all code and data on the blockchain. The data on the blockchain has bundles of records. A blockchain is secure and unchangeable, making it robust. All code on the blockchain is on smart contracts, which are the building blocks of blockchain apps. We will write a smart contract that will contain all the tasks as set by the requester. Smart contracts are written in a programming language called solidity. Blockchain is trustless, we know no one can change it, and thus, the app will behave in the same way every time. Blockchains are like a microservice that executes business logic. We will build a client-side application that will talk directly to the blockchain and be deployed on it. The process has been illustrated clearly in Fig. 1 given. Figure 2 shows that decentralized application is essentially a frontend and a smart contract backend combined in one package. It is also very popularly known as the backbone of Web 3.0. The frontend and backend together work as an integrated utility for the entire smart contract. Each user is connected to the decentralized app and is able to obtain and utilize all its features by operating through the frontend for easier accessibility.

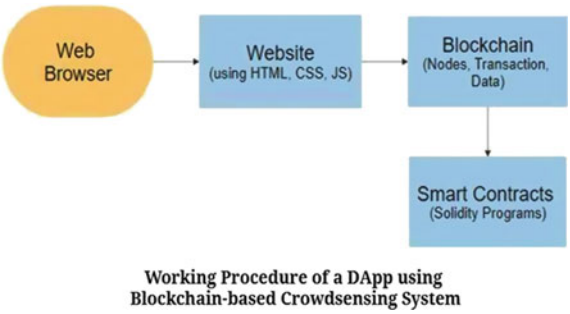


Fig. 1 Working procedure of a DApp using blockchain based crowdsensing system

Working Procedure of a DApp using Blockchain-based Crowdsensing System

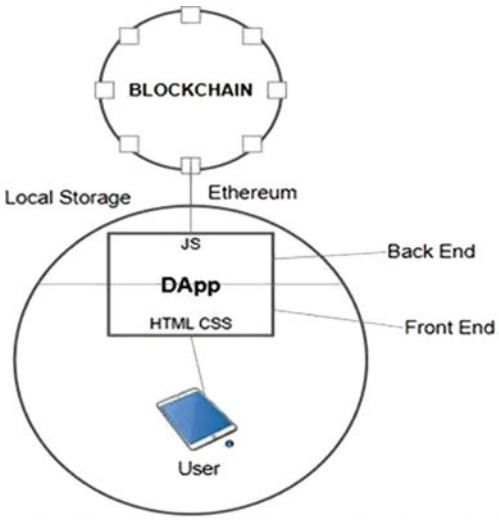


Fig. 2 DApp using blockchain with Ethereum smart contracts

DApp using Blockchain with Ethereum Smart Contracts

The backend is connected to the Ethereum smart contract in the form of local storage that incorporates into the blockchain.

4.1 Working Procedure of Smart Contracts

The working of a blockchain maybe split up into the following three categories:

a. Initiating an innovative smart contract
b. Writing a smart contract message to a blockchain
c. Public key cryptography.

4.1.1 Initiating an Innovative Smart Contract

Primarily, a public key encryption infrastructure is used in blockchain. An initiator coveting to participate in a smart contract hosted on an unauthorized blockchain can use the position to produce a key and write it on the system.

4.1.2 Writing a Smart Contract Message to a Blockchain

In a conventional blockchain implementation, when an adequate number of other members or connections arrives at the same outcome, the relevant consent rules determine that the smart contract's information is appended to the blockchain. Another way is a blockchain's administrator may conclude with it.

4.1.3 Public Key Cryptography

Decentralized applications or 'DApps' are trustless or peer to peer with the differentiating feature that there is no individual server. A DApp uses the blockchain at the core of its data storage, computation and processing, implemented by smart contracts. A traditional website model creates the user interface for the decentralized application. After the user accesses the Dapp via a web browser, it will connect to the client-side application built on the server. However, the application does not use a traditional web-based database. Instead, the database is the blockchain on which the users connected to the crowdsensing network perform various tasks, including WiFi sensing tasks.

Inside a DApp, a user's account is a digital blockchain key stored on the mobile device's memory. Popular tool that helps users manage the accounts and interact with the DApp is MetaMask, an extension for web browsers. These are used primarily by the end-users. A suite of software tools used by developers during the construction of a DApp includes integrated development environments (IDEs) such as Remix IDE and the dominant programming language for Ethereum smart contracts called solidity, as used in this use case.

4.2 Proposed Framework

We introduce smart contracts applications to reward users in our aimed framework, which sensing task requesters should first launch with particular reserve deposit. There is no centralized stage in the crowdsensing process in the decentralized system, which we recommend anymore. Alternatively, by implementing blockchain algorithms, the crowdsensing process is managed by a decentralized system. As proposed in our framework, the blockchain-based crowdsensing process has been explained below in the form of a flowchart as shown in Fig. 3.

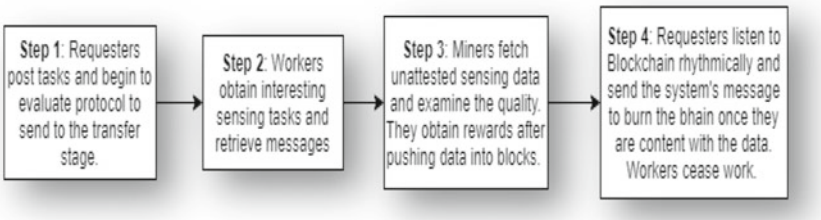


Fig. 3 Blockchain based crowdsensing process

4.3 Implementation of Proposed Model

The use case chosen for the project is roadsensing. Requesters create the intelligent contract 'roadsensing', initiate it, and encrypt sensitive data by private keys. Various parameters and features incorporated into the contract include rewards assigned to each task, the required count of data points, the obtained points, the source and destination of the data point collected and the task's requester. Several states of the requester's assigned tasks are uncreated, created and inactive. Additionally, the smart contract denotes multiple conditions for the road sensing task that includes—worst, poor, average, good and excellent. To better depict the crowdsensing process by the use of smart contracts and blockchain, we have split into three phases as follows—abbreviations used in algorithms for the sake of simplicity are R—Requester, W—Worker.

The first algorithm, **Algorithm 1**, in the implementation deals with creating our smart contract and posting the requester's required sensory tasks. The inputs taken in by the algorithm include rewards assigned for the task, the required data for the task, the source as well as the destination for the task. The output will be the status of the completion of the task. The precondition set for the algorithm is if the state is uncreated and if the assigned task's message value is either greater than or equal to the entire product of the reward allocated. Moreover, the required count and the details specified as per the smart contract will be assigned, and the state of the contract becomes 'created'. Lastly, the status of the contract is returned along with the TaskInitiated() function and log.

Algorithm 1—***Smart Contract Creation and Posting of Tasks by Requester***

Input: Rwd- R*wd*- reward designated for the task, *ReqC*- required data points for the task, S*rc*- the source of data collection of the task, *Dest*- destination location
Output: *Status*- status of the task

1: Setting of Task
2: **if** State = Uncreated **and** Msgvalue > = *Rwd* * *ReqC* **then**
3: R = *MsgSender*
4: reward = *Rwd*
5: requiredCount = *ReqC*

6: source = *Src*
7: destination = *Dest*
8: state = Created
9: **return** *Status*
10: **end if**
11: **return** TaskInitiated

In **Algorithm 2**, the TaskState and the ObtainedCount are parameters of input. The task status, source, destination and rewards assigned on the road sensing task's registered count are considered output. If the contract's state is 'created', the worker can only review the task details. When aborting the task, the requester will again check for the contract state's conditions if the amount of obtained count is less than equal to the required count of the data.

Points assigned to the task. If these pre-requisites are met, the contract's state will become 'inactive'. Finally, the transaction's remaining balance will go to the requester and the AbortTask() function is returned.

*Algorithm 2—**Worker Receiving Task Details***

Input: Task*State, ObtC* - obtained count

Output: *Status*- status of the task, *Src, Dest, Rwd, ReqC*

1: Getting of Task
2: **if** State = Created **then**
3: **return** ViewTask
4: **end if**
5: Aborting of Task
6: **if** State = Created **and** 7: **if** ObtC < = ReqC **then** 8: State = Inactive
9: Transfer balance to R 10: **return** AbortedTask
11: **end if**

The project's final algorithm, **Algorithm 3**, administers the task's committing and designates awards to the workers after data is uploaded and registers successfully in the blockchain. The source, destination and road condition is taken in as the input, whereas the average road speed, the obtained data count and the task status are the outputs. The following conditions need to be checked to commit the task—if the contract's state is 'created', if the obtained count is less than the required count of the source as well as the destination of the worker's submitted data matches the criteria set by the requester. Finally an additional condition need to check that road's average speed is greater than 10. When all the above-stated conditions are satisfied, the obtained count is incremented by 1. Lastly, when the required count becomes equal to the obtained count, the state is finally changed to 'Inactive', indicating the smart contract process's end. The balance of the transaction is transferred back to the requester and the status is returned. The rewards allocated for the task go to the worker, and the DataCommited() function is executed.

*Algorithm 3—**Committing of Task and Designating Rewards***

Input: S*rc*- the source of data collection of the task, *Dest*- destination

location, *RoC*- Road condition, *AvgS*- the average speed of the road, *ObtC* - obtained count

Output: *Status*- status of the task

1: Committing of Task
2: **if** State = Created **and**
3: **if** *ObtC* < *ReqC* **and**
4: **if** LengthCheck(source, *Src*) = **true and**
5: **if** LengthCheck(destination, *Dest*) = **true and**
6: **if** *AvgS* > 10 **then**
7: *ObtC* + = 1
8: **end if**
9: **if** *ObtC* = *ReqC* **then**
10: State = Inactive
11: Transfer balance to R 12: **return** *Status*
13: **end if**
14: Transfer rewards to W
15: **return** DataCommited

5 Results and Discussion

The main aim is to develop a decentralized crowdsensing system for the road sensing application. We outlined the process of mobile crowdsensing by implementing Ethereum smart contracts on blockchain.

Purpose	Tool used
Programming languages	Solidity, JavaScript
Frontend	HTML, CSS
Backend	JS
Blockchain	Ethereum blockchain
Framework	Web3.js

Our proposed Road sensing decentralized application was implemented by leveraging Ethereum smart contracts. We used solidity and JavaScript as the programming languages and HTML as the markup language. Solidity is a contract-oriented, object-oriented high-level programming language designed for executing smart contracts on blockchain platform. Solidity is native to Ethereum virtual machine. In addition, we leverage web3.js which allows in designing the decentralized application that interacts with the Ethereum blockchain. Web3.js is an Ethereum API based on JavaScript which provides a way to interact with the Ethereum node.

The usage of smart contracts enhances, intensifies and refurbishes the security aspect of the project because of the following grounds-

1. Smart contracts refer to the immutable computer protocols that digitally aid an agreement's execution. They are a speedier, more economical, more agile and more secure way of administering and operating agreements. Hence, the security of the mobile crowdsensing blockchain system enhances.

2. The setting of pre-conditions using smart contracts also heightens security. Various modifiers and requirements created and emphasized in the intelligent contract demand the worker's authentication before a task commits and rewards distributed.
3. Cryptography is used to secure transactions in smart contracts to prevent attacks like double spending.

Investigation of the Secure Mobile Crowdsensing System using Blockchain

In this section, we briefly examine and investigate our crowdsensing-based road sensing system from the security aspect.

1. Before participating in the sensing task, participants making the deposits can help in inhibiting several attacks efficiently.
2. By executing smart contracts, only the ETH accounts of the participants (workers and requesters) are known and thus guarantee the anonymity of the participants.
3. A fundamental defect exists in the execution of the project that arises from the loophole of its execution. Intelligent intruders and attackers may use front-run attacking to seize or withdraw partially the rewards meant to provide the workers. For instance, attackers monitoring the network may fetch the workers' committed data and proffer them instantly. The designated worker might not get the expected reward in its entirety if the data's uniqueness has been considered.
4. However, there have been several solutions that may be successful in avoiding the attack—one of those is proclaiming hoaxes.

The implementation of the project as executed successfully on the Remix IDE platform is presented as follows—Requesters create the road sensing task by specifying the input parameters like rewards, number of data required, source and destination and then call setTask() function. The execution of task publishing (setTask) phase is being depicted in Fig. 4. Requesters have the option to view the amount of data received so far by calling getDataCnt() function. The data count is incremented every time the worker submits the appropriate data. Figure 5 depicts the execution of the function getDataCnt() function. For submitting the collected data, workers make a call to commitTask() function. Once the preset conditions are met, the worker submits the data and the data count will be incremented by one on successful submission.

Transaction costs deal with the expense of transmitting data, and a total of three components account for the transaction cost of the blockchain's function. They include-

- transaction's base charge
- the smart contract's disposition charge
- the expense of the contract's zero and nonzero bytes of data

Execution costs include the value of calculation processes, which are performed as a completion of the transaction. As illustrated in Fig. 6, the comprehensive report of the transaction expenses for each task function used in the project has been portrayed.

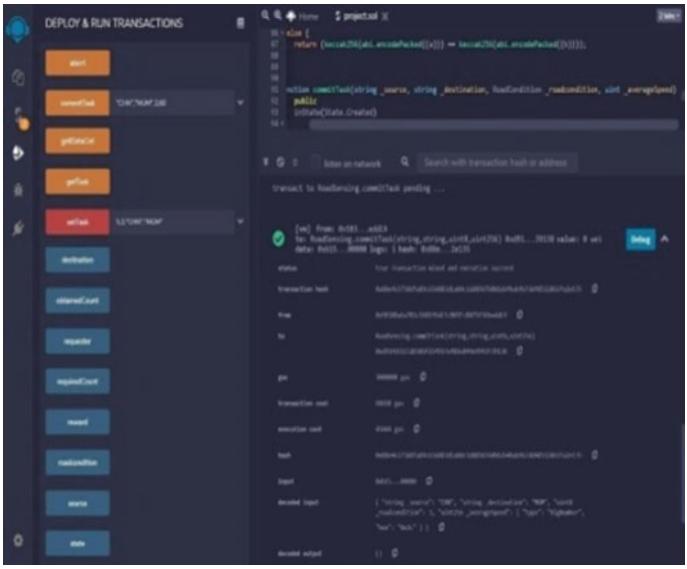


Fig. 4 Execution of setTask() function

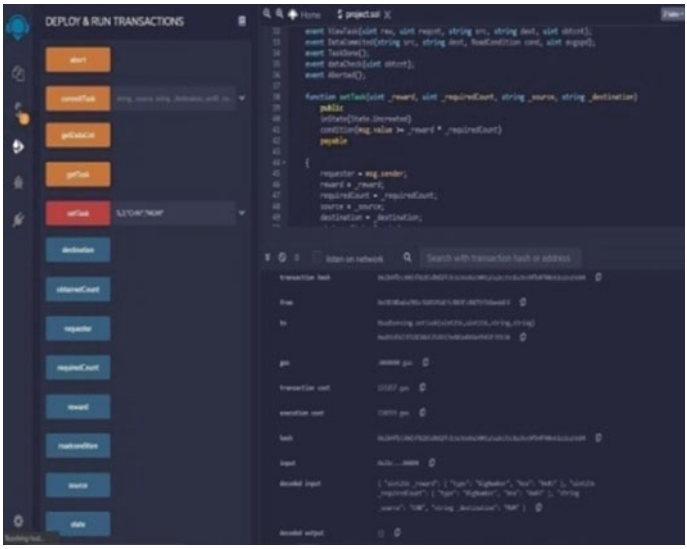


Fig. 5 Execution of getDataCnt() function

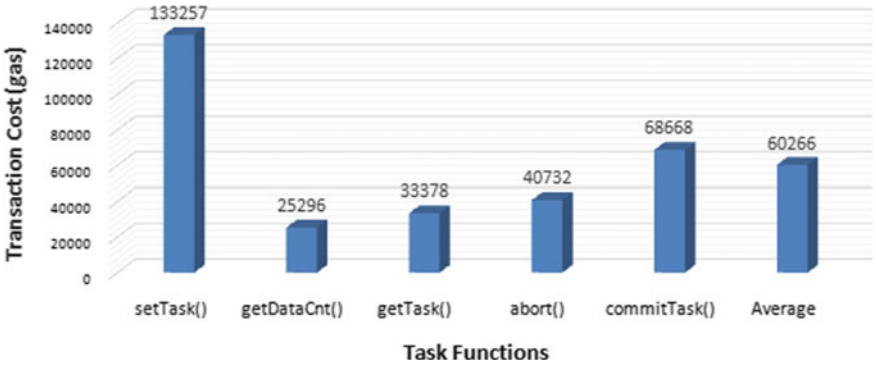


Fig. 6 Graph analysis of transaction costs for each task function

The various task functions used were setTask, getDataCnt(), getTask(), abort() and commitTask(). An average for all the costs taken into account is provided for a better estimate of our project's prospects. The transaction costs have been given in terms of the gas being spent on each function.

Thus, the total cost incurred for the data calculations is sustained, and hence, the proposed system's efficiency is enhanced. However, the transaction costs may be enhanced further by reducing the base cost of the transactions of each of the task functions. The objective of this project is to develop a decentralized crowdsensing system that replaces the conventional triangular architecture. We also aimed at leveraging the Ethereum smart contracts by which a successful trade-off between data quality and security preservation can occur. Hence, both these objectives are satisfied by the execution of our project and proposed model.

6 Conclusion

In conclusion, we proposed an architecture for a decentralized application of the mobile participatory crowdsensing paradigm-based road sensing system. We realized that the conventional crowdsensing architecture suffers from single point of failure threat. We leveraged Ethereum smart contract deployed on blockchain technology. The proposed architecture exercises smart contracts that enable crowdsensing providers to proffer their requests and run an action to ascertain the most fitting mobile users engaged in administering the crowdsensing tasks. The framework also employs a reward system for workers and deals with accessing the crowdsensing provider by successfully regulating blockchain. The proposed architecture implements an incentive-based optimal option for users who handle each crowdsensing task. Hence, we achieve the objectives of the project and enhance the focused parameters of the problem. As a future study, we focus on fully protecting the privacy for

the application to meet the real-world requirements. We also consider authenticating the data contributors to avoid data fallacies for getting the reliable data.

References

1. Huang J et al (2020) Blockchain-based mobile crowd sensing in industrial systems. IEEE Trans Ind Inf 16(10):6553–6563
2. Shien LK, Singh MM (2017) A secure mobile crowdsensing (MCS) location tracker for elderly in Smart City. In: AIP conference proceedings, vol 1. AIP Publishing, p 020085
3. ArulPrakash M (2016) Dynamic evolutionary information diffusion over mobile social networks. ARPN J Eng Appl Sci 11(19):1145711464
4. ArulPrakash M, Kamal A, Manisha A (2018) QR-Code scanner based vehicle sharing. ARPN J Eng Appl Sci 13(10):3441–3448
5. Luo C et al (2019) Predictable privacy-preserving mobile crowd sensing: a tale of two roles. IEEE/ACM Trans Networking 27(1):361–374
6. Toch E (2014) Crowdsourcing privacy preferences in context-aware applications. Personal Ubiquit Comput 18. https://doi.org/10.1007/s00779-012-0632-0
7. Kim M, Kim Y (2020) Multi-blockchain structure for a crowdsensing-based smart parking system. Future Internet 12:90
8. Li M et al (2019) CrowdBC: a blockchain-based decentralized framework for crowdsourcing. IEEE Trans Parall Distrib Syst 30(6):1251–1266
9. Zou S, Xi J, Wang H, Xu G (2020) CrowdBLPS: a blockchain-based location-privacy-preserving mobile crowdsensing system. IEEE Trans Ind Inf 16(6):4206–4218
10. Xiong J, Ma R, Chen L, Tian Y, Lin L, Jin B (2018)Achieving incentive, security, and scalable privacy protection in mobile crowdsensing services. Wireless Commun Mob Comput
11. Xiao L, Jiang D, Xu D, Su W, An N, Wang D (2018) Secure mobile crowdsensing based on deep learning. China Commun 15(10):1–11
12. Zhang Y, Kantarci B (2019) Invited paper: AI-based security design of mobile crowdsensing systems: review, challenges and case studies. In: 2019 IEEE international conference on service-oriented system engineering (SOSE), San Francisco, CA, USA, pp 17–1709
13. Pius Owoh N, Mahinderjit SM (2020) SenseCrypt: a security framework for mobile crowd sensing applications. Sensors (Basel) 20(11):3280
14. Peng T, Liu J, Wang G, Liu Q, Chen J, Zhu J (2020)A user-defined locationsharing scheme with efficiency and privacy in mobile social networks. Sci Program **2020**, Article ID 7832875:13
15. Gisdakis S, Giannetsos T, Papadimitratos P (2016) Security, privacy, and incentive provision for mobile crowd sensing systems. IEEE Internet Things J 3(5):839–853
16. AlJemy K, AlAnazi M, AlSofiry M, Baig A (2019) Improving IoT Security Using Blockchain. In: 2019 IEEE 10th GCC conference & exhibition (GCC), Kuwait, pp 1–6sss

A Comparative Analysis of Event Detection and Video Summarization

G. Balamurugan and J. Jayabharathy

Abstract In advancement of computer vision, intelligent video surveillance has led to rapid growth of video processing and providing effective requirement of advanced techniques for analysis of events in the video. Event detection plays phenomenon role in the smart and intelligent surveillance. Events such as traffic rule violations and misbehaving in common places are used detect automatically in time and must be reported as abnormal events. Abnormal event detection is a critical task in the video surveillance with challenging facts. Detecting the abnormal events and summarizing the key essential frames from surveillance video from multimedia repositories is very essential problem in video surveillance. Therefore, video summarization techniques are used for quick browsing for high quantity of video data for indexing. In these fact stockholders needs the video summarization of the abnormal event video as an abstract of the original video. In this paper, various video summarization techniques and methods are analyzed which can produce static or dynamic summaries. This analysis presents different methodologies and techniques in video summarization and also discussed some important datasets used in video summaries.

Keywords Computer vision · Intelligent video surveillance · Abnormal events · Video summarization · Event detection

1 Introduction

Video surveillance is termed as the incessant monitoring of various events and behaviour of objects in the surrounding area covered for monitoring. Intelligent video surveillance also involves monitoring the changing behaviour observed across objects prevailing in the monitoring area. Surveillance frameworks are conveyed for

G. Balamurugan (✉) · J. Jayabharathy
Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry University, Kalapet, Pondicherry, India
e-mail: gbalamurugan1991@pec.edu

J. Jayabharathy
e-mail: bharathyraja@pec.edu

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_54

observing different premises like homes, banks, workplaces and across open social occasion places, for example, air terminals, railroad stations and theatres to forestall the event of any accidents. Abnormal event detection refers to a deviation from the normal habitual [1]. They may be any abnormal events or activities identified during surveillance. This is due to the requirements of the humans, as they are not interested to watch the entire video stream which consumes more time [2]. Nowadays, lots of videos regarding sports, news, movies, surveillance, etc., are available online through various websites that requires lots of storage space and consumes more time to search, retrieve, download and watch [3]. Likewise, video summarization process includes the detection of the object or the event, extraction of the frames and the selection of high-quality key frames. It is mainly categorized into static and dynamic summarization [4, 5].

Static summarization also said to be frame based summarization, process on the frames and produce a set of frames as output. In this, the input video is registered as frames or images, and then each and every frame is processed to produce a sequence of frames as the output. The sequences of frames are selected randomly based on quality and accuracy. Dynamic summarization also said to be video based summarization, process on the frames and produce a video clip as output. In this, the input video is processed to produce a sequence of frames in a video clip as output [6]. This video clip consists of highlights or important events from the entire video stream in a sequential order as a precise summary. In this section, the event detection and its techniques are discussed [7]. Then, we outline the contributions for various summarization techniques and analyses of various datasets used in the video summarization are organized.

In this context, the techniques and methods which are employed for abnormal event detection and summarization produces an outline for the researchers for utilization of appropriate methods on desired applications. The essential summarization methods have to satisfy the higher rate of precise summarization.

2 Event Detection and Extraction

Object detection or event detection plays an important role in the video summarization process. They are processed based on feature, cluster, event and trajectory. Figure 1 states the functionalities for classification of object detection [6, 7].

In this, event detection and extraction provide and it process the important features based on their shot selection and its characteristics of same cluster.

2.1 Feature Based Detection

The objects are detected based on certain characteristics that are involved in the frames. Based on the defined characteristics, the objects are detected, and then the

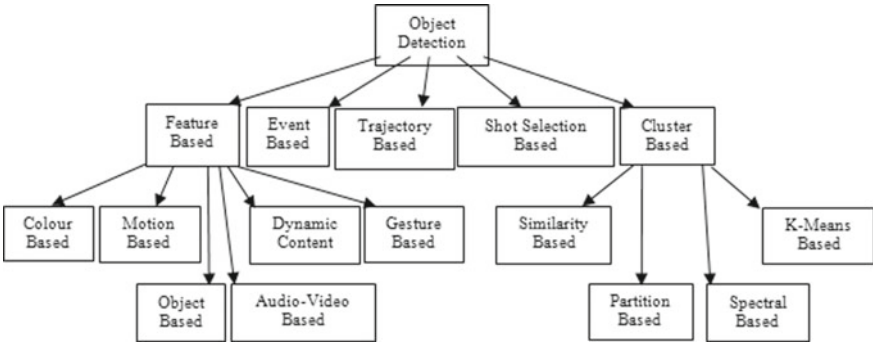


Fig. 1 Classification of object detection

frame that involves those characteristics is extracted. The digital video which is given as the input consists of many features in it. These features are classified into colour, motion, dynamic content, gesture, audio–video and object.

2.2 Motion-Based Extraction

In this, the foreground of the frame is detected based on the movements or change in the action of the objects involved in it. The frequency score is derived by taking difference between two frames. This difference is used to extract the required set of key frames. This is mainly used on sports, traffic, surveillance, movie and documentary videos.

2.3 Colour-Based Extraction

In this, the foreground object is detected based on the colours involved in the frames. It mainly concentrates on RGB (red, green, blue) model, which is said to be the primary colours. Based on the colour, histogram value required key frames are derived. This is mainly used on sports and surveillance videos.

2.4 Audio–video Based Extraction

Input video is separated into the visual sequence, and the audio sequence and is processed separately. Speech is recognized and classified from the audio sequence to generate the audio summary. On combining the visual frames and the audio summary,

the required video clip is produced. This is mainly used on music, movie, phone calls and news videos.

3 Cluster based detection

Clustering is the process of combining the similar data points or values or activities into a single data. Here, it is used to cluster the similarities that are found in the frames and make it as a single point which is used to detect the required objects from the frames. It is classified into similarity, k-means, partitioning and spectral based clustering.

4 Event based detection

In this, the frames are extracted based on an event or incident that occurred in the given input video. Pixel intensities and absolute values are calculated between the present frame and with the reference frame. It used to determine with set of frames which consists of the events that are to be summarized. This is mainly used on sports, traffic and surveillance videos.

5 Video summarization methodology

Several techniques and methods used or proposed for the video summarization. They are processed with the help of different algorithms as per the requirement of the method define. These algorithms are used for detecting the objects, object extraction, retrieval and summarization. The main objective is to produce a precise video from the entire video stream which is given as the input. Some of the methodologies are discussed below.

5.1 Face Quality Assessment Using CNN

This method detects the face in the frame from the input and derives the frames with high face quality in video [2–4]. It consists of three main parts—face detection, quality score and key frame selection (Fig. 2).

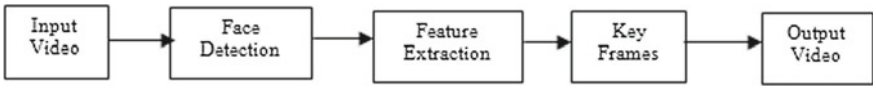


Fig. 2 Face quality assessment

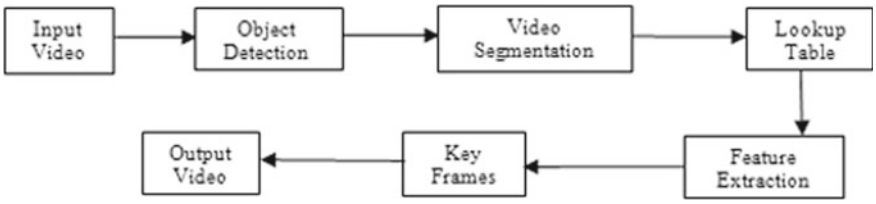


Fig. 3 CNN and Bi-LSTM

5.2 Multi-View Summarization Using CNN and Bi-LSTM

Multi-view video summarization is the process of working on different frames from multi-view videos such as from different camera for a single scenario [1, 8, 9]. This method integrates deep neural networks such as CNN (Convolution Neural Network) and Bi-LSTM (Bi-directional Long Short-Term Memory) to produce an accurate video summarization (Fig. 3).

5.3 Perceptual Video Summarization

In this, the given input video is registered as frames or images. This method process on the single frames or images and produce the final summarized video clip [10, 11]. GMM uses estimation of background and detection of object in the frames (Fig. 4).

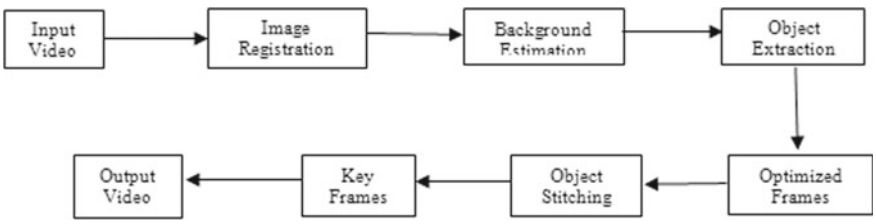


Fig. 4 Perceptual video summarization

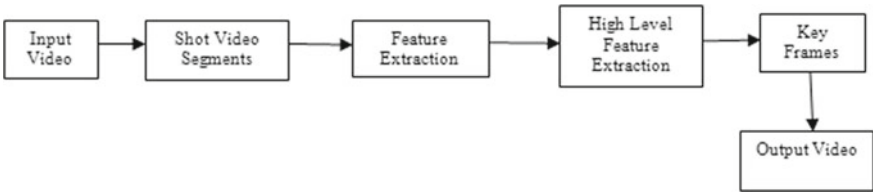


Fig. 5 LPQ-TOP based summarization

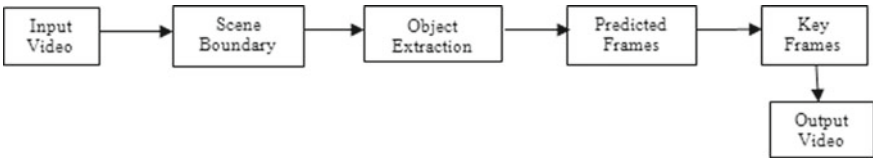


Fig. 6 Video summarization based on scene boundary detection

5.4 Dynamic Summarization of Videos Based on LPQ-TOP

The given input video is segmented into video shots. These video shots consist of subset of frames which includes a set of similar frames from the given input video [5]. Each shot will include a set of similar sequential frames, where first and the last frame defines the change of the event in the frames (Fig. 5).

5.5 Video Summarization Based on Scene Boundary

It uses summarization of videos on the scene in the frames from the given input video. The input video is divided based on the scenes involved in it. This scene detection is done with the help of motion-based feature extraction. Intensity value is produced between two successive frames based on which the changes in the scenes are determined. It is used to provide the scene-based boundary to the frames [12] (Fig. 6).

5.6 SalSum Methodology

It summarizes the key frames based on important features using SalGAN (Saliency Generative Adversarial Network) model. In this, each and every frame in the input video are processed based on the colour and the saliency features. The colour analysis is done with the help of hue histogram, and the saliency analysis is by SalGAN (Fig. 7).

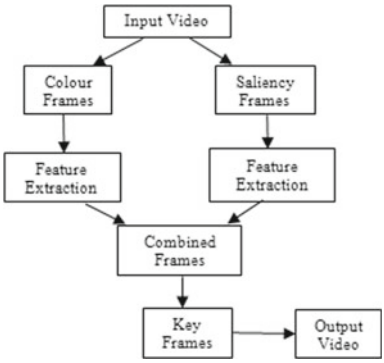


Fig. 7 Saliency generative adversarial network

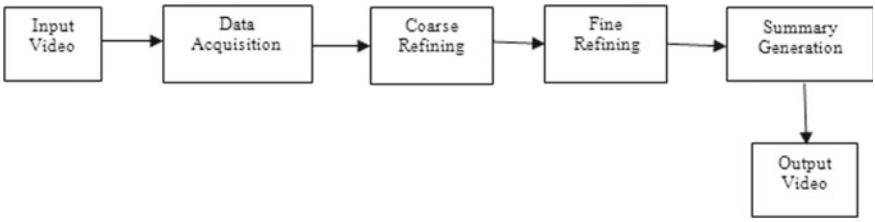


Fig. 8 Deep learning-based video summarization [1, 8, 9]

5.7 A Deep Learning-Based Video Summarization

This method is mainly divided into four parts—data acquisition, coarse refining, fine refining and summary generation. In the data acquisition, the input data is taken from the source and are registered into continues set of frames. Then, in coarse refining low-level feature extraction and comparison is made with each frame to select the required frames. This is done with the help of oriented FAST and rotated BRIEF (ORB) algorithm which compares two consecutive frames and produce the required frames (Fig. 8).

5.8 Character-Oriented Video Summarization

This method is mainly divided into three parts—detection module, re-identification module and aggregation module [3]. In this, the input video stream is processed as frames. Detection module is used to detect the frames that have the potential RoIs (Region of Interest) with any characters. This is done based on R-CNN (Fig. 9).

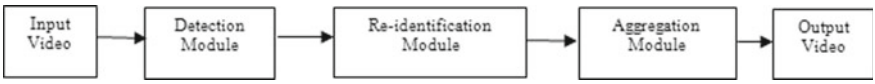


Fig. 9 Video summarization based on character-oriented [2, 4, 11]

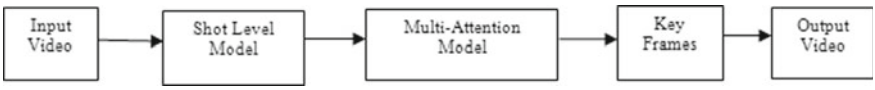


Fig. 10 Hierarchical self-attention for video summarization

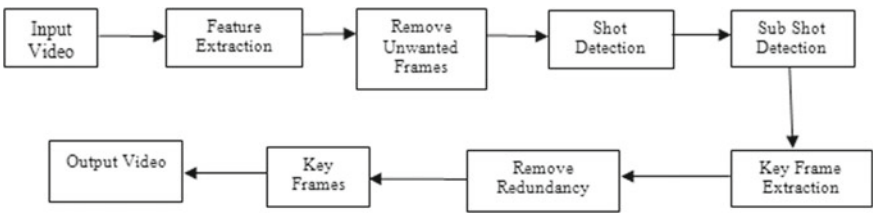


Fig. 11 Key frame video summarization

5.9 Hierarchical Self-Attention for Video Summarization

The method Hierarchical Multi-Attention Network is a 2-stage hierarchical structure which consists of shot-level reconstruction model and multi-head attention model. In shot-level reconstruction model, VAE-GAN model is used on the original video frames to combine similar frames into a video shot (Fig. 10).

5.10 A Key Frame Based Video Summarization

It is processed with the colour-based features. The colour histogram, HOG, saturation and contrast, are the features that are used to extract from frames from the given input video frames (Fig. 11).

6 Analysis of dataset and techniques of Video Summarization

It deals various analysis of video summarization techniques used for various applications. The methods like face quality assessment using CNN is used for detect the faces in the frame from input video and to extract the frames with high face quality.

MVS using CNN and Bi-LSTM is used to extract accurate video summarization. Hierarchical self-attention for video summarization method uses a two-stage hierarchical structure that consists of shot-level reconstruction model and multi-head attention model. Key frame based video summarization is processed with the colour-based features. The datasets used for abnormal event detection and localization of video anomalies are UCSD Pedestrian, UMN, CHUCK Avenue, Subway, PETS 2009, Anomalous Behaviour/York, QMUL Junction, MIT Traffic, Violent flows [16]. The most commonly used datasets for the video summarization are Own dataset and PETS2001, Open video project (OVP), VSUMM, Youtube database. SumMe dataset, Berkeley Segmentation dataset. UT Egocentric (UTE), ADL, Open video storyboard. In this context, various algorithms and techniques which are employed for video summarization processed for finding essential frame with accuracy. The datasets which are employed are processed in standard algorithms based on the algorithm functions.

7 Conclusion

Abnormal event detection like road accident and their surveillance video summarization in the field of computer vision used in intelligent transportation and secure video surveillance applications. The comparative analysis of various techniques and algorithms are used for abnormal event detection deals with advantages on higher accuracy with robust function. Video summarization has extreme functionalities based environments with optimal quality of services and parameters like F1 score. The main challenges that have to be considered in the surveillance systems are automatic detection of abnormal events in part of difficulties like occlusion, video quality, etc. In the part of video summarization, selection of the particular key frames for summarization of abnormal events will be difficult. Hence, this comparative analysis on abnormal event detection techniques with merits will provide essential ideas on the functional contribution on each algorithm with proper datasets and technical view of summarization techniques.

References

1. Muhammad K, Hussain T, Del Ser J, Palade V, de Albuquerque VHC (2020) DeepReS: A deep learning-based video summarization strategy for resource-constrained industrial surveillance scenarios. **IEEE** Trans Ind Inf 16(9)
2. Zhou P, Xu T, Yin Z, Liu D, Chen E, Lv G, Li C (2020) Character-oriented Video Summarization with Visual and Textual Cues", IEEE Trans Multimedia 22(10)
3. Thomas SS, Gupta S, Subramanian VK (2019) Context driven optimized perceptual video summarization and retrieval. In: IEEE transactions on circuits and systems for video technology, vol 29, no 10

4. Khan MZ, Jabeen S, ul Hassan S, Hassan MA, Khan Al-Khwarizmi MUG (2019) Video summarization using CNN and bidirectional LSTM by utilizing scene boundary detection. In: 2019 International conference on applied and engineering mathematics (ICAEM)
5. Kaur P, Kumar R (2018) Analysis of video summarization techniques. Int J Res Appl Sci Eng Technol (IJRASET) ISSN: 2321–9653; IC Value: 45.98; SJ Impact Factor :6.887, 6(I)
6. Jayabharthy J, Balamurugan G (2019) A survey on video summarization techniques and classification. Int J Res Anal Rev (IJRAR) 6(2)
7. Jayabharathy J, Balamurugan G, Vishnu Priya R (2021) Abnormal event summarization in video surveillance using hierarchical recurrent neural network, DE, pp 3568–3579
8. Basavarajaiah M, Sharma P (2019) Survey of compressed domain video summarization techniques. In: ACM computing surveys 52, 6, Article 116
9. Liu Y-T, Li Y-J, Yang F-E, Chen S-F, Frank Wang Y-C (2019) Learning hierarchical self-attention for video summarization. In: 2019 IEEE international conference on image processing (ICIP).
10. Nair G, Johns KE, Shyna A, John A (2019) An overview of machine learning techniques applicable for summarisation of characters in videos. In: 2019 international conference on intelligent computing and control systems (ICCS)
11. Ahmed SA, Dogra DP, Kar S, Roy PP (2019) Trajectory-based surveillance analysis: a survey. In: IEEE transactions on circuits and systems for video technology, vol 29, no 7
12. Balamurugan G, Jayabharathy J (2019) A study on moving object recognition for video surveillance applications. J Adv Res Dyn Control Syst 11(01-Special Issue)

Meanings of Machine Learning and Deepening Researches on Methods of Classifying and Recognizing Users Gender on Internet

Nguyen Dinh Trung, Trung-Hieu Le, Dinh Tran Ngoc Huy, Nguyen Thi Hang, Le Thi Thanh Huong, and Nguyen Thuy Dung

Abstract In various firms from technology, manufacturing till digital, ICT firms, etc., they have to deal with big data of clients and if we can classify users based on gender, age, and other criteria, and it will help to facilitate our businesses. SVM is representing (short abrreviation) for Support Vector Machine that used as method for classification and regression analysis, and used in fields of statistics and computer science. Standard SVM will take advantage of data (input) and classifies it into two different classes. The study will perform Research and learn about gender prediction problem, present some gender prediction methods that have been studied before. Through learning and applying input data preprocessing tools, and researching and studying about the Support Vector Machine algorithm on two layers and many layers, it builds a program to get content from users' posts on the social network Facebook.

Keywords SVM Support Vector Machine · Computer science · Internet users · Gender

N. D. Trung (✉)
National Economics University (NEU), Hanoi, Vietnam
e-mail: trungnd@neu.edu.vn

T.-H. Le · L. T. T. Huong
Dai Nam University, Hanoi, Vietnam
e-mail: hieult@dainam.edu.vn

L. T. T. Huong
e-mail: lethanhhuong@dainam.edu.vn

D. T. N. Huy (✉)
Banking University HCMC, Ho Chi Minh city, Vietnam

International University of Japan, Niigata, Japan

N. T. Hang · N. T. Dung
University of Information and Communication Technology, Thai Nguyen University, Taiyuan, Vietnam
e-mail: nthang@ictu.edu.vn

N. T. Dung
e-mail: ntdung@ictu.edu.vn

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_55

1 Introduction

Currently, advertisers are offering their entire information to all customers they have. That is why people users often face large amounts of inappropriate information such as age, occupation, culture and gender mismatch.

This information overload that does not reach the destination leads to a drop significant in online marketing. Since then, the classification of user's internet to provide statistics, advertising plans to help the system approach to provide relevant and useful information for each relative audience important. Starting from the current situation, the thesis will present about the sex determination method to classify internet users is implemented and performed using machine learning techniques, using user information of known gender and information about their web browsing history to train the receiver know the gender of other users when we only know the access history of the site and directory data that the person is interested in.

The scope of this study includes:

Issue: What are the methods to recognize user's gender on internet? What are research results?

2 Literature review

We can see that two-class classification algorithms such as SVM all have the common feature of requiring data to be represented in the form of feature vectors, but other algorithms must use parameter estimation and the optimal threshold, while the SVM algorithm can find these optimal parameters by itself. Among the methods, SVM is the method that uses the largest feature vector space (more than 10,000 dimensions) while other methods have much smaller dimensions (such as Naïve Bayes 2000, k-Nearest Neighbors is 2415…). According to Hieu et al. [1], automated learning behavior analysis of learners is becoming an essential topic in the field of education, where effective systems are needed to monitor learning progress and provide necessary feedback to instructor activities. Human activity recognition is to identify the type of human activity from the received signal stream (image, accelerometer, etc.). This is a topic that has received a lot of attention from researchers and has opened up a new trend in robotics interaction in recent times [2] performed the recognition of 20 natural gestures obtained from five body-mounted accelerometers. Features such as mean value, energy, entropy onfrequency domain, correlation features calculated on a window of 512 samples size (corresponding to 6.7 s) are passed through different classifiers such as: Bayes, decision tree. Also while female tend to join in making friends topics and advertising and express emotions more frequently, male will provide information for help and both of them express positive emotion in technology communication communication. And this will affect their activities in community. Le et al. [3] stated despite the fact of existing works on human hand gestures using wearable sensors, each focuses on a specific application and difficult

to be generalized. Experimental results show that with limited camera angles, the postures are highly distinctive and easily discriminated with the highest performance of 98.85% and 97.40% in terms of precision and recall, which motivates a wide range of applications and new research directions for human–machine interaction, wearable's, the Internet of Things (IoT) and so on. Huy et al. [4] mentioned computer played vital roles and the need of computer skills [4] and Huy [5] mentioned governance best practices needed. Huy et al. [6] stated vital banks roles in supporting this sector and confirmed by [7]. Hence, we make conclusive summary that we need to make deepening research on recognizing internet user gender via their internet usage especially with SVN method explanation and so, we propose this work should be added.

3 Methodology

The purpose of the SVM method is to find the maximum boundary distance, which is illustrated as follows (Fig. 1).

Theoretical basis

Authors use SVM method, in fact it is a math problem of optimization helping to find a space F and the decision hyperplane f over F so that the classification error is minimal.

The main steps of the SVM method need we express data under vectors of real numbers. And we need to convert non-numeric input into SVM's digital form. In the training process, we will use the algorithm to optimize the distance between hyperplanes in the classification process, determine the classification function in the feature space by mapping data into the feature space by describing kernel, solving for both data cases is separable and non-linearly separable in feature space.

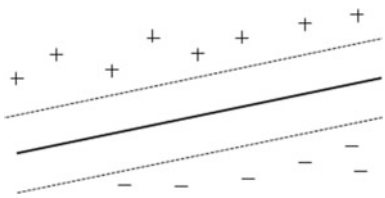


Fig. 1 The hyperplane divides the learning data into 2 classes + and – with the largest boundary distance

Table 1 Status list statistics by user and post

	Users		Status	
	Number	Ratio (%)	Number	Ratio (%)
Male	659	57.8	109,170	49.7
Female	482	42.2	107,702	50.3
Total	1141	100	216,872	100

Table 2 Statistics of the word count of the data set

Dictionary	Total rest
Unigram dictionary	12,923
Bigram dictionary	370,663
Trigram dictionary	1,230,451
Average	538,012

4 Main results

4.1 Overall Results

Description of input data.

The file full_status_filter.csv contains a list of status of many different users.

Table 1 is the statistics of the input data set by user and by status:

With per-user statistics, we consider a user to have many statuses, the set of statuses represents the gender of that user.

With statistics for each status, each status represents a user's gender, the statuses of the same user are separate when evaluated by articles.

Comment: From Table 3.8, I see that women write more status than men. On average, a user writes about 190, of which a male has 163 status and a female has 226 status (Table 2).

After I have a dictionary, I will find the corresponding weight and create a file in Liblinear format. With each dictionary will create 3 files with 3 corresponding weights, the number of occurrences, TF-IDF and binary. There are a total of 9 files as in Table 3.

Experimental results.

Run 9 files in turn on the computer with the following configuration:

Operating System: Desktop Windows 10.

Processor: Intel Core i5.

RAM memory: 16 GB.

Environment: Java 8.

Liblinear version 2.11.

The results in Table 4 show the accuracy of predicting the user's gender on each

Table 3 List of files in liblinear format

No	File	Description
1	Unigram_count.libsvm	Bộ từ điển unigram với trọng số xuất hiện của từ
2	Unigram_tfidf.libsvm	Bộ từ điển unigram với trọng số TF-IDF
3	Unigram_binary.libsvm	Bộ từ điển unigram với trọng số Binary
4	Bigram_count.libsvm	Bộ từ điển bigram với trọng số xuất hiện của từ
5	Bigram_tfidf.libsvm	Bộ từ điển bigram với trọng số TF-IDF
6	Bigram_binary.libsvm	Bộ từ điển bigram với trọng số Binary
7	Trigram_count.libsvm	Bộ từ điển trigram với trọng số xuất hiện của từ
8	Trigram_tfidf.libsvm	Bộ từ điển trigram với trọng số TF-IDF
9	Trigram_binary.libsvm	Bộ từ điển trigram với trọng số Binary

Table 4 Accuracy results of data set per user

	Count (%)	Binary (%)	Tf-Idf (%)	Average (%)
Unigram	*93.87*	90.89	77.83	87.53
Bigram	93.08	91.32	76.42	86.94
Trigram	92.38	91.85	*75.99*	86.74
Trung bình	93.11	91.35	76.75	87.07

individual status. The prediction on the entire status of the user will give the following results.

Table 4 shows the highest accuracy of 93.87% with unigram dictionaries and occurrence weights. The lowest accuracy result of 75.99% belongs to the trigram dictionary with TF-IDF weights. The difference between the highest and lowest accuracy is 17.88%. The average accuracy of 9 files is 87.07%.

According to tables above, we see that if we consider the weight, the average difference is 4.87%, in which the weight of the occurrence of the word gives the best result on average 93.11%, then the binary weight is 91.35%, and the lowest is TF-IDF 76.75% (Fig. 2).

If looking at the dictionary, the difference is quite small, only 0.62%, in which unigram gives the best results on average is 87.53%, then bigram is 86.94%, and finally bigram dictionary is 86.74% as shown.

To evaluate how the number of data sets affects the accuracy of the prediction, I will divide the original data set into small random sets with the number of Statuses of a set being 10,000, 50,000, 100,000, 150,000, respectively. Follow the same steps as the original data set, I obtained the results with tenfold cross validation method as in Table 5.

According to Table 5, the highest accuracy is 64.45% of trigram dictionaries with TF-IDF weights and the lowest is 61.57% of unigram dictionaries with weighted word occurrences, and the difference of the two accuracy is 2.88%. The average accuracy of the whole dataset is 62.73%.

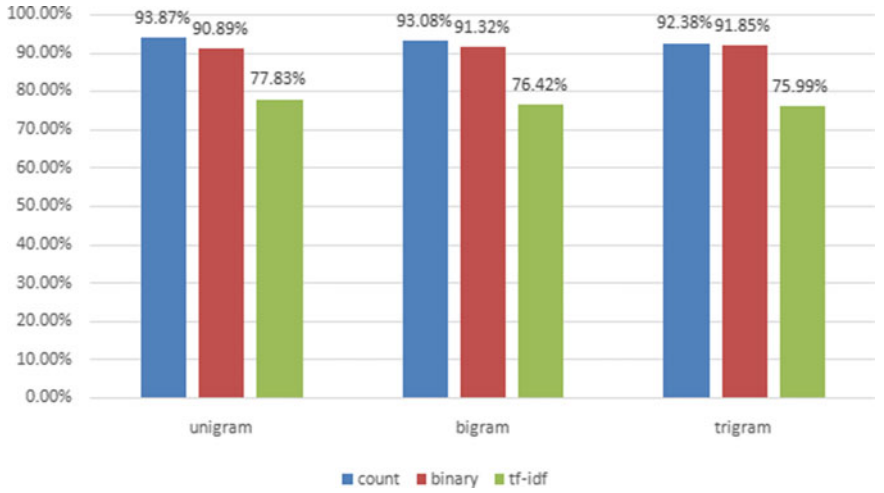


Fig. 2 The graph shows the results by dictionary of the data set for each user

Table 5 Accuracy results of data set with 10,000 Status

	Count (%)	Binary (%)	Tf-Idf (%)	Average (%)
Unigram	*61.57*	62.53	64.10	62.73
Bigram	61.66	61.96	64.15	62.59
Trigram	62.00	62.16	*64.45*	62.87
Trung bình	61.74	66.22	64.23	62.73

5 Discussion and Conclusion

Solving the problem of predicting the gender of social network users based on the article content in general and experiment with Facebook social network and the article content are in Vietnamese based on features. The problem is the foundation for many important applications to predict user gender in particular and other information in general.

The main results achieved by the thesis:

Research and learn about gender prediction problem, present some gender prediction methods that have been studied before; Analyze two characteristics of Vietnamese article content for preprocessing; Learn and apply input data preprocessing tools; Research and learn about the Support Vector Machine algorithm on two layers and many layers; Build a program to get content from users' posts on the social network Facebook; build a training and testing program with the obtained data set.

The novel of study: we generate implications of our study in which after we recognize gender of internet users, we can offer suitable services and products for male or female, based on prediction of gender and data of male or female preferences.

Limitation of research

Limit the quantity and quality of data that affect prediction results.

Acknowledgements Thank you editors, friends and brothers to assist this publishing.

References

1. Hieu LT, Hương DT, Huy DTN, Dung NTP, Trung ND (2021) Identifying learners' behavior from videos affects teaching methods of lecturers in Universities. Design Engineering, 11146–11157. Retrieved from http://www.thedesignengineering.com/index.php/DE/article/view/4117
2. Le T-H, Huy DTN, Huong LTT, Hang NT, Gwoździewicz S (2021) Recognition of user activity with a combined image and accelerometer wearable sensor. Design Eng 6407–6421. Retrieved from http://www.thedesignengineering.com/index.php/DE/article/view/3139
3. Le T-H, Huy DTN, Hang NT, Dung NT, Thanh NTP, Gwoździewicz S, Vu H (2021) Human machine interaction and meanings of machine learning—A case of hand posture recognition from wrist-worn camera. Design Eng 11174–11187. Retrieved from http://www.thedesignengineering.com/index.php/DE/article/view/4119
4. Huy DTN, Van PN, Ha NTT (2021) Education and computer skill enhancing for Vietnam laborers underindustry 4.0 and evfta agreement. Elem Educ Online 20(4)
5. Huy DTN (2015) The critical analysis of limited south asian corporategovernance standards after financial crisis. Int J Qual Res 9(4):741–764
6. Huy DTN, Dat PM, và Anh PT (2020) Building and econometricmodel of selected factors' impact on stock price: a case study. J Secur Sustain Issues 9(M):77–93. https://doi.org/10.9770/jssi.2020.9.M(7)
7. Huy DTN, Loan BT, Anh PT (2020) Impact of selected factorson stock price: a case study of Vietcombank in Vietnam. Entrepreneurship Sustain Issues 7(4):2715–2730. https://doi.org/10.9770/jesi.2020.7.4(10)

Suggesting Hotels Through Reviews Using Sentiment Analysis

S. Santhi, S. Thavasi, and N. Umakanth

Abstract Recommendation helps people in finding out the best choice based on the preferences and options of their own. Starting from the e-commerce websites, these recommendation systems play a major role in online advertisements that cannot be avoided in the Internet era today. These recommendation systems are used in variety of applications and services, and also, these systems have a remarkable place in the recommendation of hotels. A majority of people who come from other places find it difficult to find a hotel which suits their requirements in terms of money, hospitality, quality, location, and others is tedious. Hotel recommendation is required for those who come from other places and come in search for best hotels that suits their needs and own interest. This recommendation system recommends the users on basis of the comments given to the hotel previously by other customers who have visited the hotels. The customers express all their feelings and other specialties, drawback and features in comments that are helpful for the upcoming customers to judge the hotel. This system is useful for those people who often visit new places. The information extracted from the previous customers is a valuable resource that is used by the system to analyze the hotel's quality and standards and also help the new customers who have no idea about the place and the hotel.

Keywords Recommendation systems · Score calculation · Topic modelling · Text classification · LSTM · LDA

S. Santhi (✉) · S. Thavasi · N. Umakanth
Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, TamilNadu, India
e-mail: santhicse@mepcoeng.ac.in

S. Thavasi
e-mail: sktthavasi@mepcoeng.ac.in

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_57

1 Introduction

Recently, the Internet is globally wide and connects a lot of people via network, and it is becoming more and more difficult for the companies to find which sites are preferred by which group of users. As the number of customers buying products online is rapidly growing, it is really a challenging task for the companies to predict the user's interest. Companies may also require the recommender systems to retain the customers they currently have and also to attract other new users to their Web portals through attractive online advertisements and other ideas. Nowadays, social media platforms, especially Twitter [1] is the main source of recommendation system by using sentiment analysis and opinion mining. Recommendation systems give suggestions so that people nowadays depend on these services so often. It helps in making better decisions among number of choices and suggestions. Recently, it became a tedious task for the tourist and other people visiting new places. It comes out with a solution to help the confused people to get a suggestion based on their requirements. The main purpose of the recommendation systems is that the company overcomes the fear of losing customers by suggesting the correct options.

This work aims to recommend the hotels based on the ratings and reviews given by the previous users. The purpose of this work is recommending the name of top five hotels based on certain aspects to users by analyzing the other user's ratings and reviews recorded previously to improve the prediction accuracy. This work will include preprocessing of the raw dataset into a form suited for further processing like text classification, sentimental analysis score calculation and then finally grouping the hotels based on the aspects like the food, location, cleanliness and so on. The user who is new to the place can find a suitable hotel and most importantly the hotel that suits their needs and priorities. The aspect-based grouping helps the customers to find a more customized hotel for them, and they are provided with the top 5 hotels that are good at the given aspect, so the best suited hotel can be picked with assurance than to choose a new unknown hotel.

2 Literature Survey

Most of the applications of recommendation system like movies, music, studies and e-business services are based on feedback given by the user, and it can be implemented using sentiment analysis, Latent Dirichlet Allocation (LDA), text classification etc. The people's feedback is considered as a valuable source for hotel recommendation system because they are impartial to a particular hotel, and they may clearly explain a hotel's condition without hiding the demerits. So rather than advertising the hotel, this is a good idea to help the other place tourists to predict and divide in a quick manner and in no time.

(A) *Sentiment Analysis*

One of the applications like visualizing the student's opinions [2] was done through text analysis. Once the text has been selected, preprocessed and find the aspects using LDA method, then, separate the student's comments into ideas and calculating the sentiments of each idea using AFFIN Lexicon. Finally, find the satisfaction and sentiments score of each comment and idea by using group sentiment and satisfaction method. In this method, graph was generated for positive and negative sentiments against identified aspects and takes the average satisfaction of all the comments present in the single aspect.

Sentiment analysis [3] is the method of mining the people's views, feelings and opinions from the online text which will be collected by various surveys and reviews. These online reviews can have a great influence and promote the business or a product using text analysis technique. Various sentiment analysis techniques like supervised, unsupervised, cross domain and language classification have been tested widely in all fields and used to classify the positivity or negativity of movie reviews [4]. Many components of sentiments present in the movie review are similar to students' opinions about the courses [2]. The methodology used in this work is distinguished by its sentiment and visualization for ease of use by instructors.

Sentiment classification [5] was also used to classify documents rather than topics as positive or negative by using various machine or data mining technique such as support vector machine, naïve Bayes classification and maximum entropy classification. Bag-of-features framework was used to implement these machine learning algorithms for movie review corpus. The hate speech from Twitter [6] was classified by using sentiment classification with positive and negative by using different machine learning models like logistic regression, K-nearest neighbor, naïve Bayes and support vector machine.

(B) *Latent Dirichlet Allocation*

LDA is a recent topic modeling technique to separate the semantic information from the text data. LDA symbolizes each of the topics as a set of terms and then maps the set of terms with the documents by using hidden topics [7, 8]. Using this LDA method, the student's opinion [2] was analyzed and visualized through corpus or set of text documents. Each of the documents comprises of various words and comments given by a student. From this information, LDA was used to infer topics for visualizing, and each topic is referred to as an aspect of the students comment.

(C) *Text Classification*

Text classification is the method of analyzing the text automatically then assigns a set of well-defined tags according to its content in the structured and unstructured dataset. This text classification can be done in two ways, either manual or automatic classification. Many approaches are handled in automatic classification like rule-based, machine learning-based and hybrid systems. Using this traditional method for classifying the hotel based on online review data [8]. This work was started with online review data as documents

and processes the hotel classification by extracting features, weight calculation of extracted features and classifying the hotels based on hotel cluster. The trending topic of the Twitter was classified and ranked by using similarity measures and class-wise trend analysis [9]. The hotel recommendation system [10] was done by using online review. In this work, the author used word2vec, sen2vec, text2vec and doc2vec models for analyzing the online reviews and then recommend the hotels.

3 Methodology

The proposed work recommends the hotels available in a particular location based on user preferences, by considering the reviews for a particular hotel along with the calculated AFINN score to improve the prediction accuracy. The system design of hotel recommendation is shown in Fig. 1.

(A) *Text preprocessing*

Text preprocessing is a crucial step before doing any text mining task such as punctuations removal, stop words removal and text normalization for the reviews. It is done on the hotel reviews to convert the raw data into a useful and an efficient format for processing. The collected reviews are first lower cased and then white space and punctuation likes !,#,@,$,%,?. are removed, also stop words like 'to,' 'for,' 'is,' 'was,' 'are' in the reviews are removed. The reviews are stemmed and the prefixes and suffices like 'ily,' 'ing,' 'es' are removed and then lemmatized to get the root word. In lower casing, the

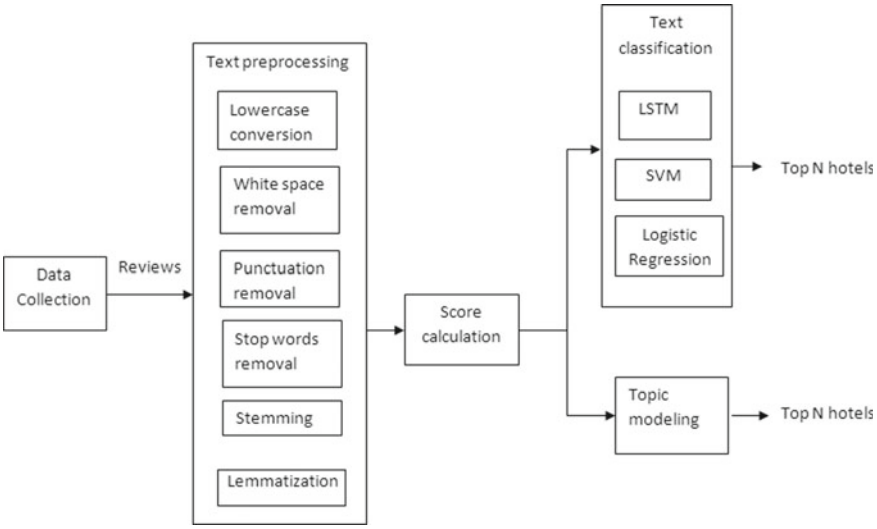


Fig. 1 Hotel recommendation system

reviews are transformed into lower case and avoid multiple copies of the same words. In punctuation removal, the punctuations in the reviews are removed, as it does not include any additional information while handling the text data.

Stop words removal includes the removal of commonly occurring words in the reviews. The NLTK package provides a list of stop words available in English and used as a reference for stop words removal. Then tokenization is performed to transform reviews into a blob and then converted them into a series of words. Stemming is a technique which is used to remove the common suffix from a word and reduce it to a single root word using Porter Stemmer algorithm. Lemmatization is a technique to reduce a token to an actual language root word using Word NetLemmatizer, and it is an effective option than stemming.

(B) *Score calculation*

Calculating the scores for reviews is necessary to identify the sentiments of the reviews. AFINN score is calculated using lexicon-based sentiment analysis approach. Here, the AFINN-en-165 lexicon is used to find the scores. AFFIN consists of above 3300 words and each word allied with a polarity score. It is a list of words rated for valence along with an integer with various positive and negative values. The score will be calculated by validating the string tokens with the AFINN list. Thus, the score calculation for a review is the sum of scores of each token. After calculating AFFIN score for each review, the scores are then normalized to range $+5$ to -5. Equation (1) is simply the comparative score for each hotel.

$$\text{Score} = \frac{\text{sum of scores of each review}}{\text{Total number of reviews}} \tag{1}$$

(C) *Text classification*

It is one of the methods of sentiment analysis which is a supervised learning technique since it processes labeled dataset. It classifies the reviews as either positive or negative. Three supervised classification algorithms are used such as support vector machine (SVM), logistic regression and neural network like long short-term memory (LSTM) which gives the highest accuracy. First, the categorical text data into machine understandable numerical data using label encoder which is an important preprocessing step in supervised learning.

Long short-term memory network is a kind of recurrent neural network which can be accomplished by learning the sequence of long-term dependency for the given problems. The foremost step is tokenization which is used to convert words into tokens and map to an integer. The embedding layer will be used to convert word tokens into specific size. The LSTM layer defines the latent state dimensions and the number of latent layers. At last, the fully connected layer uses the sigmoid function for mapping the output of the LSTM layer to a desired output size.

Support vector machine is used to classify the data points for separating a hyper plane in N-dimensional space. SVM will generate hyper planes iteratively that segregates the classes and from those hyper planes, it will choose the hyper plane with maximum margin that separates the classes correctly as positive and negative.

Logistic regression is also a classification technique that predicts a binary outcome. The preprocessed reviews are label encoded and then passed as input. Initially, the random number generator is initialized to zero. The solver is set to handle only small dataset, and they also support a minimal loss with primal formulation. The class fits across the entire probability distribution even when the data is binary.

(D) *Topic modeling*

Topic modeling is the method of recognizing most commonly used words in the reviews. This technique highlights the topics or aspect in the review. Latent Dirichlet Allocation is one of the unsupervised learning algorithms that maps the documents into bag of words. It works by the way of assigning unique ID for each word and set of words mapped with each topic and then set of topics mapped with document. Then converts the document (a list of words) into the bag of words and identifies the most commonly used words and lists it along with the probabilities.

(E) *Recommendation of hotels*

The hotels are recommended based on the text classification technique as well as the topic modeling approach. The top N hotels are recommended based on different aspects such as location, stay, staff, room, food and for general category also.

4 Results and Discussion

The hotel recommendation system results are explained in detail below. The dataset is taken from Trivago India which contains the hotels in Chennai along with the reviews given by people. Table 1 shows the sample reviews of the hotels along with the sentiment and rating.

The raw data cannot be used for recommending without cleaning it. This data is preprocessed by converting to lowercase, punctuation removal, stop words removal, tokenization, stemming and lemmatization. Consider one of the hotel reviews given by user and apply all preprocessing steps:

Table 2 shows some sample reviews after applying all preprocessing steps for Table 1.

This preprocessed data is used for further processing. The score for each hotel review is calculated using AFFIN and normalized to the scale of 5 to each hotel. Table 3 shows the normalized scores for each review.

Figure 2 shows the recommendation of top five hotels using AFFIN score.

Table 1 Chennai reviews dataset

Hotel name	Review Text	Sentiment	Rating(%)
Accord Metropolitan	Wow !!! Its truly pleasant room to visit expressly for commercial and visitor	3	100
Accord Metropolitan	Its appearances hotel doesn't check the services and comfortless in the room, before passing to the traveler	1	20
Park Hyatt	Excellent value for quality and service. Can't go wrong. Breakfast is top notch with Indian and continental options. Decor is subdued but stylish	3	86
The Park Chennai	Hotel staff are not coordinating well for the room service and Billing payment	1	40
The Residency Towers	Very clean! Great Service!	3	100

Table 2 Preprocessed reviews

Hotel name	Review text
Accord Metropolitan	Wow really nice place stay especibusi tourist purpose
Accord Metropolitan	Seem hotel check basic amen room hand room travel
Park Hyatt	Excel valuqualiti serviccant go wrong decor subdu stylish

Table 3 Score calculation

Review text	Normalized score
Wow true pleasant room visit express commerce visitor	0.83
Appearance hotel check service comfort room pass travel	0
Excel value quality service can't go wrong decor subdu stylish	−0.67

Here, the recommendation is done in two ways. First method is text classification and second method is topic modeling using LDA. In text classification, we use three classifiers SVM, logistic regression and LSTM. Table 4 shows the confusion matrix of SVM classifier, and the accuracy is 80.98%.

Table 5 shows the confusion matrix using Logistic regression classifier and the accuracy is 89.04%.

LSTM classifier is also used to classify the sentiment. Table 6 shows the model and parameters defined for LSTM classifier like 12,753,025 trainable parameters and 0 non-trainable parameters. 3171 training samples and 1563 validation samples are used in LSTM along with 10 epochs and finally achieved accuracy as 96%.

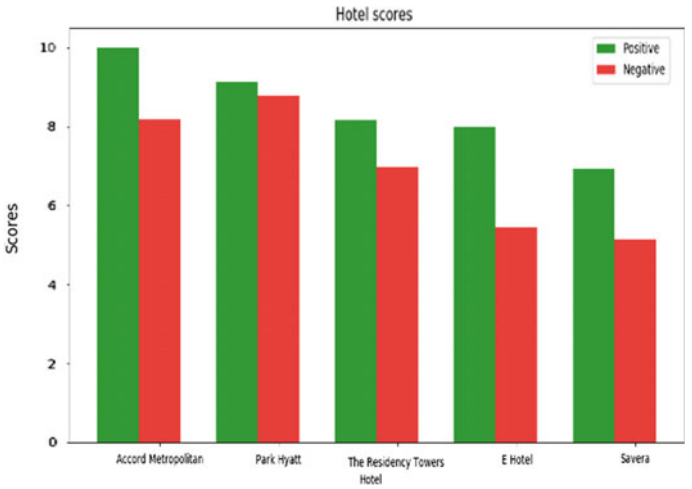


Fig. 2 Recommending top five hotels using AFFIN Score

Table 4 Confusion matrix using SVM

	Negative	Positive
Negative	21	83
Positive	1	847

Table 5 Confusion matrix using Logistic regression

	Negative	Positive
Negative	0	100
Positive	0	852

Table 6 LSTM Model Construction

Layer (type)	Output shape	Param #
embedding_1 (Embedding)	(None, 50, 128)	12,621,312
lstm_1 (LSTM)	(None, 128)	131,584
dense_1 (Dense)	(None, 1)	129

By comparing the three classifiers, LSTM is more efficient; hence, it is taken to classify the sentiments as positive (1) and negative (0) for the reviews. Table 7 shows the sentiment classified by LSTM.

By using LDA method, we will find different aspects from the reviews used by people. By using LDA in our recommendation system, we have found four aspects namely food, place/location, room, staff and service. Based on each topic, the top 5 hotels are recommended with ratings for each hotel. Table 8 shows the recommendation of the hotels based on food.

Similarly, Table 9 shows the top 5 hotels based on location along with the ratings

Table 7 LSTM sentiment

Reviews	LSTM_sentiment
seem hotel check basic amen room hand room travel	0
excel value quality service can't go wrong decor subdu stylish	0
excel room exerc is facil around atmosphere calm comfort main dine room offer excel food servic	1

Table 8 Recommendation using Food

Hotel name	Rating
Savera	5.0
Raddison, Chennai City Centre	5.0
Peace Inn	5.0
Park Hyatt	5.0
The Residency Towers	4.0

Table 9 Recommendation using Location

Hotel name	Rating
Fab Hotel Blossoms	5.0
Radisson Chennai City Centre	4.0
Accord Metropolitan	3.0
Peace Inn	3.0
The Park Chennai	3.0

on the scale of 0 as poor to 5 as best.

Table 10 shows the top five hotels based on the room and comfort in room.

The top five hotels based on staff and service are shown in Table 11.

Finally, the top 5 hotels from the reviews without considering the aspects are also found and shown in the Fig. 3.

Table 10 Recommendation using room

Hotel name	Rating
Peace Inn	5.0
ITC Grand Chola	5.0
The Residency	5.0
Savera	4.0
Raj Park	4.0

Table 11 Recommendation using staff and service

Hotel name	Rating
Radisson Chennai City Centre	5.0
Peace Inn	5.0
Park Hyatt	4.0
ITC Grand Chola	4.0
The Residency Towers	3.0

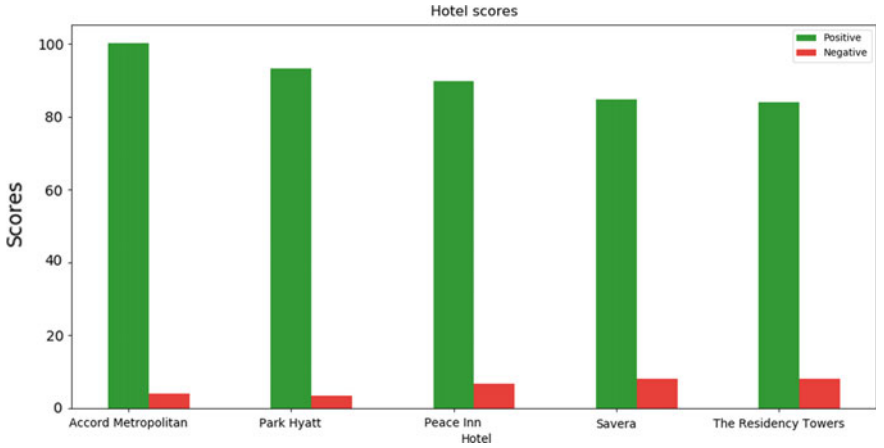


Fig. 3 Recommending Top five hotels using LDA

5 Conclusion

In this proposed system, aspect-based sentiment analysis approach is implemented by analyzing the user's behavior using hotel reviews. In this work, it mostly concentrates on designing a recommendation system for hotel industry domain where users can select the hotel according to their aspects preference and view the top hotels. The performance of recommendation system could be increased by using various text-based and sentiment classification techniques and analyzed the reviews based on accuracy. In future, this work can be improved with the help of various deep learning algorithms with sentiment analysis for suggesting the hotels, and we can recommend the hotels in different cities. This work can be further compared with more deep learning techniques for improving the accuracy and quality of hotel recommender system.

References

1. Pak, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of LREC, vol 10, no 2010, pp 1320–1326
2. Cunningham-Nelson S, Baktashmotlagh M, Boles W (2019) Visualizing student opinion through text analysis. IEEE Trans Educ 62(4)
3. Pang B, Lee L (2008) Opinion mining and sentiment analysis. In: Foundations and trends in information retrieval, vol 2
4. Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Human Lang Technol 5(1)
5. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of ACL conference empirical methods for natural language processing, vol 10, pp 79–86
6. Santhi S, Umakanth N, Hemalatha T (2019) Determining hateful and offensive terms from twitter using hate speech detection. Int J Pharm Res 11(1):867–871
7. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022
8. Song Y, Pan S, Liu S, Zhou MX, Qian W (2009) Topic and keyword re-ranking for LDA-based topic modelling. In: Proceedings of the 18th ACM conference on information knowledge management, pp 1757–1760
9. Umakanth N, Santhi S (2020) Classification and ranking of trending topics in twitter using tweets text. J Crit Rev 7(4):895–899
10. Qin H, Ye X, Zhao Y Cai X (2018) Hotel classification based on online review data. In: Proceedings of the 14th international conference on natural computation, fuzzy systems and knowledge discovery, pp 264–270

Evaluating Cost Contingency for Construction Projects: A Fuzzy Risk Analysis Approach

Pradeep K. Goyal and Savita Sharma

Abstract Controlling and managing project costs in infrastructure construction projects within budget is a matter of prime importance since these projects require a large amount of capital investment. The use of cost contingencies is found as an effective tool for reducing the cost overrun. Traditionally, the contingency is estimated using a fixed percentage of the estimated cost. Project costs' sensitivity to risk factors impacting the cost is not considered in this method resulting in underestimated or overestimated values. Therefore, in this paper, an alternate methodology is presented for developing a risk-induced model to predict the cost contingency after identifying and quantifying the risks involved in the projects. To develop the model, a rule-based fuzzy inference system has been used. The fuzzy theory can deal with incomplete, imprecise and uncertain data intrinsic to complex construction projects. This methodology provides a practical approach for estimating cost contingency by considering the frequently occurring and important risk factors impacting the cost of construction projects. Details about the development and validation of the model are presented in this research study. Project managers and decision-makers will find this model very useful for making decisions regarding various issues related to the project such as contingency estimation, bid price calculation, mark-up estimation and assessment of different projects.

Keywords Cost contingency · Risk · Construction projects · Fuzzy logic · Fuzzy inference

P. K. Goyal
Delhi Technological University, Delhi 110042, India
e-mail: pkgoyal@dtu.ac.in

S. Sharma (✉)
Government Polytechnic College, Ajmer, Rajasthan 305002, India

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_58

1 Introduction

The issue of cost overrun in the construction projects is a frequent and critical both in developed and developing countries. Controlling and managing project costs in infrastructure construction projects within budget is a matter of prime importance because of the investment of a considerable capital cost. The use of cost contingencies is found as an effective tool for reducing the cost overrun. The primary aim of contingency planning is to avoid the cost overrun problem in construction projects [5]. The contingency can be defined as the total quantity of budget, funds or time required above the estimated amount to reduce the risk of overruns of project goals to an acceptable level for the organisation [25]. Traditionally, the contingency is estimated using a fixed per cent of estimated cost. This technique has also been called as 'across-the-board' percentage addition [1] and 'Crystal ball' [23]. Using this method, the contingency estimation usually varies from 1 to 5% and it rarely goes beyond 10% [23]. The amounts are estimated on basis of the maximum cost, average cost, expert judgement and experience which can result in the underestimated or overestimated quantity [22]. This method is described as arbitrary and unscientific by various researchers [1, 6]. Project costs' sensitivity to risk factors impacting the costs is not considered in this method. However, it is observed that on account of the complexities involved with construction activities, risks and uncertainties are increasing rapidly in construction projects. Effective risk management and adopting contingency as a tool can be useful for construction project managers to control and handle these risks and uncertainties [16, 25]. As stated by Ford [12] and Marco et al. [21], contingency amount is a matter of prime importance for managing risks associated with construction projects. Therefore, the contingency modelling using knowledge-based risk assessment and incorporating integrating risk management strategy is gaining importance. Various authors [8, 18, 23, 24] adopted a risk analysis approach to estimate cost contingency in construction projects.

The risk can be analysed using quantitative and statistical methods such as Monte Carlo simulation, fault tree analysis, sensitive analysis and failure mode analysis. However, these methods require exact and accurate data. Acquiring these data for complex situations such as those associated with construction projects is a challenging task as they contain too many variables with a high degree of uncertainty and ambiguity. For dealing with the complex and subjective nature of the problems, advanced techniques such as fuzzy logic have been gaining popularity. By comparing the various techniques and theories which are employed for handling uncertainty in the construction projects, Baloi and Price [2] proposed for FST as a robust tool for analysing the uncertainties associated with construction-related activities. The fuzzy set theory (FST) manages data that are partially defined, inexact and uncertain. It is well-suited to deal with almost identical situations to those encountered in complex and large infrastructure construction projects. Hence, this study attempts to provide a framework for model development to predict project cost contingency by using the fuzzy risk analysis approach.

Section 2 of this research paper describes the relevant literature review. In Sect. 3, the methodology is presented for developing the cost contingency model. The developed model is validated in Sect. 4 of this paper, and finally, the conclusion has been drawn.

2 Literature Survey

Sections 2.1 and 2.2 present a brief literature review related to the development of a model for estimating cost contingency in construction projects and fuzzy inference systems, respectively.

2.1 Cost Contingency Models

Various methods such as the method of moments, factor rating, range estimating, artificial neural network, Monte Carlo simulation, regression, analytical hierarchy process, fuzzy set theory, CALM (computer-based), probabilistic model, PERT and standard deviation have been used by the researchers for estimating cost contingency. References [7, 23, 34] employed method of moments, whereas a probabilistic model was suggested by Touran [31] to estimate the cost contingency of a project by taking into consideration the expected number of variations and the average cost of change orders. Barraza and Bueno's [3] presented a method based on Monte Carlo simulation. Activity costs in this model were assumed to be independent and normally distributed [3]. Later on, Barraza [4] presented an empirical methodology by employing Monte Carlo Simulation, and cost contingency was allocated to project activities according to the detail of the work breakdown structure. Lorance and Wendling [17] also proposed a Monte Carlo simulation for estimating the cost contingency. Sonmez et al. [29] suggested a model based on regression analysis to forecast the cost contingencies for international projects after collecting the data from Europe, Asia, Africa and the Middle East. Thal et al. [30] also suggested a regression model to forecast the amount of contingency for a new construction project after examining 203 projects of the U.S. Air Force. Artificial neural network and multiple linear regression methods were investigated by Chen and Hartman [6] on the basis of data collected from an oil and gas company for the prediction of contingency; it was concluded by them that the artificial neural network method was more accurate than the multiple linear regression method for the estimation of contingency. Moselhi et al. [4] also found, based on an extensive literature review, that the artificial neural network model had the potential of identifying the pattern knowledge and data prediction. Artificial neural network methods were used by Jin and Zhang [15] and Williams [33] also. It is found that to learn the knowledge or the pattern by artificial neural network (ANN) is somewhat difficult, but the Fuzzy system can improve the qualitative aspect of human knowledge such as reasoning,

inference and explicit knowledge. Therefore, fuzzy methods have been used for predicting the cost contingency. The important studies have discussed here in this section. The probabilistic cost estimate of a project was predicted by Hassanein and Cherlopalle [13] using fuzzy theory by considering the risks and uncertainties. Another model was recommended by Paek et al. [24] to present a methodology for evaluating project risk and contingency for reducing the risks related to a project based on fuzzy logic. Fuzzy sets were also used by Shaheen et al. [26] for calculating project cost estimates. Many authors have suggested extracting fuzzy numbers from the expert group and analysing the data in the fuzzy range estimating analysis. Idrus et al. [14] recently proposed cost contingency estimating by using risk analysis and a fuzzy expert system. The method is flexible, rational, and it is based on fuzzy expert system. A neuro-fuzzy hybrid model, designed by Dominic et al. [9], is used to predict the final cost of infrastructure projects of water. Wang et al. [32] developed a method using neuro-fuzzy and multi-factor evaluation for estimation the cost of projects.

Based on the comprehensive literature survey, it can be concluded that a variety of models have been developed by researchers using several techniques such as Monte Carlo simulation, multiple regression, artificial neural network and probabilistic model. However, these techniques necessitate complex mathematical calculations, and predicting any information by recognising the knowledge and patterns is relatively difficult. In contrast, fuzzy systems are proven to be more effective for forecasting information since they can improve the qualitative aspect of human knowledge, such as reasoning, inference and explicit knowledge. Models based on fuzzy and fuzzy hybrid approaches have also been suggested for estimating the cost contingencies depending on the research question and problem under consideration. Despite the efficacy of fuzzy theory, there are only a few studies in the literature for evaluating cost contingency, which motivated the author to propose an alternate methodology using a rule-based fuzzy inference system for estimating cost contingency in a construction project.

2.2 Concept of Fuzzy Theory

The fundamentals and concepts of the fuzzy theory, which are used for designing the model, have been explained here in this section. The fuzzy inference system is also described briefly as the model is developed using a rule-based fuzzy inference system.

Fuzzy set

The concept of fuzzy logic was originally established by Zadah [35]. A fuzzy set A of a universe of discourse X can be defined through Eq. (1):

$$A = \{(x,\ \mu A(x))/x \in A, \mu A(x) \in [0,\ 1]\} \tag{1}$$

Here $\mu A(x)$ = Degree of membership function which provides X a membership value in the range from 0 to 1.

Membership functions

Membership functions (MF) are used for the fuzzification and defuzzification process of fuzzy inference system (FIS), for transforming the non-fuzzy or crisp input values to fuzzy values in the form of linguistic terms and vice versa. MF can be defined by a curve that represents the degree of membership of an element x in a fuzzy set in the range from 0 and 1. Triangular, Gaussian, trapezoidal, piecewise-linear, bell-shaped, sigmoid membership function, etc. are generally used as membership function. In this study, a trapezoidal membership function is taken for input and output variables based on literature review.

Fuzzy inference system

Fuzzy inference system (FIS) can be described as a rule-based reasoning system for mapping a given input to an output by applying fuzzy logic theory. Takagi–Sugeno and Mamdani processes are very important approaches for fuzzy inference. In this research, study the Mamdani type fuzzy inference process has been used as it is intuitive, well suitable to human cognition, and widely used in literature [11, 19, 20]. In the Mamdani process of FIS input members of the system are transformed into fuzzy numbers by fuzzification using membership function, by which crisp inputs are converted into a set of linguistic variables. A set of fuzzy rules in the form of if–then is then formed with the help of various fuzzy operators. The fuzzified input members are summed together as per the constructed rules, and the output distribution in the form of fuzzy sets are obtained Finally, after the defuzzification process, a crisp quantity is obtained from a fuzzy set.

3 Methodology for Developing the Model for Evaluating Cost Contingency

In this section, the methodology has been presented for developing the model for evaluating the cost contingency of a project. For this purpose, a case study has been taken from Indian construction projects.

3.1 Identification of Risk Factors Impacting the Cost of Construction Projects

For identification of the risk factors impacting the cost of construction, an extensive literature survey was carried out across the globe by Sharma and Goyal [27] for the construction industry. Fifty-five important factors impacting the construction

Table 1 Frequently Occurring Cost Escalation Factors in Construction Industry

S No.	Factors impacting cost	Category
1	Fluctuation in the prices of materials	Finance
2	Lowest possible bid procurement policy	Project and contract
3	Inflation	Finance
4	Improper govt. policy	management
5	Incorrect estimation of time and cost	construction
6	Errors and discrepancy in contract document	Project and contract
7	Quantity of Extra work done	construction
8	Frequent modification in design	construction
9	Impracticable contract duration	Project and contract
10	High project charge of labour	Finance
11	Financial difficulty experienced by contractor	Finance
12	Improper planning and scheduling of project by contractor	construction
13	Slow procedure in taking decisions	Management
14	Poor coordination between construction rticipants	Management
15	Exchange rate	Finance
16	Material Cost	Finance
17	High level of interest rate of bank loan	Finance
18	Modification in the scope of the project	Project and contract
19	Contract management	Project and contract
20	Conflicts and disputes on site	Management

cost were identified through systematic and intensive literature review (Journals, Proceedings, Web). After conducting interviews and a questionnaire survey with 50 construction practitioners from the Indian construction industry, only 20 important factors are selected for developing the cost contingency model. Table 1 indicates 20 main factors impacting the cost of the Indian construction projects. These factors are divided into four groups: 'Finance', 'Construction', 'Management' and 'Project and Contract'.

3.2 Risk Quantification of the Identified Factors Using the Risk Matrix

The risk is then quantified using fuzzy theory and risk matrix as presented by Sharma and Goyal [28]. In this method, the magnitude of the risk factors impacting cost has been calculated after considering the probability and severity level of a certain factor. The probability and severity index were estimated based on an interview conducted

Table 2 Magnitude of Risk factors Impacting Cost

Abbreviation	Factors impacting cost	Category	Quantity
R.F. 1	Fluctuation in the prices of materials	Finance related	0.694
R.F. 2	Inflation		0.687
R.F. 3	Financial difficulty experienced by contractor		0.661
R.F. 4	Exchange rate		0.642
R.F. 5	Material Cost		0.635
R.F. 6	High level of interest rate of bank loan		0.622
R.F. 7	High project charge of labour		0.664
R.F. 8	Improper planning and scheduling of project by contractor	Construction related	0.658
R.F. 9	Incorrect estimation of time and cost		0.68
R.F. 10	Frequent modification in design		0.673
R.F. 11	Quantity of Extra work done		0.675
R.F. 12	Poor coordination between construction participants	Management related	0.648
R.F. 13	Conflicts and disputes on site		0.576
R.F. 14	Slow procedure in taking decisions		0.652
R.F. 15	Improper govt. policy		0.682
R.F. 16	Contract management	Project and Contract related	0.591
R.F. 17	Impracticable contract duration		0.669
R.F. 18	Errors and discrepancy in contract document		0.678
R.F. 19	Lowest possible bid procurement policy		0.69
R.F. 20	Modification in the scope of the project		0.618

with 50 experts in the construction industry of India. The calculated magnitude of the risk factors impacting cost has been presented in Table 2.

3.3 Developing a Model for Evaluating Cost Contingency by Using Fuzzy Inference Process

To develop the model for evaluating cost contingency, a fuzzy inference process has been implemented in two phases. In the first phase of the fuzzy inference process, the

impact of the various risk factors of a group is taken as 'input variables', and 'output variable cost contingency' is determined for this group. For example, as shown in Fig. 1, the cost contingency of the group 'Finance' ('C.C. Finance') is estimated by considering the impact of risk factors (I.L.F. 1, I.L.F.2, I.L.F. 3, I.L.F. 4, I.L.F. 5, I.L.F. 6, I.L.F. 7) related to the group 'Finance' as input variables. Similarly, The output variable 'C.C. Construction', 'C.C. Management' and 'C.C. Project and Contract' is determined for 'Construction', 'Management'and 'Project and Contract' group.

In the second phase of the fuzzy inference process, 'C.C. Finance', 'C.C. Construction', 'C.C. Management' and 'C.C. Project and Contract' determined in the first

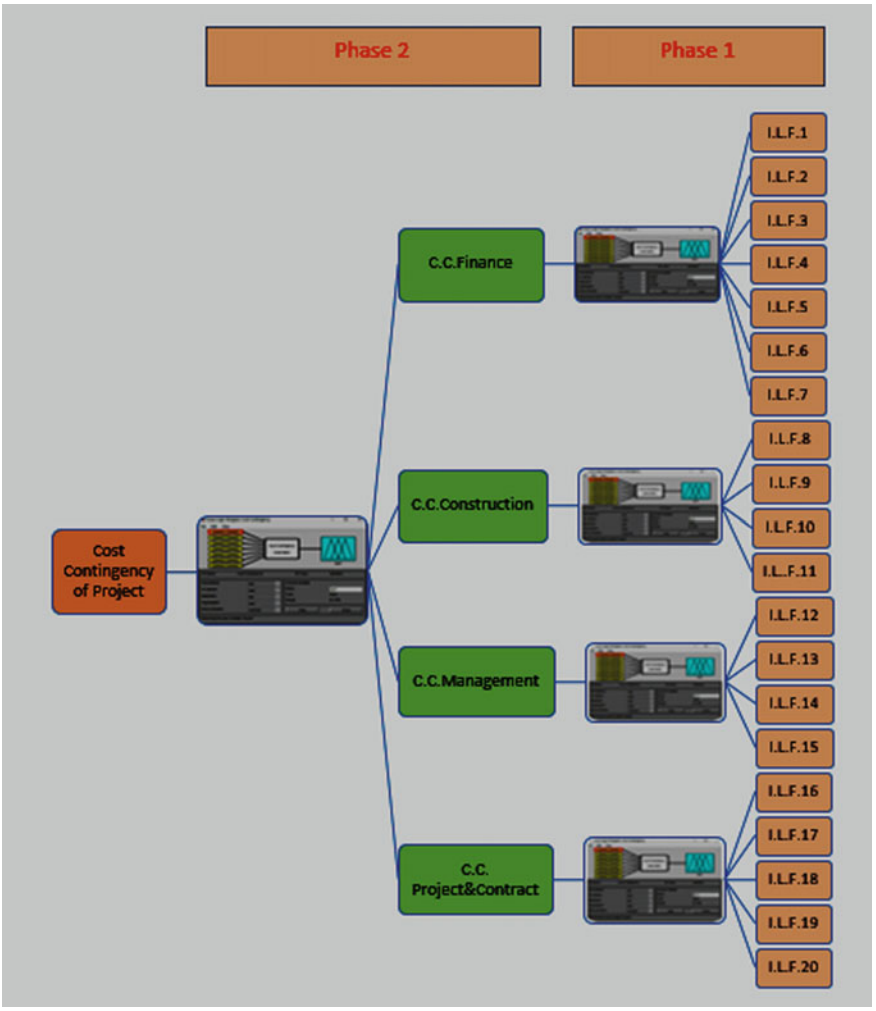


Fig.1 Two phase fuzzy inference process for model development

phase of the model are taken as 'input variables' and output variable 'cost contingency of the project' is then estimated.

3.3.1 Steps of Fuzzy Inference Process for Designing the First Phase of the Model

The process is performed using Fuzzy Logic Toolbox™ software of MATLAB Program. It consists of five primary graphical user interfaces (GUI) tools such as FIS Editor, MF Editor, Rule Editor, Rule Viewer, and Surface Viewer. The various steps of fuzzy inference system to design the model are as follows.

Defining Input and Output

The first step of model designing using the fuzzy inference process is to define the input and output variables. The impact level of the factors is taken as the input variables and cost contingency is taken as output. The various factors of group 'Finance' are Fluctuation in the prices of materials, Inflation, Financial difficulty experienced by contractor, Exchange rate, Material Cost, High level of interest rate of bank loan and High project charge of labour. These factors' impact level is considered input variables for determining the output cost contingency for the finance group ('C.C. Finance). During model design, the impact level of factors is abbreviated as I.L.F.1, I.L.F.2, I.L.F.3, I.L.F.4, and I.L.F.5, I.L.F.6 and, I.L.F.7 and the output variable is abbreviated as 'C.C.F'.

In Fig. 2 the input variables I.L. F1- to I.L.F.7, and output variables Cost Contingency Finance (C.C.F) are shown in the FIS Editor window of the Fuzzy Logic toolbox.

Similarly, input variables for group 'Construction' are I.L.F8, I.L.F9, I.L.F10, and I.L.F.11. I.L.F. 12., I.L.F.13 and I.L.F.14. I.L.F.15 are input variables for the 'Management' group. I.L.F.16, I.L.F.17, I.L.F.18, and I.L.F. 19., I.L.F.20 are considered input variables for 'Project & Contract'. Output variables for the groups are 'C.C. Construction', 'C.C. Management', and 'C.C. Project and Contract'.

Fuzzy Membership Functions

The membership function associated with all the input variables and output variables of all the factors of group 'Finance' are shown in Fig. 3.

Fuzzy Rules

The formation of fuzzy rules is an essential component of the fuzzy inference system. The rules for this study are formed with the help of experienced practitioners in Indian

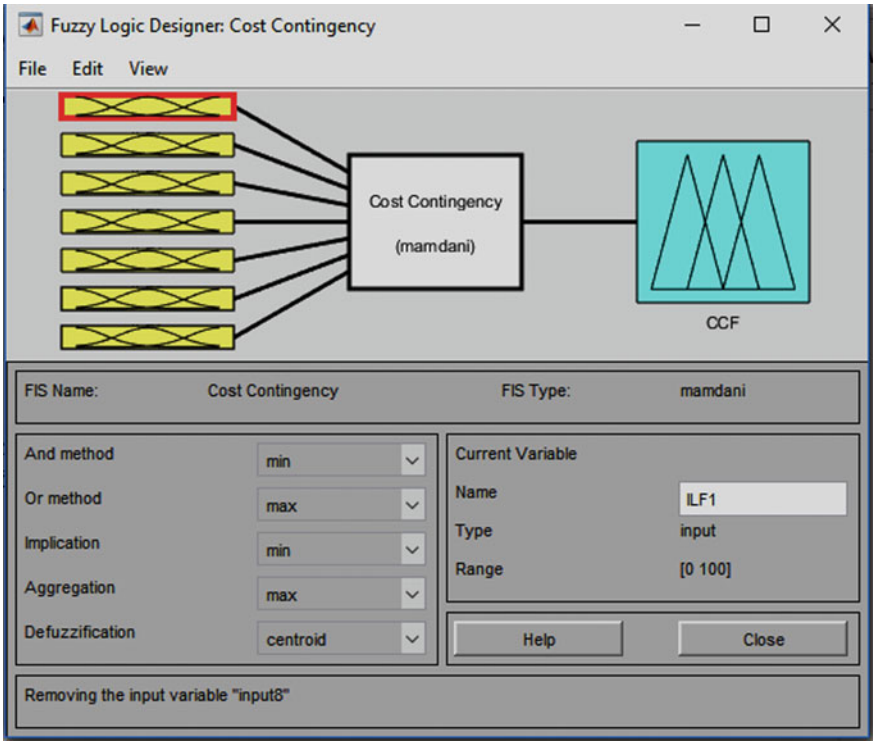


Fig. 2 Input and output variables of the group 'Finance' in FIS editor window

construction projects. Considering 20 risk factors, a total of 100 rules are formed. Samples of the fuzzy rules of group 'Finance' are presented in Table 3.

Since the cost contingency of group 'Finance' is directly related to the magnitude of impact level of factors fluctuation in the prices of materials, inflation, financial difficulty experienced by contractor, exchange rate, material cost, high level of interest rate of bank loan and high project charge of labour, therefore the cost contingency will get affected by the impact level of the factors of the group 'Finance'. The magnitude of the risk calculated in Sect. 3.2 will be taken as weighting for the rules. The sample rules constructed for the fuzzy model are shown in Table 4.

Similarly, the fuzzy rules were constructed for group 'Construction', 'Management'and 'Project and Contract'.

Defuzzification

Finally, the 'cost contingency C.C.F' for group 'Finance' is estimated by defuzzifying the rules by 'centroid of area' method as shown in Fig. 4

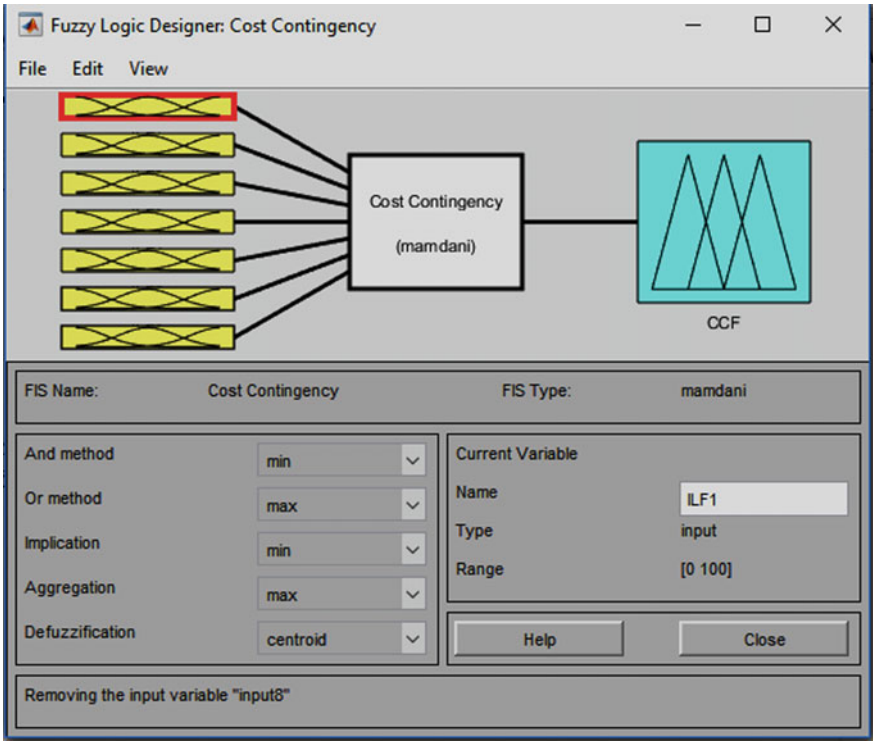


Fig. 3 Membership function for the input and output variables of the group 'Finance' in FIS editor window

Table 3 Sample of fuzzy rules group 'Finance'

Rule	Antecedent	Consequence	Weighting
1	If the impact level of risk factor 'Fluctuation in the prices of materials' is very less	Then the cost contingency of the group 'Finance' will be very less	0.694
31	If the impact of risk factor 'High project charge of labour.' is very less	Then the cost contingency of the group 'Finance' will be very less	0.664

Table 4 Sample of Fuzzy Rules of Group 'Finance' for Fuzzy Inference

Rule	Antecedent	Consequence	Weighting
1	If I.L. F. 1 is Very less	Then C.C.F is Very less	0.694
2	If I.L. F. 2 is Very less	Then C.C.F is Very less	0.687

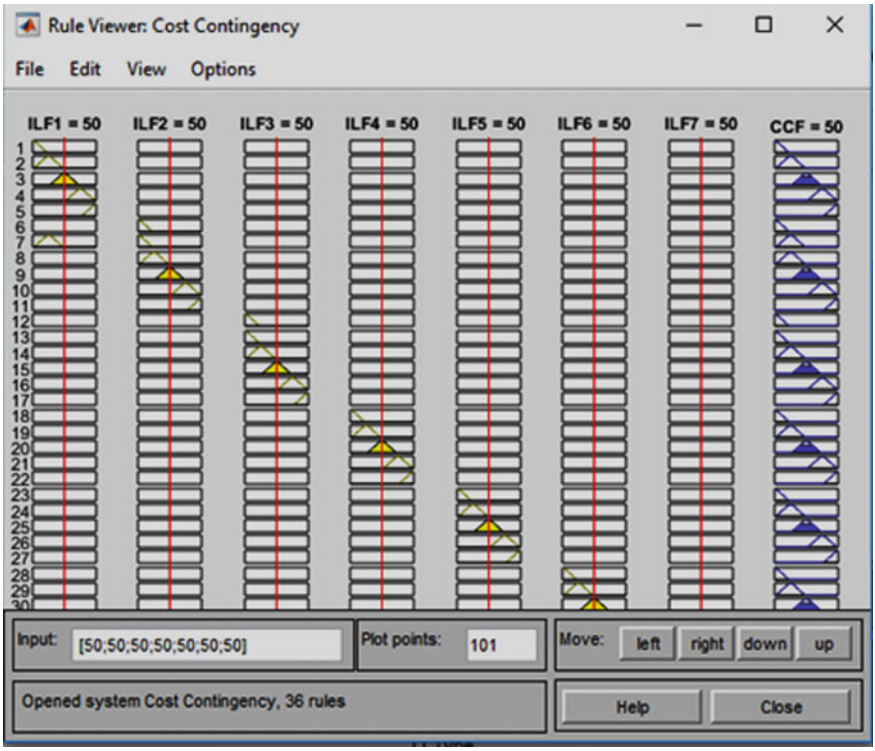


Fig. 4 Defuzzification process for cost contingency for group 'Finance'

3.3.2 Fuzzy Inference System for the Second Phase of Model Designing

For the second phase of the model designing the same steps for fuzzy inference are performed. The input members for this phase are cost contingency of finance group ('C.C. Finance'), cost contingency of construction group ('C.C. Construction'), cost contingency of management group ('C.C. Management') and cost contingency of project and contract group ('C.C. Project and Contract'). The output of the model is overall contingency of the project. It is abbreviated as O.C.P. The fuzzy inference process has been performed using Fuzzy Logic toolbox of MATLAB.

4 Testing of the Model

To test the reliability of the designed fuzzy model an interview has been conducted with a team of experts of a leading Indian construction company. The panel of the experts included the top 25 executive engineers, project managers, and site engineers of the company. The experts were requested to examine the risk factors considered for this research and filled in the required information regarding risk factors to test

the proposed model. The information provided by the group of experts is presented as shown in Table 5.

The determined impact level of the group by the model is given in Table 6. The cost contingency of the project evaluated by the proposed model was 14.4%, as shown in

Table 5 Impact level of the various factors for the selected project

S No.	Abbreviation	Factors impacting cost	Impact level in percentage
1	R.F. 1	Fluctuation in the prices of materials	20
2	R.F. 2	Inflation	10
3	R.F. 3	Financial difficulty experienced by contractor	10
4	R.F. 4	Exchange rate	5
5	R.F. 5	Material Cost	15
6	R.F. 6	High level of interest rate of bank loan	5
7	R.F. 7	High project charge of labour	5
8	R.F. 8	Improper planning and scheduling of project by contractor	10
9	R.F. 9	Incorrect estimation of time and cost	5
10	R.F. 10	Frequent modification in design	5
11	R.F. 11	Quantity of Extra work done	10
12	R.F. 12	Poor coordination between construction participants	5
13	R.F. 13	Conflicts and disputes on site	5
14	R.F. 14	Slow procedure in taking decisions	10
15	R.F. 15	Improper govt. policy	15
16	R.F. 16	Contract management	2
17	R.F. 17	Impracticable contract duration	5
18	R.F. 18	Errors and discrepancy in contract document	5
19	R.F. 19	Lowest possible bid procurement policy	10
20	R.F. 20	Modification in the scope of the project	5

Table 6 Overall Cost Contingency of Project

S No.	Group	Name of group	Impact of risk factor (%)	Overall cost contingency (%)
1	C.C.F	Finance	17.2	14.4
2	C.C.C	Construction	11.5	
3	C.C.M	Management	17.4	
4	C.C.P	Project and Contract	28.5	

Table 6, whereas the actual cost contingency was taken as 10% in this project, which shows an error of only 4.4%.

5 Conclusion

This paper has presented an alternate methodology for developing a risk-induced model to predict the cost contingency. The risk factors have been analysed using fuzzy theory as the theory is capable of dealing with incomplete, imprecise, and uncertain data Intrinsic to real-world complex problems such as those found in construction projects. The application of the proposed methodology has been illustrated for developing a model, by considering a case study of the Indian construction industry. Through an extensive literature review, 20 frequently occurring and important risk factors impacting the cost of construction projects were identified and classified into four groups as 'Finance', 'Construction', 'Management', and 'Project & Contract'. After taking into account the likelihood and severity index, the risk magnitude of these factors was then calculated by interviewing 20 experts involved in Indian construction projects and then using the Mamdani type rule-based fuzzy inference system a model was developed. Trapezoidal type membership function was defined for the input and output variables. The Fuzzy logic toolbox of MATLAB software was used for the process of the fuzzy inference system. This model was validated also for a construction project. The value obtained from the model was 14.4% and the actual cost contingency in the project was taken as 10%, which is very close to the results obtained from the model. Therefore, the proposed model for predicting the cost contingency can be used by customising according to the specific project. The methodology provides a practical approach for estimating cost contingency by taking into consideration the frequently occurring and important risk factors impacting the cost of construction projects. Project managers and decision-makers will find this model very useful for making decisions regarding various issues related to the project such as contingency estimation, bid price calculation, mark-up estimation, and assessment of different projects.

Conflict of Interest Statement On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Baccarini D (2006) The maturing concept of estimating project cost contingency—a review. In: 31st Australasian University Building Educators Association Conference (AUBEA 2006), Curtis University of Technology, Australia
2. Baloi D, Price ADF (2003) Modeling global risk factors affecting construction cost performance. Int J Proj Manag 21(4):261–269

3. Barraza G, Bueno R (2007) Cost contingency management. J Constr Eng Manag 23(3):140–146
4. Barraza G (2011) Probabilistic estimation and allocation of project time contingency. J Constr Eng Manag 137(4):259–265
5. Boukendour S (2005) A new approach to project cost overrun and contingency management. In: OCRI partnership conferences series process and project management, 22 March 2005, Ottawa
6. Chen D, Hartman FT (2000) A neural network approach to risk assessment and contingency allocation. AACE international transactions. Risk.07.01- Risk.07.06 Contingencies. J Constr Eng Manag 126(2):130–136
7. Diekmann JE (1983) Probabilistic estimating: mathematics and applications. J Constr Eng Manag 109(3):297–308
8. Dikmen I, Birgonul MT, Han S (2007) Using fuzzy risk assessment to rate cost overrun risk in international construction project. Int J Project Manage 25:495–505
9. Ahiaga-Dagbui DD, Tokede O, Smith SD, Wamuziri S (2013) A neuro-fuzzy hybrid model for predicting final cost of water infrastructure projects. In: Conference: 29th ARCOM conference at: reading, UK, September 2013. https://doi.org/10.13140/2.1.2382.6880
10. Elizabeth S, Sujatha L (2015) Project scheduling method using triangular intuitionistic fuzzy numbers and triangular fuzzy numbers. Appl Math Sci 9(4):185–198
11. Esragh F, Mamdani EH (1981) A general approach to linguistic approximation, fuzzy reasoning, and its application. Academic Press, Cambridge, MA, USA
12. Ford DN (2002) Achieving multiple project objectives through contingency management. J Constr Eng Manage 128(1):30–39
13. Hassanein AAB, Cherlopalle V (1999) Fuzzy sets theory and range estimating. AACE International Transactions, K4111. Hassanein, 1999.pdf, AACE Int Trans, K4 (1)
14. Idrus A, Nuruddin MF, Rohman MA (2011) Development of project cost contingency estimation model using risk analysis and fuzzy expert system. Expert Syst Appl 38:1501–1508
15. Jin XH, Zhang G (2011) Modelling optimal risk allocation in PPP projects using artificial neural networks. Int J Proj Manag 29(5):591–603
16. Kutsch E, Hall M (2010) Deliberate ignorance in project risk management. Int J Proj Manag 28(3):245–255
17. Lorance RB, Wendling RV (2001) Basic techniques for analysing and presenting cost risk analysis. Cost Eng 43(6):25–31
18. Mak S, Picken D (2000) Using risk analysis to determine construction project. J Constr Eng Manag. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:2(130)
19. Mamdani EH (1997) Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans Comput 26(12):1182–1191
20. Mamdani H, Assillian S (1975) An experimental in linguistic synthesis with a fuzzy logic controller. Int J Man Mach Stud 7(1):1–13
21. Marco AD, Rafele C, Thaheem MJ (2015) Dynamic management of risk contingency in complex design-build projects. J Manage Eng. ISSN 0733–9364/04015080(10)
22. McGrew JF, Bilotta JG (2000) The effectiveness of risk management: measuring what didn't happen. Manag Decis 38(4):293–301
23. Moselhi O (1997) Risk assessment and contingency estimation. AACE Int Trans 90–95
24. Paek J, Lee YW, Ock JH (1993) Pricing construction risk: fuzzy set application. J Constr Eng Manag 119(4):743–756
25. Project Management Institute (2013) A guide to the project management body of knowledge (PMBOK® Guide)
26. Shaheen A, Robinson FA, AbouRizk SM (2007) Fuzzy numbers in cost range estimating. J Constr Eng Manag 133(4):325–334
27. Sharma S, Goyal PK (2014) Cost overrun factors and project cost risk assessment in construction industry—a state of the art review. Int J Civil Eng 3(4):139–154
28. Sharma S, Goyal PK (2019) Fuzzy assessment of the risk factors causing cost overrun in construction industry. Evol Intel. https://doi.org/10.1007/s12065-019-00214-9

29. Sonmez R, Ergin A, Birgonul MT (2007) Quantitative methodology for determination of cost contingency in international projects. J Manag Eng 23(1):35–39
30. Thal AE, Cook JJ, White ED (2010) Estimation of cost contingency for air force construction projects. J Constr Eng Manage 136(11):1181–1188
31. Touran A (2003) Probabilistic model for cost contingency. J Constr Eng Manag 129(3):280–284
32. Wang WC, Bilozerov T, Dzeng RJ, Hsiao FY, Wang KC (2017) Conceptual cost estimations using neuro-fuzzy and multi-factor evaluation methods for building projects, pp 1–14. https://doi.org/10.3846/13923730.2014.948908
33. Williams TP (2003) Predicting final cost for competitively bid construction projects using regression models. Int J Project Manage 21:593–599
34. Yeo KT (1990) Risks, classification of estimates and contingency management. J Manag Eng 6(4):458–470
35. Zadah LA (1965) Fuzzy sets. Inf Control 8(3):338–353

Author Index

Y.-C. Hu et al. (eds.), *Ambient Communications and Computer Systems*, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0