Pradeep Kumar Jha · Prashant Jamwal ·
Brajesh Tripathi · Deepak Garg ·
Harish Sharma *Editors*

Proceedings of the Second Congress on Control, Robotics, and Mechatronics

CRM 2024, Volume 2

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 409

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

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

Pradeep Kumar Jha · Prashant Jamwal ·
Brajesh Tripathi · Deepak Garg · Harish Sharma
Editors

Proceedings of the Second Congress on Control, Robotics, and Mechatronics

CRM 2024, Volume 2

Editors
Pradeep Kumar Jha
Department of Mechanical Engineering
Indian Institute of Technology
Roorkee, Uttarakhand, India

Prashant Jamwal
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sułtan, Kazakhstan

Brajesh Tripathi
Department of Mechanical Engineering
Rajasthan Technical University
Kota, Rajasthan, India

Deepak Garg
SR University
Warangal, Telangana, India

Harish Sharma
Department of Computer Engineering
Rajasthan Technical University
Kota, Rajasthan, India

Organization

General Chair

Dr. Brajesh Tripathi, Department of Mechanical Engineering, Rajasthan Technical University, Kota, India
Prof. Deepak Garg, SR University, Warangal, India
Dr. Pradeep Kumar Jha, Department of Mechanical Engineering, Indian Institute of Technology, Roorkee, India
Dr. Prashant Jamwal, Nazarbayev University, Kazakhstan
Dr. Harish Sharma, Rajasthan Technical University, Kota, India

Convener

Dr. Bohra Nitin Kumar, SR University, Warangal, India
Dr. Pankaj Kumar, SR University, Warangal, India

Program Chair

Dr. Sandeep Kumar, CHRIST (Deemed to be University), Kengeri Campus, Bangalore
Dr. Himanshu Mittal, Indira Gandhi Delhi Technical University for Women, New Delhi
Dr. Ashish Tripathi, MNIT, Jaipur, India

Publicity Chair

Dr. Mukesh Saraswat, Jaypee Institute of Information Technology, Noida
Dr. Kusum Kumari Bharti, Dr B R Ambedkar National Institute of Technology Jalandhar, India

Technical Program Committee

Kedar Nath Das, National Institute of Technology Silchar (Nits) Silchar India
Manoj Thakur, IIT Mandi India
Maurice Clerc, France
Meng-Hiot Lim, Nanyang Technological University Singapore
Neil Buckley, Liverpool Hope University Uk
Nishchal K. Verma, Indian Institute of Technology Kanpur India
A. K. Verma, Western Norway University of Applied Sciences Haugesund Norway
Ali Mirjalili, Torrens University Australia
Namita Mittal, Malaviya National Institute of Technology Jaipur
Amir H. Gandomi, Niversity of Technology Sydney, Australia
Anand Nayyar, Duy Tan University Da Nang Vietnam
Anupam Yadav, National Institute of Technology Jalandhar India
Alok Kanti Deb, IIT Kharagpur India
Carlos A. Coello Coello, Investigador Cinvestav
D. L. Suthar, Wollo University Ethiopia
Vikas Bajpai, The LNM Institute of Information Technology, Jaipur
Dr. Santosh Kumar Viparthi, Indian Institute of Technology, Ropar
Joong Hoon Kim, Korea University South Korea
K. S. Nisar, Prince Sattam Bin Abdulaziz University Riyadh Saudi Arabia
Faruk Ucar, Marmara University
Dumitru Baleanu, Cankaya University Turkey
Dr. Suneel Dutt Purohit, Rajasthan Technical University, Kota
Prashant Jamwal, Nazarbayev University Kazakhstan
Satyasai Jagannath Nanda, Malaviya National Institute of Technology Jaipur
Dr. Mukesh Saraswat, JIIT, Noida
Dr. Kusum Kumari Bharti, Nit Jhalandhar, India
Dr. Himanshu Mittal, Ndira Gandhi Delhi Technical University for Women, New Delhi
Dr. Ashish Tripathi, MNIT, Jaipur, India
Dr. Vikas Bansal, Engineering College Ajmer
Dr. Vikas Vajpayee, LNMIIT, Jaipur, India
Dr. Manish Chaturvedi, Department of Mechanical Engineering, Rajasthan Technical University, Kota
Raul Campilho, Isep—School of Engineering, Porto, Portugal

Luiz Guerreiro Lopes, University of Madeira

Suleyman Nokerov, Oguz Han Engineering and Technology University of Turk-menistan, Ashgabat Turkmenistan

Naiwrita Dey, RCC Institute of Information Technology, India

Dr. Jay Katupitiya, Niversity of South Wales Sydney, Australia

Dr. Anand Veeraragavan, The University of Queensland, Australia

Dr. Ann Gaitonde, University of Bristol, Bristol, UK

Dr. Brian Monaghan, University of Wollongong, Australia

Dr. Da (David) Abbink, Delft University of Technology, Netherlands

Dr. David Wynn, University of Aukland, NZ

Dr. Michael Bermingham, The University of Queensland, Australia

Dr. Mohammad Deghat, University of South Wales Sydney, Australia

Dr. Nicholas Gilmore, University of South Wales Sydney, Australia

Dr. Tore Butlin, University of Cambridge, UK

Dr. Zhongwei Chen, The University of Queensland, Australia

Dr. Christopher J. Hernandez, Cornell Engineering Sibley School of Mechanical and Aerospace Engineering, USA

Dr. James Lloyd, Cornell Engineering Sibley School of Mechanical and Aerospace Engineering, USA

Dr. Kamran Behdinan, University of Toronto, Canada

Dr. Kevin Kochersberger, Virginia Tech, USA

Dr. Kumar A. Senthil, National University of Singapore, Singapore

Dr. Lance Collins, Cornell Engineering Sibley School of Mechanical and Aerospace Engineering, USA

Dr. Lino Guzzella, ETH Swiss Federal Institute of Technology, Switzerland

Dr. Mario Trujillo, University of Wisconsin- Madison, USA

Dr. Matthew Peter Miller, Cornell Engineering Sibley School of Mechanical and Aerospace Engineering, USA

Dr. Nasser Ashgriz, Niversity of Toronto, Canada

Dr. Palani Balaya, National University of Singapore, Singapore

Dr. Paul Bishop, Dalhousie University, Canada

Dr. Raghuram V. Pucha, Georgia Tech, USA

Dr. Reza Mirzaeifar, Virginia Tech, USA

Dr. Robert Katzschmann, ETH Swiss Federal Institute of Technology, Switzerland

Dr. Satish Kumar, Georgia Tech, USA

Dr. Sean Peterson, University of Waterloo, Canada

Dr. Shorya Awtar, University of Michigan, USA

Dr. Sridhar Kota, University of Michigan, USA

Dr. Surya Kalidindi, Georgia Tech, USA

Dr. Yogesh Gianchandani, University of Michigan, USA

Dr. Inoue Ryo, Tokyo University of Science, Japan

Dr. Jan Dragseth, Norwegian University of Science and Technology, Norway

Dr. Jonathan Jeffers, Imperial College London, London UK

Dr. Kazuhiro Nogita, The University of Queensland, Australia

Dr. Kean Aw, University of Aukland, NZ

Dr. Konstantin Konstantinov, University of Wollongong, Australia
Dr. Lihua Tang, University of Aukland, NZ
Dr. M. (Martijn) Wisse, Delft University of Technology, Netherlands
Dr. Martin Veidt, The University of Queensland, Australia
Dr. Paul Cooper, University of Wollongong, Australia
Dr. Pavlos Aleiferis, Imperial College London, London UK
Dr. Peter Madindwa Mashinini, University of Johannesburg, SA
Dr. R. (Robert) Babuska, Delft University of Technology, Netherlands
Dr. Richard Morgan, The University of Queensland, Australia
Dr. Sanjaya Vipula Bandara Danwatta, KYUSHU University, JP
Dr. Sasaki Shinya, Tokyo University of Science, JP
Dr. Shoji MORI, Kyushu University, JP
Dr. Susumu Kudo, Kyushu University, JP
Dr. Tom Allen, University of Aukland, NZ
Dr. Vikram Deshpande, University of Cambridge, UK
Dr. Yue Zhao, University of Wollongong, Australia
Dr. Ravindra S. Kulkarni, R V College of Engineering Bengaluru, India
Dr. Tejinder Kumar Jindal, PEC Chandigarh, India
Dr. Anand Kumar, BIT Mesra, India
Dr. Girish Kant Garg, BITS Pilani, India
Dr. H. C. Thakur, Gautam Buddha University Greater Noida, India
Dr. Kartik Venkatraman, IISC Bangalore, India
Dr. Rahul Dev Gupta, MMU Haryana, India
Dr. Sudhir Joshi, UPES Dehradun, India
Dr. Atul Kumar Agrawal, DTU Delhi, India
Dr. G. S. Dangayach, MNIT Jaipur, India
Dr. Mukul Shukla, MNNIT Allahabad, India
Dr. S. K. Singhal, HBTU Kanpur, India
Dr. Surendra Singh Kachhwaha, PDEU Gandhinagar, India
Dr. A. K. Sharma, DCRUDT Haryana, India
Dr. A. N. Sinha, NIT Patna, India
Dr. A. Salih, IISST Thiruvananthapuram, India
Dr. Ashish Tewari, IIT Kanpur, India
Dr. B. N. Singh, IIT Kharagpur, India
Dr. D. P. Mishra, IIT Kanpur, India
Dr. Manish Mishra, IIT Roorkee, India
Dr. Neeraj Aggarwal, DR. B.A.TU Maharashtra, India
Dr. Om Prakash, UPES Dehradun, India
Dr. P. Chandra, NIT Patna, India
Dr. Pradeep K. Sahoo, IIT Roorkee, India
Dr. R. K. Pant, IIT Bombay, India
Dr. Raja Benerjee, IIT Hyderabad, India
Dr. Rakesh kumar, PEC Chandigarh, India
Dr. Siddhartha, IIT Hamirpur, India
Dr. Vikas Kumar, JCBUST Haryana, India

Dr. Bikramjeet Basu, IISC Bangalore, India
Dr. Inderdeep Singh, IIT Roorkee, India
Dr. Kamal Kar, IIT Kanpur, India
Dr. N. Ramesh Babu, IIT Madras, India
Dr. S. G. Deshmukh, IIT Delhi, India

Preface

This book contains the proceedings of the Congress on Control, Robotics, and Mechatronics (CRM2024). SR University, Warangal, India, has organized CRM 2024 during February 03–04, 2024. CRM 2024 was technically sponsored by the Soft Computing Research Society, India. The conference was conceived as a platform for disseminating and exchanging ideas, concepts, and results of the researchers from academia and industry to develop a comprehensive understanding of the challenges of the advancements in Control, Robotics, and Mechatronics. This book will help strengthen the amiable networking between academia and industry. The conference focused on Control, Robotics, and Mechatronics and their applications.

We have tried our best to enrich the quality of the CRM 2024 through a stringent and careful peer-review process. CRM 2024 received many technical contributed articles from distinguished participants from home and abroad. After a very stringent peer-reviewing process, only 64 high-quality papers were accepted for presentation and the final proceedings.

This book presents the second volume of 31 Control, Robotics, and Mechatronics research papers and serves as reference material for advanced research.

Roorkee, India   Pradeep Kumar Jha
Nur-Sułtan, Kazakhstan   Prashant Jamwal
Kota, India   Brajesh Tripathi
Warangal, India   Deepak Garg
Kota, India   Harish Sharma

Contents

5 Smartphone-Controlled Robot Manipulator for Pick and Place Applications 69

Venkata Sushma Chinta, Sowmya Kethi Reddi,
Nagini Yarramsetty, Shaik Chand Pasha, and Palla Nishant Bhanu

6 Impact of Russia-Ukraine War on Oil and Gas Sector of Indian Economy 79

Kuldip Katiyar, Tejal Gupta, and Kundan Kumar Jha

7 Optimal Stacking Method with GA-Based Feature Selection for Multi-Label Classification 91

Hemavati and R Aparna

About the Editors

Pradeep Kumar Jha, Ph.D. Professor, Department of Mechanical & Industrial Engineering, IIT Roorkee. Dr. Pradeep Kumar Jha is presently working as a Professor in the Department of Mechanical and Industrial Engineering. After receiving bachelor's and master's degrees from M.I.T. Muzaffarpur, Bihar, and National Institute of Foundry and Forge Technology, Ranchi, Jharkhand, in the year 1995 and 1999 respectively, Professor Jha obtained his Ph.D. degree from IIT Kharagpur in the year 2004. Prior to joining as an Assistant Professor in IIT Roorkee in the year 2007, he also served as Sr. Lecturer and Assistant Professor at IIT Guwahati and IIT (ISM) Dhanbad respectively. Prof. Jha specializes in the area of Manufacturing Engineering, with special emphasis on casting operations, modeling, and simulation of continuous casting operations and metal matrix composites. He has to his credit more than 100 publications in international and national journals and conferences and two book chapters.

Prof. Prashant Jamwal earned Ph.D. degree and a post-doctoral fellowship from the University of Auckland, New Zealand. Earlier he had obtained M.Tech. from I.I.T., India, securing first position in all the disciplines and B.Tech. from MNREC, Allahabad, India. Presently, he is working as a Professor at the School of Engineering and Design Sciences, Nazarbayev University (NU), Astana, Kazakhstan, and as an Adjunct Professor at University of Canberra, Australia. He is actively pursuing research in robotics and artificial intelligence, multi-objective evolutionary optimization, biomedical engineering, and renewable energy. Over the past decade, he has applied his research in the development of medical robots for rehabilitation and surgical applications besides developing improved algorithms for cancer data analytics. He is working as an Editor for the International journal of bio-mechatronics and bio-robotics and as a Reviewer to quite a few international journals and conferences of repute. Prof. Jamwal led many internationally funded research projects and has so far received research grants worth more than $9 million including prestigious World Bank grants.

Dr. Brajesh Tripathi received his Ph.D. in Mechanical Engineering in 2008 from Indian Institute of Technology Kharagpur, India. He was a Post-doctorate Fellow at the Michigan State University, USA, for two and half years and worked on a US government sponsored project on Computational Fluid Dynamics. Presently, he holds Associate Professor position at Rajasthan Technical University Kota. Dr. Triapthi received 06 research grants for various sponsored projects from GNIDA, NAL, BRNS, MHRD, and AICTE. He has guided 08 Ph.Ds. and 26 M.Tech. scholars. He has published more than 130 research articles in peer-reviewed international journals and conferences. Dr. Tripathi's primary interest is working on practical problems using fundamental concepts of Computational Fluid Dynamics especially in diesel engine, heating, ventilation and air-conditioning, turbulent flows, unconventional energy systems, etc.

Prof. Deepak Garg is Ph.D. in Computer Science with specialization in Efficient Algorithm Design. With 25 years of experience, he has worked as Professor and Dean, School of Computer Science at Thapar Institute of Engineering and Technology, Patiala, and Bennett University, Greater Noida. He is a Distinguished Professor in Artificial Intelligence and is known as one of the top voices in the country on AI. He is a regular Blogger in Times of India with a nickname of "Breaking Shackles" and writes about Higher Education landscape and AI. He is on the Advisory Committee of AIRAWAT, Super Computing Mission of Govt. of India to create a 1 Exaflop backbone infrastructure for the country. He has represented various Committees at NAAC, NBA, UGC, and AICTE. Working with Startup Founders is his passion, he serves on tech advisory board of Drishya.ai, ByteXL, and Global AI Hub. He is the only CAC ABET Commissioner from the country and has served as PEV for last six years. He has been in the Board of Governors of IEEE Education Society, USA. He served as Chair of IEEE Computer Society, India Council, and IEEE Education Society, India Council. He is on Academic Councils, and Board of Studies of various institutions. He has 160 publications with 1600 Citations, h-index 20, and i10 index 41.

Harish Sharma is Associate professor at Rajasthan Technical University, Kota, in Department of Computer Science & Engineering. He has worked at Vardhaman Mahaveer Open University Kota and Government Engineering College Jhalawar. He received his B.Tech. and M.Tech. degrees in Computer Engineering. from Government Engineering College, Kota, and Rajasthan Technical University, Kota, in 2003 and 2009 respectively. He obtained his Ph.D. from ABV-Indian Institute of Information Technology and Management, Gwalior, India. He is the Secretary and one of the founder members of Soft Computing Research Society of India. He is a Lifetime Member of Cryptology Research Society of India, ISI, Kolkata. He is the Associate Editor of "International Journal of Swarm Intelligence (IJSI)" published by Inderscience. He has also edited special issues of the many reputed journals like

"Memetic Computing", "Journal of Experimental and Theoretical Artificial Intelligence", "Evolutionary Intelligence", etc. His primary area of interest is nature inspired optimization techniques. He has contributed to more than 125 papers published in various international journals and conferences.

Chapter 1
DDCF: Enhancing Educational Resource Recommendation in E-Learning Platforms Using Collaborating Filter Approach

Dudla Anil kumar and **M. Ezhilarasan**

Abstract An e-learning-based recommendation system suggests learning resources, courses, or materials to users based on their preferences, behavior, and historical interactions within an e-learning platform. This system aims to enhance the learning experience by providing personalized and relevant content to individual users. However, lacking feedback can make improving and fine-tuning the challenging recommendation algorithms difficult. Additionally, inaccurate or insufficient metadata can lead to less accurate recommendations. To improve recommendation systems in e-learning, we propose a novel deep-learning technique. Data is collected from the Udemy dataset and then processed—cleaning the data and normalizing it to perform well in the recommendation system. Relevant features are extracted to classify and recommend courses to the user. The Hierarchical Multi-scale Long Short-Term Memory (HMLSTM) technique is used to classify the course level into beginning, intermediate, and advanced categories. The user's ratings are then identified to recommend courses using the Improved Golden Jackal Optimization (IGJO) Algorithm. Finally, courses are recommended to the user based on user ratings using a collaborative filtering technique. The performance of the proposed method is evaluated by precision, accuracy, recall, and F1-score metrics and compared to state-of-the-art techniques.

D. A. kumar (✉) · M. Ezhilarasan
Department of Information Technology, Puducherry Technological, University, Puducherry, India
e-mail: anilkumar.d@pec.edu

M. Ezhilarasan
e-mail: mrezhil@ptuniv.edu.in

1.1 Introduction

Using keyword search strategies to find the right learning content for students has become more challenging due to the abundance of learning resources available on e-learning platforms. Finding the right learning resources has, therefore, become very difficult. Using recommender systems is one method to overcome this obstacle [1–3]. Furthermore, some scholars have discussed the necessity of personalization in online learning as recently as possible. Giving recommendations to students to encourage and assist them during the e-learning process is one of the newer kinds of customization in the e-learning environment [4–6].

Recommender systems use three recommendation methodologies, which are typically used in e-commerce businesses to represent customers' interests: collaborative filtering (CF), content-based (CB), and hybrid recommendation [7]. One of their main faults is that most present recommendation systems consider rating matrices as meaningful data and do not adequately consider contextual data, such as material attributes, to improve recommendations.

The WWW's ubiquity during the last ten years has eclipsed any previous technological advancements throughout history. Thus, we increasingly rely on a wide range of web apps [8–11]. These programs function well and are reliable. Web engineering is a developing subject that aims to satisfy these objectives. Website personalization is closely linked to recommender systems. For the bulk of newly growing web domains, including e-tourism, e-commerce, and e-education, a well-designed recommendations system helps solve web engineering problems [12].

In actuality, recommendation systems (RSS) are a type of data filtering that seeks to provide the contents of the data most likely to interest the user. Three distinct kinds of techniques—CF, CB, and Hybrid Filtering—have recently been suggested in research to remove irrelevant resources. CF is the most widely used RS out of all available ones. By comparing a frequent user to various other users, the CF algorithm seeks to provide recommendations [13–15]. There are two steps to it: (1) Establish a neighborhood with the most comparable users as its members; (2) Make a prediction based on the data these neighbors have provided. Similarity metrics like the Euclidean distance and the cosine distance are the foundation of CF. The selection of similarity metric significantly impacts how well a recommendation performs for a user. To create a community, one must search for members who share the same interests as the current user. One significant drawback of CF is that it ignores the relationships between members within the same group. To overcome the issues in the recommendation system in e-learning, propose a new deep learning technique to the recommendation system. The contribution of this research is,

- Collect the data from the Udemy dataset for processing data cleaning and normalize data to perform well in the recommendation system.
- Extract the relevant features to classify and recommend courses to the user. Then, the course level will be classified using the Hierarchical Multi-scale Long Short-Term Memory (HMLSTM) technique into beginning, intermediate, and all levels.

- After that, the user rating will be identified to recommend courses for users using the Improved Golden Jackal Optimization (IGJO) Algorithm.
- Finally, the courses based on user ratings should be recommended to the user using the collaborative filter-based Dynamic Decay Collaborative Filtering technique.

These sections make up the remainder of the article. The literature that is relevant to the paper is listed in Sect. 1.2. In Sect. 1.3, the proposed approach is described. The findings are shown in Sect. 1.4. Section 1.5 concludes with the conclusions.

1.2 Literature Survey

In this section, we mentioned previous studies on e-learning-based recommendation systems. Vedavathi et al. [16] introduced a hybrid optimization algorithm (HOA) based on an effective e-learning recommendation (EELR) system for user preferences. An HOA is built by the EELR system using IWO algorithm and a DRNN. The groups are then used to rate the various types of e-learners using DRNN, and the recommendation group provides course proposals to the clients. After that, the learners' preferences and behavior are assessed by mining the configurations the IWO computation periodically watches. Recommender systems advise students on items they would like to evaluate instead of a learner searching for information efficiently.

Tarus et al. [17] presented a hybrid recommendation strategy that combines CF algorithms, sequential pattern mining (SPM), and context awareness to suggest educational materials to students. The context awareness component of the recommendation approach incorporates contextual data about learner, such as knowledge level and learning objectives; the sequential access patterns of the learner are found through web log mining using SPM algorithm; and the forecasting and recommendation generation for the target learner is carried out by CF using contextualized data and sequential access patterns of the learner. According to the assessment, the developed hybrid recommendation strategy can provide recommendations with higher accuracy and quality than existing recommendation approaches.

According to the assessment, Ibrahim et al. [18] suggested a hybrid recommendation system can provide recommendations with higher accuracy and quality than existing recommendation approaches. By establishing three FBRS modules—the Class Identification Module (CIM), the Subclass Identification Module (SIM), and the Matchmaking Module (MM)—we address a structure to enhance and unify the EL environment. Furthermore, the FBRS approach defeats synonymy and customization while achieving high security and response time.

The temporal correlation coefficient-based TCCF recommendation system was developed by Cui et al. [19] as an improved version of the K-means with cuckoo search (CSK-means) algorithm. Similar people can be grouped using the clustering method for faster and more accurate recommendations. To improve TCCF even more, a strong and distinctive recommendation system based on preference pattern

(PTCCF) was developed. It might provide a higher-quality suggestion by analyzing user behavior. Experimental results validate the efficacy of the presented technique.

Alatrash et al. [20] presented a unique recommendation system based on sentiment evaluation and utilizing Natural Language Processing (NLP) and Convolutional Neural Network (SABCNN) techniques. Evaluating the meaning of sensations that a sentence or text has may be classed into multiple levels. The process begins with gathering and analyzing learner evaluations of the e-content and the ratings that go along with them on e-platforms. The suggested approach works better than the ones used currently.

Tarus et al. [21] proposed a hybrid recommendation technique that combines context awareness, CF algorithms, and sequential pattern mining (SPM) with the goal of recommending learning materials to every student. After using the SPM approach to mine websites to identify the student's frequent access habits, context awareness is used to incorporate contextual information about the student, including knowledge level and learning objectives; Using contextualized data and the learner's sequential access patterns, CF computes forecasts and generates suggestions for the learner.

Vedavathi and AnilKumar [22] presented a hybrid optimization algorithm (HOA)-based effective e-learning recommendations (EELR) approach to user preferences. The EELR system constructs an HOA using a DRNN and IWO computation. First, various e-learner types are grouped together into e-learner groups using DRNN, from which students can choose course recommendations. Then, by carefully mining the settings that are periodically tracked by the IWO computation, the conduct and preferences of the learners are evaluated. Rather than requiring students to actively search for information, recommender systems provide them with recommendations for publications to evaluate.

From the literature survey, there are some limitations to overcome. New users or courses may face a "cold start" problem, where the recommendation system struggles to provide accurate suggestions due to insufficient data about user preferences or course content. Some recommendation systems may not effectively consider contextual factors such as the user's current learning goals, preferences, or the specific context of a learning activity. Computational time complexity of the process is higher in each previous study, and the recommendation accuracy is lower. So, here, we propose a new deep learning-based technique to overcome these issues to handle and accurately predict the recommendation prediction system, processed by feature extraction and data cleaning performance. It also employs the collaborative filtering-based method to recommend courses based on the user's preferences.

1.3   Proposed Methodology

An e-learning-based recommendation system is a software application that utilizes algorithms and data analysis techniques to provide personalized recommendations to learners within an online education or e-learning environment. To improve the recommendation system in e-learning, propose a novel deep learning technique. Collect the data from the Udemy dataset to process data cleaning and normalize data to perform well in the recommendation system. Extract the relevant features to classify and recommend the courses to the user. Then, the course level will be classified using the Hierarchical Multi-scale Long Short-Term Memory (HMLSTM) technique into beginning, intermediate, and all levels. After that, the user rating will be identified to recommend the courses for the user using the Improved Golden Jackal Optimization (IGJO) Algorithm. Finally, the courses based on user ratings should be recommended to the user using the collaborative filter-based Dynamic Decay Collaborative Filtering technique. An e-learning-based recommendation system is a software application that utilizes algorithms and data analysis techniques to provide personalized recommendations to learners within an online education or e-learning environment. Figure 1.1 shows the architectural structure of the proposed methodology.

1.3.1   Pre-processing

Data cleaning is crucial in the data pre-processing phase of e-learning recommendation systems. It involves identifying and addressing issues in the dataset to ensure its quality and reliability. In the data pre-processing phase, clean the data by eliminating missing values, then normalize the data to perform other processes.
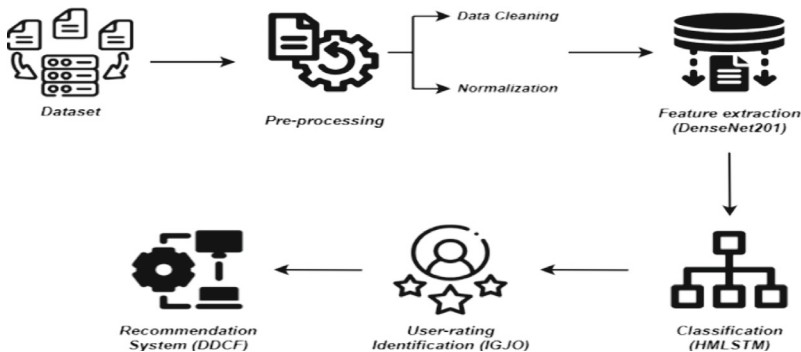
Fig. 1.1   An architecture of proposed methodology

(a) *Data Cleaning*

Identify and analyze missing values in the dataset, particularly in columns related to user profiles, course details, and user ratings. Decide on a strategy for handling missing values, such as imputation or removal of records with missing values.

(b) *Normalization*

In the normalization step, both min–max normalization and z-score-based normalization are implemented. When working with much of it, normalization effectively scales course data into a specified range. After normalization, gradient descent accelerates and gets more accurate. Scaling data between specific ranges is used every day for min–max normalization, which applies a linear trend to the starting date.

$$z_n = \frac{x - x_{\min}}{x\min_{\max}(\text{New} - \text{New}) + \text{New}} \tag{1.1}$$

where the variables $x_{\max}$ and $x_{\min}$ stand in for the highest and lowest values, respectively. The notation $\text{New}_{\min_x}$ depicted the smallest integer, whereas $\text{New}_{\max_x}$ depicted largest.

1.3.2 Feature Extraction

After pre-processing, the input data is given to the feature extraction process to extract the relevant features to complete the recommendation process accurately. The feature extraction process used by DenseNet201 technique. The DenseNet201 uses a condensed network to provide highly parametrically efficient and easily trainable models. This is because various layers may reuse features, which increases the variety in input of the succeeding layer and enhances performance. Direct connections between every layer before and after are added to the DenseNet201 architecture to increase connectivity.

A mathematical explanation of the attribute concatenation is as follows:

$$z^l = H_l\left(\left[z^0, z^1, \ldots\ldots, z^{l-1}\right]\right) \tag{1.2}$$

$H_l(.)$ is a non-linear transformation characterized as a composite function consisting of BN, ReLU, and (3×3) convolution. For ease of deployment, the map of features combination matching to layer 0 to $l - 1$ is referred to as $\left[z^0, z^1, \ldots\ldots, z^{l-1}\right]$ and is incorporated into a single tensor. Dense blocks are generated in the network structure for the objective of down-sampling, and these blocks are divided into layers known as transition layers. These layers comprise BN, a 2×2 average pooling layer, and a 1×1 convolution layer. The hyper-parameter k, representing the growth rate in DenseNet201, demonstrates how the dense design gives state-of-the-art outcomes: Because of its framework, which treats attribute maps as the network's global state, DenseNet201 performed admirably even at lower

growth rates. As an outcome, all map features from earlier layers are accessible to every subsequent layer.

$$(FM)^l = k^0 + k(l-1) \tag{1.3}$$

In this case, k^0 indicates the channels in the input layer: A 1×1 convolution layer is added before each 3×3 convolution layer to increase computational efficiency by decreasing the amount of input map features, which are usually more significant than the amount of outcome feature maps, k: The bottleneck layer, a 1×1 convolution layer, is introduced and generates 4 k attribute maps. Two dense layers containing 64 and 128 neurons are added for categorization.

$$t^{l-1} = \text{Bernoulli}(p) \tag{1.4}$$

$$\ddot{x}^{l-1} = t^{l-1} * c^{l-1} \tag{1.5}$$

$$\ddot{x}^l = f(w^k \ddot{x}^{l-1} + o^l) \tag{1.6}$$

The vector dimension is c^{l-1}: The dropout method, utilized in the first two layers of the whole connection layer, randomly blocks specific neurons based on a predetermined probability. The non-normalized outcomes are transformed into 0 or 1 binary results via sigmoid activation. The formula of the sigmoid is:

$$y = \frac{1}{1 + e^{-(\Sigma w_l x_l)}} \tag{1.7}$$

The inputs and weights are described, respectively, by w_i and x_i.

1.3.3 Course Level Classification

The course levels were then categorized into groups according to their respective categories using the HMLSTM approach. For this reason, a modified LSTM known as HMLSTM is used in this work. HMLSTM comes in a lot of forms. The HMLSTM has been modified in this study to meet the needs of the IDS model. The general LSTM uses three gates to obtain inputs, which are then processed using the sigmoid function. This HMLSTM introduces a parameterization boundary detector that learns the termination criteria to retrieve the temporal attributes in every layer by obtaining a binary output result. Dense connections are also included for a layer to create a combination of feature maps by absorbing the map features of all preceding layers as input. Enhancing the LSTM system's spatial characteristic learning process increases the efficiency of the intrusion detection classifier. The HMLSTM approach treats the boundary detector as the final segment when set to 1 at a time step of layer ℓ. Update,

Copy, or Flush are the available operations that the layers choose from based on the border states. The first step is to formulate the conventional LSTM equations.

$$\text{Gates and candidate} : \begin{bmatrix} i_t \, f_t \, u_t \, o_t \end{bmatrix} = Wx_t + Uh_{t-1} + b \tag{1.8}$$

$$\text{Cell state} : c_t = c_{t-1} \Theta \, \sigma\,(f_t) + \tanh tanh(u_t) \, \Theta \, \sigma\,(i_t) \tag{1.9}$$

$$\text{Hidden state} : h_t = \sigma\,(o_t) \, \Theta \, \tanh tanh(c_t) \tag{1.10}$$

In this case, h_{t-1} indicates the earlier hidden state, c_{t-1} indicates the previous cell state, and x_t indicates the present input to the LSTM. The input for the LSTM consists of three of these variables. The candidates for the activation forget that the symbols denote inputs and output gates i_t, f_t, u_t, and o_t, respectively. The weight matrix is represented by W, and the activation function matrix by U. The bias is indicated by b.

When these standard functions are combined with the boundary detector parameter (z_t),

$$\begin{bmatrix} i_t \, f_t \, u_t \, o_t \, z_t \end{bmatrix} = Wx_t + uh_{t-1}^1 + z_{t-1} Vh_{t-1}^2 + b \tag{1.11}$$

The presented HMLSTM framework comprises L layers ($\ell = 1, 2,\ldots,$ L), where the update procedure is carried out at time t for every layer.

$$h_t^l, c_t^l, z_t^l = f_{\text{HMLSTM}}^l \left(c_{t-1}^l, h_{t-1}^l, h_{t-1}^{l-1}, h_{t-1}^{l+1}, z_{t-1}^l, z_t^{l-1} \right) \tag{1.12}$$

z_{t-1}^l and z_t^{l-1} are the two boundary values that decide the forget gate of HMLSTM, which is represented by the function fHMLSTM ℓ. It is possible to modify the cell states as

$$\begin{aligned} c_t^l &= \{ f_t^l \, \Theta \, c_{t-1}^l + i_t^l \, \Theta \, g_t^l \text{ if } z_{t-1}^l = 0 \text{ and} \\ z_t^{l-1} &= 1 \text{ (update) } c_{t-1}^l \text{ if } z_{t-1}^l = 0 \text{ and} \\ z_t^{l-1} &= 1 \text{(copy) } i_t^l \, \Theta \, g_t^l \text{ if } z_{t-1}^l = 1 \text{ (Flush)} \end{aligned} \tag{1.13}$$

The cell proposal vector is shown by g in this instance. In contrast to the conventional LSTM, the input, forget, and output gates must be calculated at each time step in addition to g.

If the boundary z_{t-1}^l is discovered at the bottom layer but was absent in the preceding time phase, the Update function is carried out to update the layer ℓ summary depiction. Update operations are utilized infrequently because this scenario doesn't happen very often. Expressed, the copy operation is $(h_t^l, c_t^l) \leftarrow (h_{t-1}^l, c_{t-1}^l)$, meaning

that up until the bottom layer summary input is received, the top layer remains unaltered. Two subtasks make up the Flush operation. This indicates that if the Eject is not carried out, Reset deletes the bottom layer overview but permits the upper layer to absorb it.

$$\left[f_t^i i_t^l o_t^l g_t^l \bar{z}_t^l\right] = \left[\text{sigm sigm sigm tanh hardsigm}[]_{\text{slice}} \left(s_t^{\text{recurrent}(l)} + s_t^{\text{top-down}(l)} + s_t^{\text{bottom-up}(l)} + b\right)\right] \tag{1.14}$$

Here

$$s_t^{\text{recurrent}(l)} = U_l^l h_{t-1}^l \tag{1.15}$$

$$s_t^{\text{top-down}(1)} = z_{t-1}^l U_{l+h}^l h_{t-1}^{l+1} \tag{1.16}$$

$$s_t^{\text{bottom-up}(l)} = z_t^{l-1} W_{l-1}^{ll-1} ht \tag{1.17}$$

This guarantees that the layer has more long-term dependencies when it is first established.

It is determined that the binary boundary state z_t^l is

$$z_t^l = f_{\text{bound}}\left(\bar{z}_t^l\right) \tag{1.18}$$

Utilizing the deterministic step operation, it may be modeled.

$$z_t^l = \{1 \text{ if } \bar{z}_t^l > 0.50 \text{ otherwise.} \tag{1.19}$$

1.3.4 User Rating Identification

The Improved Golden Jackal Optimization (IGJO) algorithm is used to identify user ratings. The courses are recommended based on the best user ratings to the user/ student. Nitish Chopra and Mohsin Ansari introduced the golden jackal optimization (GJO) technique in 2022. This approach resembles the golden jackal's natural hunting techniques. The golden jackal is a medium-sized terrestrial predator belonging to the canine family. The areas of Southeast Asia, Europe, North and East Africa, the Middle East, and Central Asia are home to them. The jackal can run great distances for its prey because of its small size and lengthy legs. They often hunt in packs of both female and male jackals. Three steps comprise the golden jackal's hunting behavior:

Step 1: Examine the prey.
Step 2: Encircling and upsetting the target until it stops moving.

Step 3: Pounce on the target and attack.

(a) *Mathematical Model*

This section describes the mathematical structure of the golden jackal's hunt strategy, which creates the IGJO algorithm and performs optimization.

(1) *Search Space Formulation*

The golden jackal is a swarm-based method in which the first trial, or initial potential solution, is created randomly across the search zone. Equation (1.20) is the expression below for the initial solution z_0.

$$z_0 = L + \text{rand} * (U - L) \tag{1.20}$$

where rand indicates a uniform randomized integer lies inside [0, 1], and L and U denote the bottom and upper limits of the search zone, respectively. During this stage, the first matrix is created. The prey is listed in Eq. (1.21), and the jackal pair consists of the two most fit individuals.

$$\text{prey} = \begin{bmatrix} Z_{1,1} \ Z_{1,2} \ \cdots \ Z_{1,d} \ Z_{2,1} \ Z_{2,2} \ \cdots \ Z_{2,d} \ \vdots \ \vdots \ \vdots \ \vdots \ Z_{n,1} \ Z_{n,2} \ \cdots \ Z_{n,d} \end{bmatrix} \tag{1.21}$$

The prey location refers to the specifications of a specific solution. During the optimization process, every prey's fitness value is determined by an objective function. The fitness value of Eq. (1.22) is found using the following formula.

$$\text{Fitprey} = \begin{bmatrix} f(Z_{1,1}, Z_{1,2}, ...Z_{1,d})f(Z_{2,1}, Z_{2,2}, ..., Z_{2,d}) \vdots f(Z_{n,1}, Z_{n,2}, ..., Z_{n,d}) \end{bmatrix} \tag{1.22}$$

wherein Fitprey is a matrix that holds the prey's fitness values, and f represents the cost functions. A male jackal is represented by the value that fits the best, and a female jackal is represented by the value that fits the second best.

(2) *Exploration Phase*

Though occasionally the victim manages to avoid capture, jackals can identify and track their target. The jackals wait and search for new prey as a result. During hunts, a male jackal usually leads the way, with the female following after as:

$$Z_l(t) = Z_M(t) - E|Z_M(t) - rl * \text{prey}(t)| \tag{1.23}$$

$$Z_2(t) = Z_{FM}(t) - E|Z_{FM}(t) - rl * \text{prey}(t)| \tag{1.24}$$

The vector of locations is denoted as prey(t), the present iterations are represented by t, and the positions of the female and male jackals in the search space are

represented, respectively, by $Z_M(t)$ and $Z_{FM}(t)$. E, or the prey's escape energy, is calculated as follows:

$$E = E_1 * E_0 \tag{1.25}$$

E_1 and E_0 represent the prey's starting energy state and decreasing energy level, respectively.

$$E_0 = 2 * r - 1 \tag{1.26}$$

$$E_1 = c_1 * (1 - (t/T)) \tag{1.27}$$

The selected number, r, in this case, is between [0, 1]. Throughout the iterations, t and T linearly reflect the current from 1.5 to 0, with c_1 being the constant with a value of 1.5. Based on the levy flight (LF) distribution, the arbitrary random vector is represented by the rl, which is computed as follows:

$$[rl = 0.05 * LF(z) \tag{1.28}$$

To compute the LF, use Eq. (1.10)

$$LF(z) = 0.01 \times \frac{(\mu \times \sigma)}{\left(\left| v^{\left(\frac{1}{\beta}\right)} \right| \right)} ; \sigma = \left[\frac{r(1+\beta) \times \mathrm{sinsin}(\pi\beta/2)}{r\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-2}{2}\right)}} \right]^{1/\beta} \tag{1.29}$$

In this case, the constant β has a value of 1.5, while the arbitrary numbers μ and σ are between 0 and 1. In conclusion, the golden jackal's new update location is ascertained as follows:

$$z(1+t) = \frac{z_1(t) + z_2(t)}{2} \tag{1.30}$$

(3) *Exploitation Phase*

When jackals harass prey, their ability to flee is diminished, and the jackal pairs surrounding the prey are found in the previous phase. They encircle their prey, then leap upon it to consume it. The mathematical expression for this hunting activity, together with both male and female jackals, is:

$$Z_1(t) = Z_M(t) - E \cdot |rl \cdot Z_M(t) - \mathrm{prey}(t)| \tag{1.31}$$

$$Z_2(t) = Z_{FM}(t) - E \cdot |rl \cdot Z_{FM}(t) - \mathrm{prey}(t)| \tag{1.32}$$

(4) *Exploration to Exploitation Transitioning*

The E approach uses the IGJO value to move to exploitation from exploration. The prey's energy level dramatically drops during its flight behavior. E0 rises from 0 to 1, making the prey stronger, but decreasing from 0 to -1 makes the prey weaker. When |E| exceeds 1, jackal pairs search for prey in various locations. Jackals attack and exploit their prey if |E| is smaller than 1.

1.3.5 Recommendation Courses

Employ the Dynamic Decay Collaborative Filtering (DDCF) method to recommend the courses based on user preferences. Item clustering, intriguing level recognition, decay function specifications, and preference recommendation are the four stages of the DDCF process. DDCF addressed the recommendation method's cold start and sparsity problems using item clustering, which grouped comparable things without needing a predetermined parameter. Next, based on the duration and quantity of rating records within the cluster, we determine the exciting level of every cluster for every user. DDCF uses different decay functions at every stage to describe the progression of preferences. Ultimately, we use calculated decaying rates to compute the similarity between users and forecast their future preferences.

(a) *Item clustering*

In CF-based recommendations, two deadly flaws are cold start and sparsity. Recommendations for new users or things are associated with a cold start. It is challenging to provide accurate recommendations because the algorithm lacks knowledge about new users or things. A deficiency in transaction volume and feedback data leads to the sparsity issue. The usability of collaborative filtering will suffer because it is challenging for the recommendation engine to identify individuals who share similar interests. Similar to Definition 1.1, we determine the strength of the relationship between two users using the Jaccard coefficient after removing inconsequential relationships.

Definition 1.1 (*Relation Strength*)

The profile $p = \{p1, p2, \ldots, pk\}$ comprises k properties of item o, given an item o. One can determine the relationship between two elements using,

$$R(o_i, o_j) = \frac{\left| p_i \cap p_j \right|}{\sqrt{|p_i| \times |p_j|}} \tag{1.33}$$

The relation strength is determined using the user-specified threshold α as follows:

$$RS(o_i, o_j) = \{R(o_i, o_j), \text{ if } R(o_i, o_j) \geq \alpha_0, \text{ if } R(o_i, o_j) < \alpha \tag{1.34}$$

Naturally, α might regulate the density of relationships between items during clustering, impacting the process's efficiency. After determining the relation strength, we utilize a variable-free approach to cluster the product in the structure. As stated in Definition 1.2, DDCF proposes an evaluation similar to modularity for the terminating criterion of hierarchical clustering. Naturally, we can drastically reduce the time spent clustering by reducing the number of calculation iterations.

Definition 1.2 (*Strength Gain*)

In a recommendation model, the strength function is described as follows given an item set $O = \{o_1, \ldots, o_m\}$ and the clustering outcome $C = \{c_1, c_2, \ldots, c_p\}$.

$$S(C) = \sum_{k=1}^{p} \left\lceil \frac{IS_k}{TS} - \left(\frac{OS_k}{TS} \right) \right\rceil \tag{1.35}$$

$$\Delta S_{c \rightarrow c\prime} = S(C) - S(C\prime) \tag{1.36}$$

(b) *Decay function & interest level identification*

Following the item clustering process, DDCF uses every user's rating patterns to determine the interest level of every cluster. To explain the variation in preference, we take inspiration from the idea of human brain memory. Based on the Ebbinghaus forgetting curve, DDCF divides short-term users' preferences into instantaneous and long-term interest levels. The preference at the immediate level often lasts for a very short time and can quickly fade. According to the contract, the short-term preference will eventually deteriorate even though it may remain in the brain longer than the immediate level. People tend to remember preferences that are at a long-term level for extended periods.

Then, by sorting the rating document with t_{ij} in non-decreasing order, we might generate a rating sequence $\langle (r_{i1}, t_{i1}), (r_{i2}, t_{i2}), \ldots, (r_{il}, t_{il}) \rangle$. The significant set $se_{ik} = \{(r_{ij}, t_{ij}) | t_{ij+1} - t_{ij} \leq w, t_{i0} = t_{i1}, 0 \leq j \leq l\}$, given a user-specified time duration w. The Ebbinghaus memory curve states that people often do not forget a single item after revisiting it to explain preference variance. Thus, the interest level L_{ik} of c_k for u_i can be expressed as follows:

$$L_{ik} = \{\text{inse ous}_{\text{level}}, \text{ if } 0 < |se_{ik}| \leq 3 \text{ short-term}_{\text{level}}, \text{ if } 4 \leq |se_{ik}|$$
$$\leq 6 \text{ long-term}_{\text{level}}, \text{ if } 7 \leq |se_{ik}|\} \tag{1.37}$$

As was already established, instantaneous levels are often relatively brief and can quickly degrade. To simulate the preference shift, we opt for power decay. On the other hand, it indicates that a user finds these kinds of goods fascinating when they rank the items in a cluster more than four or seven times. To mimic the evolution of preferences at the short- and long-term levels, we might apply the logistical and exponential decay functions. Each level's decay function is described as follows:

$$\text{decay}_{\text{instantan t}}(\Delta t) = \Delta t^{-\lambda} \cdot \alpha \tag{1.38}$$

$$\text{decay}_{\text{short}}(\Delta t) = \frac{2}{1 + e^{\lambda - \Delta t}}o \tag{1.39}$$

$$\text{decay}_{\text{long}}(\Delta t) = e^{-\lambda \cdot \Delta t} \tag{1.40}$$

Observe that the variables λ and α, often obtained by heuristic assessment, may be used to adjust the decay degree of the function.

(c) *Preference prediction*

In contrast to conventional CF-based recommendations, DDCF predicts the item's rate by utilizing baseline estimations and similarity calculations that consider decay. The similarity between two items can be determined using various techniques, including Pearson, Jaccard coefficients, modified cosine, and cosine. In this study, the item similarity is computed using the adjusted cosine.

Definition 1.3 (**Item Similarity**)

Assume that a recommendation system consists of a set of items $O = \{o_1, \ldots, o_m\}$ and a set of users $U = \{u_1, \ldots, u_n\}$. $o_x, o_y \in O$, U_{o_x,o_y} is the set of users in U that have rated o_x and o_y given two items.

$$\text{sim}(o_{x'}, o_y) = \frac{\sum_{u_k} \in U_{o_x,o_y} (r_{u'_k o_x} - \overline{ro_x}) \times (r_{u'_k o_y} - \overline{ro_y})}{\sqrt{\sum_{k=1}^{n}(r_{u'_k o_x} - \overline{ro_x})^2} \times \sqrt{\sum_{k=1}^{n}(r_{u'_k oy} - \overline{ro_y})^2}}) \tag{1.41}$$

In the recommendation system, the average rates of items o_x and o_y are denoted by the $\overline{o_x}$ and $\overline{o_y}$, respectively.

The rate forecast of a particular item o_j in DDCF for a user u_i can be obtained as follows:

$$P_{u_i,o_j} = (1 - \rho) \times (\mu - b_{u_i} - b_{0_j}) + \rho \times \left(\overline{r_{o_j}} + \frac{\sum_{k=1}^{m} D(r_{u_{i'}o_k}, t_{u_j,o_k}) \times \text{sim}(o_{j'}, o_k)}{\sum_{k=1}^{m} |\text{sim}(o_{j'}, o_k)|}\right) \tag{1.42}$$

In which $0 \le \rho \le 1$. Baseline estimate and decay CF are Eq. (1.42) components that can be separated apart. Equation (1.34) is then used to calculate the decay CF for prediction, depending on the time t_{u_j,o_k} ok, and the associated decay function in Eq. (1.40), to obtain the decayed rate $D(r_{u'_i o_k}, t_{u_j,o_k})$.

Table 1.1 The proposed system's environment setup

Resource	Details
RAM	8 GB
CPU	Core i5 Gen6
Software	Python
GPU	4 GB

1.4 Results and Discussion

Experiments were run to evaluate the utility of the proposed recommendation approach, and a dataset of Coursera courses was used to assess the proposed algorithm. The experiment's results were compared to those of alternative recommendation systems. The main goal of the tests was to assess how well the proposed recommender method for learning materials integrated the traditional algorithm and how practical, improved, and precise the recommendations were. Learner satisfaction with the recommendations was evaluated following the experiments.

1.4.1 Experimental Settings

Multiple settings were used in the development of the system. The environment of the evolving platform is configured, as shown in Table 1.1.

1.4.2 Description of Datasets

Udemy Course https://www.kaggle.com/datasets/andrewmvd/udemy-courses datasets are modified for classification-based course recommendations to assess our system. This dataset includes 3.682 records of Udemy courses in four different subjects: musical instruments, graphic design, web design, and business finance. Massive open online courses (MOOCs) are available on Udemy for both free and a fee. The business model that enables anyone to build a course has allowed Udemy to host thousands of programs.

1.4.3 Performance Metrics

The job of an e-learning recommender system is to recommend helpful learning resources to the student. Performance metrics for the proposed approach were looked at, including Precision (P), Accuracy (A), F1-score (F), and Recall (R). These

measurements demonstrate:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1.43}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1.44}$$

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{1.45}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1.46}$$

When the terms false negative rate (FN), actual negative rate (TN), actual positive rate (TP), and false positive rate (FP) are used, accuracy is computed as the quantity of all correct expectations divided by the total amount of observations in the dataset. The degree to which the pages we recognized as having highly rated pages genuinely do have highly rated pages is indicated by precision evaluation. Both the anticipated advantages and users of the recommended resources are TP.

1.4.4   Performance Evaluation

E-learning recommendation systems aim to improve learner engagement, satisfaction, and overall learning outcomes by tailoring the learning experience to individual needs and preferences. They are widely used in online education platforms, Massive Open Online Courses (MOOCs), and other digital learning environments. When compared to the other machine learning-based techniques with the proposed method, the proposed method got higher metrics values compared to machine learning-based techniques of 59% for SVM, 64% for NN, and 79% for Fuzzy methods; the proposed method achieved 99.18% of f1-score and 47% for SVM, 53% for NN, 70% for Fuzzy methods, the proposed method achieved 99.15% of precision. Table 1.2 shows the proposed technique compared with existing machine learning-based techniques utilized in e-learning-based recommendation systems.

Table 1.2 Comparison of the proposed and existing machine learning approaches

Methods	F1-score	Accuracy	Precision	FNR	FPR
SVM	59	77	47	0.2	0.22
NN	64	81	53	0.2	0.17
Fuzzy	76	89	70	0.15	0.090
Proposed	99.18	99.23	99.15	0.11	0.075

The machine learning technique's FNR and FPR error rates compared with the proposed method, which obtained lower error rates. The results presented in Table 1.2 highlight a significant superiority of the proposed approach over existing machine learning methods, namely SVM, NN, and Fuzzy. The proposed approach's F1-score, accuracy, precision, false negative rate (FNR), and false positive rate (FPR) metrics demonstrate remarkable values of 99.18, 99.23, 99.15%, 0.11, and 0.075, respectively. In contrast, SVM, NN, and Fuzzy exhibit comparatively lower performance across these metrics. Potential factors contributing to this enhanced performance could include the optimization techniques employed, the feature set's robustness, or the model architecture's adaptability. A more detailed exploration of these aspects would not only provide a comprehensive understanding of the results.

Figure 1.2 shows the analysis of machine learning-based techniques with the proposed method's accuracy, f1-score, and precision metrics values.

Figure 1.3 compares the FNR and FPR values of the existing and proposed methods. Comparatively, when comparing the proposed method with existing methods mentioned in the literature survey, the proposed method achieved higher accuracy than others. Because collaborative filtering method performs well in the recommendation system in the e-learning model. The existing methods of DRNN obtained 97%, GSP-CA-CF obtained 48.1%, and FBRS obtained 94.45%, while compared with the proposed method obtained 99.15% for precision. Table 1.3 shows the overall comparison of existing and proposed methods in the literature survey.

Compared to existing approaches, the proposed technique achieved a higher accuracy value of 99.23%. Some techniques that are taken to compare are not evaluated with some metrics. So (–) indicates the metrics that were not taken in that research. The results presented in Table 1.3 provide a comprehensive overview of the performance metrics, including Precision, Recall, F1-score, and Accuracy, for various methods in the literature survey. Notably, the proposed method outperforms existing approaches across all evaluated metrics, achieving an exceptional Precision of 99.15%, Recall of 99.20%, F1-score of 99.18%, and Accuracy of 99.23%. While these results showcase the superiority of the proposed approach, the authors must delve into a more detailed analysis in the results section. This enhanced analysis will

Fig. 1.2 Comparison of existing machine learning-based techniques with the proposed method
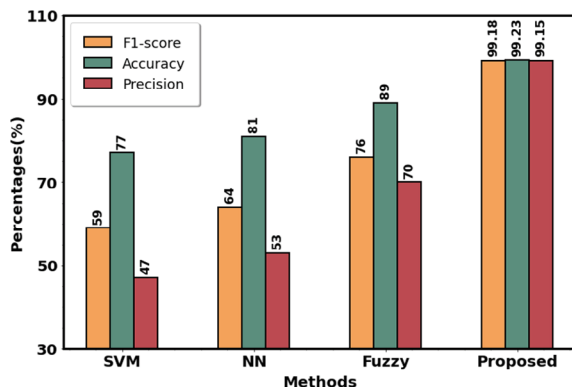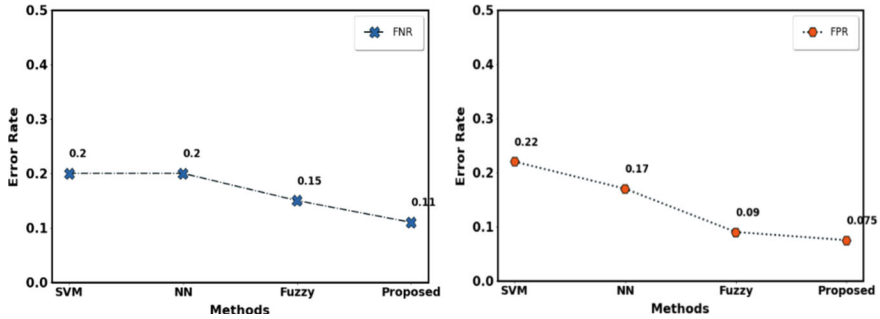
Fig. 1.3 The FNR and FPR values are compared with the proposed and existing approach

Table 1.3 Overall comparison of existing and proposed methods in literature survey

Methods	Precision	Recall	F1-score	Accuracy
DRNN [16]	97	96	–	–
GSP–CA–CF [17]	48.1	45.1	–	–
FBRS [18]	94.45	92.03	93.22	89.2
C-LSTM [20]	–	–	82.46	82.27
SPM [21]	97.16	–	97.84	97.45
HOA [22]	75.47	–	96.24	96.51
Proposed method	99.15	99.20	99.18	99.23

strengthen the credibility of the findings and contribute to the scholarly discourse on the subject.

Figure 1.4 compares proposed and previous techniques in an e-learning-based recommendation system.

1.4.5 Computational Time Complexity

Computational time complexity is a concept in computer science that measures the amount of computational time or resources an algorithm requires to solve a problem. It provides an estimate of how an algorithm's running time increases with the input size. Table 1.4 shows the comparison of computational time complexity between proposed and existing techniques.

Table 1.4 provides a comparative analysis of the computational time (in milliseconds) among various methods, including DRNN [16], GSP–CA–CF [17], FBRS [18], and the proposed method. It is evident from the results that the proposed method outperforms its counterparts, exhibiting the shortest computational time of 0.14 ms.

Fig. 1.4 Comparison of proposed and previous techniques
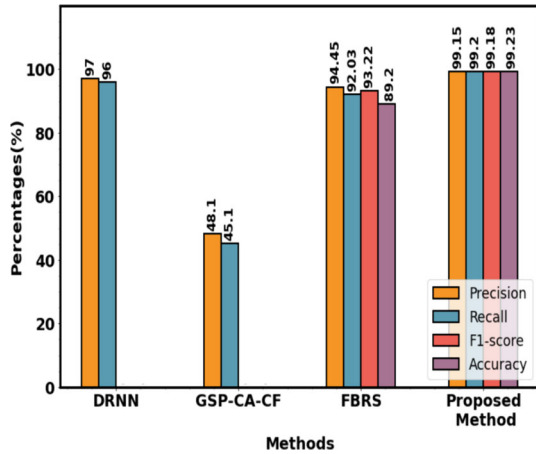
Table 1.4 Computational time of the proposed method compared with previous techniques

Methods	Computational time (ms)
DRNN [16]	0.25
GSP–CA–CF [17]	0.21
FBRS [18]	0.17
C-LSTM [20]	0.23
SPM [21]	0.20
HOA [22]	0.26
Proposed method	0.14

However, the analysis presented in the results section is currently limited, necessitating a more comprehensive examination of the factors contributing to the superior performance of the proposed approach.

Figure 1.5 compares the proposed method's computational time complexity with existing methods. Compared to the existing methods, the computational time of the proposed method is lower.

1.4.6   Evaluation of Training and Testing

Evaluation of training and testing is a critical aspect of building and assessing the performance of models. The process involves assessing how well a model has learned from training data and how effectively it can generalize to new, unseen data. The evaluation process helps understand the model's strengths, weaknesses, and overall performance.
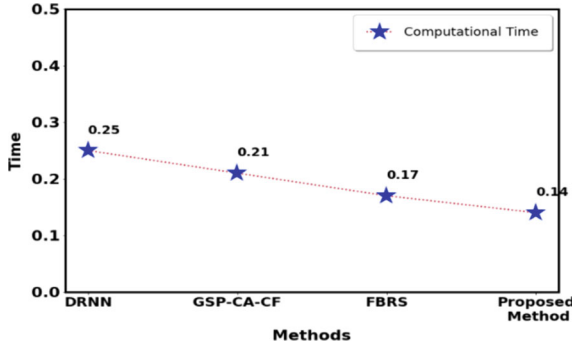
Fig. 1.5 Comparison of computational time complexity of proposed and existing methods
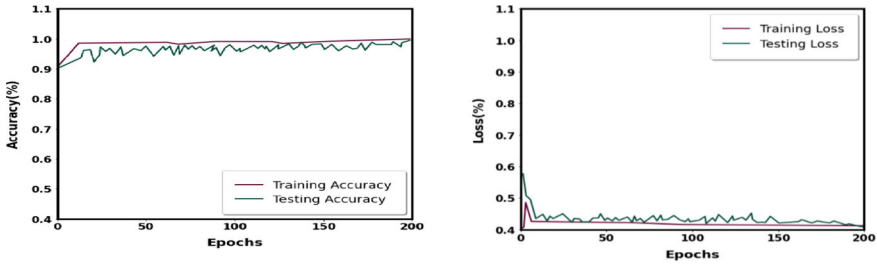
Fig. 1.6 Training and testing **a** accuracy, **b** loss

Figure 1.6 shows a graph of accuracy and loss value as the amount of iterations grew. The graph displays the advantageous impact of converging the methodology discussed in this research.

1.5 Conclusion

An e-learning-based recommendation system is a technology that provides personalized suggestions for learning resources or activities within an online education platform. The primary goal of such a system is to enhance the learning experience for individual users by offering content, courses, or resources tailored to their preferences, learning styles, and performance. This research gathers information from the Udemy dataset to clean, standardize, and analyze it so that recommendation systems can use it effectively. To categorize and recommend courses to the user, extract the relevant features. Next, divide the course level into starting, intermediate, and all levels using HMLSTM approach. After that, the IGJO algorithm will be utilized to determine user ratings and recommend courses to users. Finally, the Dynamic Decay Collaborative Filtering technique is used to recommend courses based on their ratings

to the user. The proposed method achieved a higher accuracy of 99.23% compared to existing methods with less computational time complexity. In the future, we will employ hybrid techniques to perform a recommendation system in an e-learning framework even better to achieve higher accuracy and reduce time complexity.

References

1. Ingavélez-Guerra, P., Otón-Tortosa, S., Hilera-González, J., Sánchez-Gordón, M.: The use of accessibility metadata in e-learning environments: a systematic literature review. Univ. Access Inf. Soc. **22**(2), 445–461 (2023)
2. Bhaskaran, S., Marappan, R., Santhi, B.: Design and analysis of a cluster-based intelligent hybrid recommendation system for e-learning applications. Mathematics **9**(2), 197 (2021)
3. Bagunaid, W., Chilamkurti, N., Veeraraghavan, P.: AISAR: artificial intelligence-based student assessment and recommendation system for e-learning in big data. Sustainability **14**(17), 10551 (2022)
4. Rahhali, M., Oughdir, L., Jedidi, Y., Lahmadi, Y., El Khattabi, M.Z.: E-learning recommendation system based on cloud computing. In: WITS 2020: Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, pp. 89–99. Springer Singapore (2022)
5. Ali, S., Hafeez, Y., Humayun, M., Jamail, N.S.M., Aqib, M., Nawaz, A.: Enabling recommendation system architecture in a virtualized environment for e-learning. Egypt. Inform. J. **23**(1), 33–45 (2022)
6. Bansal, J. C. (2019). Particle swarm optimization. In: Evolutionary and Swarm Intelligence Algorithms, pp. 11–23. Springer, Cham
7. Jena, K.K., Bhoi, S.K., Malik, T.K., Sahoo, K.S., Jhanjhi, N.Z., Bhatia, S., Amsaad, F.: E-learning course recommender system using collaborative filtering models. Electronics **12**(1), 157 (2023)
8. Gomede, E., de Barros, R.M., de Souza Mendes, L.: Deep autoencoders to adaptive e-learning recommender system. Comput. Educ.: Artif. Intell. **2**, 100009 (2021)
9. Jeevamol, J., Renumol, V.G.: An ontology-based hybrid e-learning content recommender system for alleviating the cold-start problem. Educ. Inf. Technol. **26**, 4993–5022 (2021)
10. Amane, M., Aissaoui, K., Berrada, M.: A multi-agent and content-based course recommender system for university e-learning platforms. In: International Conference on Digital Technologies and Applications, pp. 663–672. Springer International Publishing, Cham (2021)
11. Yan, L., Yin, C., Chen, H., Rong, W., Xiong, Z., David, B.: Learning resource recommendation in e-learning systems based on online learning style. In: International Conference on Knowledge Science, Engineering and Management, pp. 373–385. Springer International Publishing, Cham
12. Rafiq, M.S., Jianshe, X., Arif, M., Barra, P.: Intelligent query optimization and course recommendation during online lectures in e-learning system. J. Amb. Intell. Human. Comput. 1–20 (2021)
13. Shahbazi, Z., Byun, Y.C.: Agent-based recommendation in an e-learning environment using knowledge discovery and machine learning approaches. Mathematics **10**(7), 1192 (2022)
14. Ezaldeen, H., Misra, R., Bisoy, S.K., Alatrash, R., Priyadarshini, R.: A hybrid e-learning recommendation integrating adaptive profiling and sentiment analysis. J. Web Semant. **72**, 100700 (2022)
15. Tahir, S., Hafeez, Y., Abbas, M.A., Nawaz, A., Hamid, B.: Smart learning objects retrieval for e-learning with contextual recommendation based on collaborative filtering. Educ. Inf. Technol. **27**(6), 8631–8668 (2022)
16. Vedavathi, N., Anil Kumar, K.M.: An efficient e-learning recommendation system for user preferences using hybrid optimization algorithm. Soft Comput. **25**(14), 9377–9388 (2021)

17. Tarus, J.K., Niu, Z., Kalui, D.: A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. Soft. Comput. **22**, 2449–2461 (2018)
18. Ibrahim, T.S., Saleh, A.I., Elgaml, N., Abdelsalam, M.M.: A fog-based recommendation system for promoting the performance of E-Learning environments. Comput. Electr. Eng. **87**, 106791 (2020)
19. Cui, Z., Xu, X., Fei, X.U.E., Cai, X., Cao, Y., Zhang, W., Chen, J.: Personalized recommendation system based on collaborative filtering for IoT scenarios. IEEE Trans. Serv. Comput. **13**(4), 685–695 (2020)
20. Alatrash, R., Ezaldeen, H., Misra, R., Priyadarshini, R.: Sentiment analysis using deep learning for recommendation in the e-learning domain. In: Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE, pp. 123–133. Springer Singapore (2020)
21. Taurus, J.K., Niu, Z., Kalui, D.: A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. Soft. Comput. **22**, 2449–2461 (2018)
22. Vedavathi, N., Anil Kumar, K.M.: An efficient e-learning recommendation system for user preferences using hybrid optimization algorithms. Soft. Comput. **25**(14), 9377–9388 (2021)

Chapter 2
A Survey of Proximity and Range Sensing Technologies for Reliable Distance Estimation

Anuradha Thakare, Gautam Bhagat, Chetan Indulkar, Pranav Kale, and Atharva Powalkar

Abstract This comprehensive survey explores four prominent distance sensing technologies: Radar Sensor, Ultrasonic, Infrared (IR), and Lidar sensors. Each of these sensor types operates on distinct principles, offering unique advantages and applications that are invaluable across industries. Radar sensors provide exceptional precision by utilizing radio waves to detect objects' distances and velocities, making them indispensable for applications like automotive radar systems, weather monitoring, aviation, and maritime navigation. Ultrasonic sensors offer cost-effective versatility through ultrasonic sound wave reflections, finding utility in robotics, automation, object detection, and parking systems. IR sensors leverage infrared radiation to enhance motion detection and proximity sensing, playing a pivotal role in security, lighting control, and HVAC systems (Heat, Ventilation and Air Conditioning systems). Lidar sensors, utilizing laser technology, meticulously measure distances and create intricate 3D maps for forestry, environmental monitoring, archaeology, and urban planning. A comparative table encapsulates each sensor's attributes and applications, facilitating informed decisions by researchers and engineers. The survey also addresses forthcoming challenges and advancements in distance sensing technology, offering a glimpse into the field's future developments.

A. Thakare · G. Bhagat · C. Indulkar · P. Kale (✉) · A. Powalkar
Department of Computer Science & Engineering (AI & ML), Pimpri Chinchwad College of Engineering, Pune, India
e-mail: pranavkale021998@gmail.com

A. Thakare
e-mail: anuradha.thakare@pccoepune.org

2.1 Introduction

Sensors serve as a vital link between the physical and digital domains, translating real-world occurrences into actionable data. Their widespread use has revolutionized industries, science, and daily life, enabling data-driven decision-making and technological advancement. In various fields, accurate distance estimation plays a pivotal role, from robotics and autonomous navigation to environmental monitoring and industrial applications. While humans possess an innate ability to estimate distances visually, the integration of sensors for distance measurement offers unparalleled precision and reliability. This survey paper delves into the compelling reasons behind the preference for sensors over human perception in distance estimation tasks.

Sensors, such as LiDAR (Light Detection and Ranging), ultrasonic, IR Sensor and radar systems, leverage advanced technologies to measure distances with minimal error margins. Unlike human estimation, which can be influenced by lighting conditions and cognitive biases, sensors provide objective and consistent measurements, crucial for critical decision-making processes. Furthermore, sensors excel in scenarios where human perception falls short, such as in harsh environments or remote locations.

This survey paper thoroughly explores the technical underpinnings of sensor-based distance estimation, comparing their advantages over human estimation methods and pitch it down to a conclusion where users can think about when to use which sensor for projects. This literature survey contributes to a comprehensive understanding of the evolving landscape of distance estimation technologies by highlighting the indispensability of sensors in achieving accurate and dependable distance measurements.

2.2 Proximity Sensor

A proximity sensor is a device designed to identify the presence or absence of objects within its immediate vicinity, all without necessitating any form of physical contact. This technology operates on several principles, including electromagnetic fields, infrared radiation, ultrasonic waves, and capacitive sensing. The core purpose of a proximity sensor is to trigger a reaction or execute a specific action when an object enters or exits its predefined detection range. Proximity sensors have found applications across a wide spectrum of industries, encompassing industrial, automotive, consumer electronics, and security systems. The advantages they offer include rapid and reliable detection, contactless operation, and suitability for challenging environments. Nevertheless, factors such as detection range and susceptibility to external conditions may influence their efficacy. By enabling touchless interactions and enhancing control systems, proximity sensors play a pivotal role in driving automation, ensuring safety, and augmenting overall convenience.

2.3   Distance Sensor Versus Proximity Sensor

The terms "distance sensor" and "proximity sensor" are sometimes used interchangeably, however depending on the situation, they might indicate slightly different things.

A. Distance Sensor

Any sensor that detects the distance between itself and an object or surface without making physical touch is referred to as a distance sensor. Distance measurements from sensors are often stated in quantities like meters or centimeters and are provided by distance sensors. These sensors can measure distances across a wide range, from close proximity to a great distance.

B. Proximity Sensor

A particular kind of distance sensor called a proximity sensor detects the presence or absence of an object within a specified region without necessarily measuring the precise distance. In order to identify when an object gets close enough to the sensor, proximity sensors use a binary output that indicates whether the object is "near" or "far" from the sensor. These sensors are frequently utilized to detect objects, avoid obstacles, or initiate activities when an object is close to the sensor.

2.4   Prominent Sensors for Distance Estimation

2.4.1   Ultrasonic Sensor

The ultrasonic sensor works on the signal transmission by sending the ultrasonic signals and receiving the signal reflection of the gap between transmitting. The core components of this sensor are the receiver and transmitter as shown in Fig. 2.1. Ultrasonic sensors are primarily used to measure the distance, and changes in the object's position. The ultrasonic sensor works on non-contact technology. Ultrasonic sensors transform sound waves into electrical signals as mentioned [3].

(1) Working Principle:

The fundamental principle underlying the operation of an ultrasonic sensor involves the emission or generation of sound waves, followed by the measurement of the time it takes for these waves to be reflected, ultimately enabling the calculation of distance. In this process, the ultrasonic sensor comprises two essential components: a transmitter responsible for emitting sound waves through the utilization of piezoelectric crystals, and a receiver designed to capture the sound waves emitted by the transmitter, which have traversed from their source to a designated target location, as depicted in Fig. 2.2.

Fig. 2.1 An image of
ultrasonic sensor (HC-SR04)

Fig. 2.2 Working principle
of ultrasonic sensor
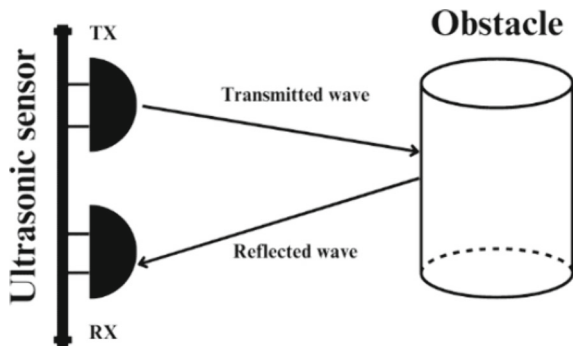
The calculation of distance, denoted as "D" in this context, is determined by the formula D = ½ T x C. Here, "D" signifies the distance being measured, "T" represents the elapsed time, and "C" denotes the constant speed of sound, which is approximately 343 m per second. This formula serves as the fundamental basis for distance calculation using ultrasonic sensors. Ultrasonic sensors are compatible in working with other sensor such as proximity sensor as mentioned in [16].

(2) Merits of Ultrasonic Sensor:

This survey paper explores ultrasonic sensor technology, underscoring its advantages detailed in Table 2.1. Utilizing sound waves beyond human hearing, these sensors offer non-contact precision, adaptability, and resilience, setting the stage for a detailed exploration of diverse applications in various fields.

(3) Limitations of Ultrasonic Sensor:

This survey paper thoroughly explores ultrasonic sensor technology, delving into its strengths and limitations detailed in Table 2.2. By providing a balanced perspective, it enhances our understanding of the practical implications of ultrasonic sensors across industries.

Table 2.1   Merits of ultrasonic sensor

Merits	Description
Non-contact operation	Ultrasonic sensors measure distance without physical contact, preventing damage to the sensor or the object
Wide detection range	Excels in detecting objects across a broad range of distances, suitable for diverse distance specifications
Color and transparency insensitivity	Unaffected by color, transparency, or surface properties of the target object, in contrast to optical sensors
Cost-effectiveness	Provides a cost-effective solution with commendable performance relative to its price point

Table 2.2   Limitation of ultrasonic sensor

Limitations	Description
Reduced precision in proximity	Diminished precision at short distances due to the time sound waves take to travel to and from the object
Disturbances and inaccuracies	Environmental variables like echoes and reflective surfaces can introduce errors or compromise ultrasonic sensor accuracy
Constrained sensing scope	Ultrasonic sensors with a cone-shaped detection area may result in imprecise measurements if objects deviate from the cone
Challenges with soft or porous objects	Ultrasonic waves may struggle to reflect effectively from soft, porous, or irregular surfaces, leading to measurement inaccuracies

(4) Applications of Ultrasonic Sensor:

(a) Obstacle Avoidance System for Unmanned Ground Vehicles Using Ultrasonic Sensors [1].

In this research, multiple SRF05 ultrasonic sensor were used in an Unmanned Ground Vehicle (UGV) to estimate the distance. These data were fed to an Artificial Neural Network (ANN) for prediction and the output accuracy was achieved up to 78.2% as shown in Table 2.3.

(b) Motion Detector with Display of Distance in the LCD Using Arduino Uno, and Ultrasonic Sensor HC-SR04 [2].

In this research, HC-SR04 ultrasonic sensor was used to determine the position of an approaching object accurately. The proposed system was used to track the human

Table 2.3   Accuracy table for obstacle avoidance system

Input	Model	Methodology	Accuracy
Input from sensor	SRF05 ultrasonic sensor	Neural network	78.2% for test data

Table 2.4 Accuracy table for motion detector

Input	Model	Methodology	Accuracy
Input from sensor	HC-SR04 ultrasonic sensor	Arduino	98.0% for test data

activity remotely and monitor the changes of distances the high accuracy offers from range 2 cm to 150 cm or 1 foot to 5 feet also the recorded distances are displayed on the LCD. Also, piezo buzzer was used to produce the sound alarm at different levels and accuracy was achieved up to 98% on test data as mentioned in Table 2.4.

2.4.2 Infrared Sensor

Infrared (IR) sensors, also known as Passive Infrared (PIR) sensors or pyroelectric sensors, have wide-ranging applications, including motion detection, automatic lighting control, and proximity sensing. These versatile devices detect changes in infrared radiation emitted or reflected by objects in their vicinity [4]. Their unique non-linear characteristics make them highly sensitive to object surface properties, enabling accurate motion and heat pattern detection. Consequently, they are extensively used in security systems for intrusion detection and occupancy sensing in automatic lighting setups. In comparison to ultrasonic sensors, infrared sensors offer fast response times and are generally more cost-effective, making them a popular choice for various applications (Fig. 2.3).

(1) Working Principle:

The working principle of infrared sensors revolves around detecting changes in infrared radiation emitted or reflected by nearby objects. These sensors incorporate pyroelectric materials capable of generating a small electric charge when exposed to temperature variations caused by infrared radiation [5]. As shown in Fig. 2.4, The sensor's circuit amplifies and processes this electric charge to identify motion or the presence of objects.

To accomplish this, an IR sensor emits infrared radiation and then detects its reflection or absorption from objects within its field of view. Typically, the sensor consists of an IR emitter and a detector. The emitted IR radiation interacts with objects, and the detector picks up the reflected IR radiation, converting it into an

Fig. 2.3 An image of Infrared (IR) sensor
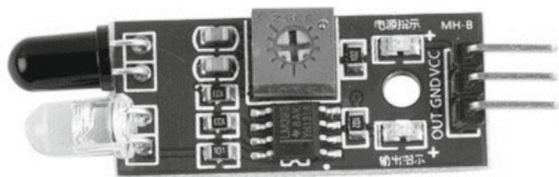
Fig. 2.4 IR sensor
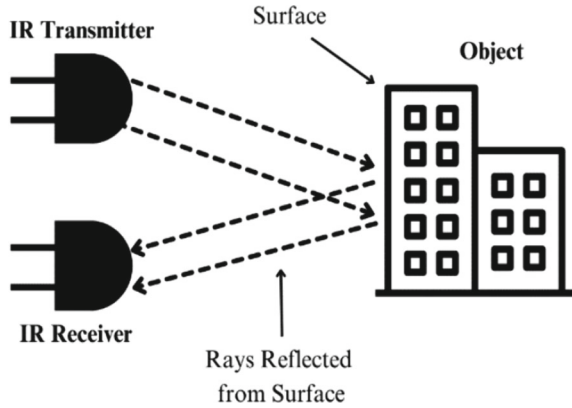
electrical signal [4]. The intensity of this signal is related to the properties of the detected objects, such as their temperature and surface characteristics. The processed electrical signal triggers specific actions, making IR sensors indispensable for motion detection, temperature measurement, and proximity sensing applications.

(2) Merits of Infrared Sensor:

This survey paper delves into IR sensor technology, highlighting its merits in Table 2.5. Harnessing infrared waves, these sensors provide non-contact precision, adaptability, and resilience. This sets the foundation for an in-depth exploration of their diverse applications across various fields.

(3) Limitations of Infrared Sensor:

For detecting accurate distances using IR sensor having prior knowledge of the surface is advised [5]. Furthermore Table 2.6 highlights the limitations of using IR sensor.

(4) Applications of IR Sensor:

(a) Using Infrared Sensors for Distance Measurement in Mobile Robots [6].

Table 2.5 Merits of infrared sensor

Merit	Description
Non-linear characteristics	Infrared sensors detect subtle heat changes with non-linear responses, enhancing motion and proximity sensing accuracy
Fast response time	With impressive response times, infrared sensors quickly react to changes in infrared radiation, crucial for time-sensitive applications
Cost-effectiveness	Cost-effective compared to alternatives like ultrasonic sensors, making them practical for various applications
Contactless sensing	Infrared sensors operate without contact, suitable for scenarios where physical contact-based sensors may be impractical or unsafe

Table 2.6 Limitation of infrared sensor

Limitations	Description
Higher cost	Although advancements in technology have gradually reduced the cost of infrared sensors, they may still be relatively more expensive than certain other sensor types
Limited range	Infrared sensors possess a limited detection range, and their effectiveness diminishes with increasing distance. Consequently, their application is constrained in scenarios requiring long-range sensing

Table 2.7 Accuracy table for distance measurement in mobile robots

Input	Model	Methodology	Accuracy
Input from sensor	SICK DME 2000 laser sensor	Sensor	50.0% for test data

Table 2.8 Accuracy table for monitoring people presence system

Input	Model	Methodology	Accuracy
Input from sensor	Panasonic grid EYE 64pixel	Pattern recognition algorithm	100.0% for test data

In this research, the SICK DME 2000 Laser Sensor was used to detect the presence of the object between two selectable measuring ranges: (1) near (10-204.7 cm) and (2) far (10-13,107 cm). [6] also suggested SUNX PXI AGV Bumper Sensor. The sensor provides a wide range from 90 cm to 250cm and accuracy was achieved up to 50.0% as shown in Table 2.7.

(b) Monitoring Indoor People Presence in Buildings Using Low-Cost Infrared Sensor Array in Doorways [8].

In this research, Panasonic Grid EYE 64-pixel IR Sensor was used to capture thermal image of user's passage with a viewing angle of 60°. The IR sensors were used to perform the detection of the people direction and with the help of the pattern recognition algorithm. The accuracy varies from environmental conditions and the average accuracy achieved as shown in the Table 2.8.

2.4.3   LiDAR Sensor

LiDAR technology, a cutting-edge innovation that harnesses the power of laser pulses, has transcended traditional boundaries to enable the creation of intricate and highly accurate three-dimensional representations of both objects and expansive environments. This sophisticated technology has triggered a seismic shift across numerous industries, propelling them into a new era of enhanced data collection, analysis, and interpretation [9].

Fig. 2.5 An image of
LiDAR sensor

(1) Working Principle of LiDAR Sensor:

Light Detection and Ranging (LiDAR) technology has made significant progress in
the field of remote sensing. LiDAR technology functions based on a fundamental
principle: a sensor sends out a beam of light towards a surface or object, then measures
the duration it takes for that light to bounce back following reflection. [10, 11]. These
time-of-flight data is then used to calculate the distance to the target accurately
(Fig. 2.5).

Both ground-mounted and airborne LiDAR have unique advantages and uses,
making them essential tools in diverse fields such as environmental monitoring, land
surveying, forestry, disaster management, and infrastructure planning [11, 12]. Their
remarkable precision in capturing detailed spatial data has greatly contributed to
advancements in scientific and engineering endeavors [12]. In the LiDAR measure-
ment process, factors like the tilt and angle of the emitted light ray also play a crucial
role in accurately calculating distances as shown in Fig. 2.6.

(2) Merits of LiDAR Sensor:

This survey paper examines LiDAR sensor technology, emphasizing its advantages
outlined in Table 2.9. By utilizing laser light, these sensors offer non-contact preci-
sion, adaptability, and resilience. This establishes the groundwork for a detailed
exploration of their varied applications across diverse fields.
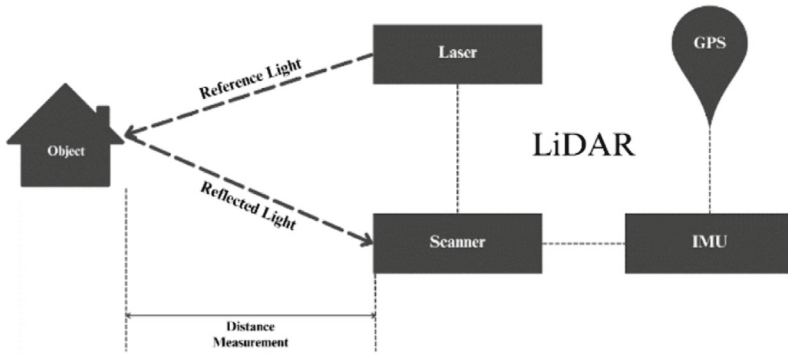
Fig. 2.6 Working principle of LiDAR sensor

Table 2.9 Merits of LiDAR sensor

Merits	Description
High measurement range and accuracy	LiDAR can accurately measure distances over significant ranges, with high precision, providing detailed and reliable data
Fast update rate	LiDAR systems can rapidly acquire data with high update rates, making them suitable for capturing real-time information
Small wavelengths and object detection	The use of short wavelengths in LiDAR enables it to detect and accurately measure small objects, which can be challenging for other remote sensing technologies like radar or sonar
Day and night operation	LiDAR works effectively in both daylight and low-light conditions, allowing for continuous data collection and ensuring consistent performance regardless of the time of day

(3) Limitations of LiDAR Sensor:

Expanding the investigation to LiDAR sensors, we concluded to a certain limitation which are mentioned in Table 2.10. If a project necessitates lightweight sensing technologies, LiDAR sensors may prove impractical due to their weight and size concerns in comparison to alternative sensors [12].

(4) Applications of LiDAR Sensor:

(a) Online Vehicle Detection Using Deep Neural Networks and Lidar Based Preselected Image Patches [7].

In this research, LiDAR sensor was used to detect the vehicle and to perform classification of vehicles in real time with the help of CNN (Convolutional Neural Network). LiDAR sensors were used to extract the patches of the image and radar sensory data

Table 2.10 Limitation of LiDAR sensor

Limitations	Description
Higher cost	Implementing LiDAR systems can be more expensive compared to other remote sensing technologies like ultrasonic or infrared (IR) sensors. This cost factor can be a limitation, especially for projects with tight budgets
Eye safety concerns	Some higher-end LiDAR devices use near-infrared light pulses, which can potentially be harmful to the human eye if proper safety precautions are not taken. Adequate safety measures are essential to protect operators and anyone else in the vicinity of LiDAR equipment

Table 2.11 Accuracy table for online vehicle detection

Input	Model	Methodology	Accuracy
Input from sensor	*LiDAR* sensor	Deep Neural Networks	80.0% for test data

Table 2.12 Accuracy table for mobile robot navigation

Input	Model	Methodology	Accuracy
Input from sensor	*LMS 151 2D LiDAR* sensor	Route planner algorithm	90.0% for test data

with accuracy up to 80% mentioned in Table 2.11. The test data for this research was up to 800 different images.

(b) Mobile Robot Navigation Using 2D LIDAR [17]:

In this research, multiple LMS 151 2D LiDAR sensors were used to achieve a field of view of 0-50 m for detecting obstacles in the device's path. The system achieved an accuracy of 90% in detecting obstacles using this setup as shown in the Table 2.12.

2.4.4 Radar Sensor

Radar sensors (Radio Detection and Ranging sensors) use radio waves to detect and identify objects. They are used when there is a need to calculate the direction, speed and distance of an object with great accuracy. This technology is used in various domains like automotive, aviation as well as military [14]. In automotive they are useful for obstacle detection, the aviation uses it to avoid collision and height measurement from surface, and the military uses it for missile guidance and target

Fig. 2.7 An image of radar sensor

tracking. One of the most valued characteristics of radar sensors is their capability to deliver precise data even in adverse weather conditions [14, 15] (Fig. 2.7).

(1) Working Principle of Radar Sensor:

Radar sensors emit radio waves or microwaves as shown in Fig. 2.8. When the emitted waves hit any object, they reflect to the sensor. Then the Time-of-Flight (time taken to return to the sensor) is calculated by the sensor. It determines the distance to the object using the established speed of light [14]. Radar sensors frequently try to remove background noise and retrieve extra details using signal processing. This technology, which provides exact distance and velocity information for navigation and danger avoidance, has essential uses in industries like aviation, vehicle safety, weather monitoring, and military surveillance [15].

(2) Merits of Radar Sensor:

While exploring Radar sensor technology, we highlighted some of the major advantages in Table 2.13. Utilizing radio waves, these sensors offer non-contact precision, adaptability, and resilience. This forms the basis for an in-depth exploration of their diverse applications across various fields.

Fig. 2.8 Working of radar sensor
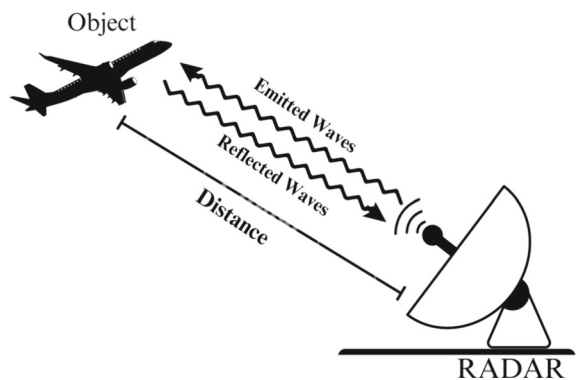
Table 2.13 Merits of radar sensor

Merits	Description
Long range	Radar sensors are excellent for long-range weather monitoring and air traffic management since they can detect objects at great distances
All-weather operation	They function well in inclement weather, such as snow, rain, and fog, when optical cameras and lidar may struggle
High accuracy	Radar sensors have a high degree of accuracy, which is essential for applications like driverless vehicles and aircraft navigation
24/7 operational	Radar is appropriate for use in surveillance, security, and defence applications since it can run continuously

(3) Limitation of Radar Sensor:

Examining radar sensors, we found multiple limitation which are mentioned in Table 2.14. By presenting a comprehensive perspective, it deepens our insights into the practical implications and challenges of radar sensor technology across diverse industries.

(4) Applications of Radar Sensor:

(a) Deep Learning-Based Indoor Distance Estimation Scheme Using FMCW Radar [14].

In this research, FMCW Radar (Frequency Modulated Continuous Wave) was used for distance estimation and 2D positioning using two radar sensors. Further, these data were processed using Deep Learning model to minimize distance estimation error approx. 0.069 m and accuracy for this research was achieved up to 93.1% as shown in Table 2.15.

Table 2.14 Limitation of radar sensor

Limitations	Description
Interference	Radar interference, which can result in misleading readings, can be caused by other radar emitters or electrical equipment
Power consumption	Some radar systems can consume a lot of power, making them unsuitable for battery-operated or portable equipment
Expensive computation	Advanced signal processing, which can be computationally demanding, is frequently needed to extract information from radar signals
Limited detection of stationary objects	Radar's limited ability to detect stationary objects results from its primary focus on the detection of moving things. As a result, radar is less efficient at identifying stationary barriers

Table 2.15 Accuracy table for indoor distance estimation

Input	Model	Methodology	Accuracy
Input from sensor	*FMCW Radar*	Deep Learning	93.1%

2.5 Results

Sensor selection is crucial for precise data acquisition. Ultrasonic sensors suit short-range proximity and obstacle detection. IR sensors excel in object detection, while LiDAR offers accuracy and range for robotics and mapping. Radar sensors provide long-range accuracy, ideal for traffic monitoring. Each sensor type has distinct advantages for specific applications. As shown in Table 2.16, it represents a brief comparison about all sensors.

Furthermore, for tasks requiring high accuracy, such as mapping and precision measurements, both LiDAR and Radar sensors outperform their counterparts by providing detailed and precise spatial information. In the realm of autonomous vehicles, where reliability and precision are critical, LiDAR and Radar sensors take precedence due to their capabilities in real-time mapping and obstacle detection. Despite limitations in range and accuracy, Ultrasonic and IR sensors still find utility in specific scenarios such as projects which are short-range, small size application and cost effective.

In summary, choosing the optimal sensor depends on the specific demands of the task at hand, with each sensor type showcasing its distinct strengths and weaknesses across different scenarios as per Table 2.17.

Table 2.16 Comparison of all sensors

Parameter	Ultrasonic sensor	IR sensor	LiDAR	Radar sensor
Principle of operation	Sound Waves	Infrared light absorption	Laser light reflection	Radio wave reflection
Field of view	Wide	Narrow to wide	Narrow to wide	Wide
Cost	Low	Low	High	High
Size	Compact	Compact to moderate	Moderate to large	Moderate to large
Power consumption	Low	Low to moderate	Moderate to high	Moderate to high
Maximum range	2 cm to 5 m	10 cm to 5 m	0.1 m to several km	10 m to several km
Accuracy	Moderate	Moderate	High	High

Table 2.17 Compatibility for all projects

Project category	Ultrasonic sensor	IR sensor	LiDAR	Radar sensor
Short range project	Compatible	Compatible	Compatible	Compatible
Proximity sensing	Compatible	Compatible	Compatible	Compatible
Weather affected	Incompatible	Compatible	Compatible	Incompatible
High accuracy	Incompatible	Incompatible	Compatible	Compatible
Autonomous vehicle	Incompatible	Incompatible	Compatible	Compatible
Mapping	Incompatible	Incompatible	Compatible	Compatible

2.6 Conclusion

In conclusion, this survey navigates through the realms of radar, ultrasonic, IR, and Lidar sensors, unravelling their distinct virtues and applications. By juxtaposing these sensors, valuable insights emerge to guide pragmatic choices for projects. Radar stands unrivalled in precision, suited for tasks like automotive, aviation, and weather monitoring. Ultrasonics cost-effectiveness fits robotics and parking systems. IR steps in for security and lighting, while Lidar's 3D prowess suits forestry and urban planning. The comparative table serves as a compass for selection. As we contemplate the path ahead, challenges and innovations beckon collaborative strides. In this sensor-driven evolution, academic collaborations promise ingenious solutions. Astute selection, based on the nuances of each sensor, thus steers us into a future woven with accurate, adaptable, and pioneering distance sensing technologies.

References

1. De Simone, M.C., Rivera, Z.B., Guida, D.: Obstacle avoidance system for unmanned ground vehicles by using ultrasonic sensors. Machines **6**(2), 18 (2018)
2. Gabriel, M.M.: Arduino Uno, Ultrasonic Sensor HC-SR04 Motion Detector with Display of Distance in the LCD. IJERT V9 (2020)
3. Andang, A., Hiron, N., Chobir, A., Busaeri, N.: Investigation of ultrasonic sensor type JSN-SRT04 performance as flood elevation detection. IOP Conf. Ser.: Mater. Sci. Eng. 550:012018 (2019)
4. Colla, V., Sabatini, A.: A composite proximity sensor for target location and color estimation. In: IMEKO Sixth International Symposium on Measurement and Control in Robotics, pp. 134–139 (1996)
5. Novotny, P., Ferrier, N.: Using infrared sensor and the Phong illumination model to measure distances. In: International Conference on Robotics and Automation, Detroit, MI vol. 2 pp. 1644–1649 (1999)
6. Korba, L., Elgazzar, S., Welch, T.: Active infrared sensors for mobile robots. IEEE Trans. Instrum. Measur. **2**(43), 283–287 (1994)
7. Benet, G., Blanes, F., Simó, J.E., Pérez, P.: Using infrared sensors for distance measurement in mobile robots. Robot. Auton. Syst. **40**, 255–266 (2002)
8. Perra, C., Kumar, A., Losito, M., et al.: Monitoring indoor people presence in buildings using low-cost infrared sensor array in doorways. Sensors **21**, 4062 (2021)

9. Bastos, D., Monteiro, P.P., Oliveira, A.S.R., Drummond, M.V.: An overview of LiDAR requirements and techniques for autonomous driving. In: 2021 Telecoms Conference (ConfTELE) (2021)
10. Fridlander, J., Sang, F., Rosborough, V., et al.: Dual laser indium phosphide photonic integrated circuit for integrated path differential absorption lidar. IEEE J. Select. Topics Quantum Electron. **28**, 1–8 (2022)
11. Poulton, C.V., Byrd, M.J., Russo, P., et al.: Coherent LiDAR with an 8,192-element optical phased array and driving laser. IEEE J. Select. Topics Quantum Electron. **28**, 1–8 (2022)
12. Jokela, M., Pyykonen, P., Kutila, M., Kauvo, K.: LiDAR performance review in arctic conditions. In: 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP) (2019)
13. Lange, S., Ulbrich, F., Goehring, D.: Online vehicle detection using deep neural networks and lidar based preselected image patches (2016)
14. Park, K.-E., Lee, J.-P., Kim, Y.: Deep learning-based indoor distance estimation scheme using FMCW radar. Information **12**, 80 (2021)
15. Hakobyan, G., Yang, B.: High-performance automotive radar: a review of signal processing algorithms and modulation schemes. IEEE Signal Process. Mag. **36**, 32–44 (2019)
16. Sabatini, A.M., Genovese, V., Guglielmelli, E., et al.: A low-cost, composite sensor array combining ultrasonic and infrared proximity sensors. In: Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems Human Robot Interaction and Cooperative Robots. https://doi.org/10.1109/iros.1995.525872
17. Csaba, G., Somlyai, L., Vamossy, Z.: Mobil robot navigation using 2D LIDAR. In: 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI) (2018)

Chapter 3
A Comprehensive Review of Sensor and Actuator Fault Detection Methods in Aerospace Systems

V. K. Athira and Manoranjan Sinha

Abstract The public has long been concerned about the safety of the aerospace system because a single commercial aircraft event can result in a significant number of fatalities. Since unmanned aerial vehicles have gained popularity over the past decade, it is also essential to be optimistic about their dependability and safety. Sensors and actuators play crucial roles in aircraft control systems, and any malfunction can lead to catastrophe. In this paper, we have gathered information on various types of common faults in sensors and actuators in aerospace systems and fault detection and diagnosis techniques with the support of an extensive literature review. Sensor and actuator malfunction detection techniques such as model-based, data-driven, artificial intelligence-based, and signal processing algorithms are evaluated in this work using substantial bibliographical data. Efficiency, fault detection time, false alarm rate, and computing complexities are investigated for the fault detection techniques. The article provides the advantages and disadvantages and recommendations for future improvement of methods suitable for aerospace systems.

Keywords Aerial vehicles · Sensor and actuator faults · Fault detection · Diagnosis and prognostics · Aircraft health monitoring

3.1 Introduction

Aviation accidents often get worldwide attention due to their catastrophic nature and result in a large number of fatalities, even though the frequency of such incidents has been decreasing in the past decade [1]. From the statistical data, it is known that these accidents are often unpredictable and can be caused as a result of inappropriate

V. K. Athira (✉) · M. Sinha
Aerospace Engineering Department, IIT Kharagpur, Kharagpur, West Bengal, India
e-mail: athiravimalan.vk@kgpian.iitkgp.ac.in

M. Sinha
e-mail: masinha@aero.iitkgp.ac.in

pilot response, system failure, environmental factors, etc. [2]. Many fatal accidents
have occurred due to faults in the sensor and actuator malfunction [3]. Similarly,
over 80% of unmanned aerial vehicle (UAV) flight-related problems are caused by
malfunctions affecting the propulsion systems, flight control surfaces, or sensors [4].
The day to day applications of UAVs are increasing day by day. Hence, the reliability
needs to be improved. Fault Detection and Diagnosis (FDD) involves detecting errors,
malfunctions, or faults in a system [5].

   As shown in Fig. 3.1, hardware redundancy is the conventional technique utilized
in the aviation industry to identify sensor and actuator problems. To measure a single
variable, many redundant sensors are used, and a voting system is used to identify
which sensor failed [6]. Reducing redundant sensors is essential to building lighter,
greener, more sustainable, and fuel-efficient airplanes in the future. For UAVs, adding
more sensors is not feasible [7]. Thus, in order to identify and isolate malfunctions and
failures in sensors and actuators, analytical redundancy needs to be included in the
sensor fusion scheme. Analytical redundancy is the process of estimating variables
and finding errors by using data-based inference or mathematical correlations [8], as
represented in Fig. 3.2. Numerous studies in this field have been conducted during the
last three decades. Particularly model-based FDD designs. Various observers, filters,
and parameter estimators are used in model-based FDD techniques [9–38]. Many
data-driven and neural network-based FDDs have been proposed recently [39–60].
There are numerous survey papers on sensor and actuator fault detection methods in
various industrial systems [7, 61–63]. However, there are very few FDD reviews on
aerospace systems available [4, 6, 64].

Fig. 3.1 Hardware
redundancy
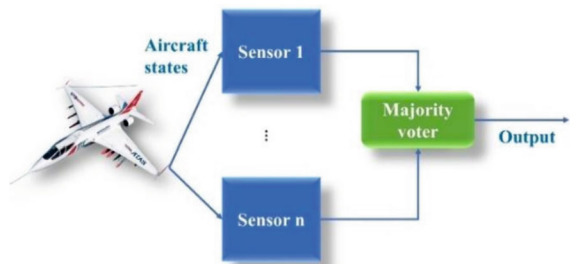
Fig. 3.2 Analytical
redundancy
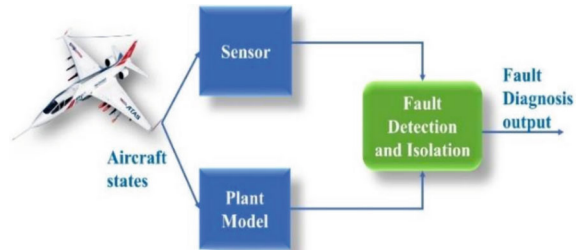
This paper is organized as follows. Sections two and three discuss various types of sensor and actuator faults in detail. Section 3.4 classifies different FDD methods. The most commonly used fault detection methods in aerospace systems are explained in sections five, six, and seven with important references. Finally, a conclusion on the comprehensive review article is made.

3.2   Sensor Faults

A sensor's intended functioning is lost in the event of a partial or total failure [65]. Air vehicles may experience serious issues if a malfunctioning sensor is left unattended. It can reciprocate into significant incidents, taking the system outside the standard flight envelope. Faults may occur in a sensor while operating for internal or external reasons. Environmental impacts include temperature changes, humidity, and corrosion [66]. Calibration errors are another significant possibility of error. These are examples of external factors. However, internal factors like overheating and short circuits can also cause malfunction. Many fatal air accidents, such as the Ethiopian Air [67] and Lion Air [68] accidents, were caused by sensor malfunction, causing the autopilot to initiate inappropriate maneuvers and leading to catastrophes. One of the most critical aspects of fault detection is the timely identification of sensor defects, which guarantees data correctness and dependability [6]. In this subsection, we have collaborated on information from different sources for common sensor faults in aerospace systems. Table 3.1 lists the typical sensor and actuator faults in aerospace systems. Figures 3.3 and 3.4 show examples of different fault types.

3.2.1   Bias Fault

Sensor errors that are caused by bias are the most frequent ones. As a result, the sensor output is continuously skewed, either higher or lower than the actual value. Much recent research has been done to identify bias errors and to represent the faults [69] mathematically. Bias faults in the inertial measurement unit (IMU), which consists of the accelerometer and three-axis gyroscope, were evaluated using actual aircraft

Table 3.1 Common types of sensor and actuator faults in aircraft

Sensor faults	Actuator faults
Bias fault	Lock-in-place fault
Drift fault	Hard-over
Increased noise faults	Loss of effectiveness
Oscillatory faults	Float
Frozen sensor fault	
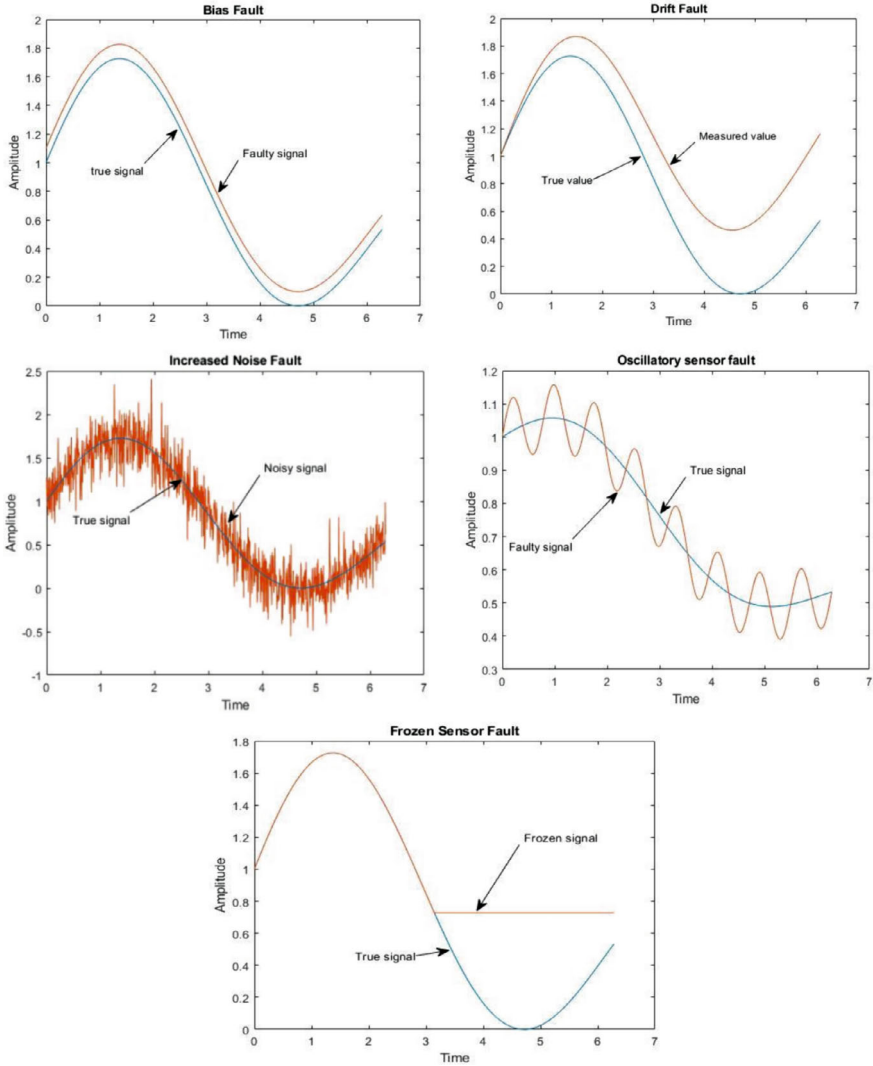
Fig. 3.3 Example of sensor fault types

data [70]. Lightweight MEMS sensors, prone to bias errors, are typically seen in UAVs [64]. A bias fault estimator for the accelerometer of an autonomous quadrotor was proposed in [71]. These malfunctions are hard to spot and might lead to many unforeseen issues. They even compromise the system's stability [72].
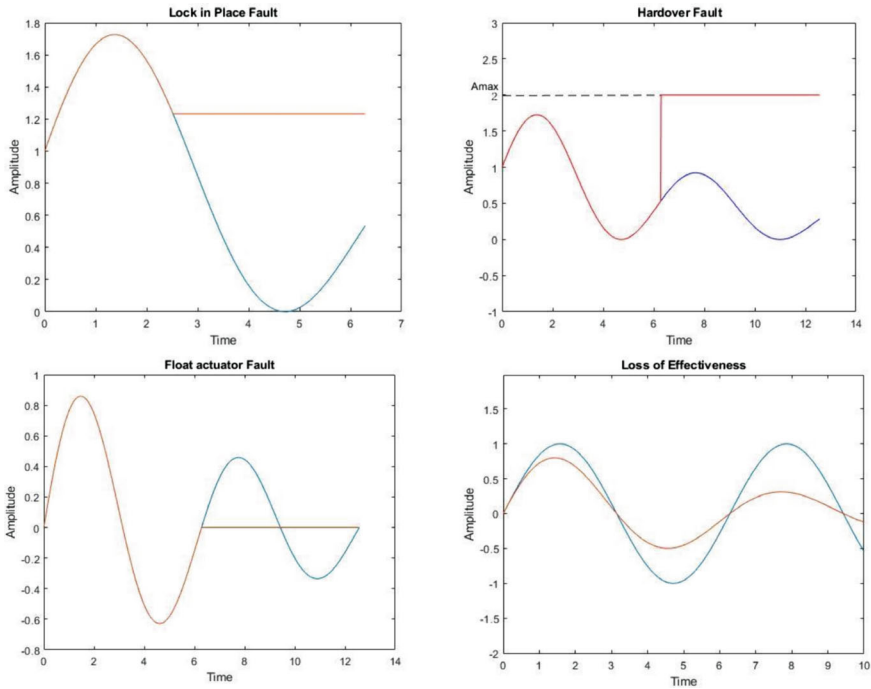
Fig. 3.4 Example of actuator fault types

3.2.2 Drift Fault

A slow shift in the real measurement over time is called a drift fault [73]. In [74], it was also referred to as "runaways," and the same was estimated in the air data sensor unit (ADS) of an aircraft. In most problems, the drift from the real value is considered linear. However, [75] provides information on the modeling of nonlinear drifts. Drift additive defects are described in a few studies as linearly deviating from the true value for a while before stabilizing at a constant bias [76]. The accelerometer readings are used in UAVs to calculate their velocity while in flight. The system's computed velocity and distance may differ significantly if the accelerometer has a drift fault [77]. These variances have an impact on the posture of the system and can cause it to lose control.

3.2.3 Increased Noise Faults

Random fluctuations in sensor readings that do not accurately reflect the signal or data being measured are referred to as noise. Most of the systems can accommodate

small amounts of sensor noise. However, a persistent rise in noise can seriously impair system performance [78]. In addition to internal noise, turbulence and wind gusts can contribute to the raised noise level in general aviation aircraft [79]. In [80], fault detection of ADS was conducted in the presence of turbulence on real aircraft data.

3.2.4 Oscillatory Faults

An oscillatory fault case is observed when there are abnormal oscillations that vary around a specific value instead of measuring the actual value. This type of fault case was used to test the FDD developed for aircraft and UAV sensors [81]. In [74], the oscillatory fault detection of ADS sensors in aircraft and other faults like bias and drift.

3.2.5 Frozen Sensor Faults

It is a situation where a sensor reading is stuck at a value or is unresponsive. This is the most severe fault case [82]. If undetected, it can severely affect the overall functionality of the system. It is also referred to as a stuck fault and can be considered as a complete failure of the sensor [83]. It is easier to detect frozen sensor faults compared to other faults, as these occur mostly due to sensor hardware failure [7].

3.3 Actuator Faults

Actuator malfunctions or defects are extremely important for controlling aircraft. It has previously resulted in a number of deadly commercial aircraft incidents. Typical actuator defect types include the following.

3.3.1 Lock in Place

When an actuator or control surface becomes immobile or fixed in one location, it loses its ability to move and react to commands. It may significantly impair the aircraft's ability to maneuver and maintain control [84]. A robust technique to diagnose lock-in-place actuator faults in the control surface was proposed in [85]. This fault is also referred to as jamming. In [86], an adaptive unknown input observer (UIO) method was implemented to identify and separate actuator failures in aircraft.

3.3.2 Hard-Over

Hard-over faults in actuators can be characterized by abrupt movement of the actuator to its maximum or minimum extreme positions irrespective of the control command provided by the controller [84]. Fault detection and localization of hard-over fault in the aileron of a fixed-wing was implemented in [87]. In addition, in-flight fault detection of hard-over fault in a fixed-wing aircraft was implemented using a neural network FDI scheme [88]. These faults are very critical and may lead to catastrophic incidents if not detected quickly.

3.3.3 Loss of Effectiveness

This describes a decrease in an actuator's or control system's capacity to operate as intended. It may not crash, but it performs worse, which affects its responsiveness and effectiveness. Rudder loss of effectiveness fault in a commercial aircraft control system was detected using an H-infinity filter [89]. With the same FDI method, loss of effectiveness in elevator and throttle was simulated [90]. A nonlinear geometric approach was used to detect aircraft actuator faults [91]. This fault appears gradually; hence, it is difficult to evaluate or detect.

3.3.4 Float

When the effector "floats" with zero moment and does not contribute to the control authority, this is known as float failure [84]. This could happen if the control surface cannot move freely as anticipated and stays close to a specific location. It impacts the aircraft's stability and control, including being locked or stuck in a specific trim setting.

3.4 Fault Detection Methods

There are numerous methods to detect and diagnose faults in sensors. These are generally classified into two categories: model-based and data-driven methods [92]. These are the most popular ways of classifying but are not limited to. Some books categorized it as signal-based, knowledge-based, and model-based [93]. A single signal and multiple signal-based classification was also used in studies [94]. Table 3.2 shows the most commonly used methods for fault detection in aerospace systems.

Table 3.2 Types of fault
detection methods

Model-based	Data-driven methods
State estimation	Neural networks
Parameter estimation	Fuzzy logic
Parity space approach	Statistical methods
Process identification	Frequency analysis
State and parameter estimation	Time–frequency analysis

3.5 Model-Based Methods

Model-based analysis uses mathematical models to examine input and output signals. It was first presented in [9], although it needs comprehensive processing details. It entails comparing the signal of the model system with the signal measured from the system, often known as "the residual," which contains a wealth of fault-related information. To determine whether the behavior of the model system is consistent with that of the real system, fault diagnostics use appropriate decision functions or rules. As such, it recognizes alterations and distinguishes malfunctions in the sensor system [10]. This approach is still important since it continues to pose problems and is still relevant in aircraft sensor fault diagnosis [11].

The nonlinear fault detection and recognition method performs exceptionally well in locating sensor failures inside the aircraft, quickly determining the precise moment at which faults arise [12]. This approach's theoretical foundations are generally well-established and primarily verified by simulation exercises [13]. Most research studies are completed after erroneous data is injected into actual or simulated data [14]. The model-based detection scheme may be impacted by uncertainties in the real system [15]. These are the few limitations of physical model-based methods. The following are some selected methods suggested for UAVs and aircraft.

3.5.1 State Estimation

Using system-measurable data to build an analytical model is the goal of the state estimation approach, sometimes referred to as the filter method. This model includes a system observer (detection filter) that isolates and identifies sensor problems by comparing observed and estimated values. This observer-based method is widely used in measuring and reconstructing measurement noise and sensor errors [16]. The general state estimation based on FDI is illustrated in Fig. 3.5. The popularly used ones are as follows:

Luenberger Observers

By tracking differences or inaccuracies between estimated states from the observer and measured states from sensors, Luenberger observers are used for fault detection

Fig. 3.5 The fault detection process using state estimation
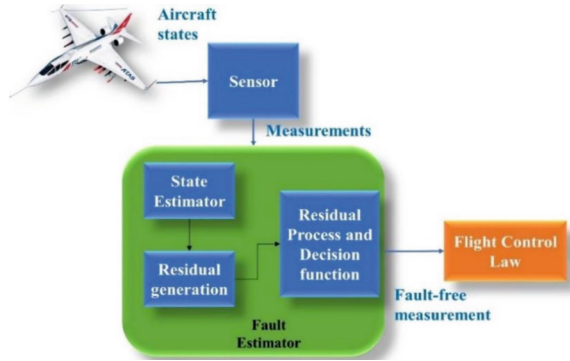
[17]. For the LTI system represented by

$$\dot{x} = Ax + Bu \tag{3.1}$$

$$y = Cx + Du \tag{3.2}$$

x is the system state vector and $\dot{x}$ represents the derivative of the state vector. A is the system matrix describing the dynamics of the states. B is the input matrix. u is the control input. C is the output matrix. D is the feedforward matrix. y is the output vector. The Luenberger observer is formulated as

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - \hat{y}) \tag{3.3}$$

$$\hat{y} = C\hat{x} + Du \tag{3.4}$$

For fault detection using the Luenberger observer, discrepancies between estimated states $(\hat{x})$ and actual measurements (y) are monitored. Fault detection using a Luenberger observer involves analyzing the residuals or error signals:

$$e = y - \hat{y} \tag{3.5}$$

Monitoring these residuals (e) allows for the detection of faults. In nonlinear systems, estimating states frequently entails linearizing the model along trajectories or around specific operational points. This modification makes it possible to apply established methods for linear systems. As a result of this methodology, the Extended Luenberger Observer (ELO) has been developed [18, 19]. Linearization would lose some characteristics; hence, a nonlinear Luenberger observer was proposed [20]. An adaptive version of the same came up later to deal with time-varying faults [21]. Sliding mode observers present an intriguing option as they enable the direct

estimation of faults that disrupt the sliding motion [22]. The benefits of sliding mode observers are their excellent dependability, resilience, and simplicity.

Kalman Filter

Although the Kalman filter [23] is mainly used for state prediction in systems, there are several techniques to modify it for fault detection, including residual analysis [24] and model augmentation [25]. Although there isn't a specific "Kalman filter for fault detection" in the conventional sense, several variations are available. One such method is to augment the fault as the unknown states of the Kalman filter. During the estimation, the fault vector is also estimated by employing the data from measured inputs and observations.

The augmentation model state equation is modified as

$$x_{k+1} = Ax_k + Bu_k + B_d d_k + B_f f_k \tag{3.6}$$

x_{k+1} is the state at time k + 1. d_k represents process noise. f_k represents the fault vector. The process noise and fault vector are estimated along with the state and can indicate the presence of faults. The equations for the Kalman filter itself, in its standard form for state estimation, involve prediction and update steps [23]. The extended Kalman filter (EKF) is applied to a nonlinear system after linearization about a point [26]. The Unscented Kalman Filter (UKF) does not require linearizing the model [27], in contrast to the EKF. By assessing the nonlinear model several times and estimating a Gaussian distribution for the state vector, it predicts the behavior of the system. Most recent research works utilize various adaptive Kalman filters for fault diagnosis [28, 29]. Particle Filtering is another method used for the estimation of faults, using a model-based approach where linearization of the state equation is not required [30].

3.5.2 Parameter Estimation

Parameter estimation is the process of estimating unknown parameters of the system with the help of available measurable quantities. In aircraft, the current state of the system can be identified using parameter estimation, and recognizing abnormalities in the estimates can be related to sensor or actuator faults [31]. Sometimes, the faults are considered as unknown parameters [32].

The parameter estimation procedure starts with a system analysis to determine how the model parameters and the outcome relate to each other. The system can then compute an output by obtaining real-time measurable data. The output obtained from the model parameters and the actual output are compared. A fault is identified if the system's estimate of the parameter is more significant than a predetermined confidence interval, signifying a notable variation in the parameter. It is essential to remember, though, that the sensor fault diagnosis filter of a nonlinear system, which is dependent on parameter estimate, may display problems with accuracy and

stability [33]. It has been used in offline procedures post-flight to detect sensor biases and flight path reconstruction [34].

3.5.3   Parity Space Approach

A parity vector is generated by collecting data from the output over a time frame. It's based on the concept of creating relationships between the inputs, outputs, and residuals of a system to detect and diagnose faults [35].

For the LTI system (1–2), the V-matrix is called the parity vector, which is the solution for the relation VH = 0. H is given by

$$H = \begin{bmatrix} C \ CA \vdots CA^s \end{bmatrix} \tag{3.7}$$

where s is the number of data points in the time frame chosen for analysis.

One major issue in fault diagnosis is the ability to distinguish and isolate defects separately [36]. The residual, a measurement of divergence from expected behavior, exhibits a regular pattern in the parity space for a particular sensor defect in cases where direct redundancy is used. It is easy to isolate faults because of this consistency. Nevertheless, finding a particular matrix (V-matrix) that restricts the impact of individual components (e.g., actuators or sensors) to a single direction in the parity space is usually difficult in temporal analytical redundancy scenarios [37]. Fault isolation, hence, becomes less straightforward and more complicated in these situations. A continuous parity space method was explored in [38].

3.6   Data-Driven Methods

Database-driven and knowledge-based techniques are among the less well-known technologies used for defect detection in aircraft and UAVs. Recently, this particular class of analytically redundant algorithms with large datasets has attracted increasing attention. In order to identify and diagnose sensor failures without depending on an exact mathematical model, this method uses an expert system. The expert system consists of a reasoning module that stores rules based on historical data and fuzzy relationships. In the context of sensor fault detection, the rules will be the relationships to examine sensor behavior. By incorporating the past data and the stored reasoning knowledge, any sensor malfunction can be detected. Data-driven methods are used when the system is highly nonlinear.

Fig. 3.6 The fault detection process using a neural network
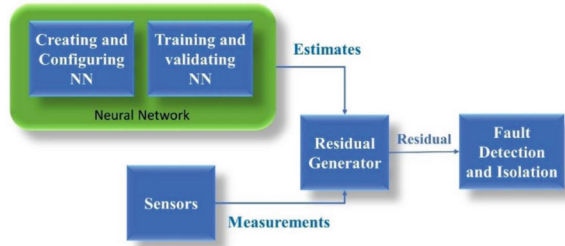
3.6.1 Neural Networks

Structures that mimic the human nervous system are called neural networks. It is made up of weights, activation functions, and nodes [39, 40]. Neurons store the associations between inputs and outputs as weights [41, 42]. As a result, these techniques do not call for system modeling on a mathematical level. The features of the system are stored in the weights of the neurons. The weights are updated during the learning process, and the stored weights are utilized to predict the states and further fault detection [43]. The neuron's structure, the learning algorithm, and the activation function are carefully chosen, or new methods are based on their variations [44]. The neural network has two phases of operation: training and prediction. During the training session, a large dataset is used to update the weights according to the input and output combinations. Together, these weights will imitate the nature of the system. Hence, during the prediction phase, with the help of learned weights, the neural network predicts the output. The estimated and measured data are compared, and a residual is produced for the fault detection application [45]. This residual signal is evaluated in order to identify sensor malfunctions. The fault detection process using a neural network is shown in Fig. 3.6.

Multilayer perceptron (MLP) [46] and radial basis function neural networks (RBFNN) [47] are the most commonly used neural networks in fault diagnosis of aerospace systems. To improve accuracy, a convolutional neural network (CNN) has recently been proposed to evaluate faults in ADS sensors [48]. Neural networks have become more popular recently due to their adaptability in real-time systems and modeling of nonlinearities. Some disadvantages of NN are the requirement for a large number of datasets to be available solely for training, overfitting, etc.

3.6.2 Fuzzy Logic

Developed as a tool to define and process inaccuracies and imprecise occurrences, fuzzy logic is a rule-based system [49]. Fuzzy logic and algorithms have been used to solve numerous control system issues [50]. Fuzzy systems, unlike other algorithms,

are intended to produce optimal or reasonable conclusions in the presence of imprecise and uncertain data [51]. It diagnoses using fuzzy rules rather than an accurate model or correlations between input and output.

Sensor fault detection and isolation using non-model-based fuzzy logic has been carried out in work [52]. Actuator or control surface malfunction estimation was also carried out using fuzzy decision logic combined with EKF [52]. A fuzzy assessment method was proposed [53] to detect surface failure in high-performance aircraft such as the F/A-18. In order to adapt many models for actuator failure detection, fuzzy logic was used in [54]. This modification allowed the technique to identify faster and reduce false alarms.

3.6.3   Supervised Machine Learning

When it comes to machine learning, support vector machines (SVMs) are supervised max-margin models that look at data for regression and classification along with related learning strategies. It comes under the statistical methods. They can handle both linear and nonlinear data and are practical tools for pattern identification [55]. The underlying rules guiding the system's behavior are taught to it through repeated training on simple data. Any departure from the accepted standard model during sensor operation is classified as an abnormality. Multiple errors can be found with a diagnostic approach that uses an SVM [56, 57]. This approach uses the ability to recognize departures from typical behavior to pinpoint different types of system errors.

For online fault detection, a Least Square-SVM(LS-SVM) has been implemented in the past [58]. There are other statistical methods used for fault detection [59]. Due to its superior classification performance, SVM is researched; however, this approach requires more complex samples and has data imbalance, which will increase expenditures [60].

3.7   Hybrid Method

A combination of detection, identification, and estimation techniques, each with advantages and disadvantages, is required for effective sensor defect diagnosis. These methods can strengthen one another when combined wisely and skillfully, improving fault detection. In order to enhance the diagnostic, a hybrid FDI for aircraft IMU was implemented, combining hardware-based FDI with an EKF [95]. In [96], a neural network was used to detect faults in sensors and actuators in small UAVs. Here, the weighting matrix was updated with the help of the Kalman filter. The computation time was improved in this way. A collaboration of fuzzy and Neural networks was used for a small helicopter [97]. A sensor FDI for UAV was presented in [98] using the Hammerstein–Wiener model and MLP neural network. The inputs to the NN are each

from the Hammerstein–Wiener model here. Similarly, a handful of research studies combined two methods. Sensor faults can be identified more efficiently by combining suitable detection, identification, and estimating algorithms. Hybrid techniques are widely preferred in recent studies.

3.8   Conclusion

The study provides an in-depth analysis of several fault detection methods that are essential to aerospace systems security and reliability. The examination encompasses a wide range of techniques designed expressly to discover issues in sensors and actuators, acknowledging the critical role these components play in maintaining operational integrity.

Sensor and actuator malfunction detection techniques such as model-based, data-driven, artificial intelligence-based, and signal processing algorithms are evaluated in this work using substantial bibliographical data. Efficiency, fault detection time, false alarm rate, and computing complexities are investigated for the fault detection techniques. Additionally, several types of actuator and sensor failures were covered. Table 3.3 lists the various fault types that have been researched in the literature, along with the techniques used to find them, some of which are covered in Sects. 3.4, 3.5, and 3.6. The reported benefits and disadvantages of each approach are also listed.

The significant drawback of sensor and actuator fault detection in aerospace systems is that the existing algorithms are evaluated in lab environments and mostly from simulated data. The applicability of the approaches employed still requires much study because it is difficult to foresee every possible circumstance in a laboratory setting. In the rapidly developing field of aerospace systems, unexpected obstacles in real-world scenarios continue to surprise researchers and designers.

Our future work will involve online fault detection in aircraft sensors using modified neuron models and an in-depth study on fault-tolerant control methods in aerospace systems.

Table 3.3 Fault types and detection methods employed in aerospace systems

Fault detection methods	Fault type	Advantages	Drawbacks
Luenberger observers	Loss of effectiveness, Locking of actuators [88] Oscillatory control surface [85]	• Reduced false alarm • Simultaneous fault detection	• Computational complexity • Unmodelled disturbances are difficult to estimate
Kalman filters	Frozen airspeed/GPS sensor [24] Oscillation, increased noise, and runaway in AOA sensor [25] Float actuator failure [26] Bias in ADS/IMU [29]	• Fast fault detection • Independence from aircraft dynamics • Gaussian measurement and process noise are considered	• Linearization of the model may cause loss of information • Gaussian noise assumptions are not always valid
Particle filter	Bias in control surface/IMU sensor [30]	• Nonlinear model considered • Non-Gaussian noise assumptions	• Very high computational complexity • Statistical distribution of noise required
Sliding mode observer	Bias in control surface/IMU sensor [22]	• Quicker convergence •Fault estimation is possible	• Computational complexity • Tuning difficulty
H-infinity filter	Loss of effectiveness in rudder [89] Loss of effectiveness in rudder and elevator [90]	• The magnitude of the fault can be estimated • Faster convergence	• Sensitive to initial noise information • Works well with linear models
Parameter estimation	Increased noise fault in actuators [31] Bias fault in IMU and ADS [34]	• Allows online detection of faults • Efficient for structure damage detection	• Less efficient in fault identification • Less accurate
Neural networks	Hard-over actuator fault [88] Bias in accelerometers sensors [47] Bias in ADS [48]	• The mathematical model is not required	• Learning convergence is not guaranteed • Learning time is high for online training
Fuzzy logic	Loss of effectiveness in control surface [52] Stabilizer and rudder lock-in fault [53]	• Lower computational burden	• Highly system dependent
SVM	Engine separation [56] Control actuator fault [57]	• Linearity in predictions	• Training data is required • Only the known class of faults can be estimated

References

1. Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations, 1959–2022, Boeing Commercial Airplanes, September 2023. Accessed 15 December 2023. https://www.faa.gov/sites/faa.gov/files/2023-10/statsum_summary_2022.pdf

2. Belcastro, C., Foster, J.: Aircraft loss-of-control accident analysis. In: AIAA Guidance, Navigation, and Control Conference. https://doi.org/10.2514/6.2010-8004 (2010)

3. Belcastro, C., Jacobson, S.: Future integrated systems concept for preventing aircraft loss-of-control accidents. In: AIAA Guidance, Navigation, and Control Conference (2010). https://doi.org/10.2514/6.2010-8142

4. Marzat, J., Piet-Lahanier, H., Damongeot, F., Walter, E.: Model-based fault diagnosis for aerospace systems: a survey. Proc. Inst. Mech. Eng., Part G: J. Aerosp. Eng. **226**(10), 1329–1360 (2012). https://doi.org/10.1177/0954410011421717

5. Ding, S.: Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools. Berlin/Heidelberg, Germany, Springer (2013). https://doi.org/10.1007/978-1-4471-4799-2

6. Allerton, D., Jia, H.: A review of multisensor fusion methodologies for aircraft navigation systems. J. Navig. **58**(3), 405–417 (2005). https://doi.org/10.1017/S0373463305003383

7. Li, D., Wang, Y., Wang, J., Wang, C., Duan, Y.: Recent advances in sensor fault diagnosis: A review. Sens. Actuators A: Phys. **111990** (2020). https://doi.org/10.1016/j.sna.2020.111990

8. Chow, E., Willsky, A.: Analytical redundancy and the design of robust failure detection systems. IEEE Trans. Autom. Control **29**(7), 603–614 (1984). https://doi.org/10.1109/tac.1984.1103593

9. Beard, R.V.: Failure accommodation in linear systems through self-reorganization. Ph.D Thesis, Department of Aeronautics Astronautics, MIT, Cambridge, MA (1971)

10. Mulder, J.A., Chu, Q.P., Sridhar, J.K., Breeman, J.H., Laban, M.: Non-linear aircraft flight path reconstruction review and new advances. Prog. Aerosp. Sci. **35**(7), 673–726 (1999). https://doi.org/10.1016/s0376-0421(99)00005-6

11. Prabhu, S., Anitha, G.: Robust fault detection and diagnosis of primary air data sensors in the presence of atmospheric turbulence. Aeronaut. J. **127**(1317), 2023–2046 (2023)

12. Ansari, A., Bernstein, D.S.: Aircraft sensor fault detection using state and input estimation. In: 2016 American Control Conference (ACC), pp. 5951−5956. Boston, MA, USA (2016). https://doi.org/10.1109/ACC.2016.7526603

13. He, Q., Zhang, W., Lu, P., Liu, J.: Performance comparison of representative model-based fault reconstruction algorithms for aircraft sensor fault detection and diagnosis. Aerosp. Sci. Technol. **105649** (2019). https://doi.org/10.1016/j.ast.2019.105649

14. Prabhu, S., Anitha, G.: An innovative analytic redundancy approach to air data sensor fault detection. Aeronaut. J. **124**(1273), 346–367 (2020)

15. Devore, M., Gandhi, N., Bateman, A.J.: Virtual redundancy for safety assurance in the presence of sensor failures. In: AIAA Guidance, Navigation, and Control Conference (2016). https://doi.org/10.2514/6.2016-0376

16. Marco, A., Massimiliano, M., Immacolata, N., Federico, C., Adolfo, S.: An SFDI observer–based scheme for a general aviation aircraft. Int. J. App. Math. Comp. Sci. **25**(1), 149–158 (2015). https://doi.org/10.1515/amcs-2015-0011

17. Alcorta Garcia, E., Frank, P.M.: Deterministic nonlinear observer-based approaches to fault diagnosis: a survey. Control. Eng. Pract. **5**(5), 663–670 (1997)

18. Zeitz, M.: The extended Luenberger observer for nonlinear systems. Syst. Control Lett. **9**(2), 149–156 (1987)

19. Nejjari, F., Puig, V., Giancristofaro, L., Koehler, S.: Extended Luenberger observer-based fault detection for an activated sludge process. In: Proceedings of the 17th IFAC World Congress, pp. 9725–9730. Seoul, Korea (2008)

20. Adjallah, K., Maquin, D., Ragot, J.: Nonlinear observer-based fault detection. In: Proceedings of the 3rd IEEE Conference on Control applications, pp. 1115–1120. Glasgow, Scotland (1994)

21. Teng, J., Li, C., Feng, Y., Yang, T., Zhou, R., Sheng, Q.Z.: Adaptive observer based fault tolerant control for sensor and actuator faults in wind turbines. Sens. **21**(24), 8170 (2021). https://doi.org/10.3390/s21248170

22. Tan, C.P., Edwards, C.: Sliding mode observers for robust fault detection & reconstruction. IFAC Proc. Vol. **35**(1), 347–352 (2002). https://doi.org/10.3182/20020721-6-es-1901.00789
23. Kalman, R.E.: A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
24. Rudin, K., Ducard, G.J.J., Siegwart, R.Y.: A sensor fault detection for aircraft using a single Kalman filter and hidden Markov models. In: 2014 IEEE Conference on Control Applications (CCA) (2014). https://doi.org/10.1109/cca.2014.6981464
25. Van Eykeren, L., Chu, Q.P., Mulder, J.A.: Sensor fault detection and isolation using adaptive extended Kalman filter. IFAC Proc. Vol. **45**(20), 1155–1160 (2012). https://doi.org/10.3182/20120829-3-mx-2028.00195
26. Gao, J., Zhang, Q., Chen, J.: EKF-based actuator fault detection and diagnosis method for tilt-rotor unmanned aerial vehicles. Math. Probl. Eng. **2020**, 8019017 (2020)
27. Liu, L., Ma, Y., Xu, B., Xiang, C., Yang, X.: Fault detection and isolation based on UKFs for a novel ducted fan UAV. In: Proceedings of the 2016 IEEE International Conference on Aircraft Utility Systems (AUS), pp. 212–218. Beijing, China (2016)
28. Saini, V.K., Maity, A.: Weighted adaptive decentralized Kalman filters for fault tolerance. In: AIAA Scitech 2020 Forum (2020). https://doi.org/10.2514/6.2020-0949
29. Lu, P., Van Eykeren, L., van Kampen, E., de Visser, C.C., Chu, Q.P.: Adaptive three-step Kalman filter for air data sensor fault detection and diagnosis. J. Guid. Control. Dyn. **39**(3), 590–604 (2016). https://doi.org/10.2514/1.g001313
30. Iglésis, E., Horri, N., Brusey, J., Dahia, K., Piet-Lahanier, H.: Simultaneous actuator and sensor faults estimation for aircraft using a jump-markov regularized particle filter. In: 2021 IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit (Romulus), pp. 1–10. MI, USA (2021). 9486593. https://doi.org/10.1109/ICPHM51084.2021
31. Korsun, O.N., Om, M.H., Latt, K.Z., Stulovskii, A.V.: Real-time aerodynamic parameter identification for the purpose of aircraft intelligent technical state monitoring. Procedia Comput. Sci. **103**, 67–74 (2017). https://doi.org/10.1016/j.procs.2017.01.014
32. Grauer, J.A.: Real-time parameter estimation using output error. In: AIAA Atmospheric Flight Mechanics Conference (2014). https://doi.org/10.2514/6.2014-2556
33. Kamali, C., Ozger, E.: Limitations of flight path reconstruction techniques. Sādhanā **44**, 32 (2019). https://doi.org/10.1007/s12046-018-1019-3
34. De Oliveira Maciel, B.C., Sandoval Góes, L.C., Hemerly, E.M., Brasil Neto, N.S.: Flight path reconstruction and parameter estimation using output-error method. Shock. Vib. **13**(4–5), 379–392 (2006). https://doi.org/10.1155/2006/960391
35. El-mezyani, T., Dustegor, D., Srivastava, S.: Parity space approach for enhanced fault detection and intelligent sensor network design in power systems. In: IEEE PES General Meeting, PES (2010). https://doi.org/10.1109/PES.2010.5590102
36. Patton, R.J., Chen, J.: A review of parity space approaches to fault diagnosis. IFAC Proc. Vol. **24**(6), 65–81, ISSN 1474–6670 (1991). https://doi.org/10.1016/S1474-6670(17)51124-6
37. Magni, J.F., Mouyon, P.: On residual generation by observer and parity space approaches. IEEE Trans. Autom. Control **39**(2), 441–447 (1994)
38. Medvedev, A.: Fault detection and isolation by a continuous parity space method. Automatica **31**(7), 1039–1044 (1995)
39. Garbarino, L., Zazzaro, G., Genito, N., Fasano, G., Accardo, D.: Neural network based architecture for fault detection and isolation in air data systems. In: 2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC) (2013). https://doi.org/10.1109/dasc.2013.6719624
40. Jager, G., Zug, S., Brade, T., Dietrich, A., Steup, C., Moewes, C., Cretu, A.M. (2014). Assessing neural networks for sensor fault detection. In: 2014 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). https://doi.org/10.1109/civemsa.2014.6841441
41. Heo, S., Lee, J.H.: Fault detection and classification using artificial neural networks. IFAC-PapersOnLine **51**(18), 470–475 (2018). https://doi.org/10.1016/j.ifacol.2018.09.380
42. Rajakarunakaran, S., Venkumar, P., Devaraj, D., Rao, K.S.P.: Artificial neural network approach for fault detection in rotary systems. Appl. Soft Comput. **8**(1), 740–748 (2008). https://doi.org/10.1016/j.asoc.2007.06.002

43. Vemuri, A.T., Polycarpou, M.M., Diakourtis, S.A.: Neural network based fault detection in robotic manipulators. IEEE Trans. Robot. Autom. **14**(2), 342–348 (1998). https://doi.org/10.1109/70.681254

44. Mrugalski, M.: Advanced Neural Network-Based Computational Schemes for Robust Fault Diagnosis. Springer, Berlin/Heidelberg, Germany (2014)

45. Chen, Y.M., Lee, M.L.: Neural networks-based scheme for system failure detection and diagnosis. Math. Comput. Simul **58**(2), 101–109 (2002)

46. Xiao-Wei, W.: A Multilayer Perceptron Neural Network Model for UAV Sensor Fault Detection, pp. 22–26 (2021). https://doi.org/10.1109/ICISCAE52414.2021.9590669

47. Samy, I., Postlethwaite, I., Gu, D.W.: Sensor fault detection and accommodation using neural networks with application to a nonlinear unmanned air vehicle model. Proc. IMechE, Part G: J. Aerosp. Eng. **224**(4), 437–447 (2010)

48. Dong, Y., Wen., Zhang, Y., Ai, J.: Deep neural networks-based air data sensors fault detection for aircraft. In: 2021 33rd Chinese Control and Decision Conference (CCDC), pp. 442−447. Kunming, China (2021). https://doi.org/10.1109/CCDC52312.2021.9601367

49. Zadeh, L.A.: Fuzzy probabilities and their role in decision analysis. IFAC Proc. **15**(1), 15−21 (1982)

50. Tong, R.M.: A control engineering review of fuzzy systems. Automatica **13**(6), 559–569 (1977). https://doi.org/10.1016/0005-1098(77)90077-2

51. Sugeno, M.: An introductory survey of fuzzy control. Inf. Sci. **36**(1–2), 59–83 (1985). https://doi.org/10.1016/0020-0255(85)90026-x

52. Savanur, S., Patel, A.: Sensor/control surface fault detection and reconfiguration using fuzzy logic. Def. Sci. J. **60**(1), 76–86 (2010). https://doi.org/10.14429/dsj.60.114

53. Raza, H., Ioannou, P., Youssef, H.M.: Surface failure detection for F/A-18 aircraft using neural networks and fuzzy logic. In: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), vol. 5, pp. 3363–3368. Orlando, FL, USA (1994). https://doi.org/10.1109/ICNN.1994.374776

54. Kim, S., Choi, J., Kim, Y.: Fault detection and diagnosis of aircraft actuators using fuzzy-tuning IMM filter. Aerosp. Electron. Syst., IEEE Trans. On. **44**, 940–952 (2008). https://doi.org/10.1109/TAES.2008.4655354

55. Suykens, J.A.K.: Support vector machines: a nonlinear modelling and control perspective. Eur. J. Control. **7**(2–3), 311–327 (2001). https://doi.org/10.3166/ejc.7.311-327

56. Zhao, Y.-P., Wang, J.-J., Li, X.-Y., Peng, G.-J., Yang, Z.: Extended least squares support vector machines with applications to fault diagnosis of aircraft engines. ISA Trans. (2019). https://doi.org/10.1016/j.isatra.2019.08.036

57. Grehan, J., Ignatyev, D., Zolotas, A.: Fault detection in aircraft flight control actuators using support vector machines. Machines **11**, 211 (2023). https://doi.org/10.3390/machines11020211

58. Khawaja, T.S., Georgoulas, G., Vachtsevanos, G.: An efficient novelty detector for online fault diagnosis based on Least Squares Support Vector Machines. In: 2008 IEEE AUTOTESTCON (2008). https://doi.org/10.1109/autest.2008.4662612

59. Samara, P.A., Fouskitakis, G.N., Sakellariou, J.S., Fassois, S.D.: A statistical method for the detection of sensor abrupt faults in aircraft control systems. IEEE Trans. Control Syst. Technol. **16**(4), 789–798 (2008). https://doi.org/10.1109/tcst.2007.903109

60. Deng, F., Guo, S., Zhou, R., Chen, J.: Sensor multi fault diagnosis with improved support vector machines. IEEE Trans. Autom. Sci. Eng. **14**(2), 1053–1063 (2017). https://doi.org/10.1109/tase.2015.2487523

61. Trapani, N., Longo, L.: Fault detection and diagnosis methods for sensors systems: a scientific literature review. IFAC-PapersOnLine. **56**(2), 1253–1263, ISSN 2405–8963 (2023). https://doi.org/10.1016/j.ifacol.2023.10.1749

62. Zhou, J., Zhu, Y., Shao, Z., Dong, Z., Yang, C.: Identification-based sensor and actuator fault diagnosis for industrial control systems and its application to HTR-PM. Control. Eng. Pract. 131, 105402, ISSN 0967–0661 (2023). https://doi.org/10.1016/j.conengprac.2022.105402

63. Soni, R., Mehta, B.: Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies. Eng. Fail. Anal. **128**, 105634, ISSN 1350–6307 (2021). https://doi.org/10.1016/j.engfailanal.2021.105634.

64. Shakhatreh, H., Sawalmeh, A.H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N.S., Khreishah, A., Guizani, M.: Unmanned aerial vehicles (UAVs): a survey on civil applications and key research challenges. IEEE Access **7**, 48572–48634 (2019)

65. Isermann, R.: Fault-Diagnosis Systems from Fault Detection to Fault Tolerance, pp. 28. Springer Berlin, Heidelberg (2006). https://doi.org/10.1007/3-540-30368-5

66. Bayar, N., Darmoul, S., Hajri-Gabouj, S., Pierreval, H.: Fault detection, diagnosis and recovery using artificial immune systems: a review. Eng. Appl. Artif. Intell. **46**, 43–57 (2015). https://doi.org/10.1016/j.engappai.2015.08.00

67. The Federal Democratic Republic of Ethiopia Ministry of Transport and Logistics.: Aircraft accident investigation Bureau Investigation report on the accident to the B737-MAX8 REG.ET-AVJ operated by Ethiopian Airlines (2022). https://bea.aero/fileadmin/user_upload/ET_302__B737-8MAX_ACCIDENT_FINAL_REPORT.pdf

68. Aircraft Accident Investigation Report. PT. Lion Airlines Boeing 737 (MAX); PK-LQP Tanjung Karawang, West Java, Republic of Indonesia 29 October 2018" (PDF). National Transportation Safety Committee. Retrieved 15 October 2023.

69. Joshi, S.M.: On Identifiability of Bias-Type Actuator-Sensor Faults in Multiple-Model-Based Fault Detection and Identification (2012)

70. Jategaonkar, R. V.: Flight Vehicle System Identification: A Time Domain Methodology. American Institute of Aeronautics and Astronautics (2006)

71. Avram, R.C., Zhang, X., Campbell, J., Muse, J.: IMU sensor fault diagnosis and estimation for quadrotor UAVs. IFAC-Papers OnLine **48**(21), 380–385 (2015). https://doi.org/10.1016/j.ifacol.2015.09.556

72. He, X., Wang, Z., Liu, Y., Qin, L., Zhou, D.: Fault-tolerant control for an internet-based three-tank system: accommodation to sensor bias faults. IEEE Trans. Industr. Electron. **64**(3), 2266–2275 (2017). https://doi.org/10.1109/tie.2016.2623582

73. Yang, J., Chen, Y., Zhang, L., Sun, Z.: Fault detection, isolation, and diagnosis of self-validating multifunctional sensors. Rev. Sci. Instrum. **87**(6), 065004 (2016). https://doi.org/10.1063/1.4954184

74. Van Eykeren, L., Chu, Q.P.: Sensor fault detection and isolation for aircraft control systems by kinematic relations. Control. Eng. Pract. **31**, 200–210 (2014). https://doi.org/10.1016/j.conengprac.2014.02

75. Balaban, E., Saxena, A., Bansal, P., Goebel, K.F., Curran, S.: Modeling, detection, and disambiguation of sensor faults for aerospace applications. IEEE Sens. J. **9**(12), 1907–1917 (2009). https://doi.org/10.1109/jsen.2009.2030284

76. Samy, I., Postlethwaite, I., Gu, D.-W.: Survey and application of sensor fault detection and isolation schemes. Control. Eng. Pract. **19**(7), 658–674 (2011). https://doi.org/10.1016/j.conengprac.2011.03.002

77. Wang, C., Liu, W., Meng, M.Q.H.: A denoising and drift-control approach for UAV trajectory tracking. In: 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014) (2014). https://doi.org/10.1109/robio.2014.7090582

78. Zheng, Y., Li, S., Xing, K., Zhang, X.: A novel noise reduction method of UAV magnetic survey data based on CEEMDAN, permutation entropy, correlation coefficient and wavelet threshold denoising. Entropy (Basel). **23**(10), 1309 (2021). https://doi.org/10.3390/e23101309.PMID:34682033;PMCID:PMC8534471

79. Bonfè, M., Castaldi, P., Geri, W., Simani, S.: Fault detection and isolation for on-board sensors of a general aviation aircraft. Int. J. Adapt. Control Signal Process. **20**(8), 381–408 (2006). https://doi.org/10.1002/acs.906

80. Lu, P., van Kampen, E.J., de Visser, C., Chu, Q.: Air data sensor fault detection and diagnosis in the presence of atmospheric turbulence: theory and experimental validation with real flight data. IEEE Trans. Control Syst. Technol. 1–9 (2020). https://doi.org/10.1109/tcst.2020.3025725

81. Silva, B., Callaham, J., Jonker, J., Goebel, N., Klemisch, J. McDonald, D., Hicks, N., Kutz, J, Brunton, S., Aravkin, A.: Physics-informed machine learning for sensor fault detection with flight test data (2020)

82. Ducard, G.: UAV sensor fault accommodation. In: Valavanis, K.P., Vachtsevanos, G.J. (eds.) Handbook of Unmanned Aerial Vehicles. Springer, Cham (2021). https://doi.org/10.1007/978-3-319-32193-6_170-1

83. Jan, S.U., Lee, Y.-D., Shin, J., Koo, I.: Sensor fault classification based on support vector machine and statistical time-domain features. IEEE Access **5**, 8682–8690 (2017). https://doi.org/10.1109/ACCESS.2017.2705644

84. Boskovic, J., Mehra, R.: Failure detection, identification and reconfiguration in flight. Control (2003). https://doi.org/10.1007/3-540-45737-2_5

85. Cieslak, J., Efimov, D., Zolghadri, A., Gheorghe, A., Goupil, P., Dayre, R.: A method for actuator lock-in-place failure detection in aircraft control surface servo-loops. IFAC Proc. Vol. **47**(3), 10549–10554 (2014). https://doi.org/10.3182/20140824-6-za-1003.02158

86. Wang, D., Niu, X., Wu, Z., Yao, Y.: Diagnosis of actuator lock-in-place for flight control systems. In: 2006 Chinese Control Conference Harbin, China, pp. 1315−1320 (2006). https://doi.org/10.1109/CHICC.2006.280648

87. Ignatyev, D.I., Shin, H.S., Tsourdos, A.: Two-layer fault detection for incremental flight control of fixed-wing UAV. In: 2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS), pp. 227−236. Cranfield, UK (2019). https://doi.org/10.1109/REDUAS47371.2019.8999692

88. Azam, M., Pattipati, K.R., Allanach, J., Poll, S., Patterson-Hine, A.: In-flight fault detection and isolation in aircraft flight control systems. In: IEEE Aerospace Conference, pp. 3555−3565 (2005)

89. Wu, N.E., Zhou, K., Salomon, G.: Control reconfigurability of linear time-invariant systems. Automatica **36**(11), 1767–1771 (2000). https://doi.org/10.1016/s0005-1098(00)00080-7

90. Szászi, I., Balas, G., Bokor, J., Marcos, A.: Linear parameter-varying detection filter design for a boeing 747–100/200 aircraft. J. Guid. Control. Dyn.-J GUID CONTROL DYNAM. **28**, 461–470 (2005). https://doi.org/10.2514/1.6689

91. Meskin, N., Jiang, T., Sobhani, E., Khorasani, K., Rabbath, C.A.: Nonlinear geometric approach to fault detection and isolation in an aircraft nonlinear longitudinal model. In: 2007 American Control Conference, pp. 5771−5776. New York, NY, USA (2007) https://doi.org/10.1109/ACC.2007.4282876

92. Puchalski, R., Giernacki, W.: UAV fault detection methods. State-of-the-Art. Drones. **6**, 330 (2022). https://doi.org/10.3390/drones6110330

93. Isermann, R.: Fault-Diagnosis Applications. Springer, Berlin/Heidelberg, Germany (2011)

94. Amorim, A.: Architecture of the battery fault detection system of the AUV Antares. Internal report; College of Technology of Jahu: Jahu, Brazil (2015)

95. Kim, S., Jung, I., Kim, Y., Park, C.: Hybrid fault detection and isolation techniques for aircraft inertial measurement sensors. In: AIAA Guidance, Navigation, and Control Conference and Exhibit (2004). https://doi.org/10.2514/6.2004-5419

96. Abbaspour, A., Aboutalebi, P., Yen, K.K., Sargolzaei, A.: Neural adaptive observer-based sensor and actuator fault detection in nonlinear systems: application in UAV. ISA Trans. **67**, 317–329 (2017)

97. Zhang, Q., Wang, X., Xiao, X., Pei, C.: Design of a fault detection and diagnosis system for intelligent unmanned aerial vehicle navigation systems. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. **233**, 2170–2176 (2019)

98. Ouadine, A.Y., Mjahed, M., Ayad, H., El Kari, A.: UAV quadrotor fault detection and isolation using artificial neural network and Hammerstein-Wiener model. Stud. Inform. Control **29**, 317–328 (2020)

Chapter 4
Predictive Model for Energy Absorption Capacity of Geopolymer Concrete

Dang-Thach Nguyen, Chi-Trung Nguyen, Diu-Huong Nguyen, and Ngoc-Thanh Tran

Abstract Geopolymer concrete (GPC) stands as a promising alternative to traditional Portland cement-based concrete in the pursuit of eco-friendly and sustainable construction materials. One of the key properties of GPC is its energy absorption capacity, which denotes the energy absorbed by the material just before failure. However, a standardized method for estimating the energy absorption capacity (EAC) of GPC is lacking, leading most studies to rely on proposed models for conventional concrete to predict this attribute. This study aims to address this gap by developing two machine learning models, artificial neural network (ANN) and random forest (RF), to predict the energy absorption capacity of GPC. The models were trained using data from 90 collected test results. Thirteen input variables were considered, relating to aggregates, alkaline solutions, water content, aluminosilicate material, silica fume, curing conditions, age, compressive strength, fiber type, fiber volume content, notch depth ratio, and maximum size of aggregate. The output variable was the EAC of GPC. The outcomes of the estimation demonstrated that the proposed RF model offered high accuracy and reliability in estimating the energy absorption capacity across both the training and testing sets. Conversely, the ANN model displayed slightly lower accuracy in the testing set. Sensitivity analysis revealed that the notch depth ratio exerted the most significant influence on the energy absorption capacity, while curing temperature emerged as the less critical variable.

Keywords Geopolymer concrete · Energy absorption capacity · Machine learning model · Sensitivity analysis

D.-T. Nguyen · C.-T. Nguyen · D.-H. Nguyen · N.-T. Tran (✉)
Institute of Civil Engineering, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam
e-mail: ngocthanh.tran@ut.edu.vn

59

4.1   Introduction

There is an urgent need to reduce CO_2 emissions from cement production, as this process currently emits approximately 4.8 billion tons of CO_2 annually [1]. Cement stands as a primary component of concrete, the most widely used construction material [2]. To mitigate cement production's environmental impact, there's a necessity for an alternative material. Geopolymer concrete (GPC) emerges as a promising option due to its use of alkali-activated binders instead of cement [2–5]. These binders are derived from aluminosilicate source materials, often industrial waste like ground granulated blast furnace slag, fly ash, rice husk ash, and metakaolin [6]. While GPC exhibits environmentally friendly traits, its mechanical properties must meet high standards for practical structural applications.

Given that energy absorption capacity (EAC), defined as the energy absorbed just before material failure, stands as a crucial property, numerous studies have investigated the fracture energy of GPC [7–13]. Ding et al. [7] demonstrated an increase in GPC's energy absorption capacity with a rise in compressive strength from 30 to 70 MPa. Similarly, Ding et al. [8] noted increased energy absorption capacity with higher alkali concentration and aluminosilicate material content, but a decrease with rising water/binder ratio. Liu et al. [9] concluded that energy absorption capacity initially decreased as silica fume content rose from 0 to 10%, but then increased between 10 to 30%. Xie et al. [10] highlighted the influence of notch depth ratio on GPC's energy absorption capacity. Mousavinejad and Gashti [11] reported an increase in energy absorption capacity with concrete age. Wang et al. [12] indicated increased energy absorption with higher fly ash replacement ratios. Conversely, Pan et al. [13] observed decreased energy absorption with increased curing time from 18 to 168 h.

These studies have demonstrated the significant influence of factors such as compressive strength, binders, notch depth ratio, concrete age, aggregates, and curing time on GPC's energy absorption capacity. However, accurately predicting this capacity remains a considerable challenge. Presently, no standardized method exists for estimating GPC's energy absorption capacity, leading most studies to rely on models designed for conventional concrete. These models often exhibit low accuracy due to their limited consideration of various factors. Despite the potential of machine learning techniques to address this issue [14, 15], their utilization in predicting GPC's energy absorption capacity remains notably limited.

To address these challenges, this research aims to develop effective machine learning-based models to accurately predict GPC's energy absorption capacity. The primary objectives include (1) to create machine learning models for estimating GPC's EAC and (2) to assess each input factor's contribution to GPC's energy absorption capacity.

4.2 Machine Learning-Based Models

4.2.1 Structure of Machine Learning Models

Two widely used machine learning-based methods, artificial neural network (ANN), and random forest (RF), were utilized to predict the energy absorption capacity of GPC. The typical structure of ANN models is depicted in Fig. 4.1, while that of the RF model is outlined in Fig. 4.2.

The ANN model's structure comprised an input layer, two hidden layers, and an output layer. The input layer accommodated 13 neurons, each representing distinct features such as aggregates and alkaline solutions, among others. Conversely, the output layer featured one node responsible for predicting the EAC of GPC. The ANN's design incorporated two hidden layers, acting as intermediary processing
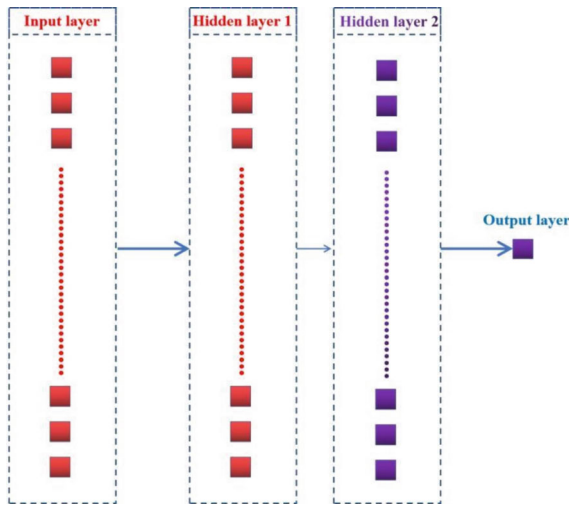
Fig. 4.1 An artificial neural network layout
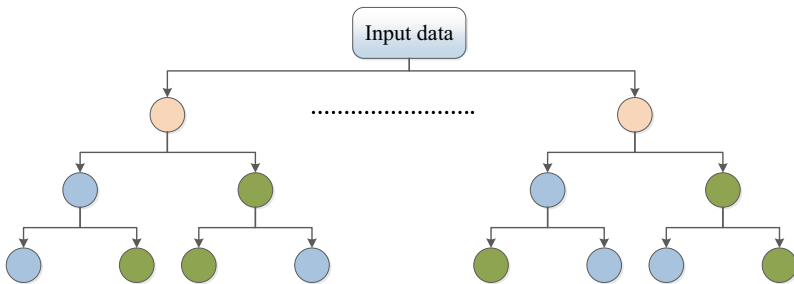
Fig. 4.2 A random forest layout

stages. The first hidden layer consisted of 40 neurons, followed by the second hidden layer comprising 50 neurons. These numerical selections were the result of systematic exploration among various configurations aimed at optimizing the model's predictive capabilities. The activation function was employed within these layers to facilitate effective communication among neurons. Additionally, the ANN utilized the algorithm and the optimizer for iterative refinement of model parameters. Furthermore, the model employed the loss function to limit the disparity between estimated and tested data.

In contrast, the RF model embodies a collaboration of multiple decision trees. Data undergoes processing through individual decision trees, and the collective predictions of these trees amalgamate to generate the final prediction. This study employed a trio of decision trees, with the number determined as 3, while setting the minimum number of samples required to divide a decision point to 2 and that of samples necessary at the leaf node to 1.

4.2.2   Experimental Database

A comprehensive compilation of 90 experimental test data was collected from several prior researches [7–13, 16–30]. Thirteen input variables with their wide ranges were considered: coarse aggregate (0–1556 kg/m3), fine aggregate (0–1174 kg/m3), sodium hydroxide (8–170 kg/m3), sodium silicate (33–445 kg/m3), water (0–276 kg/m3), aluminosilicate material (340–1249 kg/m3), silica fume (0–283 kg/m3), curing temperature (21–85 °C), curing time (1–2160 h), concrete age (3–90 days), compressive strength (14–145 MPa), notch depth ratio (0.1–0.5), and maximum aggregate size (0.12–20 mm). Additionally, the energy absorption capacity of GPC varied from 9 to 371 N/m. The dataset was divided into two segments: 80% for training and 20% for testing.

4.2.3   Statistical Measures

The accuracy of the machine learning models was investigated through two metrics which are calculated through the following equations:

$$R = \frac{n\sum_{i=1}^{n}\widehat{a}_i a_i - \left(\sum_{i=1}^{n}\widehat{a}_i\right)\left(\sum_{i=1}^{n}a_i\right)}{\sqrt{n\sum_{i=1}^{n}\widehat{a}_i^2 - \left(\sum_{i=1}^{n}\widehat{a}_i\right)^2}\sqrt{n\sum_{i=1}^{n}a_i^2 - \left(\sum_{i=1}^{n}a_i\right)^2}} \tag{4.1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{a}_i - a_i\right)^2}{n}} \tag{4.2}$$

where *R* is correlation coefficient, *RMSE* defines root mean squared error, *n* is number of data, and $\widehat{a}_i$ and a_i are the predicted and tested data, respectively.

4.3 Results and Discussion

4.3.1 Accuracy of the ANN and RF Models

Figure 3a illustrates the correlation between predicted and tested data for training the artificial neural network model, while Fig. 3b displays that of the RF model. The correlation across all models in the training dataset demonstrated strong performance, with R values ranging from 0.94 to 0.97. Additionally, the RMSE ranged from 21 to 50 N/m, indicating the accepted reliability of ANN and RF models in estimating the EAC of GPC. When comparing the machine learning-based models, the RF model showcased superior performance in the training set.

The performance of machine learning-based models in estimating EAC in the test dataset is depicted in Fig. 4.4. Figure 4a presents the prediction results of the ANN model, while Fig. 4b displays that of the RF model. The RF model demonstrated consistent and strong performance, while the ANN model exhibited slightly lower accuracy in the testing set. The R values for the ANN and RF models were 0.87 and 0.95, respectively. Furthermore, the RMSE values for the ANN and RF models in the testing set were 25 and 9, respectively. Hence, the developed RF model showcased its ability to accurately and reliably predict the EAC of GPC in both train and test sets.

a) ANN model   b) RF model
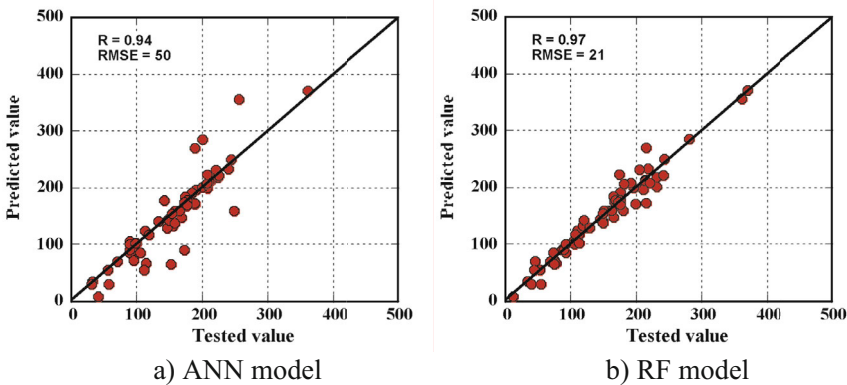
Fig. 4.3 The correlation of the machine learning-based models in training set
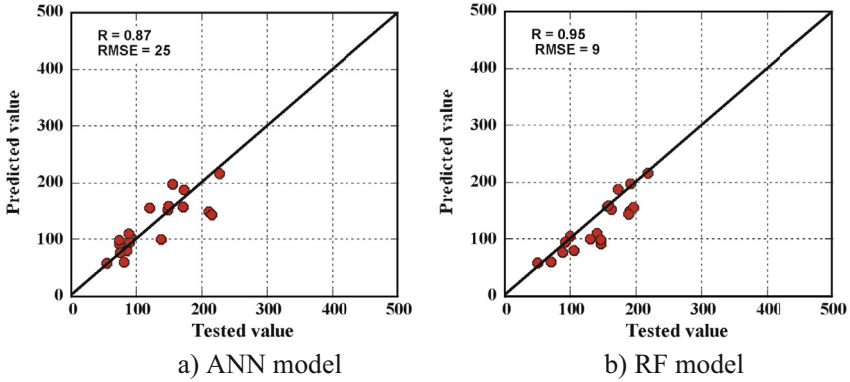
Fig. 4.4 The correlation of the machine learning-based models in testing set

4.3.2 Sensitivity Analysis

To assess the relative importance of all input parameters to the final EAC of GPC, a sensitivity analysis based on the Shapley Additive explanations (SHAP) method was used. Figure 4.5 illustrates the relative importance of all input parameters according to both the ANN and RF models. The findings from both models consistently highlight that the notch depth ratio holds the highest importance among the variables, significantly influencing the final energy absorption capacity. Conversely, the curing temperature emerged as the least influential variable.
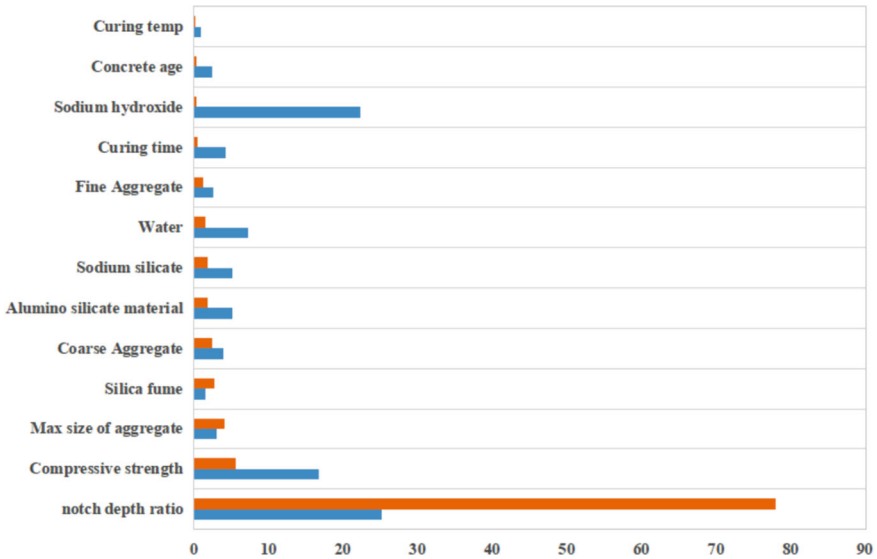
Fig. 4.5 Contribution of all input parameters

4.4   Conclusion

This study introduces the application of machine learning models to predict the energy absorption capacity of GPC. Utilizing 90 experimental test results, two widely adopted machine learning techniques, ANN and RF models, were developed, considering thirteen input parameters characterized by intricate relationships. From the prediction results and sensitivity analysis, the following conclusions emerge:

- Machine learning-based models effectively predict the energy absorption capacity of GPC by simultaneously considering multiple input parameters.
- The proposed RF model demonstrates high accuracy and reliability in estimating the energy absorption capacity across both training and testing sets. However, the ANN model exhibited slightly lower accuracy in the testing set.
- Notably, the notch depth ratio emerged as the most influential variable significantly impacting the energy absorption capacity, while curing temperature displayed the least importance among the variables.

References

1. Tran, N.T., Nguyen, D.H., Tran, Q.T., Le, H.V., Nguyen, D.L.: Experimental study and machine learning based prediction of the compressive strength of geopolymer concrete. Magazine of Concrete Research, Ahead of Print (2023)

2. Tran, N.T., Nguyen, D.L., Kim, D.J., Ngo, T.T.: Sensitivity of various fiber features on shear capacities of ultra-high-performance fiber-reinforced concrete. Mag. Concr. Res. **74**(4), 190–206 (2021)

3. Tran, N.T., Nguyen, D.L., Vu, Q.A., Kim, D.J., Ngo, T.T.: Dynamic shear response of ultra-high-performance fiber-reinforced concretes under impact loading. Structures **41**, 724–736 (2022)

4. Tran, T.K., Nguyen, T.K., Tran, N.T., Kim, D.J.: Improving the tensile resistance at high strain rates of high-performance fiber-reinforced cementitious composite with twisted fibers by modification of twist ratio. Structures **39**, 237–248 (2022)

5. Vu, T.V., Tran, T.K., Kim, D.J., Nguyen, D.L., Tran, N.T.: Strain rate sensitivity of strain-hardening fiber-reinforced concrete subjected to dynamic direct tensile loading. Struct. Concrete, Early View (2023)

6. Das, F.K., Mishra, J., Singh, S., Mustakim, S.M., Patel, A., Das, S., Behera, U.: Characterization and utilization of rice husk ash (RHA) in fly ash – Blast furnace slag based geopolymer concrete for sustainable future. Materials Today: Proceedings **33**, 5162–5167 (2020)

7. Ding, Y., Dai, J.G., Shi, C.J.: Fracture properties of alkali-activated slag and ordinary Portland cement concrete and mortar. Constr. Build. Mater. **165**, 310–320 (2018)

8. Ding, Y., Shi, C.J., Li, N.: Fracture properties of slag/fly ash-based geopolymer concrete cured in ambient temperature. Constr. Build. Mater. **190**, 787–795 (2018)

9. Liu, Y., Shi, C., Zhang, Z., Li, N., Shi, D.: Mechanical and fracture properties of ultra-high performance geopolymer concrete: Effects of steel fiber and silica fume. Cement Concr. Compos. **112**, 103665 (2020)

10. Xie, Z.L., Zhou, H.F., Lu, L.J., Chen, Z.A.: An investigation into fracture behavior of geopolymer concrete with digital image correlation technique. Constr. Build. Mater. **155**, 371–380 (2017)

11. Mousavinejad, S.H.G., Gashti, M.F.: Effects of NaOH solution concentration and aging on fracture properties and ductility of ambient-cured heavyweight geopolymer concrete. Constr. Build. Mater. **277**, 122266 (2021)

12. Wang, Y., Hu, S., He, Z.: Mechanical and fracture properties of fly ash geopolymer concrete addictive with calcium aluminate cement. Materials **12**, 2982 (2019)

13. Pan, Z., Sanjayan, J.G., Rangan, B.V.: Fracture properties of geopolymer paste and concrete. Mag. Concr. Res. **63**, 763–771 (2011)

14. Ngo, T.T., Le, Q.H., Nguyen, D.L., Kim, D.J., Tran, N.T.: Experiments and prediction of direct tensile resistance of strain-hardening steel-fiber-reinforced concrete. Mag. Concr. Res. **75**, 780–794 (2023)

15. Tran, N.T., Nguyen, T.K., Nguyen, D.L., Le, Q.H.: Assessment of fracture energy of strain-hardening fiber-reinforced cementitious composite using experiment and machine learning technique. Struct. Concr. **24**, 4185–4198 (2023)

16. Mousavinejad, S.H.G., Sammak, M.: An assessment of the fracture parameters of ultra-high-performance fiber-reinforced geopolymer concrete (UHPFRGC): the application of work of fracture and size effect methods. Theoret. Appl. Fract. Mech. **117**, 103157 (2022)

17. Mousavinejad, S.H.G., Sammak, M.: An assessment of the effect of Na2SiO3/NaOH ratio, NaOH solution concentration, and aging on the fracture properties of ultra-high-performance geopolymer concrete: the application of the work of fracture and size effect methods. Structures **39**, 434–443 (2022)

18. Mousavinejad, S.H.G., Gashti, M.F.: Effects of alkaline solution to binder ratio on fracture parameters of steel fiber reinforced heavyweight geopolymer concrete. Theoret. Appl. Fract. Mech. **113**, 102967 (2021)

19. Ding, Y., Bai, Y.L.: Fracture properties and softening curves of steel fiber-reinforced slag-based geopolymer mortar and concrete. Materials **11**, 1445 (2018)

20. Gomes, E.F., Dias, D.P., Silva, F.D.A.: Determination of the fracture parameters of steel fiber-reinforced geopolymer concrete. Theoret. Appl. Fract. Mech. **107**, 102568 (2020)

21. Nath, P., Sarker, P.K.: Fracture properties of GGBFS-blended fly ash geopolymer concrete cured in ambient temperature. Mater. Struct. **50**, 32 (2017)

22. Sarker, P.K., Haque, R., Ramgolam, K.V.: Fracture behavior of heat cured fly ash based geopolymer concrete. Mater. Des. **44**, 580–586 (2013)
23. Zhang, P., Wang, J., Li, Q., Wan, J., Ling, Y.: Mechanical and fracture properties of steel fiber-reinforced geopolymer concrete. Sci. Eng. Compos. Mater. **28**(1), 299–313 (2021)
24. Aisheh, Y.I., Atrushi, D.S., Akeed, M.H., Qaidi, S., Tayeh, B.A.: Influence of steel fibers and microsilica on the mechanical properties of ultra-high-performance geopolymer concrete (UHP-GPC). Case Stud. Construct. Mater. **17**, e01245 (2022)
25. Divya, K.K., Vidya, V.: Investigation on fracture parameters of geopolymer concrete. J. Emerg. Technol. Innovat. Res. **8**, 809–818 (2021)
26. Sabu, N.J., Thomas, J.: Fracture properties of OPC included ambient cured geopolymer concrete by size effect method. IOP Conf. Ser.: Earth Environ. Sci. **491**, 012036 (2020)
27. Zhang, N., Yan, C., Li, L., Khan, M.: Assessment of fiber factor for the fracture toughness of polyethylene fiber reinforced geopolymer. Constr. Build. Mater. **319**, 126130 (2022)
28. Wang, Y., Hu, S., Sun, X.: Experimental investigation on the elastic modulus and fracture properties of basalt fiber–reinforced fly ash geopolymer concrete. Constr. Build. Mater. **338**, 127570 (2022). https://doi.org/10.1016/j.conbuildmat.2022.127570
29. Raj, D.S., Abraham, R., Ganesan, N., Sasi, D.: Fracture properties of fibre reinforced geopolymer concrete. Int. J. Sci. Eng. Res. **4**(5), 75–80 (2013)
30. Jenifer, M.A., Kumar, S.S., Devadass, C.S.C.: Fracture behavior of fiber reinforced geopolymer concrete. Int. J. Adv Technol. Engin. Sci. **3**, 387–396 (2015)

Chapter 5
Smartphone-Controlled Robot Manipulator for Pick and Place Applications

Venkata Sushma Chinta [ID], **Sowmya Kethi Reddi** [ID], **Nagini Yarramsetty** [ID], **Shaik Chand Pasha** [ID], **and Palla Nishant Bhanu** [ID]

Abstract In contemporary society, robots have emerged as indispensable assets, capable of undertaking tasks too monotonous or perilous for humans, such as bomb defusing and handling molten metal. These automated entities significantly contribute to reducing human errors while enhancing task precision. Deployed across diverse sectors, from medical fields to locomotive industries, robots serve varied purposes, spanning repetitive tasks, analytical operations, and applications demanding high accuracy. The demand for Pick and Place robots has surged, particularly in mass production settings. This paper focuses on a robotic arm featuring 5 degrees of freedom, crafted through 3D printing technology and seamlessly integrated with an Arduino Uno controller. The innovation extends to the realm of user interaction, enabling intuitive control via a dedicated smartphone application. The synthesis of advanced technologies enhances the robot's versatility, making it adept at addressing contemporary challenges in automation, precision, and efficiency.

Keywords Pick and place robot · Five-degrees of freedom · 3D printing · End effector · Robotic arm · Modular design · Controller · Workspace

V. S. Chinta (✉) · N. Yarramsetty · S. C. Pasha · P. N. Bhanu
Mechanical Engineering Department, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
e-mail: venkatasushmachinta_mech@cbit.ac.in

S. K. Reddi
School of Management Studies, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

5.1   Introduction

Pick and place robots are used where the specific tasks are done at a place. They comprise physical entities of IoT devices like controllers-Arduino uno, Arduino AT mega and servos-MG996R, RDS 3115MG, SG90, modules-Transmitter and receivers. Now-a-days in every industry the demand for pick and place robots is drastically increasing because of their automation. This automation subsequently increases the production rate as well as precision levels. These robots are installed at a place and then the tasks are assigned to it to do repetitive work which is boring for humans to do. Humans cannot do accurate work which is done by using these robots. Meanwhile, humans can focus on complex tasks where analytical knowledge is required which results in innovations.

5.1.1   3D Printing

In the constant development of technology, it is difficult to make intricate shapes. Traditional methods like casting, forging, and sheet metal operations don't produce complex shapes which are done using 3D printing methodology. 3D printing is a computer-driven additive manufacturing technology. 3D printing works by converting the digital model into a real-time model. It works on principle fused deposition modeling. Any solid work model is converted into STL files for giving input to the 3D printer. The 3D printer consists of a nozzle which melts material like PLA and ABS using induction of heat. It melts the material and prints it layer by layer on a print bed where the final model is made. The 3D printer consists of a display screen for selection of object models to print. It has three stepper motors which are used for coordinating the nozzle. The material is fed in the form of a filament. Generally, the radius of the filament is 3.5 mm.

5.2   Literature Review

The incorporation of smartphone technology into the control systems of robotic manipulators has emerged as a significant trend, transforming the landscape of automation and broadening the scope of applications for robotic systems. This literature review aims to provide an overview of pertinent studies and advancements in the realm of smartphone-controlled robot manipulators, with a specific focus on pick and place applications.

Kumar et al. [1] conducted work in developing an image processing algorithm crucial for object detection and recognition in pick and place robots. Their emphasis on accurate feature extraction and classifiers laid the foundation for subsequent studies, highlighting the importance of precise vision-based systems. Magar et al.

[2] explored the implementation of a six-degrees-of-freedom pick and place robot using Arduino Uno ATmega328, showcasing the significance of microcontroller technology in robotic manipulators. The study underscores the relevance of Arduino platforms in achieving precise and controlled movements, contributing to the overall efficiency of the robotic arm. Surati et al. [3] provided a comprehensive review of pick and place robots using different controllers, including Arduino Uno and Arduino Mega. The authors identified a trend toward integrating object recognition algorithms into these systems, emphasizing the necessity of intelligent systems for enhanced robotic performance. Manivannan et al. [4] delved into the design and analysis of a robotic arm with Bluetooth communication, shedding light on the importance of robust communication technologies in the operation of robotic arms. Their work contributes to the establishment of reliable and efficient communication protocols. Thomas et al. [5] focused on building a robotic manipulator with three joints, incorporating an object sensor for obstacle detection. This study underscores the significance of safety features in robotic arms, ensuring their adaptability to dynamic environments and reducing the risk of collisions. Anughna et al. [6] extended the capabilities of robotic arms by utilizing Flex and Gyro sensors for design and implementation. Their emphasis on sensor technologies reflects a growing trend in enhancing robotic functionality through advanced sensor applications. Clothier et al. [7] presented a geometric approach for solving unknown joint angles in a robotic arm with five degrees of freedom. The study employed an iRobot command module and infrared range finders for object location, showcasing advancements in kinematic solutions.Wei et al. [8] developed a method for robotic object recognition and grasping using a natural background, employing cutting-edge edge detection, superpixel segmentation, and form matching algorithms. The lightweight approach without prior training demonstrated adaptability to various circumstances. Rakhimkul et al. [9] built a smarter human–machine interface for a robotic arm operated by a joystick, incorporating machine learning-based object detection and RGB-D sensor data processing. The study highlights advancements in human–robot interaction and automatic object recognition. Pan et al. [10] proposed an algorithm for generating a robotic arm's collision-free trajectory using computational fluid dynamics concepts. Although limited to static environments, the study contributes to the understanding of motion planning with fluid dynamics limitations. Yagna Et al. [11] designed a five-degree-of-freedom robotic arm using ultrasonic sensors for object distance determination. The study employed an M3 LPC1768 cortex microcontroller, demonstrating the integration of sensors for accurate object detection and manipulation. Timothyet al. [12] devised an autonomous arm for soldering fractures using six servo motors and an Arduino microcontroller. The study showcased the use of a camera for position feedback, combining software components for effective autonomous soldering. Shamsheeret al. [13] built a manipulator with three degrees of freedom controlled by hand gestures, employing an Arduino Mega 2560 and various sensors. The study demonstrated the feasibility of gesture-based control in robotic manipulation. Ghaleb et al. [14] modeled and controlled a two-link 2 DOF robotic arm, emphasizing the utilization of Permanent Magnet DC (PMDC) motors for arm

operation. The simulation in MATLAB contributed to the understanding of kinematics in a controlled environment. Sudharsan et al. [15] focused on path planning efficiency using MATLAB and ADAMS simulation tools, employing joint cubic path control algorithms for real-time operation. The study provided insights into simulation tools for assessing robotic arm functionality. Khosla et al. [16] developed a smart robotic arm with a graphical user interface (GUI), allowing users to choose specific tasks. The study showcased the versatility of control methods, including app-based interfaces, remote control, and voice commands. Qassem et al. [17] created a virtual program for controlling a 5-DOF educational robot arm, utilizing card-generated motion for communication. The study demonstrated the application of simulations for educational purposes and accurate arm positioning. Ban et al. [18] delved into real-time element identification using intelligent machines, particularly focusing on object detection with computer vision and artificial intelligence. The study highlighted the increasing role of smart robots in accurately recognizing and manipulating objects. Kumaret al. [19] used servos for powering joints in a robotic arm, emphasizing the importance of accurate manipulator design using CATIA software. The study contributed to the understanding of kinematics and servo selection for robotic arm fabrication. Kunchala et al. [20] utilized a Programmable Logic Controller and two artificial intelligence algorithms for a 5-DOF manipulator. The study focused on integrating data from camera views to direct the robot's movement, showcasing advancements in AI-based control algorithms. Sobhan et al. [21] focused on streamlining warehouse product management through a 5-DOF robotic manipulator, mitigating complexities via a GSM-based mobile app. Aladetola et al. [22] developed an economical and stable pick and place mechanism for industrial applications, employing ultrasonic sensors, servo motors, an Atmega328 Arduino microprocessor, and light steel iron frames. The study's stability validation through theoretical equations and MATLAB software establishes the microcontroller-controlled robotic arm's benefits for enhancing productivity in industrial environments. In conclusion, the literature reviewed demonstrates a trajectory of advancements in pick and place robotic systems, from vision-based algorithms and microcontroller applications to communication technologies and sensor integrations. The present work stands out by incorporating smartphone control, offering a novel dimension to robotic manipulation and paving the way for further innovations in the field.

5.3 Methodology

As shown in Fig. 5.1, the robotic manipulator is made entirely of 3D printed components, wherein the components are made purely from polylactic acid (PLA). It comprises various links like base, waist, shoulder, elbow, and wrist components like wrist pitch, wrist roll, and gripper gear parts. The parts of the robotic manipulator are printed by the 3D printer "FLASHFORGE Inventor" using a fused deposition method.

Fig. 5.1 Parts of robotic manipulator

Figure 5.2 depicts the robotic manipulator in its assembled state. The base link is connected to the waist link to form a joint which is named as waist joint. At the waist joint, a servo motor (MG-995) is connected to have rotation of the waist about the vertical axis of the base. The waist link is connected to the shoulder link to form a shoulder joint. At the shoulder joint, a servo motor (MG-995) is connected to have rotation of the shoulder about the horizontal axis of the waist.

The shoulder link is connected to the elbow link to form the elbow joint. At the elbow joint, a servo motor (MG-995) is connected to have rotation of the elbow about the horizontal axis of the shoulder. The elbow link is connected to the Wrist roll link to form a wrist roll joint. At the wrist roll joint, a servo motor (Micro SG-90) is connected to have roll of wrist. The wrist roll link is connected to wrist pitch link to form wrist pitch joint. At the wrist pitch joint, a servo motor (Micro SG-90) is connected to have pitch of the wrist. The Gripper is connected to Wrist pitch. The gripper can be opened and closed to hold the object. This can be achieved by a servo motor (Micro SG-90).

The user interface app developed for the robot manipulator using the "MIT App Inventor" platform is shown in Fig. 5.3. This app allows the user to operate the robot manipulator. The HC-05 Bluetooth module and the smartphone are paired using the two buttons at the top of the app. Figure 5.3 shows the robot manipulator (300), and to the right are six sliders (Grip, Wrist pitch, wrist roll, Elbow, Shoulder, Waist) for adjusting the servos' positions and a seventh slider (Speed) for adjusting the manip-ulator's speed. Each slider's initial, minimum, and maximum values are customized

Fig. 5.2 Assembly of all parts of robotic manipulator
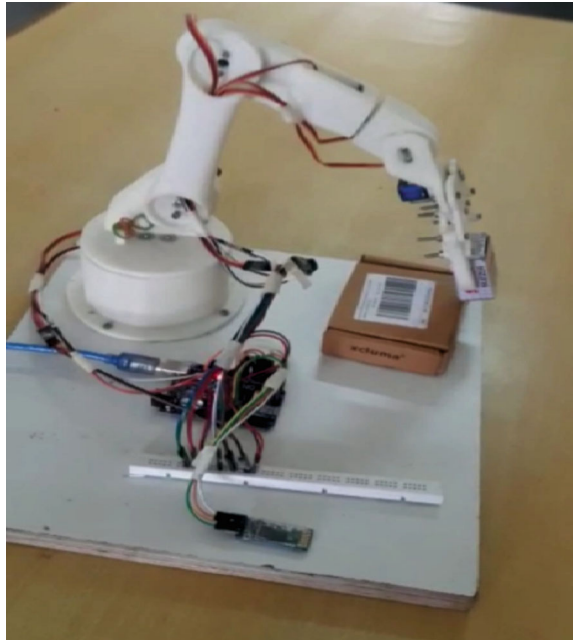
to the robot manipulator joints. The robot manipulator can be programmed to operate automatically using the SAVE, RUN, and RESET buttons at the bottom of the app. A label down below displays the total number of saved steps the inventors have made.

Figure 5.4 displays a collection of sixteen design blocks that were created specifically for the application. Blocks one to five serve the purpose of establishing a connection between the smartphone and the Bluetooth module. The blocks numbered six to eleven correspond to the slider blocks used for controlling the position of the servo. On the other hand, blocks numbered twelve to fifteen are button blocks specifically designed for programming the robot manipulator. By using the Bluetooth capability (.SendText), a text is sent to the Arduino when the position of the slider is altered. The text contains a prefix that denotes the specific slider that has been modified, together with the slider's current value. Block twelve corresponds to the SAVE button. Pressing the SAVE button results in the storage of the position of each servo motor in an array. The block labeled thirteen corresponds to the function of the RUN button. Activating the RUN button triggers the execution of the runservo() function, which carries out predetermined actions. Block fourteen demonstrates the functionality of the RUN/PAUSE button. When pressed, this button halts the robot's actions. If pressed again, it resumes the robot's automated motions. Block fifteen depicts the RESET button. The saved steps are executed repeatedly until the RESET button is pushed. Block sixteen demonstrates the functionality of the speed slider. When the position of the speed slider is altered, the corresponding value is used to modify the delay time between each repetition, so regulating the velocity of the servo motors.
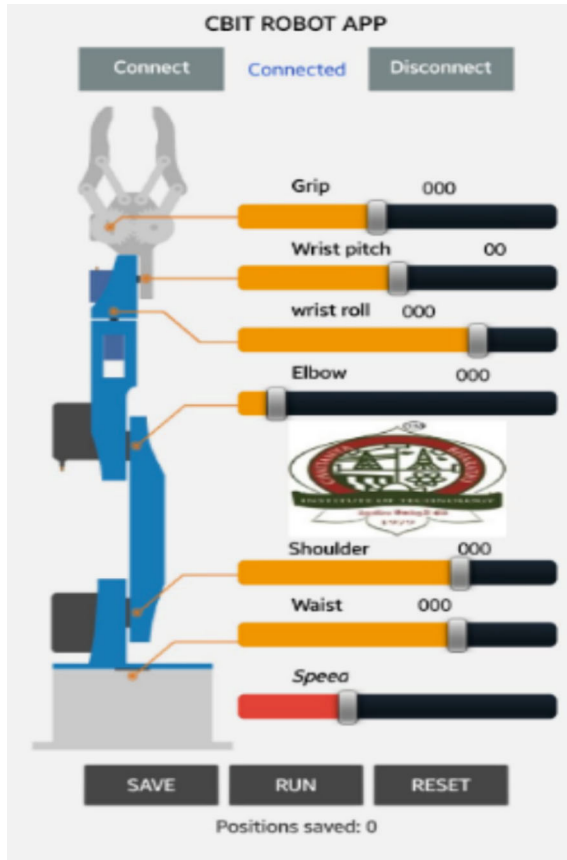
Fig. 5.3 The smartphone app interface
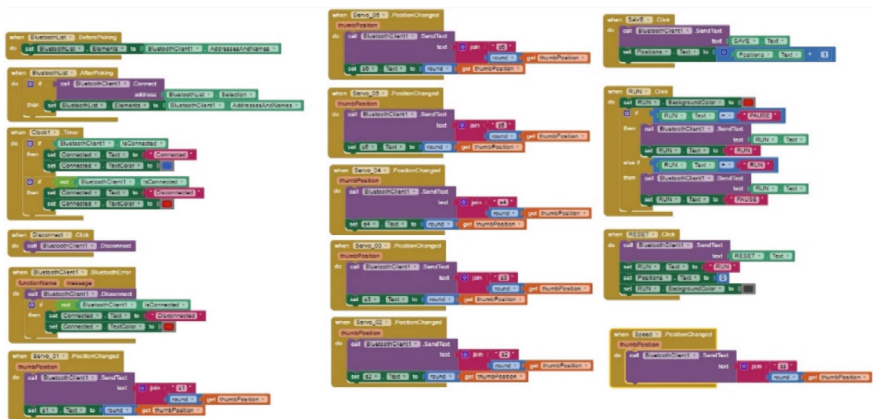
Fig. 5.4 Design blocks for the app
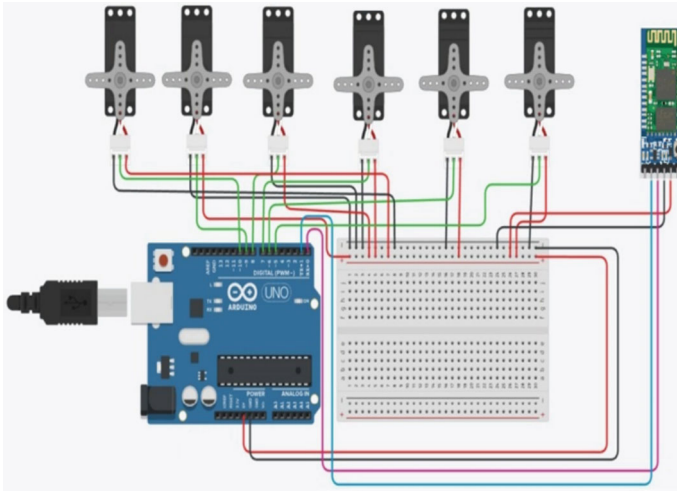
Fig. 5.5 Circuit diagram of the robotic manipulator

Figure 5.5. illustrates the circuit diagram of the robotic manipulator. It consists of Arduino board and a HC-05 Bluetooth module for communication with the smartphone. The control pins of the six servo motors are connected to six digital pins of the Arduino board. Pin 5 controls the servo attached to the waist. Pin 6 controls the servo attached to the shoulder. Pin 7 controls the servo attached to the elbow. The servo in the wrist roll is manipulated by pin 8 and the servo in the pitch motion of the wrist is controlled by pin 9. The servo coupled to the gripper's gears is activated via pin 10. Power and ground pins of each servo are connected in parallel rows on the breadboard. Those rows are connected to the power and ground pins of Arduino. The HC-05 Bluetooth module is supplied with power and grounded on a breadboard, while its transmit (Tx) and receive (Rx) pins are linked to the receive (Rx) and transmit (Tx) pins of the Arduino, respectively.

5.4 Conclusion

In summary, this paper introduces an advanced robotic arm with five degrees of freedom, controlled wirelessly through a specially designed Android app. The heart of this innovation lies in the integration of electronic components, powered by a programmed Arduino and a sophisticated Bluetooth module. The robot's design includes crucial parts like the base, waist, shoulder, elbow, and wrist, along with components like wrist pitch, wrist roll, and gripper gear parts. What sets this apart is the user-friendly interface, featuring keys like Save, Run, and Stop. The "Save" button allows the robot to remember and repeat specific movements, while the interface also allows for pausing, resetting, or deleting recorded steps, making room for

new commands. This blend of precision, adaptability, and user control positions our robotic arm as a beacon of innovation in the world of automation.

5.5 Future Scope

- Enhance the robot's task performance by integrating advanced sensors, enabling more accurate and precise movements, thus elevating its overall efficiency.
- Explore solutions to augment the robot's payload capacity, allowing it to lift and transport heavier items, thereby expanding its application across diverse industries.
- Investigate alternative communication technologies to fortify the robot's interaction with devices like smartphones. Implement robust error-handling mechanisms to address issues related to communication range and interruptions.
- Measuring Efficiency: Quantify the robot's energy consumption to comprehensively assess its operational efficiency. This not only contributes to sustainability efforts but also aids in optimizing resource usage. Additionally, adapt the robot to function effectively in diverse environments and find ways to reduce production costs for increased practical utility.
- Develop the robot's capability to detect objects independently. Enable the servos to respond to detected objects, enhancing their autonomy and usefulness in tasks that involve object recognition and handling.
- Further advance the robot's capabilities by integrating voice recognition modules and implementing NLP algorithms or technologies like Snow Boy. This addition can enhance human–robot interaction, making the robot more intuitive and versatile in its operations.

References

1. Kumar, R., Kumar, S., Lal, S., Chand, P.: Object detection and recognition for a pick and place robot. In: Asia-Pacific World Congress on Computer Science and Engineering, pp. 1–7. IEEE (2014)
2. Magar, V.M., Suroshe, K.K., Patil, H.P., Sathe, D.V., Khairnar, H.R.: Six wheel drive pick and place robot using Arduino. Int. Res. J. Eng. Technol. **07**(03), 3862–3865 (2020)
3. Surati, S., Hedaoo, S., Rotti, T., Ahuja, V., Patel, N.: Pick and place robotic arm: a review paper. Int. Res. J. Eng. Technol. **8**(02), e-ISSN: 2395–0056 (2016)
4. Manivannan, A., Hari Prasath, R., GopiNath, K., Ram Prasanna, M.: Design and development of pick and place robotic arm with path record play. J. Xi'an Univ. Arch. & Technol. **2**(5), 645–651 (2020)
5. Thomas, B.A., Michahial, S., Shreeraksha, P., Nagvi, V.B., Suresh, M.: Industry based automatic robotic arm. Int. J. Eng. Innov. Technol. **2**(11), 42–48 (2013)
6. Anughna, N., Ranjitha, V., Tanuja, G.: Design and implementation of wireless robotic arm model using flex and gyro sensor. Int. J. Recent. Technol. Eng. **8**(5), 2978–2983 (2020)

7. Clothier, K.E., Shang, Y.: A geometric approach for robotic arm kinematics with hardware design, electrical design, and implementation. Hindawi Publ. Corp. J. Robot. **984823**, 1–10 (2010)
8. Wei, A.H., Chen, B.Y.: Robotic object recognition and grasping with a natural background. Int. J. Adv. Rob. Syst. **17**(2), 1–17 (2020)
9. Rakhimkul, S., Kim, A., Pazylbekov, A., Shintemirov, A.: Autonomous object detection and grasping using deep learning for design of an intelligent assistive robot manipulation system. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 3962–3968 (2019)
10. Pan, Z., Park, C., Mnocha, D.: Robot Motion planning for pouring liquids. In: Proceedings of the International Conference on Automated Planning and Scheduling, vol. 26 (2016)
11. Jadeja, Y., Pandya, B.: Design and development of 5-DOF robotic arm manipulators. Int. J. Sci. Technol. Res. **8**(11), 2158–2167 (2019)
12. Timothy K. F.: Robotic Arm Tracing Curve Recognized by Camera. Florida Institute of Technology, Melbourne (2016)
13. Verma, S.: Hand gestures remote controlled robotic arm. Adv. Electron. Electr. Eng. **3**(5), 601–606. ISSN 2231–1297 (2013)
14. Ghaleb, N.M., Aly, A.A.: Modeling and control of 2-DOF robot arm. Int. J. Emerg. Eng. Res. Technol. **6**(11), 24–31 (2018)
15. Sudharsan, J., Karunamoorthy, L.: Path planning and co-simulation control of 8 DOF anthropomorphic robotic arm. Int. J. Simul. Model. **15**(2), 302–312 (2016)
16. Khosla, D., Sharma, M., Khanna, S.K., Khanna, P., Kaur, G.: Smart robotic arm. J. Artif. Intell. Res. & Adv. **6**(2), 58–62 (2019)
17. Qassem, M. A., Abuhadrous, I., Elaydi, H.: Modeling and simulation of 5 DOF educational robot arm. Int. J. Mech. Mechatron. Eng. **5**(3) (2011)
18. Ban, P., Desal, S., Barge, R., Chava, P.: Intelligent robotic arm. In: Proceedings of ICACC-2020, ITM Web of Conferences, vol. 32, pp. 01005 (2020)
19. Kumar, T. S., Sarath, K., Family, S. D., Bhagyesh, A. V. S., Althaf, S. K.: Design and fabrication of pick and place robotic arm. In: Proceedings of the 2nd National Conference on Recent Trends in Mechanical Engineering, GIST (2020)
20. Kunchala, H.R., Toporovsky, J.: PLC based robot manipulator control using position based and image based algorithm. Glob. J. Res. Eng.: Robot. & Nano-Tech. **17**(1) (2017)
21. Sobhan, N., Shaikat, A. S.: Implementation of pick & place robotic arm for warehouse products management. In: IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), pp. 156–161. Bandung, Indonesia (2021)
22. Aladetola, O.D., Adelakun, O.N., Ajayii, O.A.: An innovative pick and place robotic arm system using an Atmega328 microcontroller with escalating stability. Indian J. Adv. Sci. Eng. Technol. **1**(2), 39–52 (2022)

Chapter 6
Impact of Russia-Ukraine War on Oil and Gas Sector of Indian Economy

Kuldip Katiyar, Tejal Gupta, and Kundan Kumar Jha

Abstract Russia and Ukraine conflict impose the significant consequences over the global economy. It's not just a local issue; the fallout is hitting everywhere. These two countries play a big role in global trade, dealing with wheat, crude oil, and raw materials. When Russia made its move into Ukraine in February 2022, it caused a serious domino effect, crashing stock markets worldwide, including in India. My paper is diving deep into how this chaos is shaking things up in the Indian stock market, and I'm using the Nifty 50 index to guide me. Considering ASEAN economies and India are big on importing petroleum, this conflict is hitting them hard, messing with trade terms and causing a shock that's affecting consumers, businesses, and GDP growth. With oil prices skyrocketing due to the war, everyday things are getting pricier, messing with people's budgets and possibly slowing down economic growth in India. This study concentrates on evaluating how the Russia-Ukraine war has influenced Oil and Gas sector within the Indian economy.

Keywords Russia-Ukraine war · Abnormal returns · Oil and gas sector · LSTM prediction model · Indian stock market

6.1   Introduction

Russia went all in on Ukraine starting February 24, 2022, and it's not just a regular fight. This is some big-time aggression, the largest military move on a European country since World War II [3]. It kicked off Europe's biggest refugee crisis since then, with almost 3.8 million Ukrainians bailing out and millions more left without a home. This war is a major hit to the global economy, messing with growth and shooting up inflation rates. It's like a ripple effect, slowing down the whole world's economic

K. Katiyar · T. Gupta (✉) · K. K. Jha
Department of Mathematics, Chandigarh University, Mohali 140413, Punjab, India
e-mail: 2984tejalgupta@gmail.com; info@cumail.in

K. Katiyar
e-mail: kuldeepe7279@cumail.in

game and making stuff pricier. India's got some trade ties with Russia, but this mess might cost them with supply issues. The US even slapped a ban on oil and gas imports from Russia, sending Brent crude prices soaring to almost $130 per barrel—that's a 43% jump since February. Huge setback, especially since Russia's a big-time crude oil exporter . And you know what's hitting us hard? Fuel prices [14, 25]. India's economy is feeling the hit, growing slower than expected, and experts are saying growth might take a hit this time around. So, with fuel prices going through the roof, inflation is shooting up too [23]. People are feeling the pinch, cutting back on fried goodies and even veggies because the Ukraine war is hiking up edible oil and fuel costs. Companies are going all out with the oil price hikes since the Russia-Ukraine showdown, making diesel, petrol, and vegetable oils crazy expensive and squeezing consumers' wallets. With the whole crude oil scene going wild globally, India's taking a hit since we import a whopping 85% of our crude oil. Prices have shot up by almost 50% this year alone, thanks to the Russia-Ukraine drama. And with the US slapping sanctions on Russia, crude oil prices are expected to go even higher because of the ongoing tension. It's already crossed the $100 per barrel mark, hitting $108 as of May 5, 2022, the highest in 14 years and the first half of 2021.

6.2 Literature Review

The academic realm labels war as a "black swan" event, akin to financial crises, pandemics, natural disasters, elections, and terrorist attacks, all with significant global impacts on equity markets. Black swans are unpredictable and hold severe consequences. During crises, stock volatility is pronounced, particularly in emerging markets, which face immediate and enduring consequences. Military conflicts elevate investor anxiety, causing stock prices to fluctuate. Recent academic interest surged due to the unforeseen COVID-19 crisis, focusing on financial market reactions (See for reference [6–8, 15, 27]). Previous studies on events like the 2014 Ukrainian crisis demonstrated reduced co-movement between the Russian equities market and global markets. Assessing various structural breaks, such as the global financial crisis, change of government, demonetization, and COVID-19, revealed bidirectional causality between cash and futures markets [5, 28]. The impact of the Russia-Ukraine crisis on global stock returns was evident, with varying reactions in countries neighboring the conflict. Analyzing the Borsa Istanbul, stock returns initially turned positive but significantly declined in the following days. Spillovers across global financial markets sharply increased during the crisis, with French and German markets acting as major net volatility transmitters, influencing global market risk [18, 20, 21, 26]. These findings underscore the complexity and interconnectedness of financial markets during black swan events, offering valuable insights for investors and policymakers alike [11, 16].

During the Russia-Ukraine crisis, global stock markets experienced dynamic shifts in volatility roles. The Japanese stock market became the primary net volatility recipient, while the U.S. stock market transformed from a transmitter to a receiver

post-conflict. The Chinese market shifted from a receiver to a transmitter role. Alam et al. [1] investigated the crisis's impact on connectedness among commodities and G7/BRIC markets, finding gold, silver, and specific countries acted as receivers of shocks. Yousaf et al. [29] reported unfavorable impacts on international markets after the "special military operation" announcement, with Hungary, Russia, Poland, and Slovakia reacting negatively. Joshi and Aggarwal [10] studied Nancy Pelosi's visit to Taiwan's effect on Asia-Pacific stock indices, revealing limited impact, particularly on Hong Kong and China. Izzeldin et al. [9] analyzed the market and commodity reactions to the war, determining its intensity as muted compared to previous crises. Kamal et al. [13] studied the Australia stock market's response, showing significantly negative abnormal returns on the event day, especially affecting small and medium-sized firms. These studies collectively highlight the nuanced and varied impacts of the Russia-Ukraine crisis on global financial markets, offering insights into volatility shifts, regional reactions, and the differential effects on various market segments.

6.3 Methodology

This research delves into the impact of the Russia-Ukraine war on specific Indian stock indices. The methodology of choice is the event study, a solid approach to uncover how these indices are reacting to the war's vibes. We're on a mission to spill the beans on their behavior during this showdown. Event studies are like the detective work of the research world, perfect for tracking how indices perform when a major event, like the Russia-Ukraine war, takes the stage. Get ready for the data and the play-by-play on how these indices are handling the heat.

6.3.1 Data Description

We're getting down to business here. We're checking out how the whole Russia-Ukraine drama is throwing shade on the Indian stock market returns, and guess what tool we're using? The event study method—it's like our investigative sidekick. We're digging into how the index prices are doing the cha-cha, showing us how this event is shaking up the market returns, going above and beyond expectations. We marked our calendars on February 24, 2022, as the day things got real—Russia kicked off its military gig against Ukraine. While this conflict is making waves globally, we've got our eyes on how it's hitting the stock indices in the Indian playground. We've got our hands on the daily closing prices, starting from Nov 01, 2021 all the way to May 31, 2022. Now, some data needed a bit of detective work and was scooped up from the https://www.nseindia.com/reports-indices-historical-index-data **NSE Database**. Overall, we've got the lowdown for a whopping 212 days.

$$R_{it} = \log \left\{ \frac{P_{i(t+1)}}{P_{it}} \right\} \tag{6.1}$$

Equation 6.1 used to compute the logarithmic return of daily sectoral indices, where P_{it} denotes the value of ith index at time t.

6.3.2 Research Model

The event study has emerged as a captivating methodology for assessing the repercussions of significant events. This approach involves the subdivision of the overall sampling period into two distinct segments: the Event Window and the Estimation Window (as shown in Fig. 6.1). In this context, the event day is defined as the official announcement of the war, specifically February 24, 2022, denoted as t_0. The Event Window encompasses a brief timeframe before and after the Event day, where the period preceding t_0 is referred to as the anticipation window, and the duration following t_0 is termed the adjustment window (See for Refs. [19, 27]).

Various models, such as constant mean return, market-adjusted, and market models, have been employed. Each model possesses its unique limitations; for instance, a constant mean model tends to inflate returns, while a market-adjusted model represents a constrained market model. The Market Model is regarded as the benchmark model in this context [27]. According to the seminal work of Brown and Warner, the market model is considered the simplest and widely applied under diverse conditions [4]. The Event Study Method has been utilized in this research to gauge the impact of the Russia-Ukraine War on each sectoral index of the NSE. Two pivotal models, namely, the Market Model and Prediction model, have been employed for analysis in this study.
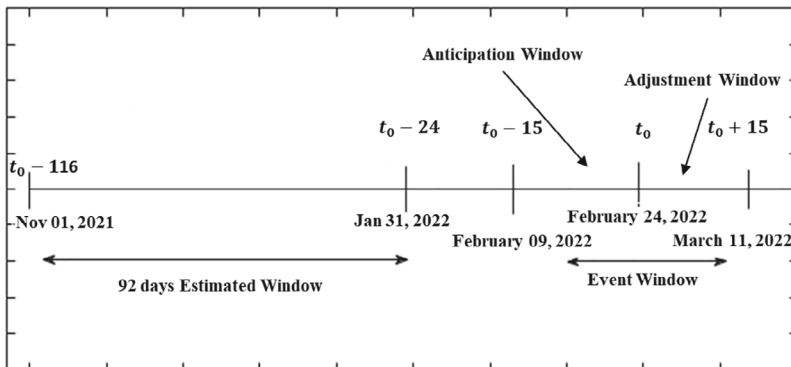
Fig. 6.1 Event time line

6.3.2.1   Market Model

The Market Model assesses the abnormal return within each sector, delineated as the variance between the sector's observed return and the calculated return derived from the market index.

The method of ordinary least squares (OLS) is employed to establish a regression line, wherein the daily return of the ith sector, R_{it}, serves as the dependent variable, and the market return, R_{mt}, serves as the independent variable during the estimation period. The anticipated return, $\hat{R}_{it}$, for each stock is calculated using Eq. 6.2.

$$\hat{R}_{it} = \alpha + \beta * R_{mt} \tag{6.2}$$

The calculation of the daily abnormal return (AR_{it}) for each sector involves determining the variance between the observed market return (R_{it}) and the estimated return ($\hat{R}_{it}$) derived from the market index, expressed in Eq. 6.3:

$$AR_{it} = R_{it} - \hat{R}_{it} \tag{6.3}$$

Moreover, we consolidate the abnormal returns (AR_{it}) over the course of the event window ($t_0 - 15, t_0 + 15$), subsequently calculating the Cumulative Abnormal Return (CAR) as follows:

$$CAR_{it(p-q)} = \sum_{t=q}^{p} AR_{it} \tag{6.4}$$

$CAR = \sum(AR_{it})$ for all time intervals between p and q, where p and q represent two distinct time points.

The calculated Cumulative Abnormal Return (CAR) offers insights into the individual stock variations during the event window. To gain a comprehensive understanding of overall market dynamics, it becomes imperative to delve into a cross-sectional investigation. This involves analyzing the collective abnormal return across all sectors. To achieve this, a cross-sectional study is conducted, computing the Average Abnormal Return (AARt) for each time instant, and subsequently deriving the Cumulative Average Abnormal Return (CAAR).

$$AAR_t = \frac{1}{k} \sum_{i=1}^{k} AR_{it} \tag{6.5}$$

$$CAAR_{(p-q)} = \sum_{t=q}^{p} AAR_t \tag{6.6}$$

In this context of Equations 6.4, 6.5, and 6.6, the variable "k" represents the total count of sectors under consideration, while "p" and "q" denote specific time instances, such that the expression ($p - q + 1$) signifies the duration of a particular time window.

Furthermore, the Cumulative Average Abnormal Return (CAAR) undergoes statistical testing to assess whether there exists a noteworthy impact on the cumulative market response as a result of the event.

6.3.2.2 Prediction Model

To prepare the data and implement a Long Short-Term Memory (LSTM) model using TensorFlow and Keras for time series forecasting, several steps are taken. Figure 6.2 demonstrates the flow diagram of the proposed Model. The MinMaxScaler from scikit-learn is employed to normalize the data, ensuring uniform scaling of all features. Utilizing Keras, a high-level neural networks API, a Sequential model is constructed, representing a linear stack of layers. The model specifically incorporates an LSTM layer with 50 units and a rectified linear unit (ReLU) activation function, chosen for its effectiveness in introducing non-linearity. The neural network's architecture is outlined with an LSTM layer to capture long-term dependencies in time series data, followed by a Dense layer to generate predicted values. To prevent overfitting during training, a Dropout layer is introduced. The dataset is split into training and testing sets, with the training set comprising 80% of the data. The LSTM layer is integrated into the model to learn patterns and dependencies in the sequential data. For evaluating the model's performance, the mean absolute error metric from scikit-learn is imported. Overall, this code establishes the groundwork for training an LSTM model on time series data, a widely adopted approach for predicting future values based on historical patterns.

Fig. 6.2 Flow diagram of LSTM prediction model
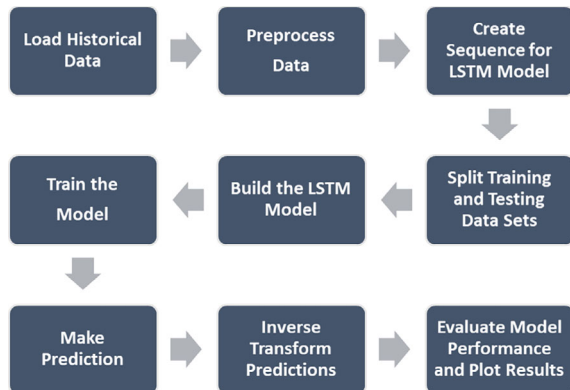
6.4 Analysis of the War's Impact on Oil and Gas Sector

In this section, we present the results of the Russia-Ukraine war on the Oil and Gas Sector of Indian economy. A big chunk of the world, including India, relies on Russia for oil because it's the second-largest crude oil exporter, right after Saudi Arabia. But, thanks to the Russia-Ukraine drama, the prices of Brent crude oil shot up to a whopping $139 per barrel in March. Luckily, it chilled out a bit and settled around $107 per barrel. Still, it's up by about 20% from the pre-war $89 per barrel. Now, with the pandemic already throwing shade, these rising oil prices could mess up manufacturing, tourism, and transportation—basically, everything. And that could trigger inflation, fiscal issues, and mess with India's external sector.

Hold up, it's not just Russia making bank here. Their oil prices dropped like it's hot due to sanctions and everyone boycotting them. To figure out how bad this whole mess is for Brent crude oil prices, we gotta look back. ICICI Securities spilled the tea, saying that with 60% of the world giving Russia the cold shoulder, we're looking at a 3 million barrels per day drop in oil-crude supply. That's keeping Brent crude prices dancing above $100 per barrel for most of the 2022 financial year. India is kinda stuck in this, importing a whopping 85% of its oil needs, and crude oil is like 9% of the whole WPI basket.

Bank of Baroda's economic whiz, Madan Sabnavis, crunched the numbers and spilled the real tea. A 10% spike in crude prices could slap on an extra 0.9% to the WPI inflation, pushing it to a hefty 11.5–12% for the financial year 2022. Oh, and every $1 more per barrel? Get ready to cough up an extra 60–70 paise for a liter of fuel. That's gonna make oil imports a whopping 25.8% of all India's imports, and that's just gonna get bigger. Plus, brace yourself for a $15 billion spike in India's CAD (around 0.4% of GDP), making that trade deficit even worse. And here's the kicker— those jacked-up fuel prices are gonna make people slam the brakes on spending. The pandemic already took a bite, dropping PFCE to Rs. 80–81 lakh crore in 2021–2022 from a cool 83.22 lakh crore in 2019–20. Rough times, my friend.

– Check out Fig. 6.3 it spills the beans. The average close price and the regular close price are doing their own thing at pretty much every spot on the chart. Now, let's dive into the close price deets. February rolls in, post-war vibes, and bam! It nose-dives to the lowest point on the scale. Fast forward to the end of March, and the close price starts shaking off that slump. It's all about that recovery game, reaching its peak performance in the middle of April. Ups, downs, and a little rollercoaster action in the stock world.
– Table 6.1 contains the Calculation of AR, CAR, AAR and CAAR of Oil and gas Sector from –15 days to +15 days on the event day.
– Check it out in Fig. 6.4 the actual and predicted prices are only vibing together at four points (1.0, 5.0, 15.0, 17.5), and everywhere else, they're doing their own thing. The actual prices are riding a rollercoaster of ups and downs too. All these ups, downs, and differences? Yeah, blame it on the war. It's like a ripple effect messing with the whole price prediction scene.

Table 6.1 Table of AR, CAR, AAR and CAAR of oil and gas sector

Date	AR	CAR	AAR	CAAR
2022-03-11	NaN	NaN	–0.002125	NaN
2022-03-10	–0.006096	–0.006096	–0.002125	–0.003048
2022-03-09	0.004877	–0.001219	–0.002125	–0.000406
2022-03-08	0.009899	0.008680	–0.002125	0.002170
2022-03-07	0.016463	0.025142	–0.002125	0.005028
2022-03-04	–0.022840	0.002302	–0.002125	0.000384
2022-03-03	–0.005404	–0.003101	–0.002125	–0.000443
2022-03-02	–0.020898	–0.023999	–0.002125	–0.003000
2022-02-28	–0.023214	–0.047212	–0.002125	–0.005246
2022-02-25	–0.017446	–0.064659	–0.002125	–0.006466
2022-02-24	0.007309	–0.057350	–0.002125	–0.005214
2022-02-23	0.004475	–0.052875	–0.002125	–0.004406
2022-02-22	–0.000954	–0.053829	–0.002125	–0.004141
2022-02-21	0.002159	–0.051670	–0.002125	–0.003691
2022-02-18	0.015841	–0.035828	–0.002125	–0.002389
2022-02-17	0.009303	–0.026525	–0.002125	–0.001658
2022-02-16	–0.008477	–0.035002	–0.002125	–0.002059
2022-02-15	–0.006388	–0.041389	–0.002125	–0.002299
2022-02-14	0.012401	–0.028988	–0.002125	–0.001526
2022-02-11	–0.007267	–0.036255	–0.002125	–0.001813
2022-02-10	–0.012123	–0.048378	–0.002125	–0.002304
2022-02-09	0.003758	–0.044620	–0.002125	–0.002028
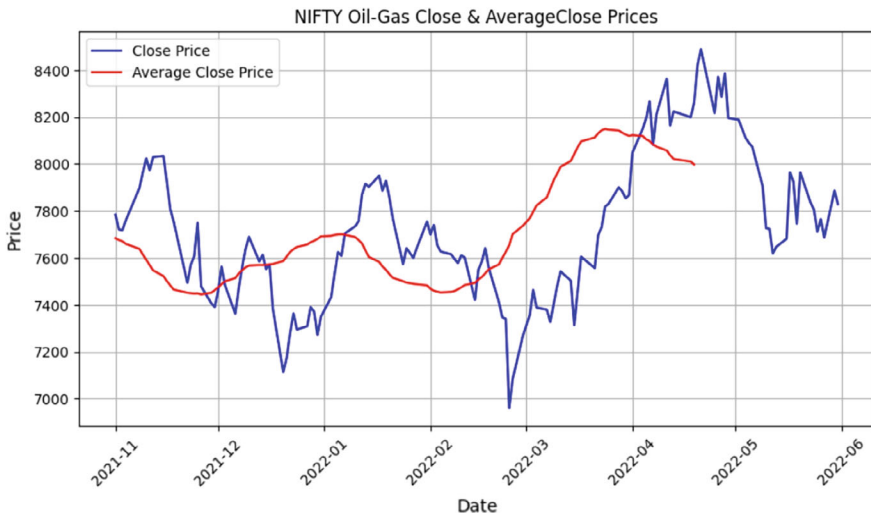
Fig. 6.3 Visualization of close and avg close price of oil and gas sector

Fig. 6.4 Performance of prediction model of oil and gas sector
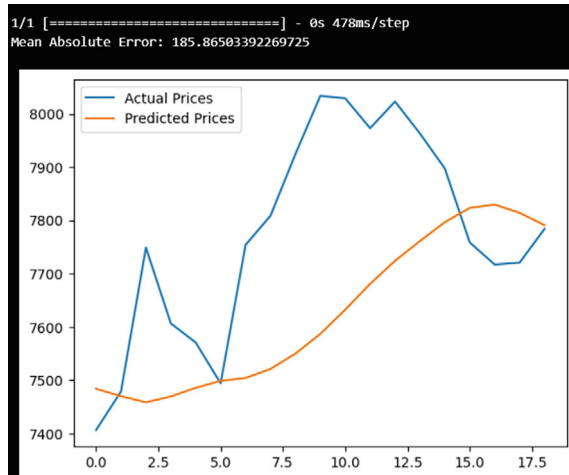
6.5   Opportunities

In the aftermath of the economic repercussions stemming from the Russia-Ukraine conflict on India, there emerges a spectrum of opportunities for strategic economic advancement [24]. The Confederation of All India Traders (CAIT) expresses apprehensions, especially in sectors heavily reliant on oil [22]. To address these concerns, India can explore fresh trade alliances, accentuating economic partnerships with nations unaffected by the conflict. Enhancing domestic manufacturing becomes pivotal, underscoring the importance of self-sufficiency amidst disruptions in global supply chains. Strategic investments in renewable energy and technological innovations open up avenues for growth, while active participation in humanitarian endeavors reflects India's dedication to global stability [2]. Giving precedence to infrastructure development, cybersecurity solutions, agricultural diversification, and environmentally sustainable initiatives aligns with evolving global patterns, presenting viable routes for economic sustainability. Effectively capitalizing on these opportunities demands meticulous strategic planning and flexibility to shape India's economic landscape post the war.

6.6   Limitations

In this study, we apply LSTM prediction model and one of the major limitations of this model is its inability to handle temporal dependencies. Therefore, if we want to apply our model with long-term dependencies model struggle to learn and predict the outputs, which decrease its performance. Also, in current study, We're zooming

in on specific stock indices in India, but there's a whole universe of other indices and markets out there waiting to be explored.

6.7 Conclusion

Despite global concerns, India's export growth is anticipated to persist in 2022, driven by factors such as reduced dependence on Russian oil and gas. Nevertheless, the ongoing Russia-Ukraine conflict poses intricate challenges for Indian policymakers due to underlying tensions. The enduring disparities between Moscow and New Delhi are unlikely to resolve swiftly, given conflicting interests. India's objectives clash with Russia's aims of strengthening China and undermining the US. To manage risks effectively, New Delhi must navigate this volatile situation with caution. A pragmatic approach is necessary, focusing on the realistic capabilities of the US rather than holding onto impractical expectations of complete withdrawal. Acknowledging the reality of a multipolar world is vital for crafting an effective strategy in this evolving geopolitical landscape. This revised version succinctly captures the original text's essence, emphasizing India's export advantage, the challenges posed by the conflict, and the need for a realistic approach towards the US. It also clarifies the nature of tensions between India and Russia.

The full-scale Russia-Ukraine war and the subsequent harsh sanctions on Russia, following its invasion of Ukraine, had a global impact on stock markets, including India's Nifty 50, which experienced a sharp decline [17]. This study empirically elucidates the rejection of the null hypothesis at a 0.1 level of significance. The surge in oil prices, coupled with the escalation of other commodity prices, exacerbated the situation, causing panic among investors [12]. The fear of a world war, rising crude oil prices, and the potential for inflation had notable effects on the economy and market performance. As per the Efficient Market Hypothesis (EMH), markets reflect available information and adapt accordingly. This illustrates how markets respond to global events, with statistical evidence indicating a negative impact on Nifty 50 index values.

References

1. Kausar Alam, Md., Tabash, M.I., Billah, M., Kumar, S., Anagreh, S.: The impacts of the russia-ukraine invasion on global markets and commodities: A dynamic connectedness among g7 and bric markets. J. Risk Financ. Manag. **15**(8), 352 (2022)
2. Banerji, A.: The impact of the war in ukraine on the indian economy. In: The Economic Impact of the Ukraine-Russia War on India (2022)
3. Bhattacharjee, A., Gaur, D., Gupta, K.: Russia-ukraine war and the impact on indian economy. J. Econ, Stud (2023)
4. Brown, S.J., Warner, J.B.: Measuring security price performance. J. Financ. Econ. **8**(3), 205–258 (1980)

5. Caporale, G.M., Gil-Alana, L.A., Poza, C.: The covid-19 pandemic and the degree of persistence of us stock prices and bond yields. Q. Rev. Econ. Finance **86**, 118–123 (2022)
6. Chen, M.H., Jang, S.S., Kim, W.G.: The impact of the sars outbreak on taiwanese hotel stock performance: an event-study approach. Int. J. Hosp. Manag. **26**(1), 200–212 (2007)
7. Gursoy, D., Chi, C.G.: Effects of covid-19 pandemic on hospitality industry: review of the current situations and a research agenda. J. Hosp. Market. Manag. **29**(5), 527–529 (2020)
8. He, P., Sun, Y., Zhang, Y., Li, T.: Covid-19's impact on stock prices across different sectors-an event study based on the chinese stock market. Emerg. Mark. Financ. Trade **56**(10), 2198–2212 (2020)
9. Izzeldin, M., Muradoğlu, Y.G., Pappas, V., Petropoulou, A., Sivaprasad, S.: The impact of the russian-ukrainian war on global financial markets. Int. Rev. Financ. Anal. **87**, 102598 (2023)
10. Joshi, N.A., Aggarwal, A.: Nancy pelosis visit to taiwan and stock indices responses: an event study and panel data analysis for asia-pacific and selected markets. J. Commerce Account. Res. **13**(1), 45–52 (2024)
11. Joshi, N.A., Baker, H.K., Aggarwal, A.: The russia-ukraine war's impact on stock markets in asia-pacific, the united states, and europe. Eurasian J. Busin. Econ. **16**(31), 39–53 (2023)
12. Lokesharri, S.K.: The russia-ukraine conflict escalates; what next for the stock market? (2022). (Accessed 24 Feb 2022)
13. Rajib Kamal, M.d., Ahmed, S., Hasan, M.M.: The impact of the russia-ukraine crisis on the stock market: evidence from australia. Pacific-Basin Financ. J. **79**, 102036 (2023)
14. Kar, S.K.: One year of russia-ukraine war: how the conflict impacted indian economy (2023). (Accessed 24 Feb 2023)
15. Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., Chang, H.: The architecture of sars-cov-2 transcriptome. Cell **181**(4), 914–921 (2020)
16. Lok Sabha Secretariat New Delhi India.: Impact of Russia-Ukraine War onIndian Economy. Research and Information Division, Parliament Library And Reference, Research, Documentation And Information Service (LARRDIS) (2022)
17. Makwana, B.: Indian markets are the second worst performers this week in asia hit by the russia-ukraine war (2022). Updated 04 March 2022)
18. Mathew, G.: Russia-ukraine war: impact on markets and risks ahead (2022). (Accessed 24 Feb 2022)
19. Maurya, P.K., Bansal, R., Mishra, A.K.: Russia-ukraine conflict and its impact on global inflation: an event study-based approach. J. Econ. Stud. ahead of print, 1–23 (2023)
20. Nazeeruddin, M.: The russia-ukraine war crisis -it's impact on indian economy 04 (2022)
21. Ray, S.S.: Russia-ukraine war impact: Indian steelmakers set to make the most of alloy's global price rally (2022). (Accessed 4 Mar 2022)
22. Sharma, A.: Indian economy: a tale of recovery amid russia-ukraine war. The economic implications for India amid the Russia-Ukraine war, including potential impacts on trade, inflation, and various sectors such as oil, manufacturing, and FMCG (2022)
23. Sharma, S.: 100 days of russia-ukraine war: How has india been affected (2022)
24. Tellis, A.J.: What is in our interest: India and the ukraine war. In: International Relations, Geopolitics, and India's Foreign Policy in Response to the Russian Invasion of Ukraine (2022)
25. Vasisht, P.: Impact of russia-ukraine war on india's oil and gas sector (2022). (Accessed 22 Dec 2022)
26. Verma, M.: How will the russia-ukraine war impact the indian economy? March 14 (2022)
27. Verma, Y., Venkataramani, R., Kayal, P., Maity, M.: Short-term impact of covid-19 on indian stock market. J. Risk Financ. Manag. **14**, 558 (2021)
28. Wajid, A., Gupta, K.: Stock market responses to the covid-19 health crisis: evidence from the world's largest economies. Int. J. Busin. Anal. **9**(1), 1–19 (2022)
29. Yousaf, I., Patel, R., Yarovaya, L.: The reaction of g20+ stock markets to the russia-ukraine conflict "black-swan" event: evidence from event study approach. J. Behav. Exp. Financ. **35**, 100723 (2022)

Chapter 7
Optimal Stacking Method with GA-Based Feature Selection for Multi-Label Classification

Hemavati and **R Aparna**

Abstract This study proposes an optimal stacking method with genetic algorithm (GA) based feature selection for multi-label classification problems. The most prevalent form of ensemble learning is stacking, which combines the results of multiple base classifiers to improve predictive performance. However, stacking requires careful selection of base classifiers and their hyperparameters, and feature selection. We proposed the use of a feature selection algorithm driven by genetic algorithms (GA) to pinpoint a subset of significant features. This subset is intended to enhance the working capability of the stacking model. The effectiveness of the model is assessed across various benchmark multi-label datasets. The performance results are showing that it outperforms several state-of-the-art multi-label classification methods. Additionally, we experiment to make the analysis of the performance of different hyperparameters and ensemble configurations on the impact of the proposed method. Overall, the proposed optimal stacking method with GA-based feature selection provides a promising solution for multi-label classification problems.

Keywords Population · Stacking · Multi-label · Multi-label · Features

7.1 Introduction

Multi-label classification problems arise in various real-world applications like picture identification, text classification, and bioinformatics. Multi-label learning is making it a more challenging problem than in binary or multi-class classification. In the past few years, several machine-learning algorithms have been suggested to solve multi-label classification problems. However, selecting an optimal set of features

and designing an effective model architecture remains a challenging task. Ensemble learning methods [2] resulted in solving multi-label classification problems by combining multiple models to achieve improved performance. The stacking method is one such ensemble method that utilizes a meta-model to get the combinations of different base models' output. However, the performance of stacking methods can be affected by the caliber of the basic models and the selected features. Therefore, it becomes an essential requirement for developing an optimal stacking method that can effectively select relevant features and combine multiple models to enhance the model's overall performance. However, the performance of stacking methods can be affected by the caliber of the basic models and the selected features. Therefore, it becomes an essential requirement for developing an optimal stacking method that can effectively select relevant features and combine multiple models to enhance the model's overall performance. However, the performance of stacking methods can be affected by the caliber of the basic models and the selected features. Therefore, it becomes an essential requirement for developing an optimal stacking method that can effectively select relevant features and combine multiple models to enhance the model's overall performance. An optimal stacking method with genetic algorithm-based feature selection for multi-label classification. This method uses a optimal algorithm to choose the most appropriate feature and combines the outputs of multiple base models to train a meta-model. An optimal stacking method with genetic algorithm-based feature selection for multi-label classification. This method uses a optimal algorithm to choose the most appropriate feature and combines the outputs of multiple base models to train a meta-model. The proposed method aims to improve the accuracy and robustness of the model by effectively selecting relevant features and combining multiple models. We assess the proposed approach across multiple multi-label classification datasets and conduct comparisons with several methods that have shown competitive results in existence concerning their effectiveness. The experimental findings indicate that our method surpasses several existing approaches in terms of both accuracy and robustness. The suggested approach is versatile and applicable across various real-world domains, encompassing bioinformatics, text classification, and image recognition.

7.2 Related Work

In recent years, multi-label classification has received significant attention in the realm of machine learning, it has become notable for its broad range of applications, encompassing various fields, with a specific emphasis on image classification, text classification, and bioinformatics. The suggestion of various methods has been provided to tackle the challenge of multi-label classification, which incorporate decision trees, support vector machines, and neural networks [16]. However, the effectiveness of these models is significantly influenced by the feature selection procedure. One

promising approach to address the feature selection problem in multi-label classification is the optimal stacking method with genetic algorithm-based feature selection. This method combines multiple classification models and optimally weights their predictions to improve the overall accuracy of the classification. Moreover, it employs a genetic algorithm-based feature selection technique to identify the most informative features for the classification task. Granular MLFS [1]: This technique aims to choose a more condensed feature subset by utilizing information granules associated with the labels rather than assessing the complete label set. In contrast, it focuses on the relationship between features and label distributions to make feature selection decisions. In [2], authors employed Multi-label for online criterion selection. It considers and selects features that are dominant, meaning they are better than others in at least one aspect without being worse in any other aspect. In [3], authors discussed about Distributed Models with Mutual Information. These models are created for distributed computing utilizing Apache Spark. The information which quantifies the statistical dependence between features and labels. This method leverages parallel computing to handle large datasets efficiently. In [4], authors introduced MLFS, focusing on entropy approximation cardinality generalization. Users can choose the level of approximation precision depending on how accuracy and computation expenses get traded off. These filter-based methods have a limitation. that they operate on different methods, which means they may not consider the specific requirements and characteristics of the learning algorithm that will be applied after feature selection. This can result in sub-optimal feature subsets for the actual classification task. Wrapper-based methods [15], in contrast to the existing, typically consider the learning algorithm and use it to guide the process for feature consideration, potentially leading to better classification performance. The Quality assessment potential feature subsets generated by specific learning algorithms, like a multi-label Naive Bayes classifier [5]. These approaches evaluate the effectiveness of potential feature subsets produced by specific learning algorithms, such as NB classifier (Naive Bayes for multi-label classification) [6]. Typically, they outperform filter-based techniques in relation to categorization precision [7]. These approaches evaluate the effectiveness of potential feature subsets produced by specific learning algorithms, such as NB classifier (Naive Bayes for multi-label classification) [6]. Typically, they outperform filter-based techniques in relation to categorization precision [7]. These approaches evaluate the effectiveness of potential feature subsets produced by specific learning algorithms, such as NB classifier (Naive Bayes for multi-label classification) [6]. Typically, they outperform filter-based techniques in relation to categorization precision [7]. In the domain of wrapper-based techniques, genetic feature generation method, due to their ability to conduct stochastic global searches. In the domain of wrapper-based techniques, sample-based algorithms for searching techniques are commonly utilized for feature selection due to their ability to conduct stochastic global searches. Sample-based algorithms for searching techniques are commonly utilized for feature selection due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are commonly utilized for feature selection due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are typically

applied when choosing features due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are typically applied when choosing features due to their ability to conduct stochastic global searches. Sample based algorithms for techniques are typically applied when choosing feature due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are typically applied when choosing features due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are typically applied when choosing features due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are typically applied when choosing features due to their ability to conduct stochastic global searches, searching techniques are typically applied when choosing features, techniques are typically applied when choosing features, due to their ability to conduct stochastic global searches. Sample based algorithms for searching methods that are usually considered for selecting features due to their ability to conduct stochastic global searches. Sample based algorithms for searching techniques are commonly utilized for feature selection due to their ability to conduct stochastic global searches. sample based algorithms for searching methods which are usually used for selecting features due to their ability to conduct stochastic global searches. For instance, [8] introduced a novel method. Additionally, in [9] introduced different variants of binary. However, these single searching methods which are usually used for selecting features are susceptible to premature convergence within the population, which limits their search capabilities. Although methods like the multi-label crowding genetic algorithm can help prevent early convergence, but rely on the initial setup of the population. In order to address these challenges, binary classification has turned to multi-label that involves using multiple distinct sub-populations. In [10], authors have introduced a genetic algorithm rooted in the concept of tribe competition, which strives to maintain solution diversity by permitting to produce sub-sample set of features with varying numbers of features. It also ensures a comprehensive search space exploration by considering computational resources among the samples.

7.3   Proposed Method

Figure 7.1 shows the architecture of proposed method.

7.3.1   Genetic Algorithm(GA)

This algorithm is a type of meta heuristic algorithm based on the ideas of genetics and natural selection. It is frequently applied to issues related to optimization, in which identifying the optimum answer from a wide range of options is the main objective. The GA algorithm starts with a population of candidate solutions, presented as a group of strings or binary codes. The population is evaluated using a fitness function
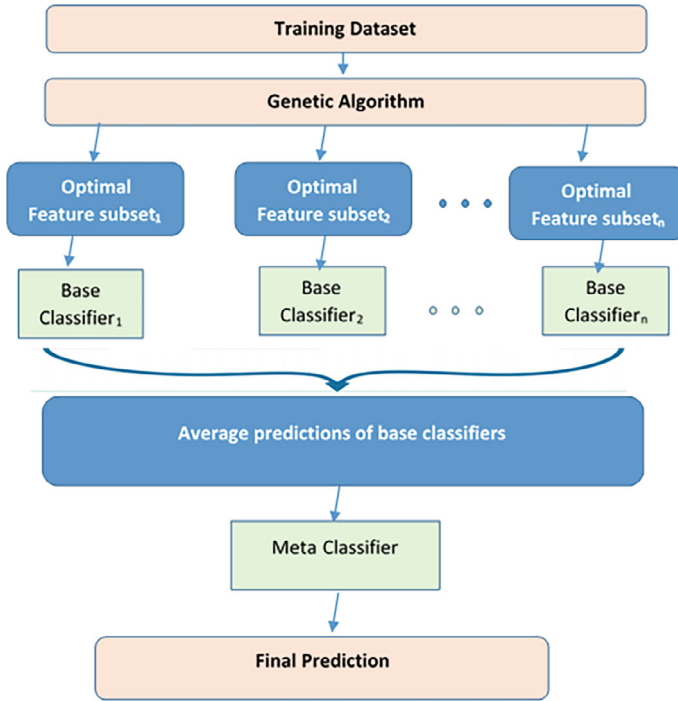
Fig. 7.1 Architecture of proposed method

that measures the quality of each solution. The fitter solutions are more likely to survive and reproduce, passing their genetic information to the new generation. The algorithm consists of several key steps:

1. Initialization: Create an initial population of candidate solutions randomly.
2. Evaluation: Evaluate the fitness of each solution using a fitness function.
3. Selection: Select the fittest individuals from the population to be the parents of the next generation.
4. Crossover: Combine the genetic information of the parents to create new offspring solutions.
5. Mutation: Randomly change some of the genetic information of the off-spring solutions to introduce diversity.
6. Replacement: Replace the least fit pattern in the population with the new offspring solutions.
7. Termination: Stop the algorithm when a stopping criterion is met, such as a maximum number of generations or a satisfactory level of fitness.

Genetic algorithms can be utilized in a wide range of problems, including optimization, feature selection, clustering, and classification. In machine learning and data mining the GA algorithm has been utilized widely to find the best feature subset

or hyperparameters for a model. One of the advantages of a genetic algorithm is its ability to handle complex and high-dimensional search spaces, where other optimization methods may struggle. However, it can also suffer from slow convergence and premature convergence if not properly tuned. The most optimistic with huge solution set with less amount of time for overall performance of the genetic algorithm.

7.4 Mathematical Model

Let X be the input matrix of size NxM, where N is the number of samples and M is the number of features. Let Y be the output matrix of size N x L, where L is the number of labels. Each element of Y is either 0 or 1, indicating the absence or presence of the corresponding label, respectively.

Step 1: Feature Selection using Genetic Algorithm

Let F be the set of all possible feature subsets. Let f be a feature subset of size K, where $K << M$. The fitness function of f can be defined as

$$fitness(f) = accuracy(f) - \alpha * complexity(f) \tag{7.1}$$

where $accuracy(f)$ is the accuracy of the classifier trained on f, and $complexity(f)$ is the number of features in f. The parameter α controls the trade-off between accuracy and complexity. The genetic algorithm can be used to search for the feature subset with the highest fitness score. The population of feature subsets can be initialized randomly, and the genetic operators can be used to generate new subsets.

Step 2: Training Base Classifiers

Let C be the set of base classifiers, each trained on a different feature subset. Let c_i be the ith base classifier, trained on feature subset f_i. The output of c_i for a sample x_j can be denoted as

$$y_{ji} = c_i(x_j) \tag{7.2}$$

The set of predictions for sample x_j across all base classifiers can be denoted as

$$Yj = y_{j1}, y_{j2}, ..., y_{jk} \tag{7.3}$$

Step 1: Stacking Ensemble Method

Let $w = [w_1, w_2, ..., w_k]$ be the weight vector learned using the validation set. The predicted labels for a sample x_j can be calculated as

$$y_j = sign(sum(wi * y_{ji)}) \tag{7.4}$$

where $sign()$ is the sign function, returning 1 if the input is positive and -1 if the input is negative.

The weights w can be learned using a variety of methods, such as linear regression, logistic regression, or support vector regression. The validation set can be used to tune

the hyperparameters of the weight learning algorithm. Overall, the optimal stacking method with genetic algorithm-based feature selection for multi-label classification involves searching for the optimal feature subset using a genetic algorithm, training multiple base classifiers on different feature subsets, and combining their predictions using a stacking ensemble method with learned weights.

7.4.1 Algorithm for Classifying Multi-Label Data Using The Stacked Ensemble Approach

---

Algorithm 7.1 : Algorithm for classifying multi-label data using the stacked ensemble approach

---

1: **Input:** $D = \{X_i, Y_i\}_{i=1}^{n}$, where $X_i \in R^n$, $Y_i \subseteq Y$ (set of all labels).
2: **Output:** Stacked Ensemble classifier
begin
3: **Step 1:** Apply the Genetic Algorithm GA(X) as as stated in Section 4.
4: **Step 2:** Convert the multi-label dataset into a binary classification problem using the
5:       Binary Relevance (BR) technique.
6:       $D_{BR} \leftarrow BR(D_{GA})$
7: **Step 3:** Utilize the fundamental classifiers on D_{BR}
8:       **for i ← 1 to m**
9:          Develop the foundational classifier BC_i based on D_{BR}.
10:      **end for**
11: **Step 4:** Create a new dataset based on the results produced by the base classifier.
12:       **for j ← 1 to m**
13:          Create a new dataset D_2 with the following contents
14:          $D_2 = (X_i', Y_i)$
15:          where $X_i' = \{BC_1(X_i), BC_2(X_i), ..., BC_m\}$
16:          and Y_i from initial dataset D.
17:      **end for**
18: **Step 5:** Prepare the meta learner using the newly created dataset D_2 obtained in Step 4 to make final predictions.
19:          This entails learning the meta learner (ML) on D_2.
20: **Step 6:** return ML
end

7.5 Experimental Analysis

We carried out experiments using 10 multi-label datasets from diverse domains, which are described in Table 7.1, and can be accessed at the following URL: http://mulan.sourceforge.net/datasets-mlc.html.

Table 7.1 Standard multi-label datasets of different domain

Name	Domain	Instances	Nominal	Numeric	Labels	Cardinality	Density	Distinct
Bibtex	Text	7395	1836	0	159	2.402	0.015	2856
Birds	Audio	645	2	258	19	1.014	0.053	133
Delicious	Text (web)	16105	500	0	983	19.020	0.019	15806
Emotions	Music	593	0	72	6	1.869	0.311	27
Enron	Text	1702	1001	0	53	3.378	0.064	753
Flags	Images (toy)	194	9	10	7	3.392	0.485	54
Genbase	Biology	662	1186	0	27	1.252	0.046	32
Mediamill	Video	43907	0	120	101	4.376	0.043	6555
Yeast	Biology	2417	0	103	14	4.237	0.303	198

7.5.1 Datasets

7.5.2 Performance Metrics

Performance metrics play a crucial role in evaluating the effectiveness of machine learning models. In this report, we will discuss and analyze three key performance metrics: One Error, Hamming Loss, and Accuracy. These metrics are particularly relevant in the context of classification problems, where the goal is to correctly assign instances to predefined classes.

One Error: It is also known as Zero-One Loss, is a binary classification metric that measures the fraction of misclassified instances. It is a straightforward and intuitive metric, as it assigns a value of 1 to misclassifications and 0 to correct classifications.

$$one_error = \frac{Number\ of\ missclassification}{Total\ Number\ of\ Instances} \tag{7.5}$$

Hamming Loss: Hamming Loss is a metric commonly used for multi-label classification problems, where each instance can belong to multiple classes. It measures the fraction of labels that are incorrectly predicted. Unlike One Error, Hamming Loss considers partial matches, making it suitable for scenarios where instances may belong to more than one class.

$$hamming_{loss} = \frac{1}{N}\Sigma_{i=1}^{N}\frac{XOR(y_{i,true}\ y_{i,pred})}{L} \tag{7.6}$$

Accuracy: Accuracy is a widely used and intuitive metric that measures the proportion of correctly classified instances among all instances. It is suitable for balanced datasets and binary classification problems.

$$accuracy = \frac{Number\ of\ Correct\ Classifications}{Total\ number\ of\ Instances} \tag{7.7}$$

Table 7.2 Comparative results of hamming loss for different methods

Dataset	Proposed	TCbGA	EMPNGA	BCO-MDP
Bibtex	0.0629	0.0635	0.0642	0.0638
Birds	0.0342	0.0359	0.0358	0.0358
Delicious	0.0626	0.0638	0.0638	0.0641
Emotions	0.2336	0.2370	0.2376	0.2366
Enron	0.0663	0.0628	0.0892	0.0840
Flags	0.0443	0.0444	0.0449	0.0447
Genbase	0.0074	0.0338	0.0315	0.0277
Mediamill	0.1341	0.1372	0.1416	0.1396
Yeast	0.2208	0.2233	0.2253	0.2241

7.5.3 Experimental Results

We performed a comparative study of the proposed method against three leading multi-population-based techniques known for their effective performance in tackling collection of features challenges: TCBGA [11], EMPnGA [12], and BCOMDP [13]. We configured the specifications for every technique to match those utilized in their respective initial studies [14]. To ensure justice in the evaluation process, we established a limit of 300 for the highest quantity of permitted FFCs (Feature Fitness Calls) and the highest quantity of selected features at 50. The total population size was consistently maintained at 50. We employed the MLNB (Multi-Label Naive Bayes) classifier along with a holdout cross-validation method to assess the quality of the feature subsets generated by every technique. For this purpose, we divided each dataset into training (80%) and test (20%) sets. The size of the population as a whole was consistently maintained at 50. We employed the MLNB (Multi-Label Naive Bayes) classifier along with a holdout cross-validation method to assess the characteristic subsets' quality generated by each method. For this purpose, we divided each dataset into training (80%) and test (20%) sets. Each experiment was conducted 10 times, and the results were averaged for analysis. Each experiment was conducted 10 times, and the results were averaged for analysis. In our proposed method, we implemented five sub-populations, with each sub-population comprising 10 individuals. Tables 7.2, 7.3 and 7.4 show the comparative results of Hamming Loss, One Error, and Accuracy respectively. For this purpose, we divided each dataset into training (80%) and test (20%) sets. Each experiment was conducted 10 times, and the results were averaged for analysis. Each experiment was conducted 10 times, and the results were averaged for analysis. In Table 7.3, The suggested approach demonstrated superior performance compared to TCbGA and BCO-MDP in relation to the one-error measure. Tables 7.2 and 7.4 show the significant performance of the proposed algorithm. The evaluation of the three metrics relies on the predicted label subsets. Therefore, the new method has the potential to surpass existing methods. Figs. 7.2, 7.3 and 7.4 shows the comparison graph.

Table 7.3 Comparative results of one error for different methods

Dataset	Proposed	TCbGA	EMPNGA	BCO-MDP
Bibtex	0.7354	0.7254	0.7684	0.7640
Birds	0.3980	0.3985	0.3833	0.3935
Delicious	0.5874	0.5986	0.6074	0.5982
Emotions	0.2336	0.2370	0.2376	0.2366
Enron	0.663	0.6528	0.592	0.540
Flags	0.3187	0.2954	0.866	0.4237
Genbase	0.764	0.338	0.5315	0.6277
Mediamill	0.3541	0.372	0.1416	0.3396
Yeast	0.2382	0.2440	0.2672	0.2513

Table 7.4 Comparative results of accuracy for different methods

Dataset	Proposed	TCbGA	EMPNGA	BCO-MDP
Bibtex	0.945	0.337	0.478	0.6518
Birds	0.342	0.359	0.358	0.358
Delicious	0.626	0.638	0.638	0.641
Emotions	0.2336	0.2370	0.2376	0.2366
Enron	0.3466	0.3391	0.3183	0.3398
Flags	0.5768	0.5685	0.5717	0.5746
Genbase	0.8765	0.338	0.3615	0.2577
Mediamill	0.6641	0.4972	0.5656	0.6796
Yeast	0.6208	0.4233	0.5253	0.5341

Fig. 7.2 Hamming loss on different datasets
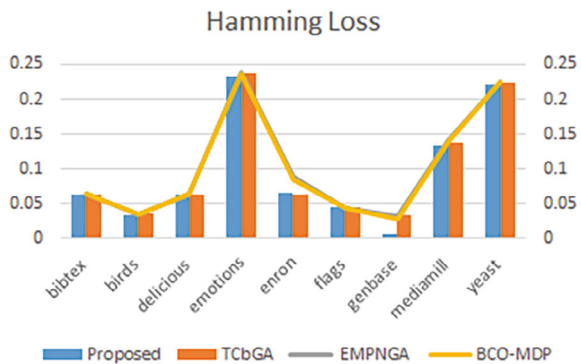
Fig. 7.3 One error on different datasets

One Error
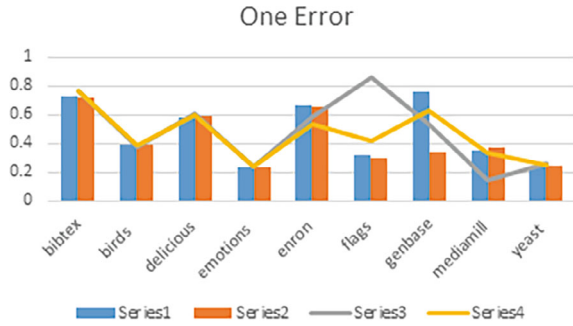
Fig. 7.4 Accuracy on different datasets

Accuracy
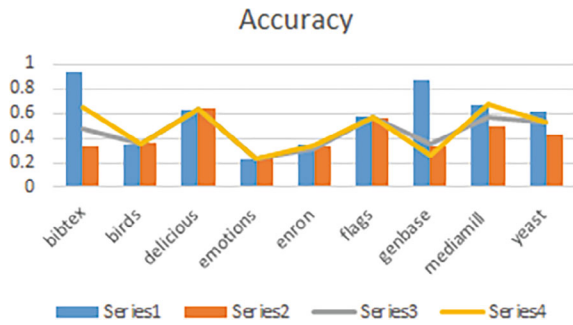
7.6   Conclusion

In our research, we introduced a novel Genetic Algorithm with a focus on label complementary communication, designed specifically for addressing the Multi-Label Feature Selection (MLFS) problem. The core aim is to enhance the search capabilities of sub-populations by enabling them to communicate and leverage their complementary discrimination abilities across multiple labels. Our empirical findings and statistical analyses have unequivocally demonstrated that our method significantly outperforms three contemporary Selecting features based on multiple populations approaches across 10 multi-label datasets. Future research endeavors could concentrate on mitigating one limitation inherent in our suggested technique: Initially, we have fixed the number of complementing indicators to be fifty percent of the overall number of labels. This value can be dynamically changed based on how each label's discriminatory strength changes as the search moves on. For instance, the label complementary communication could be implemented only for labels with discrimination performance not exceeding that of the previous generation.

References

1. Author, F.: Article title. Journal **2**(5), 99–110 (2016)
2. Hemavati, et al.: J. Phys: Conf. Ser. **2161**, 012074 (2022). https://doi.org/10.1088/1742-6596/2161/1/012074
3. Li, F. et al.: Granular multi-label feature selection based on mutual information. Pattern Recognit. **67**, 410–423 (2017)
4. Kashef, S., Nezamabadi-pour, H.: A label-specific multi-label feature selection algorithm based on the Pareto dominance concept. Pattern Recognit. **88**, 654–667 (2019)
5. González-López, J. et al.: A distributed selection of continuous features in multilabel classification using mutual information. IEEE Trans. Neural Netw. Learn. Syst. **31**, 2280–2293 (2019)
6. Zhang, M.L et al.: Feature selection for multi-label naive Bayes classification. Inf. Sci. **179**, 3218–3229 (2009)
7. Lu, Y., et al.: Improved particle swarm optimization algorithm and its application in text feature selection. Appl. Soft Comput. **35**, 629–636 (2015)
8. Mafarja, M., et al.: Whale optimization approaches for wrapper feature selection. Appl. Soft Comput. **62**, 441–453 (2018)
9. Dong, H., et al.: A novel hybrid genetic algorithm with granular information for feature selection and optimization. Appl. Soft Comput. **65**, 33–46 (2018)
10. Lim, H., et al.: MFC: initialization method for multi-label feature selection based on conditional mutual information. Neurocomputing **382**, 40–51 (2020)
11. Breaban, M., Luchian, H.: A unifying criterion for unsupervised clustering and feature selection. Pattern Recognit. **44**, 854–865 (2011)
12. Ma, B., Xia, Y.: A tribe competition-based genetic algorithm for feature selection in pattern classification. Appl. Soft Comput. **58**, 328–338 (2017)
13. Zhang, W., et al.: A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring. Expert Syst. Appl. **121**, 221–232 (2019)
14. Wang, et al.: Feature selection for classification of microarray gene expression cancers using Bacterial Colony Optimization with multi-dimensional population. Swarm Evol. Comput. **48**, 172–181 (2019)
15. Maldonado, B., et al.: A review of recent approaches on wrapper feature selection for intrusion detection. Expert Syst. Appl. **198**, 116822 (2022). https://doi.org/10.1016/j.eswa.2022.116822
16. Binkhonain, et al.: A review of machine learning algorithms for identification and classification of non-functional requirements. Expert Syst. Appl.: X. **1**, 100001 (2019). https://doi.org/10.1016/j.eswax.2019.100001

Chapter 8
Deep Learning-Based Facial Expression Recognition System for Age, Gender, and Emotions

C. Sheeba Joice ⓘ**, S. Hitha Shanthini** ⓘ**, R. Chandru, T. Devadharshan, and D. Eswar**

Abstract Human–Computer Interaction (HCI) is an intelligent way, which aims at creating scalable and flexible solutions. Large corporations and technological firms support HCI because it enables them to benefit from infrastructure and technology that is available on demand for information-centric applications without requiring the usage of public clouds. Facial expression recognition and software-based facial expression recognition systems are essential due to their ability to mimic human coding abilities. This study provides two approaches for estimating age and gender characteristics from human faces and suggests a system for identifying the emotional state of individuals given facial expressions. This study aims to further understand how human age and gender affect face expressions. Currently, the model can distinguish between seven different emotions based on a person's facial data: anger, disgust, happiness, fear, sadness, surprise, and neutral. The proposed system is divided into three segments: (a) Gender Detection. (b) Age Detection. (c) Emotion Recognition. Additionally, we have made use of pre-trained models (Transfer Learning) and the designs of various deep learning models, like CNN. The model performance in terms of the Recognition system's accuracy is displayed by the real-time evaluation metrics.

Keywords Facial Expression Recognition (FER) · Deep learning · Age · Gender · Emotions

C. Sheeba Joice · S. Hitha Shanthini (✉) · R. Chandru · T. Devadharshan · D. Eswar
Department of Electronics and Communication Engineering, Saveetha Engineering College, Thandalam, Chennai, India
e-mail: hithashanthini@saveetha.ac.in

C. Sheeba Joice
e-mail: sheebajoice@saveetha.ac.in

8.1 Introduction

Facial analysis has gained much recognition in the computer vision community in the recent past. Facial features can provide details about an individual's identity, race, age, gender, and emotional condition. Applications that employ age, gender, and emotion to classify people include forensic art, cosmetology, electronic customer relationship management, security and video surveillance, biometrics, electronic vending machines, and human–computer interaction [1]. Human emotions cover a wide range, from joy and excitement to grief and fear, and are the result of a dynamic interaction between physiological and psychological reactions. Emotions are essential for mental health in general, memory formation, and decision-making. Fostering empathy, establishing connections, and negotiating the complexities of interpersonal relationships all depend on our ability to recognize and comprehend this complex tapestry of emotions. There are still some unsolved concerns about age, gender, and emotion classification. Estimates of the age, gender, and emotion of unfiltered real-life faces are still not up to par with the requirements of commercial and real-world applications, even with continuous improvements in computer vision algorithms. Many approaches have been put out in recent years to address the classification challenge.

Recently, algorithms based on deep learning have demonstrated promising results in this domain, particularly when it comes to classifying the age, gender, and emotion of unfiltered facial photos. Therefore, we propose a unique end-to-end deep learning-based classification model that predicts the age, gender, and emotion of unfiltered in-the-wild facial photos [2]. This model is based on the current efforts in age and gender, as well as emotion classification and encouraging signals of improvement in CNN and deep learning. To train the CNN model [3] to predict age and gender from a face image, we structure the age and gender classification challenge as a classification issue.

The photos include a wealth of information. The data in the picture will be changed for research. The articulations that make up a face have a lot of information when used in a facial identification technique. Additionally, the gender specification is illustrated in Fig. 8.1.

Developing concepts helps define certain boundaries. Grouping is used to calculate age and emotion. Each year 7 labeled emotions are replaced in a different category. It is challenging to connect the faces in the pictures because they are so different from one another. Different methods are applied to ascertain the gender and age, as well as the emotions of various faces. The neural network's features are extracted via the convolution network. Deep learning is a subset of machine learning. The manual extraction of relevant features from photos is the initial stage in a machine learning approach [4]. The features are then used to create a model that categorizes the age and gender, as well as the emotion in the image. With a deep learning workflow, relevant features are automatically extracted from images. In addition, deep learning performs "end-to-end learning"—where a network is given raw data and a task to perform, such as classification, and it learns how to do this automatically.
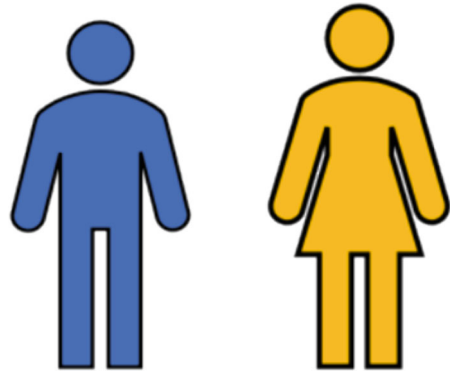
This paper is organized according to its content. Literature reviews are covered in the second section. Block diagram design is covered in the third section. The methodology is described in the fourth section. The fifth section discusses the results. The sixth section discusses the Conclusion. Section Seven discusses the future scope, while section eight covers the acknowledgments.

8.2 Literature Review

The authors of this study create a specialized network to identify emotions in pictures. To enhance recognition, they integrate context with local emotional regions. They outperform conventional techniques by encoding features using a kernel-based graph attention network. On a variety of datasets [5], their method performs better than earlier models in sentiment classification and emotion recognition. They make use of multiscale feature learning modules and a multiscale learning network with region detection. This entails creating effective regions, using a kernel-based graph network to encode features, and using Faster RCNN to detect them. High accuracy is achieved on datasets such as CAER-S and FI-8, indicating the effectiveness of the approach.

The significance of facial feature-based pain expression recognition for diagnosis and recovery in healthcare is discussed in this paper. Because of practical constraints, automated techniques are essential. The study focuses on the UNBC-McMaster database with labeled video data using facial action coding, brain imaging, and computer vision [6]. Using deep CNNs and LSTM for temporal insight, the method improves pain detection and possibly even emotion recognition. Preprocessing, standardizing labels, and optimizing a VGG-16 CNN pre-trained with Faces are highlighted in the research. The study addresses data imbalance and compares results with previous methods for both processed and raw faces.

For facial expression recognition (FER), a novel method known as "Deep Comprehensive Multi-Patches Aggregation Convolutional Neural Networks (DCMA-CNNs)" is presented in this paper. Two well-known datasets are used to assess the

method: Japanese Female Facial Expression (JAFFE) and Extended Cohn-Kanade (CK+). Two CNN branches—one for holistic features from the entire image and another for local features from overlapping patches—are used by the framework [7]. The combination of these features improves performance in both representation and classification. To handle variations such as illumination and rotation, the model makes use of Expressional Transformation-Invariant (ETI) pooling. The suggested approach demonstrates its efficacy in FER with an average recognition accuracy of 93.46% on CK+ and 94.75% on JAFFE. The study employs tenfold cross-validation for a comprehensive evaluation and emphasizes the significance of salient patch learning.

With more than a million real-world facial photos, AffectNet is a huge collection that highlights arousal, valence, and expressions. It was put together by using six different languages' emotional keyword searches on the Internet. In contrast to earlier databases that concentrated on posed expressions, AffectNet records actual, impromptu facial expressions [8]. It classifies images and assesses their valence and arousal levels using deep neural networks. When comparing human annotation errors to automated methods in other datasets, the complexity of the dataset was a contributing factor. Additional facial attributes were collected for a subset of images using the Microsoft Cognitive Face API. It may be because people perceive positive and negative emotions more easily, but valence was simpler to identify than arousal. During annotations, predetermined valence and arousal ranges were established to guarantee quality. Introducing simple Convolutional Neural Network (CNN) baselines, which deviate from conventional techniques.

The author discusses difficulties in recognizing emotions in the real world because of shifting postures and points of view [9]. They show the efficacy of their method by testing it on three real-world datasets. Conventional methods use different loss functions and alternate training to address the problem of different labels in different facial image datasets. Face landmarks are located using geometric-based techniques for emotion classification. Procruste's transformation was applied to 117 landmarks by Jeni et al., whereas Kotsia and Pitas used 119 key points. Features such as Gabor wavelets, HOGs, and LBPs are extracted from pixel images using appearance-based recognition. Patches, mid-level features, or Gaussian processes are used in advanced models. A selective loss function is introduced by the Selective Joint Multitask (SJMT) approach. The architectures of VGG-16 and ResNet-50 are contrasted with single CNNs and traditional multitasking techniques. The SFEW2.0 dataset includes labeled photos from movie clips.

8.3 Block Diagram

Figure 8.2 presents a block diagram that provides a visual representation of the overall structure and design of the proposed model.

Fig. 8.2 Overview
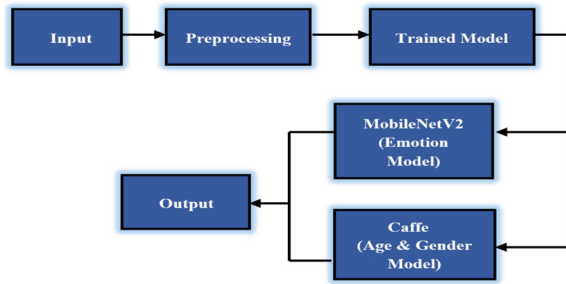
8.4 Methodology

8.4.1 Haar Cascade Classifier

The Haar cascade classifier is a realization of the Viola–Jones algorithm [10], employing a cascade structure comprising Haar-like features. This architecture includes a series of classifiers, each adept at discerning positive and negative samples based on distinctive features. Primarily utilized in OpenCV, especially for tasks like face detection.

The training phase involves optimizing the weights of these features to minimize classification errors. Once trained, the classifier efficiently examines image regions to swiftly identify the presence of the target object.

8.4.2 Data Collection

The FER2013 dataset, sourced from Kaggle, comprises more than 35,000 facial images categorized by the displayed emotion. The dataset encompasses seven emotion types: happiness, neutral, sadness, anger, surprise, disgust, and fear. This dataset originates from the ICML 2013 Representation Learning Workshop's face expression recognition challenge [11]. Extensive research in emotion and facial expression identification has heavily relied on the FER2013 dataset. This dataset has been instrumental in the creation and validation of novel algorithms for facial expression analysis, as well as in the training and evaluation of machine learning models dedicated to emotion recognition.

Every image within the compilation is a 48 × 48 pixel grayscale representation of a face. Preprocessing procedures have been applied to reduce noise, eliminate backgrounds, and optimize lighting conditions. To streamline the focus exclusively on facial features, the images undergo cropping to eliminate any extraneous elements. Figure 8.3 depicts visual representations of the FER2013 dataset images and the distribution of training images, providing a graphical overview of the dataset and its composition.

Fig. 8.3 Collection of training images
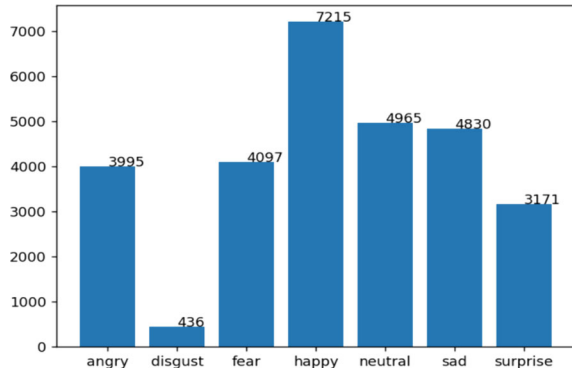
8.4.3 Data Preprocessing

A large dataset is essential for evaluating a DCNN model's performance. Furthermore, partially relevant or irrelevant features can hinder the performance of a DCNN model [12]. Various image processing parameters, including shear range, horizontal range, and zoom range, are employed in data augmentation to eliminate irrelevant features. Key arguments used in data augmentation include the rescale argument, which normalizes pixel values from the original range of [0, 255] to a standardized range of [0, 1]. This normalization process is a prevalent preprocessing step in deep learning for image data, contributing to the stability of the training process.

Another critical argument is a preprocessing function, facilitating the application of a tailored preprocessing function to the dataset images. For instance, the grayscale RGB and upsizing function is utilized to convert RGB images to grayscale and resize them to a larger dimension. This preprocessing step aims to enhance the model's performance by reducing the number of features it needs to learn while augmenting the informational content in each image.

8.4.4 Model Training

8.4.4.1 Age and Gender Model

In the domain of age and gender detection, the creation of a CAFFE model [13] involves a series of consecutive procedures, commencing with the preparation of data. A crucial requirement is a labeled dataset comprising facial images annotated with age and gender information, exemplified by datasets like Adience or IMDB-WIKI [14]. An imperative stage in this process is image preprocessing, encompassing facial detection, cropping to highlight facial regions, resizing for uniformity, and normalizing pixel values.

The subsequent phase pertains to the development of a suitable model architecture. It is imperative to design a convolutional neural network (CNN) architecture tailored explicitly for concurrent age and gender prediction. Leveraging transfer learning, the model should be initialized with pre-existing weights from a substantial dataset such as ImageNet [15]. Following this initialization, fine-tuning the designated age and gender dataset becomes crucial, facilitating effective adaptation to the specific task at hand.

The training phase necessitates the establishment of appropriate loss functions for age and gender predictions, such as Mean Squared Error for age and Softmax Cross-Entropy for gender. Throughout the training process, ongoing refinement of hyperparameters is crucial to mitigate the risk of overfitting. Vigilant monitoring of performance on both training and validation sets is imperative to ensure optimal model development.

In the subsequent real-time implementation phase, seamless integration with a webcam feed is achieved. The dynamic detection of age and gender using the CAFFE model alongside a webcam feed is realized through the utilization of libraries such as OpenCV. During this process, each frame from the webcam undergoes inference passing through the CAFFE model, enabling simultaneous prediction of both age and gender in real-time scenarios.

8.4.4.2 Emotion Model

The approach for recognizing human emotions utilizes a two-channel architecture specifically crafted for this purpose. Emotional features are isolated and extracted from the FER2013 dataset, serving as inputs to the inception layer of a convolutional neural network (CNN). The training phase involves the extraction of features and subsequent classification facilitated by the CNN.

MobileNetV2, a tailored deep learning model architecture for mobile and embedded devices, enhances its predecessor [16], MobileNet, through a combination of depth-wise separable convolution and linear bottleneck layers. Its adaptability enables deployment on various devices, including smartphones, tablets, and microcontrollers, making it a preferred choice for real-time emotion recognition applications.

In face emotion recognition with MobileNetV2, the standard procedure includes training the network on a comprehensive dataset of labeled facial images, associating each image with a specific emotion label. Throughout the training, the network's weights and biases are optimized to minimize the difference between predicted and true emotion labels for each image.

The training process employs a suitable optimization algorithm, such as Adam [19], with batches of training data. Parameters are iteratively adjusted to minimize the loss between predicted and true labels, using a loss function like mean squared error or cross-entropy. Figure 8.4 illustrates the architecture of MobileNetV2, providing a visual representation of the model's structural design. Comparative Analysis is detailed in above Table 8.1.
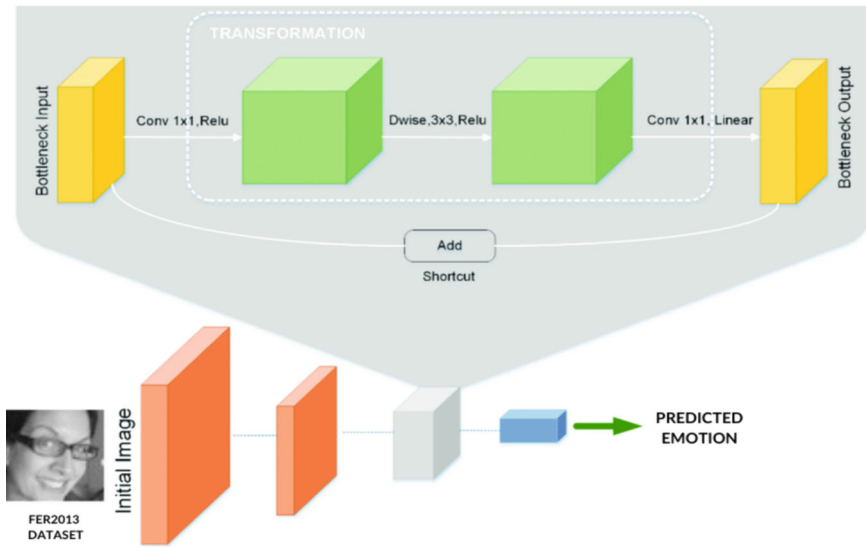
Fig. 8.4 Architecture of MobileNetV2

The optimization algorithm, such as Adam [19], crucially updates model parameters based on gradients of the loss function concerning the model parameters. Backpropagation computes these gradients, propagating errors from the output layer back through the model's layers, ensuring effective adjustment of weights and biases in each layer for optimal model optimization.

Post-training, the model is applied to classify emotions in new facial images. Preprocessing steps, including size and color normalization, are performed on input images before feeding them into the MobileNet V2 network for feature extraction. Extracted features then pass through one or more fully connected layers to predict the corresponding emotion label.

8.5 Results

The proposed model demonstrated remarkable results during experimentation with a diverse group of volunteers of varying ages, genders, and emotions as depicted in Fig. 8.7. The volunteers willingly participated in the study, where their facial expressions were evaluated. Additionally, the model's performance, as reflected in the loss and accuracy graph, was visualized using the MobileNetV2 architecture, as depicted in Figs. 8.5 and 8.6. Furthermore, the model's proficiency in predicting both ages and gender was analyzed, with the age predictions detailed in Table 8.2. This table provides a lucid comparison between the predicted and actual ages, offering

Table 8.1 Comparative analysis

CNN architectures	Computational efficiency	Real-time processing	Versatility	CPU/GPU compatibility	TensorFlow	PyTorch	Keras	CAFFE
ResNet [17]	Moderate efficiency, more parameters than MobileNetV2	Depends on model size and hardware	General-purpose architecture	Requires GPU for efficient training and inference	✓	✓	✓	✓
VGG [18]	Less Computationally efficient due to more parameters	Maybe less suitable for real-time applications	General-purpose architecture	Requires GPU for efficient training and inference	✓	✓	✓	
MobileNet V2	Efficient on mobile and edge devices, lower parameters	Suitable for real-time applications	Versatile, applicable to a wide range of tasks	Efficient on both CPU and GPU	✓	✓	✓	✓

valuable insights into the model's accuracy in estimating individuals' ages based on their facial features.

Fig. 8.5 Accuracy by using MobileNetV2
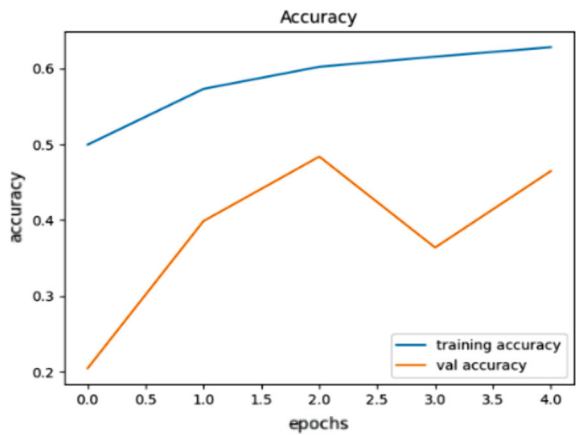
Fig. 8.6 Loss by using MobileNetV2
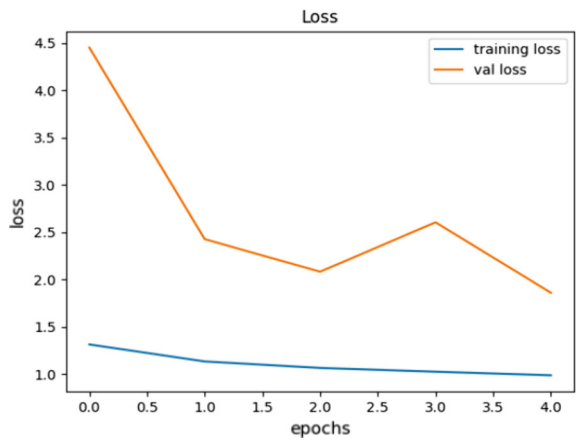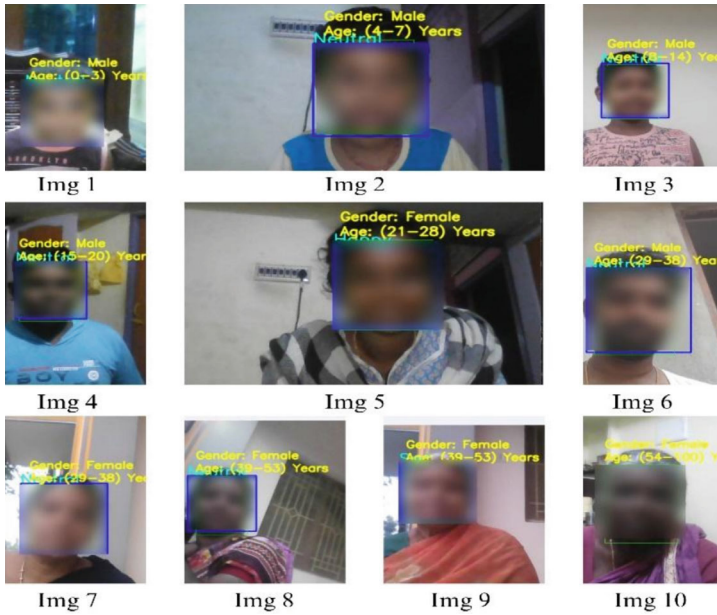
Fig. 8.7 Prediction of age, gender, and emotions in real time from a webcam

Table 8.2 Comparison between predicted and actual age, gender, and emotions

Images	Actual age	Predicted age	Actual emotion	Predicted emotion	Actual gender	Predicted gender
Img1	2	0–3	Neutral	Neutral	Male	Male
Img2	5	4–7	Neutral	Neutral	Male	Male
Img3	11	8–14	Neutral	Neutral	Male	Male
Img4	15	15–20	Neutral	Neutral	Male	Male
Img5	24	21–28	Happy	Happy	Female	Female
Img6	38	29–38	Neutral	Neutral	Male	Male
Img7	52	39–53	Sad	Sad	Female	Female
Img8	45	39–53	Neutral	Neutral	Female	Female
Img9	38	29–38	Neutral	Neutral	Female	Female
Img10	70	54–100	Neutral	Neutral	Female	Female

8.6 Conclusion

The proposed system seamlessly integrates with real-time webcam images, offering a robust solution for accurately discerning human characteristics through facial recognition. Utilizing a pre-trained deep learning model, the system excels in real-time emotion recognition with commendable accuracy. However, challenges may arise in

accurately identifying emotions for certain individuals or in scenarios where facial expressions are less distinct. These challenges highlight the necessity for continuous research and development in emotion recognition to enhance system accuracy and reliability.

To broaden its capabilities, the system incorporates age and gender recognition, leveraging prediction models, potentially implemented through frameworks like CAFFE. This holistic approach contributes to a comprehensive understanding of the user's emotional state. The inclusion of age and gender recognition not only enhances the system's functionality but also opens avenues for applications in various domains such as mental health, human–computer interaction, and personalized services. Despite its proficiency, Challenges in age prediction may arise due to issues related to brightness, potentially impacting the accuracy of the predictions. Ongoing efforts in research and development remain essential for refining and advancing the capabilities of recognition systems.

References

1. Zhang, K., Liu, N., Yuan, X., Guo, X., Gao, C., Zhao, Z., Ma, Z.: Fine-grained age estimation in the wild with attention LSTM networks. IEEE Trans. Circ. Syst. Video Technol. **30**, 3140–3152 (2020). https://doi.org/10.1109/TCSVT.2019.2936410
2. Mahendran, K., Abishek, B., Balaji, R., Dhilip, T.: Analysis of face feature recognition using MATLAB. In: Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023. pp. 908–914. Institute of Electrical and Electronics Engineers Inc. (2023). https://doi.org/10.1109/ICAIS56108.2023.10073828
3. Veena, V.N., Maria Antony, C., Scholar, M.: Age and gender recognition using deep learning (2022)
4. Li, Y., Wang, S., Zhao, Y., Ji, Q.: Simultaneous facial feature tracking and facial expression recognition. IEEE Trans. Image Process. **22**, 2559–2573 (2013). https://doi.org/10.1109/TIP.2013.2253477
5. Zhang, H., Xu, M.: Multiscale emotion representation learning for affective image recognition. IEEE Trans. Multimedia **25**, 2203–2212 (2023). https://doi.org/10.1109/TMM.2022.3144804
6. Rodriguez, P., Cucurull, G., Gonzalez, J., Gonfaus, J.M., Nasrollahi, K., Moeslund, T.B., Roca, F.X.: Deep pain: exploiting long short-term memory networks for facial expression classification. IEEE Trans. Cybern. **52**, 3314–3324 (2022). https://doi.org/10.1109/TCYB.2017.2662199
7. Xie, S., Hu, H.: Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans. Multimedia **21**, 211–220 (2019). https://doi.org/10.1109/TMM.2018.2844085
8. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Comput. **10**, 18–31 (2019). https://doi.org/10.1109/TAFFC.2017.2740923
9. Pons, G., Masip, D.: Multitask, multilabel, and multidomain learning with convolutional networks for emotion recognition. IEEE Trans Cybern. **52**, 4764–4771 (2022). https://doi.org/10.1109/TCYB.2020.3036935
10. Viola, P., Jones, M.J.: Robust real-time face detection (2004)
11. Goodfellow, I.J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J.,

Xie, J., Romaszko, L., Xu, B., Chuang, Z., Bengio, Y.: Challenges in representation learning: a report on three machine learning contests. Neural Netw. **64**, 59–63 (2015). https://doi.org/10.1016/j.neunet.2014.09.005

12. Institute of Electrical and Electronics Engineers: Data augmentation for improving deep learning in image classification problem (2018)
13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding (2014)
14. Rothe, R., Timofte, R., Van Gool, L.: Deep expectation of real and apparent age from a single image without facial landmarks real age 20 years DEX age prediction (2016)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**, 84–90 (2017). https://doi.org/10.1145/3065386
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: inverted residuals and linear bottlenecks (2018)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014)
19. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014)

Chapter 9
Design and Implementation of Adaptive Switched-Capacitor DC-DC Converter for PV-Fed Variable Step Value-Based P&O MPPT Controller

V. Prashanth, CH Hussaian Basha⬛, **Fawad Ur Rahman, B. K. Chandan Gowda, and H. S. Kruthik**

Abstract Based on the current literature survey, the Solar Photovoltaic (SPV) modules are playing the principal role in the current standalone power distribution systems because their merits are optimal atmospheric pollution, less noise pollution, freely available in nature, mostly suitable for all rural areas, and creating more job opportunities. However, the drawback of this system is less functioning efficiency and nonlinear power availability. So, the advanced Perturb & Observe (P&O) concept is used in this work to overcome the nonlinear behavior of the SPV system. The features of this Maximum Power Point Tracking (MPPT) concept are fast MPP identifying speed, quick dynamic response, and few oscillations across the functioning point of the SPV system. Here, the adjustable step value is utilized in this proposed powerpoint identifying controller. Due to that the convergence speed of the system is enhanced. In the second stage, the Switched-Capacitor Boost Converter (SCBC) is interfaced in the middle of the SPV and resistive network to optimize the SPV power generation cost. The merits of this proposed DC-DC converter are less ripple in the SPV output voltage, good voltage conversion ratio of the system, plus fast switching speed. The overall system is investigated by the use of the MATLAB tool.

Keywords Switched-capacitor circuit · Low ripples · More voltage gain · Good steady-state performance · Plus quick MPP tracing response of the SPV

V. Prashanth · F. U. Rahman · B. K. Chandan Gowda · H. S. Kruthik
NITTE Meenakshi Institute of Technology, Karnataka 560064, India

CH Hussaian Basha (✉)
SR University, Anantasagar, Hasanparthy, Warangal 506371, India
e-mail: sbasha238@gmail.com

9.1   Introduction

From the previous literature review, the Renewable Power Sources (RPS) are geothermal energy, biomass energy, hydro power, wind power systems, and solar systems. In Azizi et al. [1], the authors took the geothermal systems for supplying power to the urban people. Here, the geothermal power plants collect the fuels from the underground systems for generating the steam. The heated steam is sent to the geothermal power turbines with high pressure for generating electrical power. The merits of geothermal systems are that they require very little space for installation, noise-less energy, create a greater number of jobs, have allowable double recycling, plus a little maintenance cost [2]. But it has many drawbacks which are greenhouse gas emissions, air, and water pollution. To limit the drawbacks of geothermal systems, biomass energy is used for electricity generation [3]. In this system, the biomass is burned in a particular boiler to produce the high pressurized steam. This steam passes through the series of connected blades for the power production process. The features of biomass systems are highly reliable, more flexible, highly abundant, and extensively reduce human wastage. However, it needs a high catchment area and a highly expensive system, and burning waste may release hazardous gasses [4].

So, most of the power supply industries are focusing on the hydro networks. In this power generation network, the water velocity is captured and it's sent to the hydropower turbines which are interconnected with the generator for transmission of the electricity. The merits of hydro systems are clean source, flexible source, highly sustainable, and more protection toward diversity [5]. Also, it reduces the risk of floods. However, the disadvantages of this type of power network are moderate impact on the environment, limited availability of reservoirs, a highly expensive system, and not safe for all weather conditions. The limitations of hydro networks are limited by utilizing the wind systems. In this system, the wind kinetic energy is passed through the wind blades to create the lift. As a result, the wind blades start rotating and are coupled with the drive shaft for running the generator [6]. The merits of wind systems are creating more jobs, providing good economic growth, clean sources, more suitable for local communities, and a cost-effective system. The disadvantages of this system are more dangerous to wildlife, creates noise, and is suitable only for limited places.

Now, the SPV methodology is used to limit the drawbacks of wind power networks. Photo energy is available in excess and has very low atmospheric pollution. This energy system gives more jobs to human beings and is very easy to maintain. Here, the SPV cell is designed by combining the P-type and N-type semiconductor materials. Once, the P and N-type materials combination form a junction which is broken by using the sunlight energy [7]. So, the electricity is generated from the SPV cell which is collected at the load side resistor. The maximum available SPV cell voltage is in the range of 0.75 to 0.82 which is not acceptable for domestic applications. Here, the SPV cell integration has been made by using the series and parallel connection to develop the PV module. In addition, the SPV cells are designed by

applying different types of methodologies which are polycrystalline and monocrystalline. In some cases, the SPV cells are manufactured by utilizing the thin film [8]. Among the SPV cell technologies, the monocrystalline methodology is selected for supplying power to the bidirectional batteries.

Another issue of SPV systems is nonlinear voltage generation which is overcome by using the MPPT concept. The proposed standalone SPV network is shown in Fig. 9.1, and its related power curves are shown in Fig. 9.2a, and b [9]. From the current literature survey, there are various categories of MPPT networks available in nature which are identified as nature-inspired, artificial intelligence, and soft computing controllers. The basic normalized MPPT controllers are fractional voltage and current techniques. The fractional voltage is an approximated controller, and it may not give the exact value of the SPV voltage. In addition, this technique gives a less accurate MPP position of the SPV network. Similar to this method, the fractional current controller is also an approximate method. This fractional current method is not utilized at rapid changes in temperature conditions of the SPV network. Here, the advanced P&O controller is developed to find the functioning point of the SPV system. The features of this controller are fast convergence speed, good system response, and provides optimum switching pulses to the power converter.

The final issue of the SPV system is more installation cost which is optimized by applying the various power DC-DC converters. From the latest published articles, the converters are differentiated as transformer-dependent converters and without transformer converter [10]. The transformer-based power converters need one more rectifier for the power conversion. So, the overall system manufacturing cost is high, and more space is required for developing the system. In addition to these drawbacks, these converters are not helpful for a standalone SPV system. So, a switched-capacitor
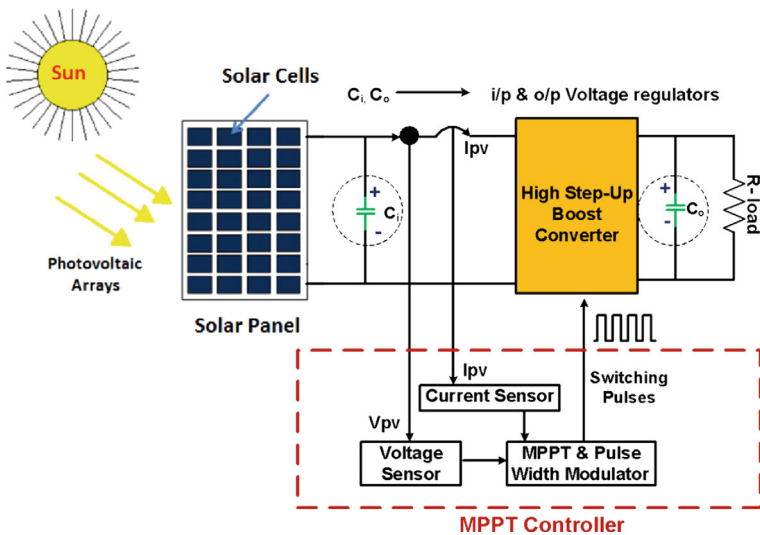
Fig. 9.1 Proposed SPV fed switching capacitor power DC-DC converter
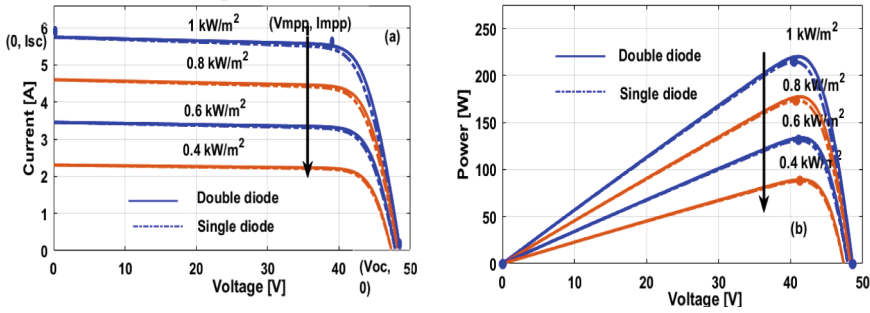
Fig. 9.2 **a** Current versus voltage and **b** Voltage versus power curves

DC-DC power converter is used to enhance the source power conversion efficiency. The features of this converter are a high voltage conversion ratio, stable output voltage, and few converter output power distortions.

9.2　Mathematical Analysis of SPV Cells

The SPV cell works based on the solar incident photon energy. When the incident photons on solar modules are very high then the power captured from the modules is high. Otherwise, the availability of SPV output voltage is very low. As a result, the lifespan and working efficiency of the SPV network is reduced. In addition, the SPV module's functioning point may settle any one of the local operating positions [11]. So, the maximum power supply from the overall system is affected. Here, the SPV module is implemented by utilizing the 2-diode circuit as given in Fig. 9.3. Basically, most of the articles refer to the 1-diode circuit for the design of the SPV array. The 1-diode circuit is an approximated SPV cell, and it may not give accurate output characteristics. So, the present solar systems are designed by utilizing the 2-diode circuit as given in Fig. 9.3.

Fig. 9.3 2-diode circuit for developing the SPV modules
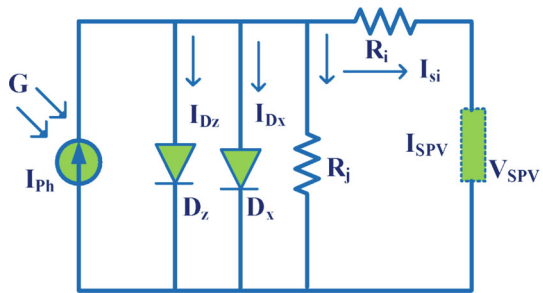
The analysis of the 1-diode circuit has been made by utilizing the SPV open-circuited voltage, diode factor, peak voltage, peak available power, and diode available current at saturation conditions. Here, for the 2-diode circuit, there are a few more variables needed which are the second-diode reverse saturation current and ideality constant. The 2-diode model is highly accurate because all the SPV cell constraints are utilized over here. The nominal existing power of the SPV is evaluated by utilizing the cells short-circuit current and open circuit voltage. Here, the nominal values of SPV systems depend on the atmospheric temperature and partial shading effect on the solar modules. The available SPV system current is obtained as

$$I_{SPV} = I_{ph} - I_{0z}\left(e^{\frac{V_{SPV}+I_{SPV}R_i}{a_z V_{Tz}}} - 1\right) - I_{0x}\left(e^{\frac{V_{SPV}+I_{SPV}R_i}{a_x V_{Tx}}} - 1\right) - \frac{V_{SPV} + (I_{SPV} \times R_i)}{R_j}$$

(9.1)

$$I_{0z} = I_0 = I_{0x} = \frac{I_{sc\_STC} + k_i \Delta T}{e^{\left(\frac{(V_{oc\_SC}+k_v*\Delta T)}{\{\frac{a_z+a_x}{p}\}V_{teq}}\right)} - 1}$$

(9.2)

$$V_{Teq} = V_{Tz} = V_{Tx} = \frac{N_s * K * T}{q}$$

(9.3)

$$I_{ph} = I_{0z}\left(e^{\left(\frac{V_{oc}}{a_z*V_{Tz}}\right)} - 1\right) - I_{0x}\left(e^{\left(\frac{V_{oc}}{a_x*V_{Tx}}\right)} - 1\right) - \frac{V_{oc}}{R_j}$$

(9.4)

$$I_{sc} = I_{ph} - I_{0z}\left(e^{\left(\frac{I_{sc}*R_i}{a*V_{Tz}}\right)} - 1\right) - I_{0x}\left(e^{\left(\frac{I_{sc}*R_j}{a*V_{Tx}}\right)} - 1\right) - \frac{I_{sc}*R_i}{R_j}$$

(9.5)

$$I_{MPP} = I_{ph} - I_{0z}\left(e^{\frac{(V_{MPP}+I_{MPP}*R_s)}{a*V_{Tz}}} - 1\right) - I_m$$

(9.6)

$$I_m = I_{0x}\left(e^{\left(\frac{V_{MPP}+I_{MPP}*R_S}{a*V_{Tx}}\right)} - 1\right) + \frac{V_{MPP} + I_{MPP} * R_s}{R_p}$$

(9.7)

Here, the variables I_{ph}, I_{0z}, I_{0x}, V_{SPV}, I_{SPV}, R_i, and R_j are the cell current, reverse current of diode one, reverse current of diode-2, cell load voltage, SPV load current, series resistor, and parallel resistor. If the cell is short-circuited then the available cell current (I_{Sc}) is given in Eq. (9.5). The evaluated thermal voltages are represented as V_{Teq}, V_{Tz}, and V_{Tx}. In addition, the parameters q, k, N_s, ΔT, a_z, and a_x are identified as electron charge, Boltzmann constant, utilized cells, variation of temperature, and selected ideality factors of diodes. The design constraints of the SPV module are given in Table 9.1.

Table 9.1 Selected values of SPV system at various sunlight conditions

Parameter	Full form	Values
P_{rated}	Rated power of SPV system	220.17W
V_{rated}	Available rated power of SPV system	41.22 V
N_p	Available cells in parallel linked arrays	1
N_s	Connected cells in the way of series	72
I_{SC}	Current of SPV system at short circuit	5.75A
R_i	Resistance appeared in series of the cell	0.508Ω
R_j	Resistance appeared across the SPV cell	208.88Ω

9.3 Analysis of Switched-Capacitor Power Converter

Based on the present literature survey, the flyback isolated DC-DC power conversion circuits give fluctuated power to the local consumers [12]. Also, it requires a bridge-type four-diode rectifier for transforming the alternative signal to oscillated voltage. Similar to this converter, the forward converter may need a few more components to enhance the system power profile. As a result, the manufacturing price of the entire system has improved. In addition, the handling of the isolated-based SPV network is quite a difficult task [13]. So, the present researchers are focusing on transformerless power conversion technologies which give low power transmission losses and more efficiency when equated to the other power conversion technologies [14].

In this work, an advanced switching capacitive boost converter is selected for the SPV network in order to improve the power supply rating of the source. The design structure and its working stages are given in Fig. 9.4a-c. From Fig. 9.4a, the source impedance value is compared with the load equivalent impedance value for generating the needed duty signal to the switching stage capacitor-based boost converter. Here, based on Fig. 9.4b, the converter should have to work in a continuous power supply mode of operation. So, the inductor L_z is selected with a very high value for the converter making it work in continuous power supply mode of operation.

The derivation of voltage gain has been done by assuming that most of the switches are working in an ideal state of operation and that voltages that appear across the capacitors are constant. Finally, the current passing through the inductive components is linear. From Fig. 9.5, based on the converter waveforms, the element Q is in a conduction state then the C_z, C_x are in an energy storage state. From Fig. 9.4c, the voltage conversion ratio of the converter is obtained by using Eqs. (9.8) and (9.9). The charge state equation is applied to Eq. (9.10).

$$\{V_{PV} = V_{LZ} V_{CC} = V_{CZ} + V_{CX} V_{CV} = V_{Cb} \tag{9.8}$$

$$\{V_{Lz} = V_{PV} - V_{Cz} V_{Cz} = V_{Cx} V_0 = V_{Ca} + V_{Cc} + V_{L0} \tag{9.9}$$
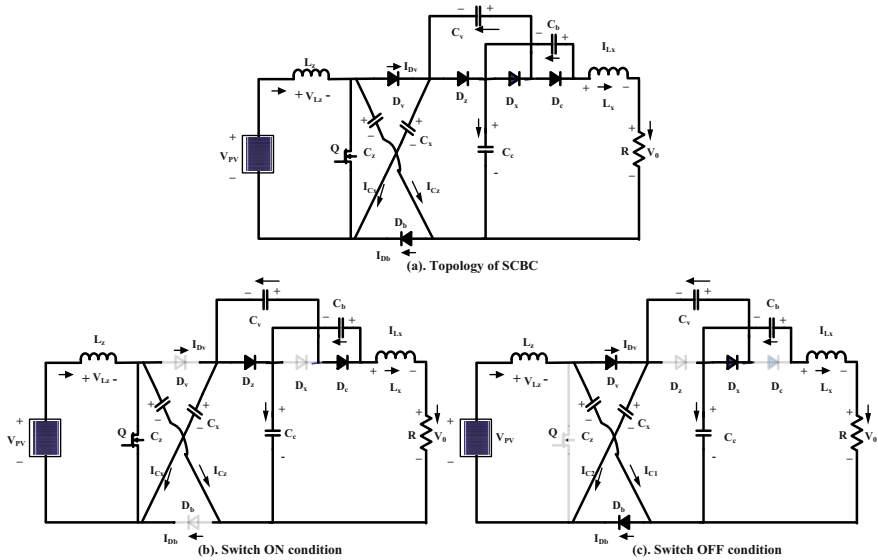
Fig. 9.4 Proposed converter, **a** General structure, **b** Switch conduction, and **c** Switch block state

$$V_{PV}DT_S + ((V_{PV} - V_{Cz}) * (1 - D)T_S = 0 \qquad (9.10)$$

$$\{V_C = V_{Cz} = V_{Cx} = V_{Cv} = V_{Cb} = \frac{1}{(1-D)}V_{PV} \ V_{Cc} = 2V_C = \frac{2}{(1-D)} * V_{PV} \ V_0$$
$$= V_{Cz} + V_{Cv} + V_{Cb} = 3V_C = \frac{3}{(1-D)} * V_{PV}$$
$$\qquad (9.11)$$

9.4 Development of Advanced P&O-Created MPPT Methodology

Based on the present renewable power production systems, the power point tracking is useful for getting the peak power from the renewable systems [15]. Here, the adaptive conventional different step-based P&O methodology is applied for extracting the peak voltage of the SPV system. Here, the step selection on the V-I curve of the SPV network is quite difficult. So, at the algorithm starting stage, the step constant value is very high for running the SPV near the actual MPP place. After reaching the functioning point of the fuel stack to the required MPP state the step constant value will be very low [16]. Due to this step value minimization, the fluctuations of converter voltage are moderate. Here, the present available power value is subtracted from the past power value for varying the duty signal of the DC-DC converter.
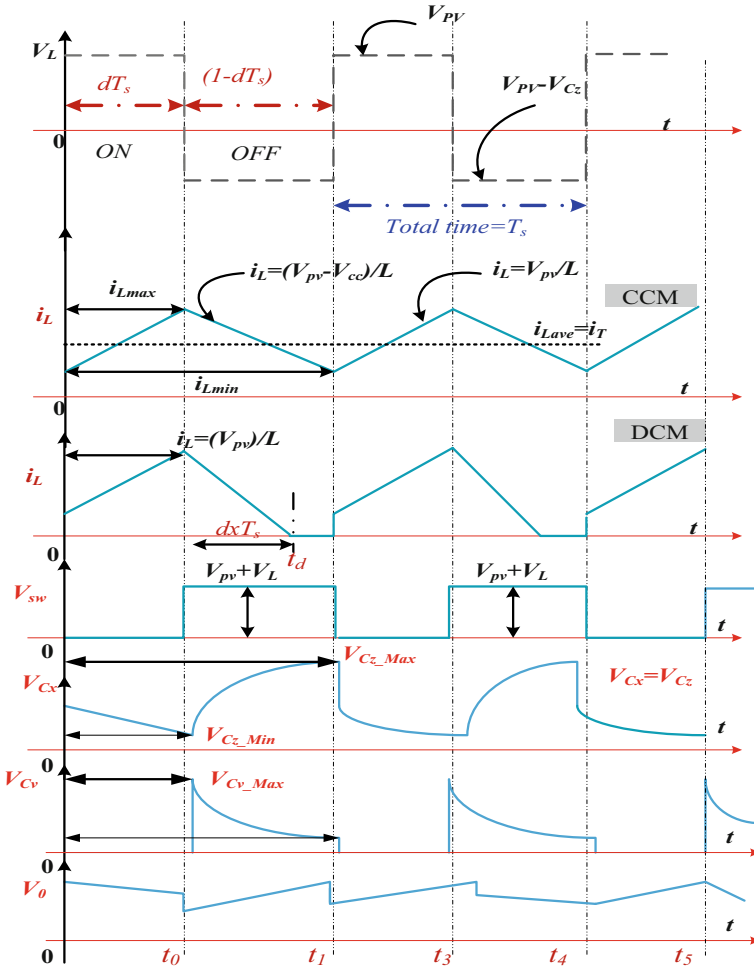
Fig. 9.5 Proposed switching capacitive-dependent power DC-DC converter

The working diagram of the proposed MPPT methodology is given in Fig. 9.6. Based on Fig. 9.6, the updating of the duty signal to the DC-DC power converter is given in Eq. (9.12).

$$D(b) = D(b-1) + z * \left(\frac{P(b) - P(b-1)}{V(b) - V(b-1)} \right) \tag{9.12}$$

$$D(b) = D(b-1) - z * \left(\frac{P(b) - P(b-1)}{V(b) - V(b-1)} \right) \tag{9.13}$$
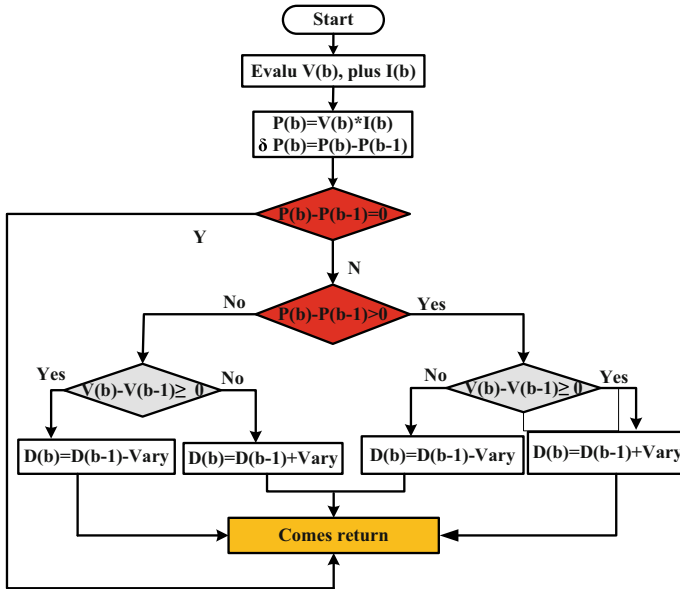
Fig. 9.6 Utilized differentiated step constant-based P&O controller

The variables D(b), P(b), z, and V(b) are the duty value of SCBC, power of PV, slope constant of V-I curve, and available voltage of PV. Similarly, the D(b-1), P(b-1), z, and V(b-1) are the past stored duty and power constants.

9.5 Results Analysis of Switching Capacitive DC-DC Converter

The selected converter with differential step constant P&O methodology study has been done by putting on the MATLAB Simulink. Here, the differential step constant P&O powerpoint identifier is used to run the overall system at quick changes in atmospheric conditions of the PV. The selected supply side inductive (L_z) element value is 95.8mH which is useful for optimizing the distortions of the supply power. Also, it reduces the sudden variation of the source voltage. The utilized capacitive values across source and load are $C_z = C_x = C_c = C_v = 370\,\mu$ F and $C_b = 370\mu$F, respectively. Finally, the utilized load resistor is 82Ω. At the load side, the manufacturers used the inductive element value L_x is 100mH.
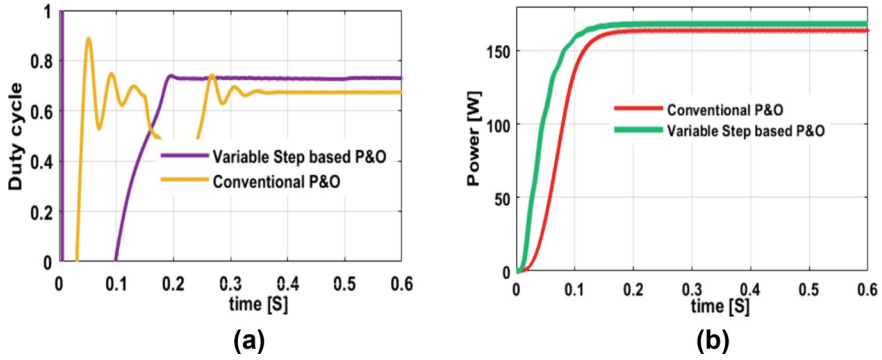
Fig. 9.7 **a** Obtained duty signal of switching power DC-DC converter at 800W/m^2 **b**. Obtained converter power of switching power DC-DC converter at 800W/m^2

Here, initially, the overall system is investigated at 800W/m^2. The converter functioning duty value by applying the conventional P&O method at static irradiation value is 0.67 and its available load power is 174.06W. Similarly, the available power by utilizing the proposed MPPT controller is 178.20W, and its duty is selected as 0.65. So, at constant atmospheric irradiation conditions, the variable step parameter-based conventional P&O method gives superior performance when equated with the general conventional MPPT controllers. The obtained waveforms of the power converter duty signal and converter power signal are given in Fig. 9.7a and b. Similarly, at quick changes of sunlight irradiation conditions (1000W/m^2), the determined duty signal of the DC-DC converter by applying both the MPPT methodologies is 0.72 and 0.74 and its related generated converter powers are 208.11W and 205.33W, respectively. The overall system output parameters are shown in Table 9.2. For fast changes of sunlight conditions, the available duty and power of the converter are shown in Fig. 9.8a and b.

Table 9.2 Proposed switching capacitive power DC-DC converter output values

Type of controller	Proposed power DC-DC converter											
	Sunlight irradiations (0 to 0.6 s) _ 800W/m²				Sunlight irradiations (0.6 to 1.2 s) _ 400W/m²				Sunlight irradiations (1.2 to 1.8 s) _ 1000W/m²			
	V_{PV} (V)	duty	V_0 (V)	P_0 (W)	V_{PV} (V)	duty	V_0 (V)	P_0 (W)	V_{PV} (V)	duty	V_0 (V)	P_0 (kW)
P&O	25.9	0.47	150.19	174.9	13.7	0.43	78.22	90.19	36.18	0.72	170.11	208.11
Variable step P&O	30.01	0.48	157.98	174.77	13.15	0.472	79.99	92.25	37.44	0.74	155.11	205.33
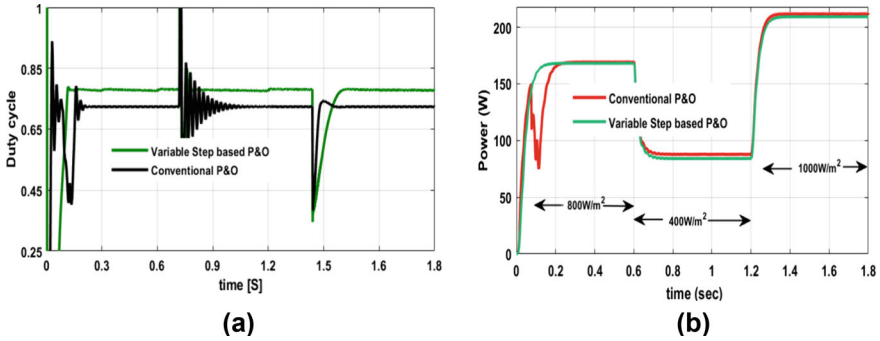
Fig. 9.8 a Obtained duty signal of switching power DC-DC converter at dynamic atmospheric condition **b** Obtained converter power of switching power DC-DC converter at dynamic atmospheric

9.6 Conclusion

The switching capacitive-based power converter with adjustable step constant P&O method is analyzed by using the MATLAB tool. Here, the 2-diode model SPV module gives more supply power with optimum duty signal. The merits of this SPV array are highly efficient, and more reliable, plus gives high accuracy for tracking the functioning point of the SPV. In addition, the power converter enhances the supply power transformation efficiency of the SPV with a good voltage conversion ratio. Here, the switching signal of the converter is generated by using the variable step-based P&O controller. The features of this MPPT method are fewer fluctuations of converter voltage, plus good dynamic response.

References

1. Azizi, N., Esmaeilion, F., Moosavian, S.F., Yaghoubirad, M., Ahmadi, A., Aliehyaei, M., Soltani, M.: Critical review of multigeneration systems powered by geothermal energy resources from the energy, exergy, and economic point of views. Energy Sci. & Eng. **10**(12), 4859–4889 (2022)
2. Basha, C.H.H., Rani, C.: Different conventional and soft computing MPPT techniques for solar PV systems with high step-up boost converters: a comprehensive analysis. Energies **13**(2), 371
3. Kiran, S.R., Basha, C.H., Singh, V.P., Dhanamjayulu, C., Prusty, B.R., Khan, B.: Reduced simulative performance analysis of variable step size ANN based MPPT techniques for partially shaded solar PV systems. IEEE Access **10**, 48875–48889 (2022)
4. Basha, C.H., Rani, C.: Design and analysis of transformerless, high step-up, boost DC-DC converter with an improved VSS-RBFA based MPPT controller. Int. Trans. Electr. Energy Syst. **30**(12), e12633 (2020)
5. Cho, K., Kim, Y.: Improving streamflow prediction in the WRF-Hydro model with LSTM networks. J. Hydrol. **605**, 127297 (2022)
6. Fathabadi, H.: Possibility of utilizing wind turbines to recover a portion of the kinetic energy losses of a car. IEEE Trans. Veh. Technol. **68**(9), 8663–8670 (2019)

7. Saif, O.M., Shaker, A., Abouelatta, M., Zekry, A., Elogail, Y.: Numerical simulation and design of all-thin-film Homojunction Perovskite/c-Si tandem solar cells. Silicon. 1–17 (2023)
8. Basha, C.H., Murali, M.: A new design of transformerless, non-isolated, high step-up DC-DC converter with hybrid fuzzy logic MPPT controller. Int. J. Circuit Theory Appl. **50**(1), 272–297 (2022)
9. Pathy, S., Subramani, C., Sridhar, R., Thamizh Thentral, T.M., Padmanaban, S.: Nature-inspired MPPT algorithms for partially shaded PV systems: a comparative study. Energies **12**(8), 1451 (2019)
10. Rafikiran, S., Basha, C.H., Devadasu, G., Tom, P.M., Fathima, F., Prashanth, V.: Design of high voltage gain converter for fuel cell based EV application with hybrid optimization MPPT controller. Mater. Today: Proc. (2023)
11. Basha, C.H.H, Rani, C., Brisilla, R.M., Odofin, S.: Mathematical design and analysis of photovoltaic cells using MATLAB/Simulink. In: Soft Computing for Problem Solving: SocProS 2018, vol. 1, pp. 711–726. Springer Singapore (2020)
12. Henriksen, M.M., Larsen, D.Ø., Muntal, P.L.: A design methodology for high-voltage, highly-integrated switched-capacitor power converters, and implementation at 48 V-12 V, 23 W/cm^3 and 93.5% peak efficiency. IEEE Trans. Power Electron (2023)
13. Govinda Chowdary, V., Udhay Sankar, V., Mathew, D., Hussaian Basha, C.H., Rani, C.: Hybrid fuzzy logic-based MPPT for wind energy conversion system. In: Soft Computing for Problem Solving: SocProS 2018, vol. 2, pp. 951–968. Springer Singapore (2020)
14. Nadimuthu, L.P.R., Victor, K., Basha, C.H., Mariprasath, T., Dhanamjayulu, C., Padmanaban, S., Khan, B.: Energy conservation approach for continuous power quality improvement: a case study. IEEE Access **9**, 146959–146969 (2021)
15. Basha, C.H.H, Rani, C., Brisilla, R.M., Odofin, S.: Simulation of metaheuristic intelligence MPPT techniques for solar PV under partial shading conditions. In: Soft Computing for Problem Solving: SocProS 2018, vol. 1, pp. 773–785. Springer Singapore (2020)
16. Udhay Sankar, V., Bhanutej, Basha, C.H.H, Mathew, D., Rani, C., Busawon, K.: Application of WDO for decision-making in combined economic and emission dispatch problems. In: Soft Computing for Problem Solving: SocProS 2018, vol. 1, pp. 907–923. Springer Singapore (2020)

Chapter 10
Hybrid Swarm Intelligence Approach for Energy Efficient Clustering and Routing in Wireless Sensor Networks

Bandi Rambabu, B. Vikranth, Medikonda Asha Kiran, Satyanarayana Nimmala, and L. Swathi

Abstract The aim of wireless sensor networks (WSNs) is to gather and transmit data from the environment to the base station. However, this can lead to excessive energy consumption, which can reduce the network's lifespan. To mitigate this issue, clustering is a popular technique that can enable energy-efficient data transmission. This study proposes a novel hybrid technique called the "Hybrid Artificial Bee Colony Algorithm and Particle Swarm Optimization (HABCPSO)" to enhance the cluster head selection in the LEACH algorithm. By incorporating PSO, the hybrid algorithm improves the global search behavior of the ABC and achieves an optimal cluster head location. This synergistic approach results in a 40% increase in residual battery power and maintains over 50% alive nodes even in later network rounds, outperforming ABC, PSO, and C-LEACH by substantial margins. Experimental results demonstrate the remarkable energy efficiency of HABCPSO. It selects an average of 48 cluster heads in initial rounds, compared to 65 for ABC, 68 for PSO, and 115 for C-LEACH. This optimized cluster configuration contributes to a notable reduction in energy consumption and a prolonged network lifetime.

Keywords Cluster head · Swarm intelligence · Artificial bee colony algorithm · Particle swarm optimization · LEACH

B. Rambabu (✉) · B. Vikranth · S. Nimmala
CVR College of Engineering, Hyderabad, India
e-mail: rambabubandi@gmail.com

M. A. Kiran
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India

L. Swathi
Sphoorthy Engineering College Hyderabad, Hyderabad, India

10.1 Introduction

Wireless sensor networks have become an important technology in recent years, particularly in the development of "Micro Electrical Mechanical Systems (MEMS)" [1]. These networks find application in a variety of domains, including military and civil applications, agriculture, animal tracking, and habitat monitoring. Sensor nodes are randomly distributed throughout the environment and collect data that is transmitted to an intermediate node [2]. Because these nodes are battery-powered, they have limited computing and processing capabilities. In regions that are far away or hostile, replacing or recharging batteries is challenging, necessitating the development of network layouts and operations that maximize the use of available energy by each node [3].

One popular technique for optimizing wireless sensor networks is clustering, which minimizes the energy consumed during data exchange [4]. The network is divided into clusters, each of which is managed by a node acting as a "cluster head". The cluster head is responsible for managing data transmission activities and aggregating information for transmission to the base station. This node requires additional energy for data transfer and is responsible for coordinating processing events. The primary challenge in designing efficient wireless sensor networks is identifying the most effective cluster head [5]. The LEACH protocol is a clustering-based hierarchy technique for wireless sensor networks, but it has certain limitations. For example, it ignores the node's position and remaining energy, making it less effective in certain situations. To address these limitations, the C-LEACH algorithm was developed [6], which uses Simulated Annealing optimization to construct clusters and select a higher energy level node as Cluster Head [7]. This typically results in a cluster head with an energy level higher than the average energy level of other nodes in the cluster, improving the efficiency of the overall network.

The Artificial Bee Colony (ABC) algorithm has demonstrated significant competitiveness against other population-based algorithms in the literature [8]. However, one limitation of ABC is its inadequacy in exploiting solutions, as it excels in exploration [9]. To address this issue, we propose a novel approach for artificial bee colonies that incorporates the particle swarm search mechanism. Good point set theory is used to generate the initial population rather than random selection, which accelerates convergence. The employed bees, onlookers, and scouts engage the PSO mechanism to search for new candidate solutions, thereby enhancing exploitation capability. Furthermore, to further improve the algorithm's searching capabilities, the current iteration's best answer is subjected to a chaotic search operator.

The primary objective of this research is to integrate a "Hybrid ABC Algorithm with Particle Swarm Optimization (HABCPSO)" into the C-LEACH algorithm to improve the network's lifetime and reduce energy consumption. The proposed methodology considers both the residual energy and the distance between the cluster head and non-cluster members in determining the optimal cluster head.

10.2 Related Work

Various clustering strategies, including "Low-Energy Adaptive Clustering Hierarchy (LEACH), Harmony Search Algorithm (HSA), and Particle Swarm Optimization (PSO)", have been proposed [10]. However, these algorithms are limited by trade-offs between exploration and exploitation and local search restriction [11]. To overcome these limitations, a "hybrid approach called Hybrid Harmony Search Technique and Particle Swarm Optimization (HSA-PSO)" [12] is proposed for energy-efficient cluster head selection. While HSA is region-specific, PSO is better at traversing regions in search of optimal solutions. However, in high-dimensional problems, PSO can be constrained by optimization constraints, making it challenging to explore every potential area of the search space. Therefore, the proposed hybrid approach combines the dynamic potential of HSA with the substantial search efficiency of PSO to select competent sensor nodes in the network as cluster heads. The hybrid HSA-PSO algorithm outperforms solo PSO and HSA algorithms in terms of throughput and residual energy by 83.21% and 29.14%, respectively, and is effective in balancing the network's energy efficiency to extend its lifespan [13].

"The Firefly Optimization Cyclic Redundancy (FFOCR)" [14] protocol is a hybrid firefly-based clustering method that aims to select cluster heads in wireless sensor networks (WSNs) to extend the network's lifespan. In comparison to the LEACH protocol, the FFOCR approach avoids the NP issue and delivers faster convergence with fewer local optima. It outperforms both LEACH and energy-efficient hierarchical clustering in reducing packet loss and extending the network's lifespan. Another approach, the "Hybrid Harmony Search algorithm and Cuckoo search optimization (HSA-CSO) [15] based Cluster Head Selection Scheme (CHS-CHSS)", prioritizes cluster head uniformity to balance node energy and extend the network's lifespan. The HSA-CSO approach was found to outperform ACO and PSO techniques in terms of network longevity, energy usage, and node survival rates. "The Gravitational Search Firefly Algorithm (GSFA)" [16] based clustered routing strategy uses a fitness function for cluster head selection and optimal route generation based on the cluster head balancing factor, residual energy, and intra-cluster distance. Finally, a proposed scheme combines the Artificial Bee Colony (ABC) algorithm with PSO capabilities to select the optimal cluster head in wireless sensor networks.

In another published work, it cleverly combines elements of two existing algorithms—Artificial Bee Colony (ABC) and Monarchy Butterfly Optimization (MBOA) to overcome their individual limitations. HABC-MBOA [17] leverages MBOA's "mutated butterfly adjusting operator" during the crucial "employee bee phase" of ABC. This injects fresh exploration potential into ABC, preventing early trapping in local optima and maintaining a healthy balance between searching for new options and refining promising ones. Simulations demonstrate the effectiveness of HABC-MBOA, with a remarkable 18.92% increase in the number of alive nodes compared to traditional methods.

The paper "Self-Adapting Differential Search Strategies Improved Artificial Bee Colony Algorithm-based Cluster Head Selection Scheme for WSNs" proposes a

creative solution: SADSS-IABCA [18], a hybrid algorithm that merges the strengths of Artificial Bee Colony and Differential Search. The crucial twist comes in the form of self-adapting parameters. SADSS-IABCA adjusts its search behavior on the fly, preventing it from getting stuck in suboptimal solutions like lesser algorithms. This dynamic approach translates into impressive results. Compared to traditional methods, SADSS-IABCA boasts a staggering 54.2% increase in the number of alive nodes, essentially doubling the network's lifespan. Additionally, it delivers a 22.7% improvement in throughput, ensuring faster data transmission with less energy consumption. These impressive quantitative results solidify SADSS-IABCA as a promising contender for extending WSN longevity and optimizing their performance.

A Hybrid ML and Fuzzy Logic Approach for Energy-Efficient Routing in WSNs [19] proposes a clever routing algorithm that blends fuzzy logic and machine learning. It picks the most energy-efficient cluster heads and charts optimal routes, boosting network lifespan by 50% and data delivery by 20%, all while keeping energy consumption low.

The following Table 10.1 shows the comparative study of the above-mentioned related works as discussed in this section.

10.3   Hybrid Artificial Bee Colony and Particle Swam Optimization Based Cluster Head Selection Scheme

The C-LEACH protocol is an advanced version of the LEACH algorithm designed for building optimal clusters. This protocol consists of two phases: "(1) setup phase and (2) steady-state". In the setup phase, each sensor node gathers its location and energy level and sends them to the base station, which computes the average energy level. If a sensor node's energy level exceeds the average, it is selected as the current round's cluster head. The C-LEACH algorithm utilizes simulated annealing at the base station to generate the optimal clusters. This paper proposes a novel approach that incorporates the Artificial Bee Colony algorithm with PSO into the C-LEACH algorithm. The proposed method evaluates the fitness function using residual energy of the nodes and the distance between sensor nodes and cluster heads to determine the optimal placement of cluster heads. This approach enhances network longevity, promotes efficient energy usage and conservation, and improves data transfer.

10.3.1   Optimal Cluster Head Selection in C-LEACH-Using Hybrid ABC and PSO Algorithm

The proposed solution extends the network's lifespan and reduces its energy use. In the C-LEACH algorithm, the following steps are utilized to build clusters and select

Table 10.1 Comparative study

Approach	Algorithm(s)	Strengths	Weaknesses	Performance metrics
HSA-PSO	Hybrid Harmony Search and Particle Swarm Optimization	Balances dynamic potential and search efficiency, extends network lifespan	High-dimensional complexity	Throughput, residual energy
FF OCR	Hybrid Firefly Optimization and Cyclic Redundancy	Fast convergence, avoids NP issue, reduces packet loss	Limited details on optimization	Network lifespan, packet loss
HSA-CSO	Hybrid Harmony Search and Cuckoo Search	Prioritizes cluster head uniformity, extends network lifespan	Complexity of combined algorithms	Network longevity, energy usage, survival rate
GSFA	Gravitational Search Firefly Algorithm	Fitness-based cluster head selection and routing, considers balancing factors	Complex parameter tuning	Cluster head selection, route optimization
HABC-MBOA	Hybrid Artificial Bee Colony and Monarchy Butterfly Optimization	Mutated operator prevents early stagnation, balances exploration and exploitation	Complex parameter tuning	Network longevity, energy usage, survival rate
SADSS-IABCA	Self-Adapting Differential Search and Artificial Bee Colony	Self-adapting parameters avoid suboptimal solutions, extends network lifespan and improves throughput	Complexity of combined algorithms	Alive nodes, throughput
EF-FIBR	Machine Learning and Fuzzy Logic	Efficient cluster head selection, optimal routing, extends lifespan and data delivery	High complexity	Network lifespan, data delivery, energy consumption

the ideal cluster head. Figure 10.1 depicts the HABCPSO methodology presented and embedded in the C-LEACH algorithm.

10.3.2 Hybrid ABC-PSO Algorithm for Optimization

The Hybrid ABC-PSO algorithm combines the strengths of Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO) meta-heuristics to achieve enhanced performance in optimization problems. This algorithm leverages the exploration capabilities of ABC and the efficient exploitation abilities of PSO, leading to

```
┌─────────────────────────────────────────────────────────────┐
│  Wireless sensor network(N-number of sensor nodes, Total Population │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  initialize candidate solutions with kopt number of optimal CH │
│                  randomly from the population WSN              │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│         Initialized the parameters of Hybrid ABC with PSO      │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Apply the Hybrid ABC with PSO for the updating the candidate  │
│ solutions by efficient exploration and exploitation of search in the total │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Calculate the fitness value of each new candidate solution based on │
│ leftover energy, distance and then apply greedy algorithm to update the │
│ solutions                                                      │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Store the highest fitness candidate solution as an optimal solution in │
└─────────────────────────────────────────────────────────────┘
                              ↓
       ┌──────────┐      ◇ Tc > Tn ◇
       │          │ ←────
       └──────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Choose highest fitness value candidate solution as a the optimal set of │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Started Steady State: data gathered in each cluster by head and transmitted │
└─────────────────────────────────────────────────────────────┘
                              ↓
       ┌──────────┐      ◇        ◇
       │          │ ←────
       └──────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│  Collect all the statistics related WSN like life time (number of rounds) , │
│              throughput and energyconsumption etc.             │
└─────────────────────────────────────────────────────────────┘
```

Fig. 10.1 Flowchart for the Hybrid ABC-PSO

improved convergence speed and solution quality. Here's a detailed breakdown of the algorithm.

1. Initialization:

Generate a population of N particles (solutions) within the defined search space. Initialize employed bees (solutions currently exploiting promising regions) and onlooker bees (unemployed bees searching for new resources).Set parameters such as maximum iterations (max_iter), population size (N), learning rates for ABC and PSO components (α, ω), and stagnation thresholds for employed bees ($limit$).

2. Employed Bee Phase (ABC):

Food Source Modification: Each employed bee randomly modifies its solution (xi) based on the global best solution ($gbest$) found so far. This modification involves a perturbation vector (Δxi) drawn from a Gaussian distribution centered around $gbest$ with a scaling factor (φ).

Fitness Evaluation: Calculate the fitness value of the modified solution ($f(xi + \Delta xi)$) using the fitness function.

Greedy Selection: If the new solution has a better fitness than the current one, replace the original solution in the population.

3. Onlooker Bee Phase (Hybrid ABC-PSO):

Probability Distribution: Onlooker bees choose potential solutions based on a probability distribution weighted by the fitness values ($f(xi)$) and a PSO-inspired velocity term (vi).

Velocity Update: The velocity term guides bees towards promising regions influenced by personal best ($pbesti$) and global best positions ($gbesti$) from both PSO and ABC components.

Solution Selection: Onlooker bees select solutions according to the calculated probabilities. This probabilistic selection balances exploration and exploitation, encouraging bees to move towards promising regions while still exploring the search space.

4. Scout Bee Phase (ABC):

If a food source hasn't been improved for limit consecutive iterations, it is abandoned, and the employed bee associated with it becomes a scout bee.

Scout bees randomly generate new solutions within the search space, promoting exploration and preventing premature convergence.

5. PSO Update (Hybrid ABC-PSO):

Update the velocity of each particle (vi) based on its current velocity, personal best position ($pbesti$), and the global best positions ($gbesti$) from both PSO and ABC components. Update the particle's position (xi) using the updated velocity and a constriction factor (χ) to control the search space exploration.

6. Global Best Update:

Update the global best solution ($gbest$) based on the best solution found during the current iteration, considering both ABC and PSO components.

7. Loop and Termination:

Repeat steps 2–6 for *max_iter* iterations or until a specified termination criterion is met (e.g., convergence threshold).

10.3.3 Advantages of Hybrid ABC-PSO

1. Improved exploration–exploitation balance compared to individual ABC or PSO algorithms.
2. Enhanced convergence speed and solution quality due to the synergistic effect of both algorithms.
3. Robustness against premature convergence through scout bees and PSO's velocity update.

10.4 Results Analysis

We examined the performance of the proposed scheme through simulations using MATLAB, and we compared the results to C-LEACH, PSO, and ABC, the sensor nodes are dispersed at random in a field that is 1000×1000 m, and the initial energy of each sensor node is set to 10 J. The number of nodes in the network is 500. The number of clusters is based on the optimal number of clusters given a field with a number of sensors and the first-order energy model. The experimental setup is presented in Table 10.2, which includes the parameter values.

In a wireless sensor network, cluster head selection plays a crucial role in optimizing energy consumption and prolonging the network's lifetime. the performance of four different clustering algorithms (HABCPSO, ABC, PSO, and C-LEACH) on a wireless sensor network for a varying number of rounds (100–1000). The performance of the algorithms has been measured based on three metrics: percentage of leftover battery, percentage of live nodes, and number of optimal cluster heads.

In this study we compare the performance of four algorithms, HABCPSO, ABC, PSO, and C-LEACH, in terms of the optimal number of cluster heads selected for different numbers of rounds. The number of rounds in the wireless sensor network is considered the independent variable, and the optimal cluster head values are considered the dependent variable. As the lower number of optimal cluster heads is considered to be better performance of the algorithm, we will analyze the results accordingly.

Table 10.2 Simulation parameters

Parameter	Value
Network area	1000 m × 1000 m
Number of nodes	500
Energy model	First-order energy model
Initial energy	10 Joules
Packet size	1024 bytes
Routing protocol	HABC-PSO

Fig. 10.2 Number of cluster heads
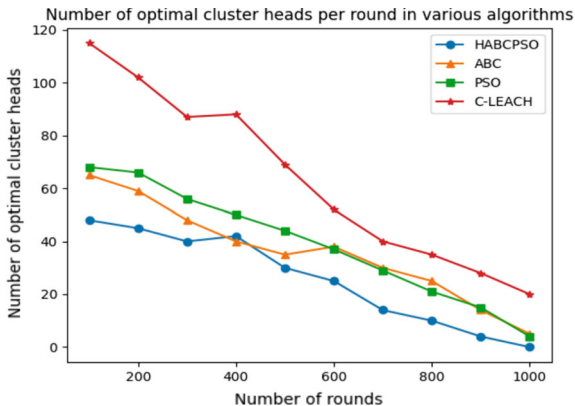
As shown in Fig. 10.2, the HABCPSO algorithm outperforms the other three algorithms for most of the rounds. In the initial rounds, HABCPSO selects only 48 cluster heads, while ABC and PSO select 65 and 68 cluster heads, respectively, and C-LEACH selects 115 cluster heads. As the number of rounds increases, the optimal number of cluster heads selected by the HABCPSO algorithm also increases but still remains lower than the other three algorithms. For instance, at 1000 rounds, HABCPSO selects 0 cluster heads, while the other three algorithms still select a significant number of cluster heads. PSO algorithm selects 4 cluster heads, ABC algorithm selects 5 cluster heads, and C-LEACH algorithm selects 20 cluster heads. This result suggests that the HABCPSO algorithm can achieve better energy efficiency and prolong the network lifetime compared to the other three algorithms in a wireless sensor network.

Fig. 10.3 represents the performance of four different algorithms, HABCPSO, ABC, PSO, and C-LEACH, on a wireless sensor network with different numbers of rounds. The performance is measured in terms of the percentage of leftover battery, which is an important parameter for energy-efficient algorithms. According to Fig. 10.3, the HABCPSO algorithm outperforms the other three algorithms in terms of the percentage of leftover battery for all the number of rounds. However, the difference in performance decreases with an increase in the number of rounds. The HABCPSO algorithm shows 100% leftover battery for the first round and maintains a high percentage of leftover battery for all the other rounds. On the other hand, the other algorithms show a significant decrease in the percentage of leftover batteries with an increase in the number of rounds.

Figure 10.4 depicts the performance in terms of percentage of alive nodes, HABCPSO performs the best, with values consistently above 50%, even in the later rounds. ABC and PSO both show a decrease in performance as the rounds increase, with values dropping below 30% in the later rounds. C-LEACH performs the worst in this metric, with values consistently below 30%, and dropping to 5% in the later rounds. When considering the percentage of leftover battery, HABCPSO again shows the best performance, with values consistently above 40% even in the later rounds.

Fig. 10.3 Percentage of remaining battery
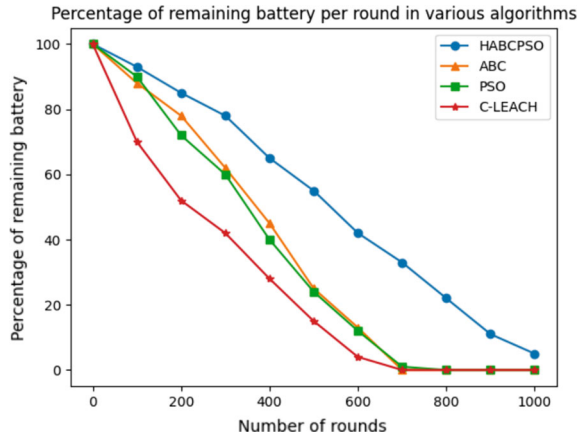
Fig. 10.4 Number of live nodes
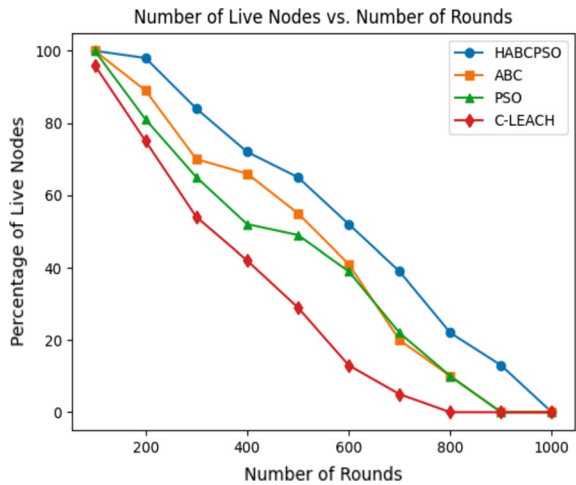
ABC and PSO both show a decrease in performance as the rounds increase, with values dropping to 0% in the later rounds. C-LEACH performs the worst in this metric, with values consistently below 30%, and dropping to 0% in the later rounds.

10.5 Conclusions

In conclusion, this research paper evaluated the performance of four different clustering algorithms (HABCPSO, ABC, PSO, and C-LEACH) in a wireless sensor network based on three metrics: percentage of leftover battery, percentage of live nodes, and number of optimal cluster heads. The results showed that the HABCPSO

algorithm outperformed the other three algorithms in terms of energy efficiency, network lifetime, and node survival. The algorithm selected a lower number of optimal cluster heads, maintained a high percentage of leftover battery, and kept a high percentage of alive nodes throughout all rounds. On the other hand, the other three algorithms showed a significant decrease in performance as the number of rounds increased. These findings suggest that the HABCPSO algorithm is an effective solution for cluster head selection in a wireless sensor network, as it optimizes energy consumption and prolongs the network's lifetime, making it a valuable contribution to the field of wireless sensor networks.

References

1. Quy, V.K., Ban, N.T., Nam, V.H., Tuan, D.M., Han, N.D.: Survey of recent routing metrics and protocols for mobile ad-hoc networks. J. Commun. **14**, 110–120 (2019)
2. Shi, B. & Zhang, Y. A novel algorithm to optimize the energy consumption using iot and based on ant colony algorithm. Energies 14, (2021).
3. Mekala, S., Chatrapati, K.S.: A hybrid approach to neighbour discovery in wireless sensor networks. Intell. Autom. Soft Comput. **35**, 581–593 (2022)
4. Yarinezhad, R., Hashemi, S.N.: Increasing the lifetime of sensor networks by a data dissemination model based on a new approximation algorithm. Ad Hoc Netw. 100 (2020)
5. Fatemidokht, H., Rafsanjani, M.K., Gupta, B.B., Hsu, C.H.: Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular Ad Hoc networks in intelligent transportation systems. IEEE Trans. Intell. Transp. Syst. **22**, 4757–4769 (2021)
6. Singh, S.P., Sharma, S.C.: An improved cluster-based routing algorithm for energy optimisation in wireless sensor networks. Int. J. Wirel. Mob. Comput. **14**, 82 (2018)
7. Potthuri, S., Shankar, T., Rajesh, A.: Lifetime improvement in wireless sensor networks using hybrid differential evolution and simulated annealing (DESA). Ain Shams Eng. J. **9**, 655–663 (2018)
8. Rambabu, B., Venugopal Reddy, A. Janakiraman, S.: A hybrid artificial bee colony and bacterial foraging algorithm for optimized clustering in wireless sensor networks. Int. J. Innov. Technol. Explor. Eng. **8**, 2186–2190 (2019)
9. Yu, X., Wang, H., Lu, Y.: An adaptive ranking moth flame optimizer for feature selection. Math. Comput. Simul **219**, 164–184 (2024)
10. Janakiraman, S., Deva Priya, M.: An energy-proficient clustering-inspired routing protocol using improved Bkd-tree for enhanced node stability and network lifetime in wireless sensor networks. Int. J. Commun. Syst. **33**, e4575 (2020)
11. Gheisari, M. et al.: A survey on clustering algorithms in wireless sensor networks: challenges, research, and trends. In: Proceedings of the- 2020 International Computer Symposium. ICS 2020, pp. 294–299 (2020) https://doi.org/10.1109/ICS51289.2020.00065
12. Shankar, T., Shanmugavel, S., Rajesh, A.: Hybrid HSA and PSO algorithm for energy efficient cluster head selection in wireless sensor networks. Swarm Evol. Comput. **30**, 1–10 (2016)
13. Janakiraman, S., Jayasingh, B.B.: A hyper-exponential factor-based semi-markov prediction mechanism for selfish rendezvous nodes in MANETs. Wirel. Pers. Commun. **108**, 1493–1511 (2019)
14. Sarkar, A., Senthil Murugan, T.: Cluster head selection for energy efficient and delay-less routing in wireless sensor networks. Wirel. Netw. **25**, 303–320 (2017)
15. Gupta, G.P.: Improved cuckoo search-based clustering protocol for wireless sensor networks. Procedia Comput. Sci. **125**, 234–240 (2018)

16. Lalwani, P., Banka, H., Kumar, C.: GSA-CHSR: Gravitational search algorithm for cluster head selection and routing in wireless sensor networks. Appl. Soft Comput. Web **1**, 225–252 (2017)
17. B Rambabu, A.V.R.S.J.: Hybrid artificial bee colony and monarchy butterfly optimization algorithm (HABC-MBOA)-based cluster head selection for WSNs. J. King Saud Univ. - Comput. Inf. Sci. **3**, 67–79 (2019)
18. Bandi, R., Ananthula, V.R., Janakiraman, S.: Self adapting differential search strategies improved artificial bee colony algorithm-based cluster head selection scheme for WSNs. Wirel. Pers. Commun. **121**, 2251–2272 (2021)
19. Mekala, S., Mallareddy, A., Tandu, R.R., Radhika, K.: Machine Learning and Fuzzy Logic Based Intelligent Algorithm for Energy Efficient Routing in Wireless Sensor Networks. Lecture Notes in Computer Science (including Subser. Lecture Notes in Artificial Intelligence. Lecture Notes Bioinformatics), vil. 14078 LNAI, pp. 523–533 (2023)

Chapter 11
Machine Learning Pipeline for Multi-grade Classification in Pancreatic Cancer Detection Using Urinary Biomarkers

Pragya Pandey, Param Verma, Garima Aggarwal, and Malay Kishore Dutta

Abstract Pancreatic cancer's aggressive nature has driven the exploration of advanced machine learning techniques. Nevertheless, the difficulty arises from the fact that pancreatic ductal adenocarcinoma (PDAC) lesions identified through biopsies often occur at advanced stages, which impedes early detection initiatives. In response to this limitation, the integration of innovative and validated urinary biomarkers emerges as a solution, addressing the shortcomings associated with traditional biopsy methods. These urinary biomarkers prove instrumental in overcoming the delay in detection commonly associated with PDAC lesions, presenting a promising avenue for early diagnosis and improved patient outcomes. This study introduces an innovative two-stage machine learning pipeline for pancreatic cancer diagnosis, utilizing urinary biomarkers. The initial binary classification differentiates pancreatic ductal adenocarcinoma (PDAC) from a cohort comprising benign hepatobiliary disease and control cases, incorporating robust data preprocessing, and addressing class imbalance. The subsequent fine-grained classification within the PDAC group categorizes cases into grades I–IV, employing a Support Vector Machine (SVM) classifier optimized through grid search. The approach achieves outstanding accuracy, with the best SVM hyperparameters yielding 99.7% accuracy in validation and testing sets. This method advances early detection and precise classification of pancreatic cancer, with ongoing research comparing other methods for comprehensive assessment.

Keywords Pancreatic cancer · Machine learning pipeline · Urinary biomarkers · Hyperparameter optimization · Early detection

P. Pandey (✉) · P. Verma · G. Aggarwal
Department of Computer Science Engineering, Amity University, Noida, India
e-mail: pragyapandey2709@gmail.com

G. Aggarwal
e-mail: gmehta@amity.edu

M. K. Dutta
Amity Centre for Artificial Intelligence, Noida, India

11.1   Introduction

With one of the worst prognosis of any malignancy, pancreatic ductal adenocarcinoma (PDAC) poses a formidable challenge in the field of oncology. Although surgical resection remains the only potentially curative treatment for this aggressive cancer, the reported 5-year survival rates stutter at an upsetting range of 2–10% as highlighted in the paper by Nakagohri et al. [1]. The importance of an early and accurate diagnosis is crucial for increasing these survival rates.

There are three crucial stages in the diagnosis of pancreatic cancer that determine the subsequent course of treatment. The initial phase involves the detection of the tumor itself. The enormous difficulty of achieving an early diagnosis is one of the critical factors contributing to the low survival rates. It is possible to identify a greater percentage of patients with pancreatic cancer in its early stages if detection of these tumors is made more sensitive. The process that includes diagnosing pancreatic adenocarcinoma involves separating it from other pancreatic conditions, such as chronic pancreatitis, benign or malignant islet cell tumors, and intraductal papillary mucinous neoplasms. To customize the most effective therapeutic strategies, this differentiation must be done with extreme precision. The ability of imaging techniques to enable precise tumor staging is the last important aspect of diagnosis. Any signs of vascular infiltration, lymph node involvement, or potential distant metastases assume exceptional importance when they are related to pancreatic cancer. The paper by Kamisawa et al. [2] marks pancreatic intraepithelial neoplasia, tiny non-invasive cellular proliferations that first appeared in pancreatic ducts, as the cause of the disease.

The development of effective treatment plans for pancreatic cancer relies heavily on accurate staging, utilizing various data sources such as cross-sectional imaging studies. Studies, like the one by Appel et al. [3], emphasize the importance of pretreatment evaluations and standardized pathology reviews, providing insights into resectability, imaging modalities, and prognostic markers in pancreatic cancer.

A groundbreaking study by Debernardi et al. [4] has made significant strides in pancreatic cancer detection by examining pre-diagnostic samples taken up to five years before detection. This study, using urinary biomarkers (LYVE1, REG1B, TFF1) and the blood marker CA19-9, shows promising potential, capable of identifying the disease up to two years before the typical diagnosis (72% accuracy at 90% specificity).

The major contribution of the present research lies in the thoughtful design of a machine learning pipeline, addressing the urgent need for heightened diagnostic accuracy by incorporating essential urinary biomarkers. The study showcases remarkable performance, surpassing alternative classifiers and emphasizing the pressing importance of early detection in pancreatic cancer. This research pioneers advanced methods for early detection, particularly through the integration of urinary biomarkers within a meticulously designed two-stage machine learning pipeline.

1. The present model introduces a pioneering machine learning pipeline, meticulously designed to leverage demographic data and crucial urinary biomarkers for a holistic approach to pancreatic cancer detection.
2. The current model demonstrates outstanding performance, surpassing alternative classifiers and reaffirming the imperative for early detection in pancreatic cancer.
3. The study optimizes SVM hyperparameters, achieving exceptional accuracy and reinforcing the potential for improved patient outcomes.
4. The model achieves fine-grained classification within the PDAC group, categorizing cases into grades I, II, III, and IV, employing a Support Vector Machine (SVM) classifier.

The remainder of the paper is organized in following sections—Part 2 gives a brief literature review of pre-existing methods, followed by description of methodology in Part 3, leading up to results and discussions in Part 4 and ending with conclusion in Part 5.

11.2 Literature Review

Research on PDAC is changing because of the critical contributions made by machine learning techniques in decision support, risk assessment, and prediction for early diagnosis and customized interventions as outlined in paper by Yokoyama et al. [5] and Liu et al. [6]

Yokoyama et al. [5] provide insights into mucin gene methylation's prognostic implications in pancreatic cancer, aligning with the trajectory outlined in the innovative machine learning pipeline and its approach leveraging urinary biomarkers. This harmonizes with the multifaceted strategy proposed in the advanced machine learning pipeline. Zhang et al.'s [7] work emphasizes the challenges in achieving timely and accurate diagnoses, connecting with the significance of updated diagnostic strategies discussed in the prior review. Karley et al.'s [8] paper on biomarkers provides information on their importance in cancer prediction. Wang et al.'s [9] machine-learning model for pancreatic cancer prognosis, including the 32-gene consensus prognostic signature (AIDPS) can be linked with Hayashi et al.'s [10] histology-based machine learning model for predicting recurrence patterns. This comprehensive approach resonates with Dalal et al.'s [11] proposal of radiomics as a cost-effective solution for categorizing pancreatic cystic lesions. Chen et al.'s [12] risk prediction model for stratifying pancreatic cancer risk, highlights the pivotal role of molecular markers, imaging technologies, and innovative machine learning in enhancing pancreatic cancer detection's accuracy and efficacy.

Present study builds up on works highlighted above to shed light on the importance of urinary biomarkers that aid early detection with high accuracy, aiming to cover the research gap in pre-existing methods.

11.3 Materials and Methods

11.3.1 Dataset Details

A dataset with a total of 590 data points was analyzed for this study. These informational points were divided into three distinct diagnostic classes: benign, pancreatic ductal adenocarcinoma (PDAC), and control. The planned study involved the collection of urinary biomarkers from three different patient groups, being healthy controls (183 samples), individuals suffering from non-cancerous pancreatic diseases, such as chronic pancreatitis (208 samples) and people who have been given a PDAC diagnosis (199 samples) The dataset obtained from an open-access paper published by Debernardi et al. [4]. It is emphasized that this dataset is completely open-access and covered by the Creative Commons Attribution (CC-BY) license. The dataset focuses on four key urinary biomarkers—creatinine, LYVE1, REG1B, and TFF1—alongside demographic information such as age and sex. Biomarkers, defined by NIH, objectively measure health, disease onset, or treatment responses. Biomarkers encompass various markers like proteins, SNP haplotypes, mRNA expression, etc., providing crucial insights for disease diagnosis, prognosis, and treatment outcomes, especially in early cancer detection. The analysis heavily relies on creatinine as a measure of kidney function. LYVE1, linked to Lymphatic Vessel Endothelial Hyaluronan Receptor 1, may impact tumor metastasis. REG1B and TFF1 are associated with urinary tract repair and pancreatic regeneration. The dataset includes benign condition diagnosis, CA 19-9 plasma levels (a pancreatic cancer biomarker), and disease stage for PDAC patients. CA 19-9 levels were examined in a subset to compare urinary and plasma biomarkers. LYVE-1, REG1A, and TFF1's significance lies in their ability to identify early-stage PDAC in urine samples, providing a non-invasive means for early cancer detection. Currently, PDAC is often diagnosed at an advanced stage, complicating effective treatment. Proteomic analysis of urine samples from healthy individuals, those with chronic pancreatitis, and PDAC patients led to the selection of these three biomarkers. CA19-9, a glycoprotein found in various organs, is a marker for pancreatic carcinoma and other cancers, according to studies by Wang et al. [13] and Humphris et al. [14]. Creatinine is a key measure of kidney function, LYVE1 is linked to tumor metastasis, while REG1B and TFF1 are associated with urinary tract repair and pancreatic regeneration.

There are given 14 important columns in this dataset, each of which adds to present knowledge of early pancreatic cancer detection. Table 11.1 represents a summary of all data columns via their minimum, maximum and mean value.

1. Sample ID: Individual identification number for each subject.
2. Patient's Cohort: Grouping into Cohort 1 (samples already used) and Cohort 2 (newly added samples).
3. Sample Origin: Information on sample origin, including Barts Pancreas Tissue Bank (BPTB), Spanish National Cancer Research Center (ESP), University College London (UCL), and Liverpool University (LIV).

Table 11.1 Parametric information about dataset

Parameter	Control subjects			Benign diagnosis			PDAC diagnosis		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Age	55.00	58.00	56.66	52.00	77.00	68.08	53.00	82.00	65.71
Plasma CA19_9	10.00	11.00	10.66	0.60	5335	1070.434	654.00	654.00	654.00
Creatinine	0.44	1.73	0.88	0.13	1.62	0.61	0.56	2.23	1.01
LYVE1	0.63	2.83	2.03	0.56	12.24	3.06	7.84	11.99	9.97
REG1B	33.40	188.25	87.42	7.08	535.28	128.81	66.83	1293.81	503.06
TFF1	138.63	529.98	330.59	0.04	1529.18	564.47	138.32	2921.50	1467.85
REG1A	155.27	735.28	366.91	39.911	7301.55	1183.79	110.98	4066.13	1417.72

4. Age: Subjects' ages in years.
5. Sex: Gender classification (M = male, F = female).
6. Diagnosis: Categorical variable: 1 for control group, 2 for benign hepato-biliary disease (including chronic pancreatitis), and 3 for pancreatic ductal adenocarcinoma (PDAC).
7. Stage: Grading system for PDAC, with stages IA, IB, IIA, IIIB, III, and IV.
8. Benign Samples Diagnosis: Precise diagnosis for individuals with non-cancerous conditions in the benign category.
9. Plasma CA19-9 U/ml: Concentration of the monoclonal antibody CA 19-9 in blood plasma, compared to urinary samples in the model.
10. Creatinine mg/ml: Indicator of kidney health found in urine.
11. LYVE1 ng/ml: Protein linked to tumor metastasis.
12. REG1B ng/ml: Urinary concentrations of a protein related to pancreas regeneration.
13. TFF1 levels in urine: Linked to urinary tract regeneration and repair.
14. REG1A ng/ml: Urinary concentrations of a protein related to pancreas regeneration.

11.3.2 Proposed Methodology

The methodology employed for advancing early detection of pancreatic cancer was designed, commencing with a rigorous series of steps. The initial phase involved comprehensive data collection, entailing close collaboration with medical professionals to acquire a dependable dataset. Stringent validation procedures were implemented to ensure the accuracy and reliability of the dataset. The focal point of the study was the investigation of crucial biomarkers with potential for pancreatic cancer detection. Biomarkers such as Creatinine, LYVE1, REG1B, and TFF1 were carefully identified and examined to elucidate their significance and diagnostic capabilities.

Data preprocessing, a pivotal step in the methodology, was executed through several essential stages. The dataset was loaded and meticulously examined using Pandas for further analysis. Distinctions between numerical and non-numerical features were established, and missing data in numerical columns were addressed using the SimpleImputer method with the mean strategy. Non-numeric features were encoded with LabelEncoder to ensure compatibility with machine learning algorithms. Stratified data partitioning was employed to generate distinct sets for training, validation, and testing. A particular emphasis was placed on separating training and validation data to facilitate a comprehensive model evaluation process and mitigate the risk of data leakage.

The primary focus then shifted to model construction and hyperparameter tuning. Feature variables (X) and the target variable (y) were defined for the machine learning model, setting the stage for the introduction of the Support Vector Machine (SVM) classifier. Hyperparameters of the SVM classifier, including the regularization parameter 'C' and kernel types (linear, RBF, and polynomial), were systematically explored through GridSearchCV. The identification of the best hyperparameters aimed to optimize model performance. In the subsequent model evaluation phase, the best SVM classifier was deployed to make predictions on the validation set, and its performance was rigorously assessed, predominantly through accuracy and classification reports. The model was then implemented on the test set to evaluate its performance in real world scenarios, confirming its suitability for practical use.

As an innovative strategy to enhance the model's performance and robustness, voting classifiers were introduced. A thorough understanding of both Soft and Hard Voting Classifiers paved the way for the incorporation of ensemble methods before the development of k-fold cross-validation, providing a dynamic approach to prediction and classification. The methodology concluded with the visualization of a confusion matrix for the testing set using a heatmap, offering clarity and insight into the model's performance. Following the successful implementation of the Support Vector Machine (SVM) and the introduction of voting classifiers, the study extended its scope by comparing the model's performance with several cutting-edge machine learning algorithms. Metrics such as accuracy, precision, recall, and F1-score were used to thoroughly compare the performance of algorithms like Random Forest, Logistic Regression, and Gradient Boosting with the SVM-based model. Additionally, Fig. 11.1 illustrated the proposed methodology for the pipeline, serving as a visual summary of the research results. The binary classification of the dataset, distinguishing between cancerous and benign cases, facilitated a swift and accurate diagnostic procedure. The research's second phase then shifted its emphasis to a more complex assessment of cancerous cases, specifically targeting the identification of the grade of cancerous cells.
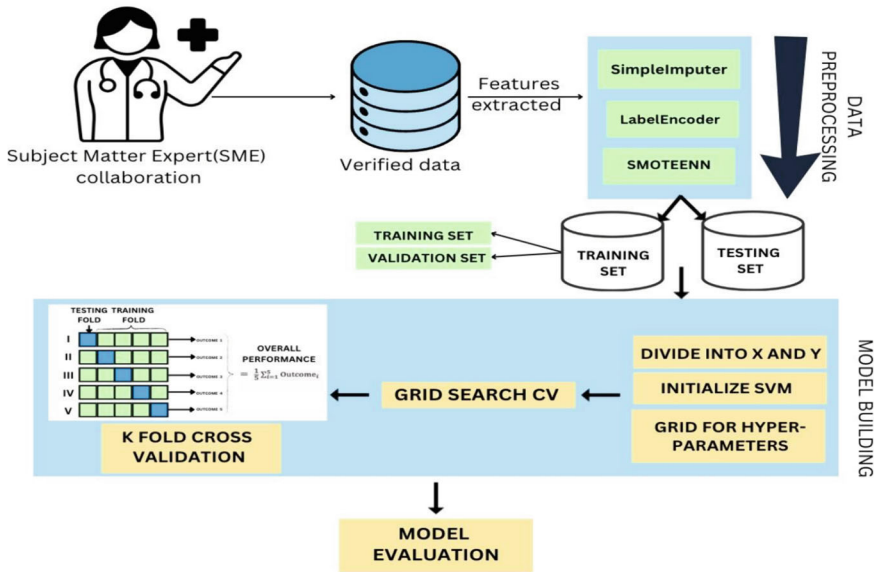
Fig. 11.1 Proposed methodology for pipeline

11.3.3 Description of Major Techniques Being Used

SimpleImputer and LabelEncoder—Preprocessing is crucial for early pancreatic cancer detection. The model utilizes SimpleImputer to handle missing values, maintaining dataset integrity. SimpleImputer's versatility extends to different domains, the same is highlighted in a research by Lin and Tsai [15]. LabelEncoder complements it, assigning numerical labels to categorical attributes for improved machine learning. Beyond research, LabelEncoder is valuable in text-related fields. Its operation, as mentioned by work done by Jia and Zhang [16] and Zeng et al. [17] involves assigning sequential integers to categories, unlocking qualitative information's potential in various scientific domains.

Handling class imbalance using SMOTEENN—SMOTEENN (Synthetic Minority Over-sampling Technique Edited Nearest Neighbors) is a cutting-edge technique that addresses class imbalance in machine learning, particularly beneficial for medical diagnostics such as early pancreatic cancer detection. Developed for improving dataset balance and model accuracy, SMOTEENN combines synthetic sample generation for minority classes with noise removal. Its versatility extends beyond cancer detection, proving valuable in tasks like credit scoring and fraud detection, as demonstrated by Lin et al.'s [18] work showcasing its effectiveness in ionospheric scintillation detection.

SVM Classifier and GridSearchCV—The Support Vector Machine (SVM) stands out as a crucial algorithm for early pancreatic cancer detection, excelling in high-dimensional data complexity and minimizing misclassification. Its adaptability, demonstrated by Mathur and Foody [19], is further optimized through methods like GridSearchCV, as highlighted by Mangkunegara and Purwono [20], ensuring peak performance and precision in medical diagnostics.

K Cross validation and Voting Classifier—K-Fold Cross-Validation facilitates robust model performance assessment by segmenting the dataset into "k" equal-sized subsets. Each fold switches off as the testing set and the training set over the course of "k" iterations. The performance of SVM classifiers is improved by K-Fold Cross-Validation, which is essential for early pancreatic cancer detection. Voting Classifiers, such as Soft and Hard Voting, increase prediction accuracy and dependability by combining knowledge from various base models. They combine SVM and other model outputs in the early pancreatic cancer detection case, resulting in a thorough decision-making process. The work presented by Anguita et al. [21], Fushiki [22] and Jung [23] highlight the importance of SVM as used in various machine learning applications. Voting Classifiers are used in a variety of fields where combining different model predictions increases the reliability of the results. Voting Classifiers mathematically combine predictions from various models, while K-Fold Cross-Validation averages metrics over "k" folds, ensuring accurate model evaluation. Kumar et al. [24] research applies a voting classifier for breast cancer prediction contributing to the broader understanding and application of these techniques across diverse domains.

11.4   Experimental Results and Discussion

Metrics like accuracy, recall, precision, F1 score, and support play a vital role in evaluating a machine learning model for early pancreatic cancer detection. Accuracy gauges overall effectiveness, while recall assesses the model's ability to detect true cancer cases, reducing the risk of false negatives. Precision measures the accuracy of positive predictions, with a perfect value close to 1. The F1 score balances recall and precision, offering a comprehensive assessment considering false positives and false negatives. Support reveals class distribution in the dataset, crucial for context-specific evaluation. The confusion matrix is an indispensable tool in assessing machine learning models for early pancreatic cancer detection. Categorizing predictions into True Positives, True Negatives, False Positives, and False Negatives, it provides a precise evaluation, improving clinical outcomes and patient care by reducing missed diagnoses and unnecessary alarms.

Table 11.2 Binary classification metrics

Diagnosis	Total samples	TP	FN	FP	TN	Recall	Specificity	F1 score
Control	86	36	0	0	50	1.00	1.00	1.00
Benign	76	25	0	0	61	1.00	1.00	1.00
PDAC	76	25	0	0	61	1.00	1.00	

11.4.1   Binary Classification Metrics

In the current binary classification, classes 1 and 2 are combined to represent control and benign hepatobiliary disease, while class 3 (pancreatic cancer) is kept separate.

Table 11.2 provides comprehensive binary classification metrics, detailing the total samples, true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for both "Control and Benign" and "PDAC" categories. Furthermore, the table includes overall diagnosis metrics for "Control," "Benign," and "PDAC" categories. Each category shows perfect performance with 100% recall (sensitivity), specificity, and F1 score, indicating that the model accurately identifies cases of control, benign conditions, and pancreatic cancer.

11.4.2   Multi Classification Metrics

Table 11.3 provides comprehensive metrics for multi-classification, focusing on the stages of Pancreatic Ductal Adenocarcinoma (PDAC) and the overall diagnosis. For PDAC stages (Grade I, II, III, IV), the table outlines the performance metrics such as true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN).

Table 11.3 Multi classification metrics

Stage for PDAC	Total samples	TP	FN	FP	TN	Recall	Specificity	F1 score	Overall accuracy
Grade I	21	8	3	1	9	0.89	0.75	0.81	Fold 1—0.76193
Grade II	21	1	2	0	18	1.00	0.91	0.95	
Grade III	21	0	0	3	18	0.00	1.00	0.00	Fold 2—0.8095
Grade IV	21	7	0	1	13	0.88	1.00	0.93	

11.4.3 Data Analytics and Graphical Representations

Figure 11.2 displays the dataset's inter-column relationships using heatmap, visually conveying associations between columns. The heatmap enhances understanding by emphasizing correlation strength with color gradients. Furthermore, it features confusion matrices, visually representing model performance in terms of true positives, false positives, false negatives, and true negatives. These matrices assess classification accuracy, aiding in model evaluation. Figure 11.3 presents SVM decision boundaries, visually delineating classification regions and showcasing the model's discriminative capabilities in separating and categorizing data points.
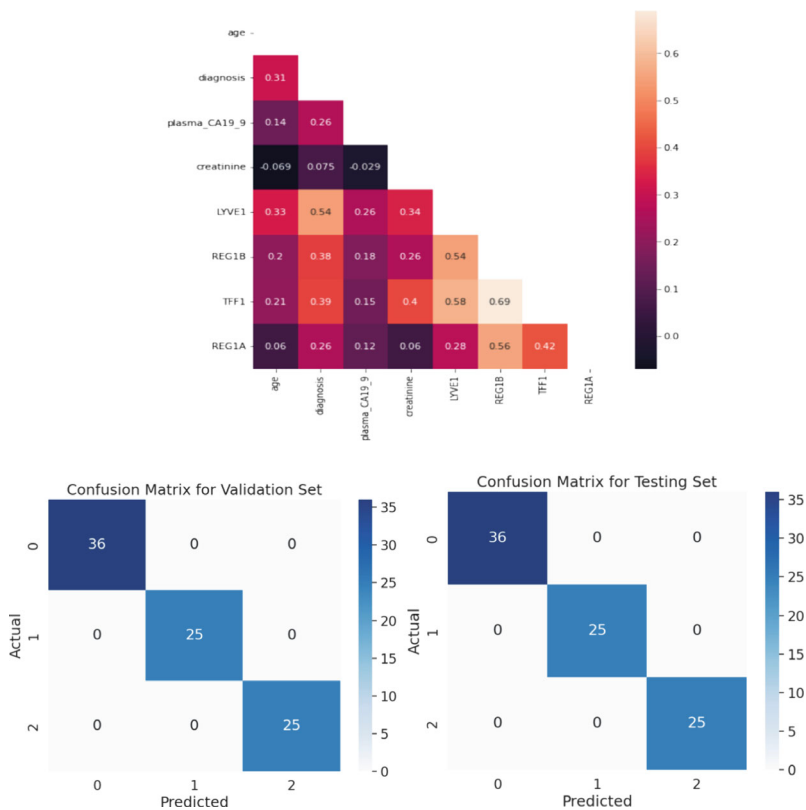
Fig. 11.2 Heatmap between various columns and confusion matrices
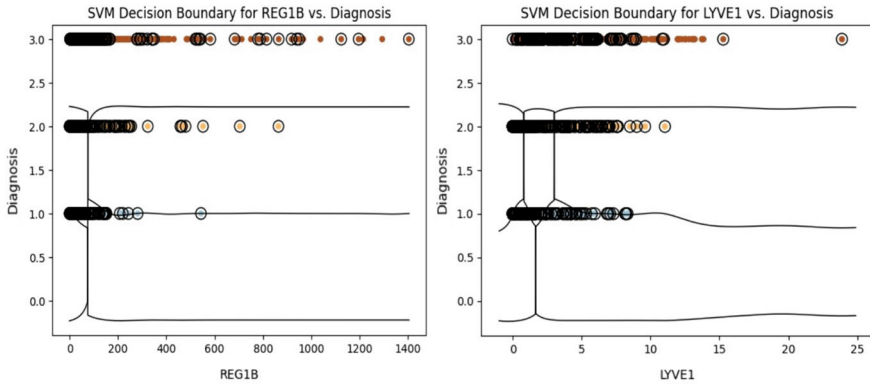
Fig. 11.3 SVM decision boundaries

11.4.4   Comparison with Existing Methods

Table 11.4 compares our framework to established machine learning methods. KNN and SVM excel with mRNA and DNA methylation data, respectively. Bayesian neural networks predict molecular subtypes with radiomic features, Logistic Regression performs well with clinical data, and the proposed SVM with a linear kernel achieves 100% accuracy using urinary biomarkers. Table 11.5 summarizes model performance in diagnosing control, benign, and pancreatic ductal adenocarcinoma.

Table 11.4   Comparison of proposed framework with SOTA machine learning methods

Model	Metric used	Metric value	Parameter used
SVM classifier by Sinkala et al. [25]	Accuracy	98%	DNA methylation-based
Bayesian regularization backpropagation neural network by Nasief et al. [26]	CV-AUC	92%	Kurtosis–coarseness combination
	Accuracy	94%	Kurtosis–coarseness–NESTD
Logistic regression by Sala Elarre et al. [27]	Accuracy	71%	Granulocyte colony-stimulating factors
Gradient-boosted survival regression by Kaissis et al. [28]	Concordance Index (DF)	0.76 ± 0.05	Radiomic parameters
SVM with linear kernel (proposed method)	**Accuracy**	**99.7%**	**Urinary biomarkers**

Table 11.5 Comparison of proposed framework with other machine learning methods

Model	Diagnosis	Precision	Recall	F1 score	Overall accuracy
Ridge classifier	Control	0.88	1.00	0.94	0.94
	Benign	1.00	0.80	0.89	
	PDAC	1.00	1.00	1.00	
SVM classifier with RBF kernel	Control	0.62	0.94	0.75	0.71
	Benign	0.75	0.36	0.49	
	PDAC	0.95	0.72	0.82	
Logistic regression	Control	0.92	1.00	0.92	0.93
	Benign	1.00	0.96	0.98	
	PDAC	1.00	1.00	0.91	
SGD classifier	Control	0.84	0.75	0.79	0.63
	Benign	0.52	0.52	0.52	
	PDAC	0.48	0.52	0.52	
SVM with linear kernel (proposed method)	**Control**	**0.92**	**0.92**	**0.92**	**1.00**
	Benign	**0.86**	**1.00**	**0.93**	
	PDAC	**1.00**	**0.84**		

11.5 Conclusion

This research aims to advance early pancreatic cancer detection through an integrated approach of medical data and advanced machine learning techniques. Rigorous dataset selection and biomarker analysis form a robust machine learning pipeline, including data preprocessing, model development, and hyperparameter optimization. Support Vector Machine (SVM) classifiers, combined with voting ensembles, exhibit significant discriminative potential between cancerous and benign cases, demonstrating exceptional results in accuracy, precision, recall, F1 score, and support metrics. The presentation of confusion matrices enhances understanding, and the study extends beyond binary classification to evaluate the grade of cancerous cells, enhancing clinical relevance. This study has the potential to revolutionize early pancreatic cancer detection and patient management, underlining the promise of this integrated approach. Future avenues include the integration of advanced machine learning algorithms and broader applications in early cancer detection.

References

1. Nakagohri, T., Kinoshita, T., Konishi, M., Inoue, K., Takahashi, S.: Survival benefits of portal vein resection for pancreatic cancer. Am. J. Surg. **186**(2), 149–153 (2003)
2. Kamisawa, T., Wood, L.D., Itoi, T., Takaori, K.: Pancreatic cancer. Lancet **388**(10039), 73–85 (2016)

3. Appel, B.L., Tolat, P., Evans, D.B., Tsai, S.: Current staging systems for pancreatic cancer. Cancer J. **18**(6), 539–549 (2012)

4. Debernardi, S., Blyuss, O., Rycyk, D., Srivastava, K., Jeon, C.Y., Cai, H., et al.: Urine biomarkers enable pancreatic cancer detection up to 2 years before diagnosis. Int. J. Cancer **152**(4), 769–780 (2023)

5. Yokoyama, S., Hamada, T., Higashi, M., Matsuo, K., Maemura, K., Kurahara, H., et al.: Predicted prognosis of patients with pancreatic cancer by machine learning. Clin. Cancer Res. **26**(10), 2411–2421 (2020)

6. Liu, K.L., Wu, T., Chen, P.T., Tsai, Y.M., Roth, H., Wu, M.S., et al.: Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digit. Health **2**(6), e303–e313 (2020)

7. Zhang, L., Sanagapalli, S., Stoita, A.: Challenges in diagnosis of pancreatic cancer. World J. Gastroenterol. **24**(19), 2047 (2018)

8. Karley, D., Gupta, D., Tiwari, A.: Biomarker for cancer: a great promise for the future. World J. Oncol. **2**(4), 151 (2011)

9. Wang, L., et al.: Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. Elife **11**, e80150 (2022)

10. Hayashi, K., et al.: Prediction of recurrence pattern of pancreatic cancer post-pancreatic surgery using histology-based supervised machine learning algorithms: a single-center retrospective study. Ann. Surg. Oncol. **29**(7), 4624–4634 (2022)

11. Dalal, V., et al.: Radiomics in stratification of pancreatic cystic lesions: machine learning in action. Cancer Lett. **469**, 228–237 (2020)

12. Chen, W., et al.: Risk prediction of pancreatic cancer in patients with abnormal morphologic findings related to chronic pancreatitis: a machine learning approach. Gastro Hep Adv. **1**(6), 1014–1026 (2022)

13. Wang, Q.H., Ji, Z.G., Chen, Z.G., Li, H.Z., Fan, H., Fan, X.R., et al.: Serum CA 19-9 as a good prognostic biomarker in patients with bladder cancer. Int. J. Surg. **15**, 113–116 (2015)

14. Humphris, J.L., Chang, D.K., Johns, A.L., Scarlett, C.J., Pajic, M., Jones, M.D., et al.: The prognostic and predictive value of serum CA19. 9 in pancreatic cancer. Ann. Oncol. **23**(7), 1713–1722 (2012)

15. Lin, W.C., Tsai, C.F.: Missing value imputation: a review and analysis of the literature (2006–2017). Artif. Intell. Rev. **53**, 1487–1509 (2020)

16. Jia, B.B., Zhang, M.L.: Multi-dimensional classification via decomposed label encoding. IEEE Trans. Knowl. Data Eng. (2021)

17. Zeng, X., Huang, J., Ding, C.: Soft-ranking label encoding for robust facial age estimation. IEEE Access **8**, 134209–134218 (2020)

18. Lin, M., Zhu, X., Hua, T., Tang, X., Tu, G., Chen, X.: Detection of ionospheric scintillation based on xgboost model improved by smote-enn technique. Remote Sens. **13**(13), 2577 (2021)

19. Mathur, A., Foody, G.M.: Multiclass and binary SVM classification: Implications for training and classification users. IEEE Geosci. Remote Sens. Lett. **5**(2), 241245 (2008)

20. Mangkunegara, I.S., Purwono, P.: Analysis of DNA sequence classification using SVM model with Hyperparameter Tuning Grid Search CV. In: 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), pp. 427–432. IEEE (2022)

21. Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S.: The 'K' in Kfold cross validation. In: ESANN, pp. 441–446 (2012)

22. Fushiki, T.: Estimation of prediction error by using K-fold cross-validation. Stat. Comput. **21**, 137–146 (2011)

23. Jung, Y.: Multiple predicting K-fold cross-validation for model selection. J. Nonparametric Stat. **30**(1), 197–215 (2018)

24. Kumar, U.K., Nikhil, M.S., Sumangali, K.: Prediction of breast cancer using voting classifier technique. In: 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy, and Materials (ICSTM), pp. 108–114. IEEE (2017)

25. Sinkala, M., Mulder, N., Martin, D.: Machine learning and network analyses reveal disease subtypes of pancreatic cancer and their molecular characteristics. Sci. Rep. **10**(1), 1212 (2020)
26. Nasief, H., et al. A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer. NPJ Precis. Oncol. **3**(1), 25 (2019)
27. Sala Elarre, P., et al.: Use of machine-learning algorithms in intensified preoperative therapy of pancreatic cancer to predict individual risk of relapse. Cancers **11**(5), 606 (2019)
28. Kaissis, G., et al.: A machine learning algorithm predicts molecular subtypes in pancreatic ductal adenocarcinoma with differential response to gemcitabine-based versus FOLFIRINOX chemotherapy. PLoS One **14**(10), e0218642 (2019)

Chapter 12
Automated Student Assignment Evaluation Based on Information Retrieval and Statistical Techniques

Banoth Samya , **M. Archana, T. Venkata Ramana** ,
Kachapuram Basava Raju, and Kamalakar Ramineni

Abstract Efficient automated evaluation is urgently needed in programming courses within the field of online education. This study introduces a model that focuses on the TF-IDF algorithm, a reliable technique for information retrieval, to simplify the assessment procedure in programming courses with ten distinct categories. TF-IDF is a numerical statistical technique that measures the relevance of tokens. When combined with cosine similarity, it yields impressive outcomes in document scoring and ranking. The suggested paradigm greatly decreases the amount of time and effort teachers need to spend on manual assessment. The experimental results confirm the effectiveness of the model, demonstrating its strong ability to automatically assess programming tasks.

Keywords TF-IDF · Cosine similarity · Tokenization · Token vector · Solution feature · Token error

B. Samya · M. Archana (✉) · T. V. Ramana
Department of CSE, CVR College of Engineering, Hyderabad, India
e-mail: m.archana@cvr.ac.in

K. B. Raju
Department of AI, Anurag University, Hyderabad, India
e-mail: kbrajuai@anurag.edu.in

K. Ramineni
School of Computer Science & Artificial Intelligence, SR University, Warangal, Telangana, India

12.1   Introduction

Because of its simplicity and ease of use, electronic education is the most potential instructional approach. Students usually use a virtual school to study programming lessons. One of the main challenges with the online educational system is the e-assessment of programming projects. Generally speaking, instructors of programming training have a lot of work to review. The process of manually reviewing assignments would be laborious and ineffective [1, 4]. It is quite laborious to assess a program since one must verify that every element, including assertions, conditions of use, and their scopes, constant initialization, and so on, is accurate [2, 3]. By combining the TF-IDF and cosine analogy algorithms, we have created a self-learning evaluation method that addresses the evaluation difficulty in e-assessment. The arrangement of this piece of writing is as follows: An overview is given in Sect. 12.1, followed by the structure of the suggested system in Sect. 12.2, examination methods, solution attributes, and system performance in Sect. 12.3, program evaluation process and tokens mistakes calculation in Sect. 12.4, and efficiency research in Sect. 12.5. Lastly, we reviewed the paper's contribution, flaws, and future directions in Sect. 12.6.

12.2   Proposed Methodology

The evaluation of this model made use of the cosine analogy and token frequency-inverse record frequency techniques. We started by compiling every possible set of solutions to programming issues into a solution corpus.

Tokenizing each program is the next stage. Tokenization was then followed by the removal of unnecessary tokens. Next, by inversely offsetting the weight, we were able to determine the inverse document frequency of every token. By producing the TF-IDF score vectors of the corresponding program solutions, the solution matrix was produced. Squaring the cosine of each submission to each matrix's solution vector allowed for the determination of reliability. The two main parts of the complete system design are the supplied program assessment and the generated solution matrix.

12.2.1   Solution Matrix Generation

The resultant component of the system that was suggested is shown in Fig. 12.1. Following the passage of path protection, multi-condition service, branch coverage, and statement coverage [10], a solution program set was chosen. The programs are subsequently tested using a set of test cases. Following the completion of those tests, the following processes were used to build a solution matrix:
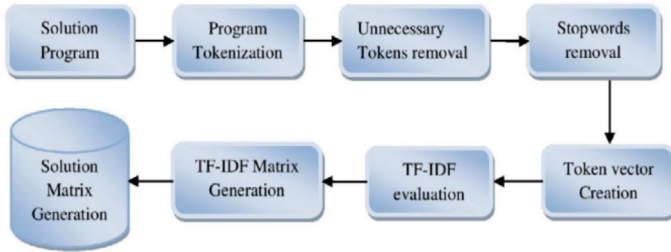
Fig. 12.1 Solution matrix generation

Preprocessing Applications: Program tokenization was used to remove stop words and superfluous tokens during preprocessing. Before that, we eliminated any superfluous comments and spaces from the program to ensure proper tokenization.

Program Tokenization: At this point, tokens were created from the declaration of variables, their setting up, claims, under certain circumstances, etc.

Stop talking and take away superfluous tokens: Writing comments, preprocessor orders, and user interface instructions is a common practice when programming. A program's comprehension and completeness are greatly enhanced by the use of library functions, even if these have little impact on programming performance evaluation. As a result, we have chosen a selection of popular phrases, library function names, and preprocessor directives that are frequently used in program writing.

IDF-TF Evaluation: The amount of times a token t appears in script p is its token frequency. The linear recurrence weight of the token t in program p is

$$W_{t,p} = 1 + tf_{t,p} \, if \, tf_{t,p} > 0 \, otherwise \, 0 \tag{12.1}$$

Inverse document frequency is the measure of a token's rarity as seen from the perspective of a fixed corpus. As it gets harder to find, the IDF weight increases. A token has zero IDF if it appears in any program in the corpus. If a token is missing from the corps, there is going to be a divide-by-zero event. We therefore inserted one with a fraction. To appropriately scale up the weight, we also used a logarithm in both terms.

$$idf_{t,p} = log_{10} \frac{|SolutionCorpus|}{1 + df_t} \tag{12.2}$$

TF-IDF matrix generation: An order m x n result matrix was produced by applying the following formula. Each answer's length (token count) is denoted by m, while the total amount of responses is indicated by n. A solution vector was thought to exist in every row of the matrix. This matrix was used to assess a submitted program's accuracy.

$$tf.idf_{t,p} = \left(1 + log_{10} tff_{t,p}\right) x log_{10} \frac{|Solution \, Corpus|}{1 + df_t} \tag{12.3}$$

12.2.2   Submission Evaluation

The evaluation section of the proposed model is shown in Fig. 12.2. The provided program was first assessed by human graders. The percentage of software accuracy was the definition of grade in this concept. A fresh entry is considered a query vector, and the cosine relationship is used to compare it to every other vector in the outcome, Utilizing the greatest similarity, and accuracy. In the sine value of the measure of similarity formula, we defined every entry as q and every solution array as s.

Figure 12.3 illustrates the whole structure. First, the trainer assessed a fresh submission. After that, the representation rated the software. The instructor retains the model's examination record for each assessment to track progress and quantify progress.

Using Algorithm 1, we examined the token ratio and in verse article frequency of the solution set in our analysis of the suggested method for solution matrix generation. We merged all vectors in a matrix and named it the solution matrix after vector creation for each solution.

> **Algorithm 1**: TF-IDF vector generation
> Input: program code
> Output: tfidf vector
> Tokenize program
> Remove stopwords and unnecessary tokens
> for token $\in$ {1,....,n} do
>   a.   Compute $tf_{t,p}$
>   b.   Compute $idf_{t,p}$
> end for
> Compute $tf_{t,p}$ x $idf_{t,p}$

We estimated the token frequency of presentation in the assessment section after tokenization and deleting stop words. Each of the tokens was given its matching IDF
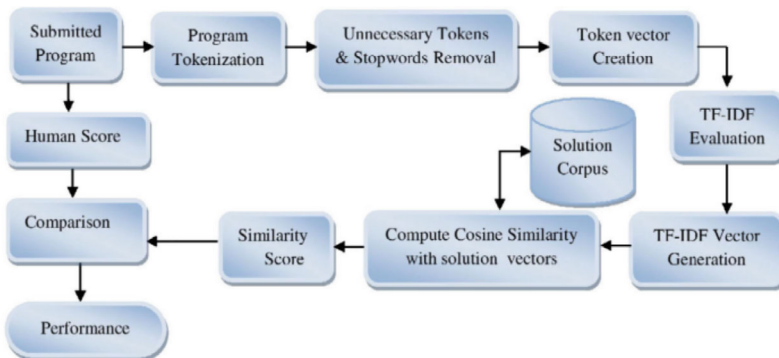
Fig. 12.2 Evaluation of submitted program
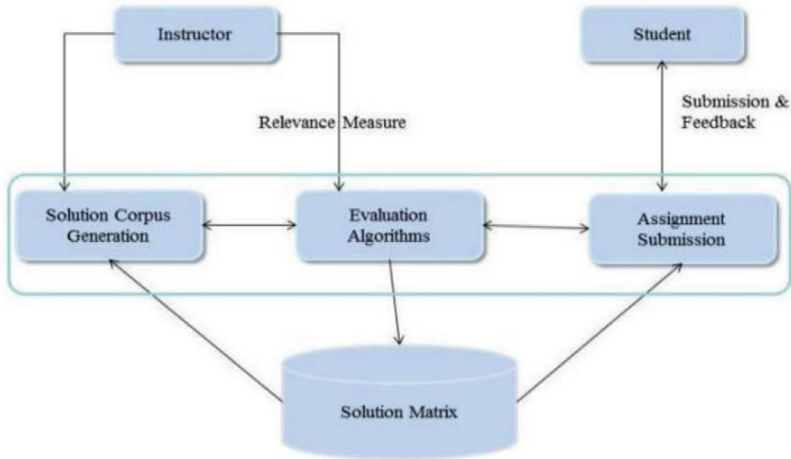
Fig. 12.3 Architecture of proposed model

value based on the sample. The search vector q is then created by using the input's TF and IDF product values. The second approach has been broken down into its assessment component.

Algorithm 2: Submission assessment
　　　　Input: submission code
　　　　Output: submission correctness
　　　　Assign correctness of submission by expert human
　　　　Generate a tfidf vector of submission
　　　　Determine the cosine correlation between an array input and the result.
　　　　Measure the maximum cosine similarity
　　　　Measure the accuracy of the model, comparing the closeness between human and
model evaluation

12.3   Evaluation Methodology

There are numerous valid solutions to a C programming challenge [5]. Using the for loop, do while loop, and while loop, any programming problem may be written. There may be many conditional statements within a loop [7–9]. Another type of method is function-based. Therefore, there could be an approach utilizing any one of the previously mentioned techniques for one computer problem [6].

12.3.1 Solution Feature

Every potential solution should be used to generate a model to accurately measure programming assignments. We considered every potential solution aspect when developing our model. We describe the entire model in a hierarchical graph structure, as illustrated in Fig. 12.4, accounting for every possibility. To confirm loops and conditional correctness, a submission must satisfy any feasible graph route. During the review process, we viewed a contribution as a hierarchical graph path. We produced a query vector from the given program after tokenizing and filtering out extraneous tokens. A proper submission token was completely matched with a graph route. Correct writing conditions and loops mean good program writing performance. So, if we simply look at conditions and loop tokens during assessment it's not a horrible idea. To show, we determined that Case 1's submission was error-free. We altered the input vector once more to assess the validity of the loop and dependent tokens. The filtered vector was completely matched with the third top route of the hierarchical network, ensuring that the loop and conditional statement were correctly written in the submission.

12.3.2 Performance Analysis

During the evaluation process, we evaluated four different scenarios to assess performance. We have highlighted an especially common mindset of beginners while creating programs in these situations.

Case 1: A submitted program may contain some incorrect tokens. Students frequently write incorrect tokens due to a lack of comprehension or memorization of programs. A syntactic error is made by the trainee in this instance. Such an error suggests a lower level of care for example tolerance. Case 1's submittal has six incorrect tokens. As a result, the formula gives these incorrect tokens a zero as follows: [1.3, 0.5, 3.3, 1.0, 0.1, 0, 2.3, 0, 0,0, 0.7, 0.5, 0, 0.6].

Currently, a cosine relationship indicates an inadequate fit between the program that was sent in and the right answers. In this instance, the model evaluated the submission with success.
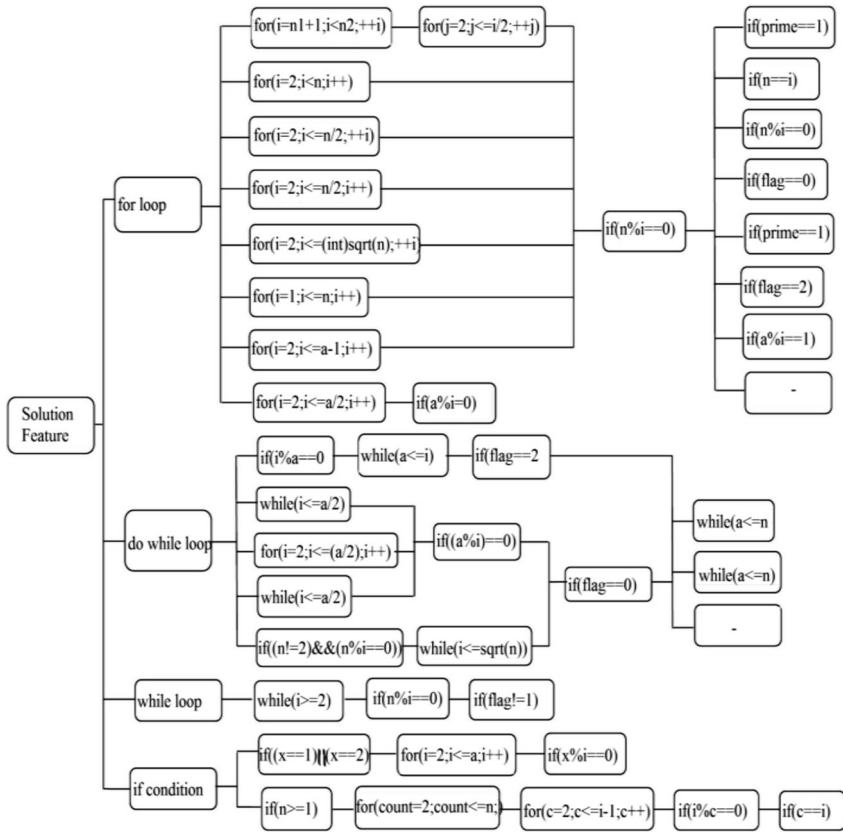
Fig. 12.4 Hierarchical graph of a solution model

Case 1:
```
#include<stdio.h>
int main
{
Int n, i, flag=0;
print("Enter a positive integer");
scanf("%d", &n);
for (i=2;i<=n/2;++i){
if(n%i==0){
flag=1;
break;}}
if(flag==0)
printf("%d is a prime number",n);
else
printf("%d is not a prime number",n);
return(0);}
```
Case 2:
```
#include<stdio.h>
int main() {
int year;
printf("Enter a year:");
scanf("%d",&year);
if(year%4==0){
if(year%100==0){
if(year%400==0)
printf("%d is a leap year.",year);
```

Case 2: A program submitted may not be related to the proper solution. Improper issue comprehension, a proclivity to avoid the instructor, and random guessing all imply case 2 category errors. We discovered four irrelevant tokens in this contribution about our solution corpus. As a result, the IDF weight is set to 0 for these tokens, as in the previous example. As a result, cosine similarity yields a very poor match. In this scenario, the model was also able to correctly execute evaluation.

12.4   Results and Discussions

For the automatic evaluation of numbers with primes, the use of binary searches, and bubble sorting programs, we developed three distinct examples, with each offering thirty solutions. Every model answer was contrasted with a new suggestion. Every novel entry has been evaluated thirty times since each variant was constructed with thirty solutions. The sine connection rate between every possible program and a supplied program is shown in Fig. 12.5. The highest possible similarity was used to assess the correctness of the submissions as well as the model performance.

To assess the correlation between the number of error statements and cosine similarity, we examined 57 uploaded algorithms. The relationship between error tokens and similarity is wholly unfair. Tokens are uncommon, so choosing the incorrect
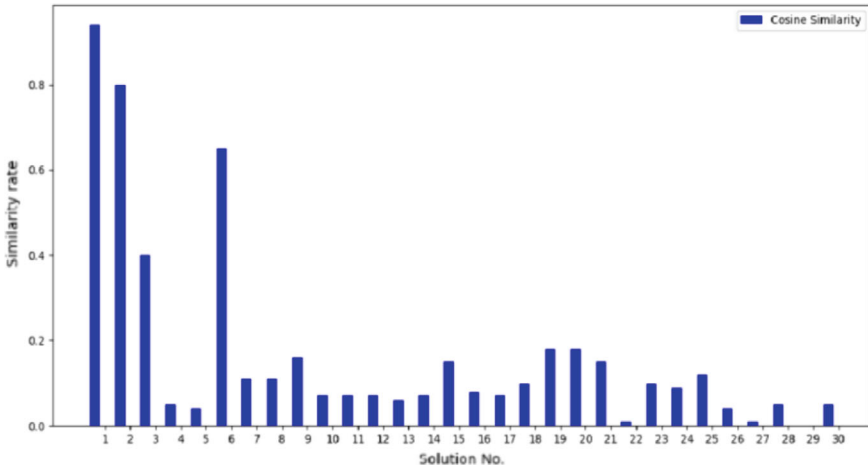
Fig. 12.5 Cosine analogy between the resultant set and the one provided

one could lead to variations in closeness. Nevertheless, we succeeded in roughly establishing a link between them. For estimation, we employed the linear regression model. Figure 12.6 illustrates the relationship between error rate and cosine agreement. In the cosine similarity range of 0.99 to 0.95, Fig. 12.7 shows a minor error. This is frequently caused by a startup error or a single incorrect token. Submission has at most two incorrect tokens in the 0.95 to 0.85 range. We had difficulty characterizing the range of 0.84 to 0.65. Figure 12.6 depicts how the number of error tokens varies within this range. The readily accessible error symbols fell between one and five under this range. Submissions that have an identity range of less than 0.5 are typically deemed inaccurate or irrelevant.

Fig. 12.6 Error symbols' impact on the cosine relationship experiment
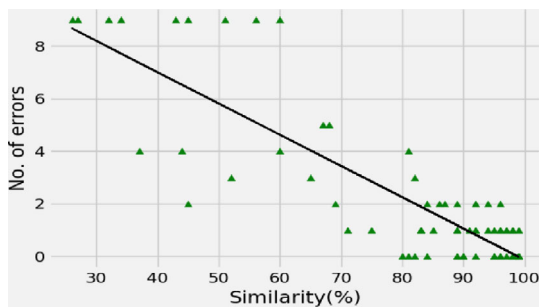
Fig. 12.7 Relationship between error token and similarity range (approximate)
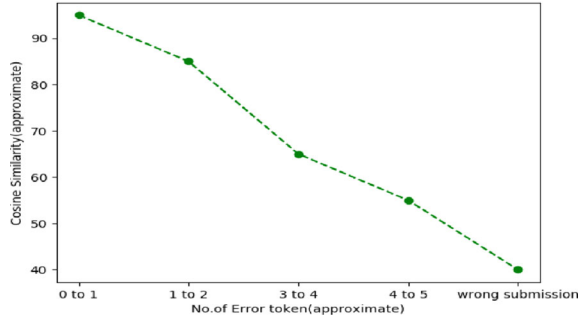
Fig. 12.8 Model performance
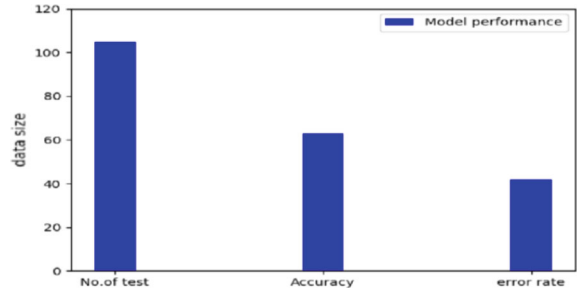
12.5 Model Performance

The TF-IDF approach performs well in data retrieval and content ranking tasks overall. This technique's core principle is that ideas are randomly assessed against the document corpus, and a word's weight grows based on its rarity. There is no term repetition during the search. Therefore, the obtained document seems to be less significant in the second or third rank. Except for random and meaningless tokens, TF-IDF performs exceptionally well on the automatic review of programs. We employed 105 programs that were submitted to assess model performance. Except for incorrect token recognition and incorrect order, TF-IDF performs exceptionally well. To analyze the performance of the TF-IDF-based model, we employed 105 submitted programs as a test set. 63% of the programs were scored correctly by professional graders, whereas 33% were misgraded (as shown in Fig. 12.8).

12.6 Conclusion and Future Work

In this study, we proposed a model based on TF-IDF and the cosine similarity technique for the automatic evaluation of programming assignments. When there are no special conditions, the model functions well. This model's performance can be improved by incorporating some extensions. We're already working on improving the

model's performance. To improve efficiency, we are making changes to the control flow graph isomorphic relationships and token distance modification algorithms. Tests demonstrate the proposed approach's ability to get around the TF-IDF-based model's consequence occurrences. We are now evaluating the dangers of the algorithm against additional likely scenarios. We think that our model will be able to perform more like an actual expert reviewer replica if we change our hypothesis.

References

1. Zougari, S., Tanana, M., Lyhyaoui, A.: Towards an automatic assessment system in introductory and programming courses. In: International Conference on Electrical Information Technologies, pp.496–499. IEEE (2016)
2. Zougari, S., Tanana, M., Lyhyaoui, A.: Hybrid assessment method for programming assignments. In: Fourth IEEE International Colloquium on Information Science and Technology. IEEE (2016)
3. Pape, S., Flake, J., Beckmann, A.: STAGE A software tool for automatic grading of testing exercises. In: thirty eighth IEEE International Conference on Software Engineering Companion, pp. 491–500. IEEE (2016)
4. Srikant, S., Aggarwal, V.: A system to grade computer programming skills using machine learning. In: Twentieth International Conference on Knowledge Discovery and Data Mining, pp. 1887–1896. ACM (2014)
5. Sharma, K., Banerjee, K., Mandal.: A scheme for automated evaluation of programming assignments using FSMD based equivalence checking. In: Sixth IBM Collaborative Academia Research Exchange Conference on I-CARE, pp. 1–4. ACM (2014)
6. Dadic, T., Glavinic, V., Rosic, M.: Automatic evaluation of student programs. In: Conference on Innovation Technology in Computer Science Education, pp. 328–328. ACM (2014)
7. Pieterse, V.: Automated assessment of programming assignments. In: Third Computer Science Education Research Conference, pp. 45–56. ACM (2013)
8. Kanmani, S., Radhakrishnan, P.: A simple Journal methodology to grade c program automatically. In: International of Advances in Embedded System Research, pp.73–90. IJAESR (2011)
9. Huang, Chen, Luo, Chen, X., Chuang.: Developing an intelligent diagnosis and Assessment tool for introductory programming. In: International Forum of Educational Technology Society, pp. 139–157 (2008)
10. Lakumarapu, S., Agarwal, R.: Time-based connotation rule mining based on T-apriori algorithm using Weka tool slants. In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 261–264. https://doi.org/10.1109/CTEMS.2018.8769122

Chapter 13
Research on the Purchase Intention Analysis Algorithm of Virtual Goods Based on Neural Network and SEM Model

Gaobiao Cao [ORCID]

Abstract With the development of the times, virtual product shopping has become more and more integrated into our daily lives, affecting consumers' choices and decisions. This not only provides merchants with a broader space for development, but also provides consumers with more convenience and diversity. Therefore, we conduct in-depth research on virtual product purchase intentions to better understand consumers' psychological construction and shopping decision-making process in the digital age. Combined with the method of structural equation modeling (SEM), special attention is paid to the motivations and behavioral factors of consumers purchasing virtual goods in the digital environment. Through advanced neural network technology, we constructed an SEM model to carefully analyze the relationship between each potential variable. In-depth analysis of real-time shopping data reveals the critical paths and interactions in the formation of virtual product purchase intentions. In this research framework, we not only provide a new perspective on the virtual goods market and provide a basis for formulating more effective marketing strategies, but also provide a more comprehensive understanding of the nature of consumer shopping behavior in the digital era.

Keywords Neural network · SEM model · BP algorithm

13.1 Introduction

In today's digital and networked era, virtual goods have become an indispensable part of consumers' lives. Research on the purchase intention of virtual goods is of great significance to both e-commerce companies and academia [1].

G. Cao (✉)
Beijing Normal University-Hong Kong Baptist University, United International College, Zhuhai 519087, Guangdong, P R China
e-mail: caogaobiao@163.com

First of all, for e-commerce companies, understanding consumers' purchasing intentions is the key to formulating marketing strategies and improving sales performance. The characteristics of virtual goods and consumer psychological factors have a direct impact on purchase intentions. Through in-depth research on consumer demands and preferences for virtual goods, companies can better grasp market trends, accurately locate target consumer groups, develop virtual goods that are more in line with market demand, and formulate corresponding pricing strategies, promotion strategies, and service strategies, thereby improving product competitiveness and market share [2]. In addition, an in-depth understanding of the market environment can also help companies formulate more targeted marketing strategies, improve market response speed, and enhance market competitiveness.

Secondly, the significance of research on virtual product purchase intention to the academic community cannot be ignored. Through the study of virtual product purchase intentions, we can enrich and improve consumer behavior decision-making models, conduct in-depth analysis of consumer behavior, and explore the regularity of consumer psychology and behavior. This will help expand consumer behavior theory, improve the depth and breadth of academic research, and provide more theoretical support and empirical evidence for related subject areas. At the same time, research on virtual product purchase intentions can also help reveal consumer trends and characteristics in the digital era, and play a positive role in promoting the development of e-commerce, digital marketing, and other related fields.

In short, research on virtual product purchase intention is of great significance to both e-commerce companies and academia. Through in-depth research on the impact of factors such as the characteristics of virtual goods, consumer psychology, and market environment on purchase intentions, we can provide guidance for companies to formulate more effective marketing strategies and improve sales performance; it also provides new evidence for the academic community. Evidence, enrich and improve relevant theories, and promotes the development of academic research. Therefore, strengthening the research on virtual product purchase intention has important practical significance and academic value for promoting consumer behavior research and e-commerce development in the digital era.

SEM and neural network models [3] are commonly used data analysis methods that can be used to explore the influencing factors of virtual product purchase intention. The SEM model can deeply explore the influence and relationship of various factors on the purchase intention of virtual goods through methods such as path analysis [4]; while the neural network model can automatically learn and identify the patterns in the data by simulating the working mode of neurons in the human brain. Patterns and laws to more accurately predict virtual product purchase intentions.

Therefore, it is of great significance to use SEM and neural network models to conduct research on virtual product intentions. First of all, this method can more comprehensively explore the factors that affect the purchase intention of virtual goods, including the characteristics of the virtual goods themselves, consumer psychology, market environment, and other factors, and help reveal the mechanism and rules of the formation of purchase intention of virtual goods. Secondly, using

SEM and neural network models for research can improve the accuracy and credibility of the research, avoid the subjectivity and bias that may exist in traditional research methods, and provide guarantee for the scientificity and reliability of the research results. Finally, using SEM and neural network models to conduct virtual product intention research can provide e-commerce companies with more accurate market analysis and marketing decision support, help companies better grasp market opportunities and challenges, and improve product competitiveness and market share.

In summary, using SEM and neural network models to conduct research on virtual product intention has important practical significance and academic value. It will help to deeply explore the formation mechanism and rules of virtual product purchase intention, and improve the accuracy and credibility of the research. Provide more accurate market analysis and marketing decision support for e-commerce companies.

13.2   Related Work

In order to do a good job in user intention work, many scholars have conducted related research in the field of prediction and recommendation through word embedding and graph neural networks.

A novel latent representation model, GLR, is introduced, taking user preferences into account and addressing the continuous transition effects between interest points and regions. This model integrates categories and check-in frequencies of interest points and regions into the latent vector representation, capturing the continuous transformation between them to learn user preference vectors at different times. Tan et al. [5] employed social and geographical information, employing a matrix filling model with matrix decomposition, embedding category information, and kernel density estimation to model geographical influence. Manotumruksa et al. [6] presented a Contextual Recurrent Collaborative Filtering (CRCF) framework for short-term user preferences, considering the relevance of recently visited points. Tang and Wang proposed a CNN-based method for modelling user location and contextual preferences, using word embedding vector space to capture users' switching modelling. Wang Lu [7] word2Vec and RNN to capture interest point conversion modelling, introducing a normalization layer and a time distance gate to enhance recommendation performance. Previous models, such as Markov chain, RNN, and word2vec, overlooked variations in check-in sequences on different dates, such as the distinction between "Work" and "Fun" on weekdays and weekends. Zhao et al. [8] developed a method considering temporal features, introducing a temporal interest point embedding model to capture sequence context and temporal characteristics. Additionally, a geographical hierarchical pairwise ranking model was proposed to enhance recommendation performance by considering geographical influence. The Geo-Triser model serves as a unified framework for point-of-interest recommendations. Some researchers explored user travel purposes, with Chen et al. [9] presenting the Trip2Vec method based on deep embedding. This method extracts time context, considers start and travel times, and incorporates information about nearby points

of interest and activity popularity, inferring the evolution pattern of travel purposes based on occurrences at different times.

The surge in graph-structured data, particularly in social network graphs, has led to the development of various graph neural network (GNN) models. Notable variants like the graph convolutional network (GCN) and graph recurrent network (GRN) have found applications in diverse tasks, including prediction and recommendation. Scarselli et al. [10] introduced a GNN model featuring an information propagation mechanism, where each node updates its state through iterative information exchange until reaching stability. This approach generates embedding vector representations for users and items. While many GNN methods for point-of-interest recommendation leverage user or point characteristics, such as age, gender, or interests, Wang et al. [11] innovatively represented user-interest interactions as a high-order connection graph, capturing complex nonlinear interactions and proving the effectiveness of incorporating graph structures.

Previous graph feature learning models focused on tasks with single outputs, like graph-level classification. However, several graph input problems require sequences of outputs. Yu et al. [12] proposed NGPR for point-of-interest recommendation, constructing a heterogeneous social network graph and using Node2Vec to generate latent vectors for interests and users. Wu et al. [13] introduced SR-GNN, modelling user conversation sequences as graphs and utilizing an attention mechanism to capture historical preferences' impact on current preferences. Song et al. [14] presented a graph convolutional network model for social recommendation, leveraging an attention mechanism to aggregate user representations with friends' representations. Existing point-of-interest recommendation algorithms often fall short of fully integrating various influencing factors and providing dynamic recommendations over time. To address this, Lang et al. [15] proposed a multi-bipartite graph network model (MBR) for point-of-interest recommendations. This model constructs a large heterogeneous network, comprising six weighted bipartite graphs based on user clustering, to depict the relationships among users' social connections, geographical locations of points of interest, and time information. This innovative approach aims to offer more accurate and dynamic recommendations to users.

13.3 Model

13.3.1 Artificial Neural Networks

Artificial neural network (ANN) simulates the propagation process of human brain neurons through basic linear weighting calculation and activation function (Activation Function). In the artificial neural network model, one neuron constitutes a network node, and a large number of neurons form a neural network through weighted connections. There is a weight between every two neurons in the network, which means that all signals passing through this path in the network need to be weighted
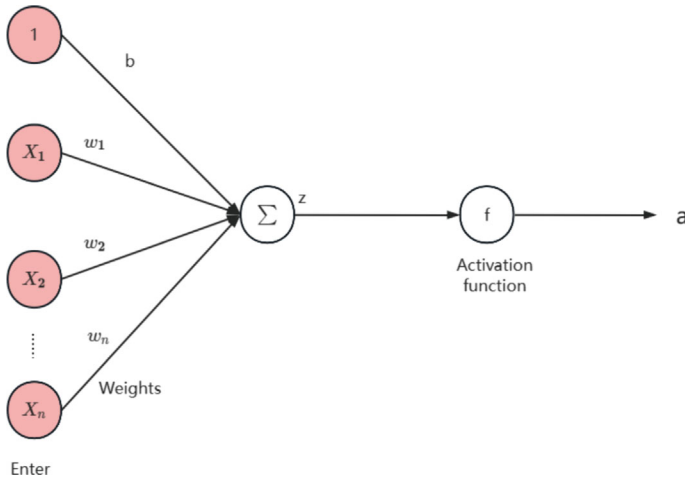
Fig. 13.1 Typical neuron model

before they can continue to be transmitted forward. There is also an activation function in the network, which can map the input of the network through nonlinear mapping. The output, activation function is the key to the nonlinear properties of, artificial neural networks. The connection methods, weights, and activation functions between neurons jointly determine the output of the neural network. Fig. 13.1 shows a typical neuron model with n input data.

A single neuron structure has n input data $\times$ 1, $\times$ 2, $\times$ 3$\cdots$ xn, and this set of data can be represented by a vector group [$\times$1, $\times$ 2, $\times$ 3$\cdots$ xn]. First, the neuron will perform a weighted summation of the n data to obtain the net input $z \in R$. The function expression is shown in Eq. 13.1:

$$Z = \sum_{i=1}^{n} w_i x_i + b = w^T + b \tag{13.1}$$

Among them, $w_T = [x_1, x_2, x_3 \cdots x_n] \in R$ is the n-dimensional weight input vector of the neuron, and $b \in R$ is the threshold.

Next, the net input z is processed by the nonlinear function $f(\cdot)$ to obtain the activation value, which is the output value of the neuron. The function expression is shown in the following Eq. 13.2:

$$a = f(z) \tag{13.2}$$

The nonlinear function $f(\cdot)$ is the activation function, which can distinguish neurons into activation states and inhibition states, giving neurons nonlinear output capabilities, and is the key to the entire neuron model.

13.3.2   SEM Model

Structural equation modeling (SEM) is a powerful statistical analysis method widely used in fields such as social sciences, psychology, and economics to study complex relationships between multiple variables. Through tools such as path diagrams, factor analysis, and covariance matrices, the model enables researchers to consider both observed and unobserved variables to gain a more complete understanding of the data structure.

In SEM, basic mathematical expressions include measurement models and structural models. A measurement model describes the relationship between observed variables and latent variables:

$$X = \lambda \xi + \varepsilon \tag{13.3}$$

Among them, X is the observed variable, ξ is the latent variable, λ is the loading factor, indicating the strength of the relationship between the observed variable and the latent variable, and ε is the measurement error of the observed variable.

Structural models reveal the direct influence relationships between latent variables:

$$\xi = B\xi + \xi \tag{13.4}$$

where B is the path coefficient matrix, representing the direct effects between latent variables, while ξ is the error term of the structural model, which takes into account the variation not explained by the model.

The entire model forms a comprehensive theoretical framework by combining the measurement model and the structural model. The covariance matrix provides information on the correlation between variables, while the fit index is used to evaluate the degree of fit between the model and the observed data. Its specific structural model diagram is shown in Fig. 13.2.

An important application of SEM is factor analysis, which achieves simplification and understanding of the data structure by determining the relationship between observed variables and latent variables. The flexibility and comprehensiveness of SEM make it an important tool for researchers to deeply understand the relationships between variables, verify theories, and conduct model inferences. Therefore, SEM not only provides the ability to model complex data patterns, but also provides scientific researchers with rich tools to gain a deeper understanding of the potential relationships behind the data.
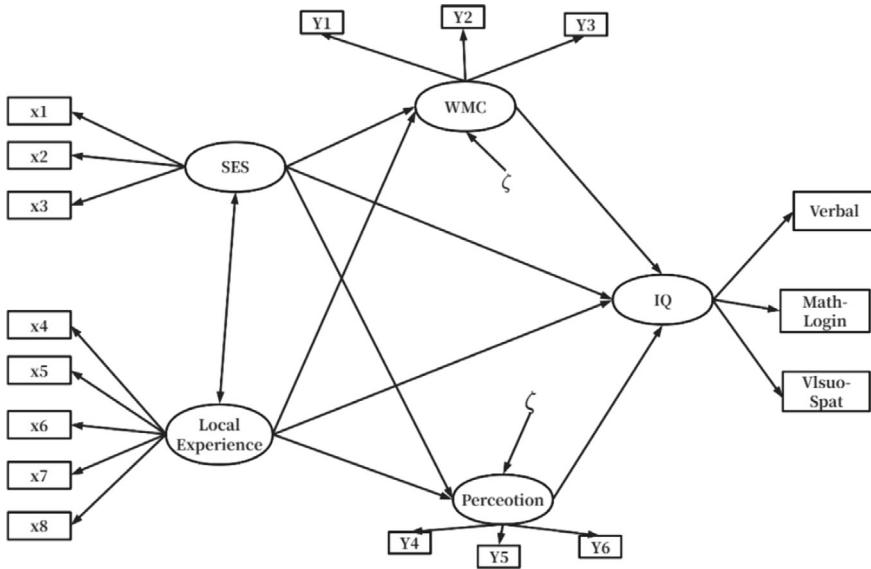
Fig. 13.2 SEM structural model diagram

13.3.3 SANN Model

The idea of constructing the SANN model in this article is as follows: First, use SEM to determine 8 exogenous latent variables (perceived cost, value uncertainty, information loss, self-efficacy, social influence, information needs, emotional sustenance, perceived benefits) and 2 endogenous latent variables. Linear causal relationship between latent variables (virtual product purchase intention, virtual product purchase behavior); secondly, convert the causal path of SEM into the topological structure of the ANN model to obtain the SANN model, and implement model hierarchical division and initialization of weights and thresholds; Finally, an algorithm is selected to train the SANN model and calculate the relationship between exogenous latent variables and endogenous latent variables in the factor model that affects virtual product purchasing behavior after adding nonlinear effects.

Compared with the traditional ANN model, the SANN model has two differences: The first point is that the connections between neurons in the SANN model are partial connections instead of full connections. The connection method is significantly different from the causal path analysis in the SEM model. The path of the coherence level is the same; the second point is that the initial connection weights between adjacent levels of neurons in the SANN model have definite values rather than being randomly generated during the training process. This article is based on the standardized factor loading values calculated in the SEM model and Path coefficients determine the initial connection weights between neurons.

SANN input layer neurons are evolved from SEM exogenous measurement variables, represented by $x_i (i = 1,2,3,\cdots,I)$. The neuron structure of the hidden layer of SANN evolves from the exogenous latent variables and endogenous latent variables of SEM. The neurons of the first hidden layer are composed of exogenous latent variables of SEM, expressed by $\xi_n (n = 1,2,3,\cdots,N)$; the second layer of hidden layer neurons is composed of endogenous latent variables of SEM, represented by η_r (r = $1,2,3,\cdots,R$).

SANN output layer neurons are determined by SEM endogenous measurement variables, represented by y_j ($j = 1,2,3,\cdots,J$), which are the external reflection of virtual product purchase intention and virtual product purchase behavior. The causal relationship between the variables formed by the measurement model and the structural model in the SEM model is the basis for determining the connection mode between the neurons in each layer of the SANN. Therefore, the factor loadings of the measurement model and the path coefficients of the structural model in the SEM model become the corresponding SANN neural networks. The connection weight between elements. The weight vector group L between two adjacent layers of SANN is composed of the connection weight between each neuron. The specific rules are: the initial weight between the input layer neurons and the first layer hidden layer neurons are corresponding to the SEM model. The factor loadings of the weight vector group L^H between the second hidden layer and the second layer is an R $\times$ N-dimensional vector; the weight vector group L^O between the second hidden layer and the output layer is a J $\times$ R dimensional vector.

According to the above-mentioned SANN model construction idea, the topology structure of the SANN model of virtual commodity purchase intention is determined, as shown in Fig. 13.3.

13.4   Experiment

13.4.1   Data Set

In supervised learning algorithms in machine learning, in order to prevent the model from overfitting or underfitting, the original sample data is generally divided into a training set (TrainSet) and a test set (TestSet), with the former used for model training and the latter used for model effect testing. Commonly used division ratios include 70%:30%, 80%:20%, 75%:25%, etc. In this paper, Taobao data set of Ali Tianchi Competition is used for the experiment, and the data set connection is https://tianchi.aliyun.com/competition/activeList. The original sample data volume of this article is 507, and 355 training set samples and 152 test set samples are randomly distributed according to the ratio of 70% and 30%. The training set is used to fit the model and calculate model parameters such as connection weight values and thresholds; the test set is used to evaluate the performance of the model trained through the training set data.
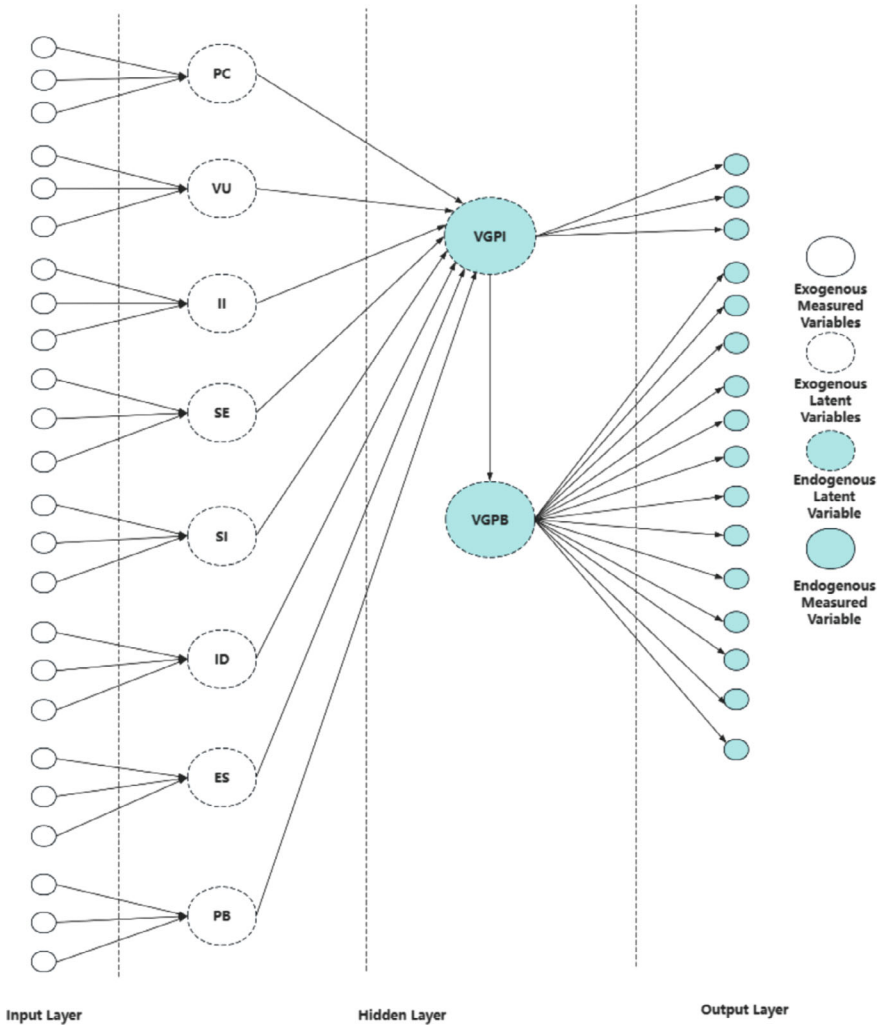
Fig. 13.3   SANN model structure diagram

13.4.2   Construction of Neural Network Model

The structure of the SANN model of virtual product purchase intention is a multi-level feed-forward network. The multi-level feed-forward network is usually trained using the Back Propagation (BP) algorithm. The essence of the BP algorithm is to minimize the error based on the gradient descent (Gradient Descent) strategy and the chain rule. By conducting error analysis on the expected results obtained from model training and the actual output results of the sample data, the network weights and thresholds are gradually modified, and finally minimize the error between the

training output results and the expected results. The BP algorithm can not only be applied to feed-forward networks, but also can be used to train feedback networks, and is widely used in ANN training.

The learning and training process of the BP algorithm can generally be divided into parameter setting, feed-forward calculation stage, and reverse learning stage: (1) Parameter setting: Before the BP algorithm starts training the model, the number of hidden layers, the number and distribution of neurons of the model are preset., training times, activation function, training target, learning rate, etc. The settings of these parameters will determine the direction of the final weights and thresholds of the neural network. (2) Feed-forward calculation stage: The feed-forward calculation process is also called the excitation propagation process. The input data reaches the hidden layer through the input layer, and then repeated linear weighted summation and nonlinear transformation processing of the activation function are performed layer by layer until Finally, the final result of the model is output in the output layer. (3) Reverse learning stage: The whole process of reverse learning is based on the Delta learning rule and is reversely propagated through the output error, that is, the error is first propagated forward from the output layer to the hidden layer, and the connection weights and thresholds of the hidden layer and other related parameters are Adjust, then continue the reverse transfer, repeat the adjustment process of the layer parameters until the error is transferred to the input layer, end this round of adjustment, and the final result is to minimize the model error.

13.4.3   Result Analysis

This study uses 507 sample data, divided into training sets and test sets according to the ratio of 70% and 30%, to quantitatively evaluate the effectiveness of the SANN model of factors influencing personal digital hoarding behavior, that is, to test the digital hoarding willingness and digital hoarding output of the model. The predicted output of the measured variables of behavior $Y_P = [y_{p1}, y_{p2}, y_{p3}, \cdots, y_{pJ}]$(p $= 1,2,3,\cdots$,P) is compared with the expected output $D_p = [d_{p1}, d_{p2}, d_{p3}, \cdots, d_{pJ}]$(p $= 1,2,3,\cdots$,P) accuracy. The root mean square error (RMSE) is very sensitive to subtle changes in values, so it can be used to evaluate the changes in the steady-state error between the input variable and the output variable. The specific calculation formula is shown in Eq. 13.5. The value range of RMSE is $(0, +\infty)$. The closer it is to 0, the higher the model fitting effect and prediction accuracy, and vice versa. The RMSE calculation results of the output neurons of the SANN model in this study are shown in Table 13.1.

$$RMES = \sqrt{\frac{1}{P} \sum_{p=1}^{p} \left(y_{pj} - d_{pj}\right)^2} \tag{13.5}$$

Table 13.1 Experimental results of the model under different damage ratios

Measurement standards		RMSE
Virtual goods purchase intention	Y1	0.432
	Y2	0.454
	Y3	0.421
	Y4	0.370
	Y5	0.440
	Y6	0.433
	Y7	0.396
	Y8	0.413
	Y9	0.469
Virtual goods purchasing behavior	Y10	0.415
	Y11	0.410
	Y12	0.372
	Y13	0.461
	Y14	0.443
	Y15	0.432
	Y16	0.428

According to Table 13.1, it can be seen that the RMSE of the endogenous measurement variables in the SANN model of the influencing factors of virtual commodity purchasing behavior in this article is less than 0.5. There have been many studies where the RMSE is less than 0.5, which means that the model has good convergence and the model prediction accuracy. is higher, so it can be considered that the EM-ANN model in this article has good performance and has high prediction accuracy.

The coefficient of determination R^2 is an evaluation index of the goodness of fit, which represents the extent to which the endogenous latent variables in the model can explain the endogenous measured variables. The calculation formulas are shown in Eqs. 13.6 and 13.7. The evaluation standard for R^2 is that the closer the value is to 1, the better the fitting effect of the model is. This article evaluates the overall fitting effect of the SANN model by comparing the R^2 values of the measurement variables ($y_1 \sim y_{16}$) corresponding to virtual product purchase intention and virtual product purchase behavior in the SANN model and SEM model. The comparison of R^2 values is shown in Table 13.2.

$$R^2 = 1 - \frac{\sum_{p=1}^{p}(d_{pj} - y_{pj})^2}{\sum_{p=1}^{p}(d_{pj} - \overline{d_j})^2} \tag{13.6}$$

$\overline{d_j}$ is the sample mean of the endogenous measurement variable d_j, and the calculation formula of $\overline{d_j}$ is shown in Eq. 13.7.

Table 13.2 Experimental results of the model under different damage ratios

A model of factors influencing virtual product purchasing behavior		SEM	LSTM	SANN
Virtual goods purchase intention	Y1	0.667	0.701	0.731
	Y2	0.661	0.659	0.681
	Y3	0.646	0.652	0.687
	Mean	0.658	0.668	0.700
Virtual goods purchasing behavior	Y4	0.670	0.712	0.731
	Y5	0.685	0.695	0.749
	Y6	0.727	0.713	0.726
	Y7	0.636	0.726	0.728
	Y8	0.667	0.714	0.723
	Y9	0.721	0.733	0.745
	Y10	0.659	0.726	0.732
	Y11	0.687	0.711	0.755
	Y12	0.730	0.796	0.819
	Y13	0.715	0.735	0.784
	Y14	0.718	0.733	0.747
	Y15	0.684	0.692	0.739
	Y16	0.684	0.694	0.751
	Mean	0.691	0.712	0.748

$$\overline{d_j} = \frac{1}{P} \sum_{P=1}^{P} d_{pj} \qquad (13.7)$$

From the comparison of the R^2 values of the factors influencing factors of virtual product purchase behavior in the SEM model, LSTM model, and SANN model in Table 13.2, it can be seen that: According to the experimental results in the above table, we can compare the models SEM, LSTM and SANN that affect virtual product purchasing behavior. In terms of virtual goods purchase intention prediction: Under the SEM model, the average R2 value of variables Y1 to Y3 is 0.658. Under the LSTM model, the average R2 value of the corresponding variable is increased to 0.668. The SANN model further improves the prediction accuracy, and the corresponding average R2 value reaches 0.700. In terms of virtual goods purchase behavior prediction: Under the SEM model, the average R2 value of variables Y4 to Y16 is 0.691, showing a certain fitting effect. The LSTM model has improved on this, with the average R2 rising to 0.712. SANN model has the best performance, with an average R2 value of 0.748, which is significantly better than the other two models. In general, the SANN model shows relatively better performance in the prediction ability of virtual goods purchase intention and purchase behavior, especially in the prediction of purchase behavior, which is significantly improved compared with

Fig. 13.4 The change of different model loss with training
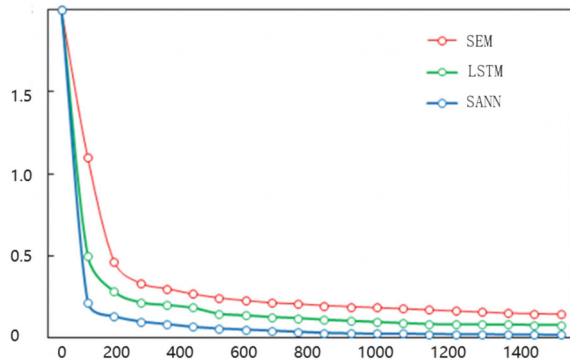
SEM and LSTM models. First, the average R2 of the SANN output virtual product purchase intention measurement variable ($y_1 \sim y_3$) is 0.658, and the virtual product purchase behavior measurement variable (The average R^2 value of $y_4 \sim y_{16}$) is 0.748, all greater than 0.5, indicating that the goodness of fit of the model is good; secondly, in the SANN model, except for the R2 value of y_6 which is slightly lower than the SEM model, the R^2 values of other model output values are all the same. Higher than the SEM model, which shows that the introduction of the nonlinear mapping capability of the SANN model not only gives the model of influencing factors of virtual commodity purchasing behavior better prediction accuracy, but also improves the overall fitting effect of the model compared with the SEM model. This shows that the SANN model can effectively improve the prediction accuracy and the overall fit of the model of influencing factors of virtual goods purchase behavior by introducing nonlinear mapping capability.

As shown in Fig. 13.4, we look at the change of the loss of different models with epochs, we can find that SANN has a better rate of decline, SANN begins to converge at 100 epochs, LSTM begins to converge at 200 epochs, and SEM begins to converge at 300 epochs. At the same time, the decline rate of SANN is significantly higher than that of LSTM and SEM, so it can be seen that SANN has better performance.

13.5 Summary

This article deeply explores the importance of virtual product purchase intention research to e-commerce companies and academia in the digital and networked era. First of all, from the perspective of e-commerce companies, understanding consumers' purchasing intentions is crucial to formulating precise marketing strategies and improving sales performance. By studying the characteristics of virtual goods, consumer psychology, and market environment, companies can better grasp market trends, locate target consumer groups, and formulate corresponding strategies to enhance competitiveness. Secondly, for academics, research on virtual product

purchase intentions has enriched and improved consumer behavior decision-making models and provided theoretical support for in-depth analysis of consumer behavior. The article emphasizes SEM and neural network models as commonly used data analysis methods, which can more comprehensively and accurately explore the influencing factors of virtual product purchase intention. The SEM model deeply explores the relationship between factors through path analysis and other methods, while the neural network model automatically learns and identifies patterns and rules in the data by simulating the working mode of neurons in the human brain, thereby more accurately predicting purchase intentions. Using these two models for research not only improves the accuracy and credibility of the research, but also provides e-commerce companies with more accurate market analysis and decision-making support, helping them better respond to market opportunities and challenges. Overall, this article highlights the dual value of research on virtual goods purchase intentions in both practical and academic fields. Through in-depth analysis of the characteristics of virtual goods, consumer psychology, and market environment, as well as the use of advanced data analysis methods, we can provide scientific basis for e-commerce companies and improve their market competitiveness, while also expanding knowledge on consumer behavior for the academic community theoretical basis.

References

1. Guenther, P., et al.: Improving PLS-SEM use for business marketing research. Ind. Mark. Manage. **111**, 127–142 (2023)
2. van Erp, Sara.: Bayesian regularized SEM: Current capabilities and constraints. Psych **5**(3), 814–835(2023).
3. Adler, Susanne Jana, Pratyush Nidhi Sharma, Lăcrămioara Radomir.: Toward open science in PLS-SEM: Assessing the state of the art and future perspectives. J. Bus. Res. **169**, 114291(2023).
4. Iroha, Nkem B., et al.: Linagliptin drug molecule as corrosion inhibitor for mild steel in 1 M HCl solution: Electrochemical, SEM/XPS, DFT and MC/MD simulation approach. Colloids Surf.S A: Phys.Chemical Eng. Asp. 660, 130885(2023).
5. de Tan, Mark, et al. (2023).: SEM-based out-of-sample predictions. Struct. Equ. Model.: Multidiscip. J. **30**(1), 132–148.
6. Manotumruksa ,Kock, Ned.: Assessing multiple reciprocal relationships in PLS-SEM. Data Anal. Perspect. J. **4**(3), 1–8 (2023).
7. Wang, Lu.: The influence of organizational culture on teacher innovation capability and tacit knowledge: A CB-SEM AMOS Analysis. Available at SSRN 4379069 (2023).
8. Zhao, Mirjana, Amir Topalović, and Lejla Turulja. (2023).: Data mining usage in Italian SMEs: an integrated SEM-ANN approach. Cent.L Eur. J. Oper. Res. **31**(3), 941–973 (2023).
9. Chen, Meg L., et al.: A comparison of fixation methods for SEM analysis of self-assembling peptide hydrogel nanoarchitecture. Nanoscale 15 (3), 1431–1440.8 (2023).
10. Scarselli ,Rawashdeh, Awni.: A deep learning-based SEM-ANN analysis of the impact of AI-based audit services on client trust. J. Appl. Account. Res. (2023).
11. Wang, Anping, et al.: Enhancing museum visiting experience: Investigating the relationships between augmented reality quality, immersion, and TAM using PLS-SEM. Int. J. Human–Computer Interact, 1–12 (2023).
12. Yu, N., Khosla, M., Dobs, K.: Using artificial neural networks to ask 'why'questions of minds and brains. Trends Neurosci. **46**(3), 240–254 (2023)

13. Wu, J., et al.: A survey of uncertainty in deep neural networks. Artif. Intell. Rev. **56**(Suppl 1), 1513–1589 (2023)
14. Song, Dale W., et al.: Can Neural Networks Predict Stock Market?. AC Invest. Res. J. 220.44 (2023).
15. Lang, J.S., et al.: Deep problems with neural network models of human vision. Behav. Brain Sci. **46**, e385 (2023)

Chapter 14
The Impact of Contextual Word Embeddings in Natural Language Processing

Ayush Mehta and Anuja Bokhare

Abstract Contextual word embeddings represent a cutting-edge advancement in Natural Language Processing (NLP). This technique facilitates machines to achieve a deeper understanding of language, proposing dynamic context awareness that resolves word ambiguities and polysemy. In this study, we provide an introduction to contextual word embeddings, explore its fundamental principles, and discuss its superior performance compared to traditional static embeddings. We emphasize its significant impact on a variety of NLP tasks, including machine translation, sentiment analysis, and question-answering systems. Although there are considerable advantages, challenges related to interpretability and ethical concerns must be addressed to fully harness the potential of contextual word embeddings in NLP. This research sheds light on the transformative potential of contextual word embeddings and the requirement for further exploration and responsible utilization in the field.

Keywords Natural Language Processing (NLP) · Contextual word embeddings · Sentiment analysis · Machine translation

14.1 Introduction

Imagine talking to a phone or computer that really recognized in the same way people are. Contextual Word Embeddings is an innovative NLP tool that has opened up the exciting possibility of allowing machines to comprehend the meaning of words in the context in which they are used. This cutting-edge method enables machines to understand the subtle nuances of language, enhancing their ability to process and

A. Mehta
Symbiosis International (Deemed University), Symbiosis Institute of Computer Studies and Research, Atur Centre, Gokhale Cross Road, Model Colony, Pune 411016, Maharashtra, India

A. Bokhare (✉)
Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Paud Road, Kothrud, Pune 411018, Maharashtra, India
e-mail: anuja.bokhare@mitwpu.edu.in

analyze text. The outcome of this impacts the superiority and accuracy in processing language.

Traditionally, computers using traditional language processing techniques have overlooked complex meanings and nuances arising from the interactions of words in telegrams or paragraphs. On the other hand, by exploiting Deep Learning techniques to create word representations which adapt to its surroundings as humans do, contextual Word Embedding circumvents this problem.

Contextual Word Embeddings are a simple idea that lets computers infer a word's different meanings based on the words that surround it. Contextual Word Embeddings, for instance, help computers in identifying the appropriate meaning based on the sentence's context; when the word "bank" can refer to both a financial organization and the edge of a river. To determine if a sentence is in a positive or negative context, authors have used sentiment analysis and POS tagging in the experiment. Authors have investigated well-known models like BERT and others during this process, seeing its potential to produce even deeper contextual awareness in the future. Authors anticipate that by applying these cutting-edge models, it will be able to extract more context from text, producing analyses in various NLP applications that are more precise and insightful.

Contextual word embeddings have a wide range of uses that have the potential to alter how one can engage with technology, from interpreting customer feelings to assisting doctors with medical reports. Contextual Word Embeddings present problems, such as comprehending how these models arrive at their results and resolving ethical issues, like any strong technology. Contextual Word Embeddings, however, are poised to make significant future strides in reshaping the AI landscape and improving the naturalness and intuitiveness of our interactions with machines as we explore this fascinating field of language understanding.

14.2   Previous Study

The previous study is based on key advancements and research areas in Natural Language Processing (NLP) with a focus on contextual word embeddings:

Devlin et al.'s [1] work on BERT represents a pivotal advancement in natural language understanding. The standard for each of the NLP tasks was set by their preclusion training approach using deep bidirectional transformers, which pushed boundaries in terms of language models. The main points of the paper should be summarized in a conclusion to the Research Paper, so that readers can understand them better [2]. Although conclusions do not usually provide new information which has not been mentioned in the article, they often recast issues or offer a different view on this subject. The basis for today's methods of inserting words was Mikolov et al.'s paper on Word Representation in Vector Space [3]. This has been a key factor in the field of natural language processing, which has had an impact on numerous applications [4]. The most important contribution to word representation is the proposed method for vectors of Global Words. Their approach played a key role in improving

the understanding of linguistic terminology [5]. A strong text analysis framework is provided by the author's recurrent convolutional Neural Networks for the classification of texts. For a wide range of applications in the field of text classification, their work has proved to be valuable [6]. Author's explanation of Mikolov et al.'s negative-sampling word embedding method sheds light on an essential technique in NLP. It's been a useful tool in the training of word embedding [7]. Language models were greatly enriched by authors' deeply contextualized representations of words. Their approach has led to better contextual understanding, which has led to improved NLP systems. [8] For the evaluation and advancement of natural language understanding, GLUE benchmarks and analysis tools are an invaluable resource. The transformer architecture, which revolutionized NLP, was introduced by "Attention Is All You Need." The landscape of modern language models has been shaped by their work [9]. The concept of Language Models as Unsupervised Multitasking Learners has had a significant influence on the development of Versatile Models [10]. In NLP, it opened the door to a variety of applications. The authors represent a strong extension of autoregressive training [11]. Their approach has enhanced the ability to understand languages, which is a critical development in Linguistic Programming. The authors have developed a robust optimization approach for BERT training [12]. Their work has led to better stability and performance of language models, thereby enabling a more coherent outcome.

Conneau et al. [13] work on unsupervised cross-lingual representation learning at scale has advanced the field of multilingual NLP. Their approach facilitates language understanding across diverse languages. Bhardwaj, Biadsy, and Hovy's [14] structured word embeddings for semantic role labeling contribute to the understanding of text structures. Their approach has applications in semantic role labeling tasks. Wang et al. [15] Super GLUE benchmark provides a more challenging evaluation framework for general-purpose language understanding systems. It has driven the development of more sophisticated models in the NLP community. Table 14.1 describes a brief summary of the literature.

14.3   Problem Statement

The purpose of this study is to investigate contextual word embeddings in natural language processing (NLP) and their implications. In contrast to traditional static embeddings, contextual word embeddings offer deeper language comprehension by allowing robots to understand word meanings based on context. The study looks at the core ideas behind these embeddings, emphasizing how dynamic context awareness, resolving word ambiguities, and improving language processing are all exhibited. It also looks at how important an impact they have on different NLP tasks like sentiment analysis, machine translation, and question-answering systems. The study addresses interpretability and ethical issues while also discussing practical applications in consumer attitude analysis and medical report help. The study highlights the

Table 14.1 Summary of previous study

Reference	Findings	Limitations/Challenges
[1]	-BERT significantly advances NLP by introducing deep bidirectional transformers -It establishes a new standard for various NLP tasks	-Limited to bidirectional context, making it computationally intensive - May struggle with handling very long documents
[2]	-Generative pre-training substantially improves language understanding -It has a profound impact on the development of language models	-Pre-training data size and quality affect model performance -Transferability to diverse do Mains and languages are not always guaranteed
[3]	-Mikolov's work forms the foundation for modern word embedding techniques -It influences a wide range of NLP applications	-Difficulty in handling polysemy and word ambiguity - Limited ability to capture semantic relationships
[4]	-Pennington, Socher, and Manning's GloVe method contributes significantly to word representation - It enhances the understanding of word semantics	- Lack of transparency in the vector generation process -May not perform well on tasks requiring fine-grained word representations
[5]	-Lai et al.'s recurrent convolutional neural networks provide a robust framework for text analysis -Their work has valuable applications in text classification	-RNN-CNN models may be computationally expensive -Not ideal for real-time applications
[6]	-Goldberg and Levy's explanation of negative-sampling word embeddings clarifies an essential NLP technique - It facilitates effective training of word embeddings	-Negative-sampling method might not be optimal for all NLP tasks -Some embeddings may struggle with rare words
[7]	-Peters et al.'s deep contextualized word representations enrich language models significantly -They improve contextual understanding in NLP systems	-May require significant computational resources -Limited performance on low-resource languages
[8]	-Wang et al.'s GLUE benchmark and analysis platform are valuable for evaluating and advancing NLU systems -It spurs research in developing more capable models	-Focuses on English tasks; performance in non-English languages is underrepresented -Benchmark tasks may not reflect real-world application scenarios
[9]	-Vaswani et al. 's 'Attention Is All You Need' revolutionizes NLP with the introduction of the transformer architecture -It shapes the modern landscape of language models	-Large model sizes can be computationally demanding -Need for extensive data for pre-training

Table 14.1 (continued)

Reference	Findings	Limitations/Challenges
[10]	-Radford et al.'s concept of language models as unsupervised multitask learners opens doors for versatile model development -It leads to various applications in NLP	-Challenges in selecting the appropriate set of tasks for training -May not always result in improved performance on all tasks
[11]	-Improved performance over BERT -Enhanced language understanding -Bidirectional and autoregressive pre-training benefits	-Computationally intensive training -Complex autoregressive training
[12]	-Optimization improvements yield enhanced performance -Pre-training on a larger dataset leads to better models -Task-specific data augments generalization	-Complex optimization techniques -Demands substantial task-specific data
[13]	-Multilingual representations are learnable from monolingual data -Model performance varies across languages -Benefits from large-scale multilingual training data	-Requires significant multilingual data -Performance varies with language similarity
[14]	-Word representations capture semantic structures -Useful for semantic role labeling tasks -Improved understanding of word relationships	-May not consistently outperform traditional embeddings -Performance is influenced by data quality and size
[15]	-Provides a challenging bench mark for evaluating language understanding models -Promotes the development of more capable models -Highlights varying task complexities in NLP	-Varying task difficulties in the benchmark -High scores may not correlate with real-world performance

potential for Contextual Word Embeddings to change the AI by detecting the need for more research.

14.4   Proposed Methodology

In current research, authors have used POS tagging in addition to sentiment analysis to determine whether the context of statements is positive or negative. Each word in a sentence is given a part of speech (such as a noun, verb, or adjective) as part of the POS tagging process, which adds crucial linguistic information to the sentiment analysis findings.

Process Flow
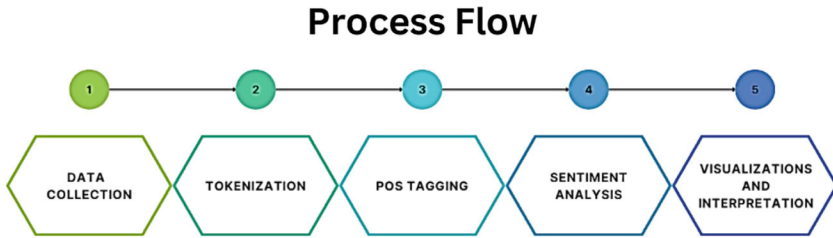
Fig. 14.1 Diagrammatic representation of proposed model

With this combination strategy, it is possible to perform a more in-depth analysis that can distinguish between positive and negative adjectives and capture minute differences in sentiment expression. As well as handling negations and modifiers, it ensures a more accurate assessment of sentiment by avoiding potential misunderstandings. The incorporation of POS tagging and sentiment analysis in my research has advanced our understanding and application of sentiment analysis techniques in a variety of fields, including natural language processing, social media analysis, and customer feedback evaluation. This has given us valuable insights into the sentiment of statements. Figure 14.1 shows diagrammatic Representation of the proposed model.

14.5 Dataset and Experiment Discussion

The dataset employed in this study comprises textual reviews, likely from diverse sources, reflecting sentiments expressed by users. The dataset's content is subject to Part of Speech (POS) tagging and sentiment analysis, allowing for a detailed examination of sentiment patterns. The study leverages this dataset to showcase the effectiveness of contextual word embeddings in tasks such as sentiment analysis, demonstrating the practical implications of the research findings.

Word Cloud of POS Tags is shown in Fig. 14.2.The Fig. 14.2 is a word cloud of POS (Part of Speech) tags in the reviews. POS tagging is a process of assigning grammatical tags to each word in a sentence (e.g., Noun, Verb, Adjective, etc.). The word cloud visually represents the frequency of different POS tags in the reviews. The most frequent POS tags will appear larger in the word cloud.

Bar Graph for Predicted Sentiment is shown in Fig. 14.3. Which displays the count of predicted sentiment categories (Positive and Negative). The sentiment analysis is performed using the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analyzer, which calculates a sentiment score for each review. Based on the compound score, the reviews are classified as Positive or Negative.

The "sentiment" column in the dataset represents the ground truth or manually labeled sentiment values for each entry. This column contains categorical values such as "positive" and "negative," which indicate the sentiment expressed in the

Fig. 14.2   w or d cloud of POS.

Fig. 14.3   Bar Graph for predicted sentiment
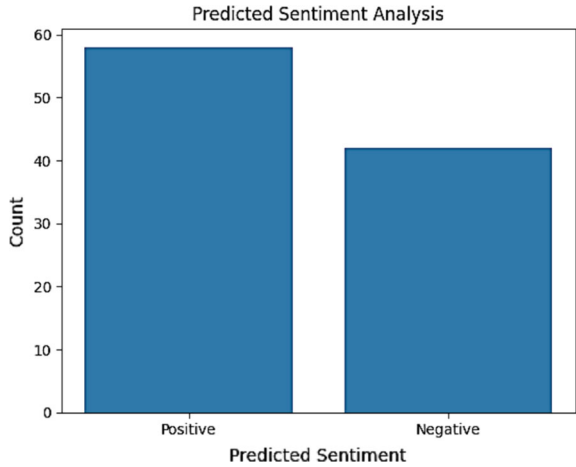
corresponding reviews. To gain insights from this data, we have created a bar graph to visualize the distribution of sentiments. The bar graph (see Fig. 14.4) displays the counts of "positive" and "negative" sentiments, allowing us to observe the balance between positive and negative sentiments in the dataset. This visualization helps us understand the overall sentiment patterns and proportions within the sampled data.

The stacked bar chart in Fig. 14.5 compares the predicted sentiment and actual sentiment analysis results for a dataset containing 100 entries. The chart displays two bars for each sentiment category: "Positive" and "Negative." The lower bar represents the count of predicted sentiment, while the upper bar shows the count of actual sentiment obtained from the dataset. Observing the chart, we can see that the VADER sentiment analyzer's predictions generally align with the actual sentiments

Fig. 14.4 Distribution of sentiments from dataset
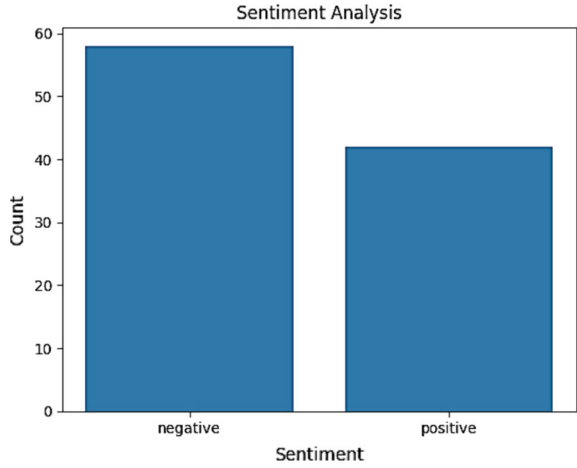
Fig. 14.5 Distribution of predicted sentiment versus actual sentiment
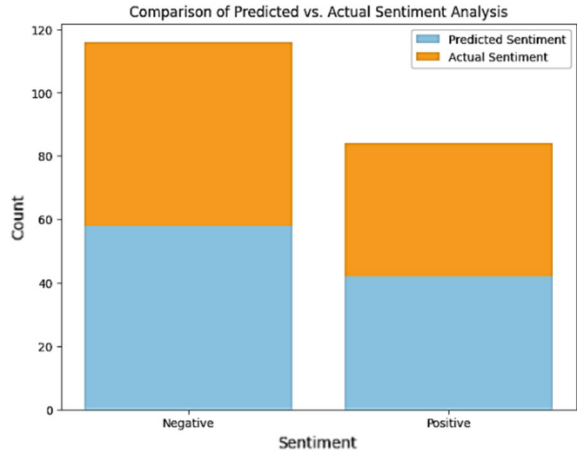
provided in the dataset, as both bars are relatively close in height for each sentiment category.

14.6   Result Analysis

In the context of sentiment analysis, the model's performance metrics were evaluated with the following results shown in Table 14.2.

Precision, representing the proportion of correctly predicted positive sentiment instances among all predicted positive instances, achieved 68%. Recall, indicating the ability to identify all positive sentiment instances correctly, exhibited a higher level of

Table 14.2 Performance metrics

Performance metrics	Precision
Precision	68%
Recall	86%
F1-Score	76%
Support	501
Accuracy	72%

86%. The F1-Score, striking a balance between Precision and Recall, reached 76%, indicating a reasonably good overall performance in correctly identifying positive sentiment instances while minimizing false positives.

The model was tested on a total of 501 instances (Support). Additionally, the model's Accuracy, measuring the proportion of correctly predicted sentiment instances overall, stands at 72%.

Furthermore, it is noteworthy that the model's accuracy is expected to substantially increase with the incorporation of more testing data. As the volume of diverse data grows, the model's ability to generalize and comprehend various sentiment patterns and nuances will improve. This improvement is expected to lead to more reliable transmission analysis results for various text inputs, highlighting the model's evolution in terms of performance over time.

14.7   Discussion

The study, which proposes a paradigm shift from traditional linguistic processing techniques, examines the groundbreaking impact of contextual words embedded in natural language processing NLP. Machine language comprehension is enhanced by semantic word embedding. The study highlights the significant impact of NLP on diverse tasks, including sentiment analysis, machine translation, and question-answering systems.

Despite their advantages, challenges in interpretability and ethical concerns are acknowledged, urging responsible utilization. Practical applications, ranging from customer attitude analysis to medical report assistance, underscore the potential of contextual word embeddings. The study reflects on the need for further exploration and responsible integration in the field. The investigation draws from previous studies, citing key advancements in NLP, such as BERT, generative pre-training, GloVe, and others. Notable contributions and limitations of these models are summarized, providing a foundation for the current research.

The proposed methodology incorporates Part of Speech (POS) tagging and sentiment analysis, enhancing sentiment assessment accuracy. The dataset and experimental findings, illustrated through word clouds, bar graphs, and sentiment distribution charts, contribute to the understanding of sentiment patterns. Result analysis reveals promising performance metrics in sentiment analysis, with precision at 68%,

recall at 86%, and an F1-Score of 76%. With additional varied testing data, the 72% accuracy of the model could be increased. The study's result emphasizes how contextual word embeddings in NLP have a revolutionary effect on language processing accuracy.

14.8 Conclusion

This work demonstrates the superiority of contextual word embeddings over conventional approaches by exploring their practical application in Natural Language Processing in an innovative way. In addition to adding depth, the novel combination of sentiment analysis with Part of Speech tagging addresses ethical issues and interpretability problems. The study offers a forward-looking viewpoint, highlighting conscientious growth and the transformational power of contextual words. Finally, contextual word embeddings—a revolutionary advancement in Natural Language Processing (NLP)—were investigated in this study. Computers now comprehend human language differently because of the usage of contextual word embeddings, which consider the meaning and context of words inside textual data. These embeddings improve language processing accuracy by helping computers comprehend the complexities of language through the use of deep learning techniques.

References

1. Devlin, Jacob, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
2. Radford, Alec, KarthikNarasimhan, and Ilya Sutskever. Improving Language Understanding by Generative Pre-Training. OpenAI Blog
3. Mikolov, Tomas, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
4. Pennington, Jeffrey, Richard Socher, Christopher, D.: Manning. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543
5. Lai, Siwei, et al.: Recurrent Convolutional Neural Networks for Text Classification. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2267–2273
6. Goldberg, Yoav, Omer Levy.: Word2Vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. arXiv preprint arXiv:1402.3722
7. Peters, Matthew E., et al.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 2227–2237
8. Wang, Alex, et al.: GLUE: a multi-task benchmark and analysis platform for natural language understanding. In: Proceedings of the 2019 International Conference on Learning Representations (ICLR)
9. Vaswani, Ashish, et al.: Attention is all you need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), 6000–6010
10. Radford, Alec, et al.: Language Models are Unsupervised Multitask Learners. OpenAI Blog

11. Yang, Zhiheng, et al.: XLNet: Generalized Autoregressive Pre Training for Language Understanding. In: Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), 5754–5764.
12. Liu, Yinhan, et al.: RoBERTa: a robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692
13. Conneau, Alexis, et al. Unsupervised Cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 8440–8451.
14. Bhardwaj, Ashutosh, FahimBiadsy, and Eduard Hovy.: Structured word embeddings for semantic role labeling. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2022–2035
15. Wang, Alex, et al.: SuperGLUE: a stickier benchmark for general-purpose language understanding systems. In: Proceedings of the 2019 International Conference on Learning Representations (ICLR)

Chapter 15
FPGA Implementation of Vector Reduction Algorithm for LFSR

Check for updates

Rashmi Gavimath, Suhas Shirol, H. M. Vijay, V. S. Saroja, and M. Rajeshwari

Abstract When logic is tested on an IC, heat generated by power dissipation causes damage to the semiconductor. The IC's life will be shortened due to this power waste. Therefore, to extend the life of an integrated circuit (IC), power dissipation must be reduced as much as feasible. Additionally, testing every potential test vector for proper IC operation is a limitation. To lower the power consumption during testing of a circuit under test (CUT), we have devised a bit-swapping linear feedback shift register (BS-LFSR). The BS-LFSR reorders the test vectors by swapping out a bit for its next closest neighbor bit. The vectors' fault coverage capability is maintained, but the total Hamming distance (THD) is lowered to lower power consumption during the shifting operation.

Keywords Linear Feedback Shift Register (LFSR) · bit · Swapping Linear Feedback Shift Register (BS-LFSR) · Built-In Self-test (BIST) · Circuit Under Test (CUT)

R. Gavimath · S. Shirol (✉) · H. M. Vijay · V. S. Saroja · M. Rajeshwari
Department of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka 580031, India
e-mail: suhasshirol@kletech.ac.in

H. M. Vijay
e-mail: vijay.hm@kletech.ac.in

V. S. Saroja
e-mail: sarojavs@kletech.ac.in

M. Rajeshwari
e-mail: rajeshwari_m@kletech.ac.in

15.1   Introduction

Power dissipation is a significant challenge during System-on-Chips (SoC) design and testing. Typically, a system uses more power during testing than in normal mode. This happens because consecutive test vectors in regular operation often correlate strongly, but test mode vectors lack such correlation. This low correlation between test vectors increases switching activity, leading to increased power dissipation [1, 2].

BIST is a testing method that uses a LFSR as a test pattern generator. It is considered the most effective method for low-power testing. The LFSR generates a comprehensive range of test vectors using a carefully selected tap sequence. Its pseudo-random nature reduces the correlation among test vectors, facilitating high fault coverage in a brief testing period. However, the reduced correlation between test vectors increases the Hamming distance between them, leading to higher switching activity in the CUT and power dissipation during test mode [3, 4].

To optimize the LFSR for power efficiency, it is crucial to produce power-efficient test vectors. The BS-LFSR is a refined version of the conventional LFSR that uses multiplexers driven by the nth bit to generate pseudo-random patterns. This design minimizes transitions in the outputs, which reduces power consumption [4–6].

These optimizations are particularly important in digital devices like computers, mobile phones, processors, and MAC units, where minimizing delay and power dissipation is essential.

15.2   Literature Survey

The testing cycle results in a higher power dissipation than the chip's typical operation. One explanation is that the circuit's power dispersion decreases in tandem with the chip size reduction. BIST is a technique that involves adding a small amount of overhead to the circuit in order to test and eliminate the need for complicated external testing components. The LFSR is used in BIST to generate the irregular patterns needed for VLSI circuit testing. This work describes multiple WRTPG, BS-LFSR, and a traditional linear feedback shift register [1, 7].

In this paper, BS-LFSRs are used to create extremely power-efficient pattern generators for VLSI design based on BIST. The hardware device known as BIST is capable of doing both in-situ and production testing on the circuit. BIST can create test vectors with little fault coverage by utilizing several random generation techniques [4–6].

In this paper, the suggested technology, a LFSR, aims to lower the power consumption of the BIST itself. It lowers power consumption in two steps 9 when testing a CUT. To minimize unnecessary flip-flop switching, the Control Logic (CL) first temporarily deactivates the clocks of the register's switching units when their output remains unchanged from the previous state. Subsequently, in the second stage, the Linear Feedback Shift Register (LFSR) swaps each bit with its adjacent neighbor

to reorder the test vectors. This reordering maintains the vectors' capacity for fault coverage while reducing the Total Hamming Distance (THD), thus decreasing power consumption during shifting operations [2–4, 8].

15.3 Design Methodology

- Step 1: START.
- Step 2: Initialize the seed esteem.
- Step 3: Initialize the clock and other signals.
- Step 4: Random test vectors will be generated.
- Step 5: If the nth Bit (last Bit) is 0, then Bit interchanging will be performed and if the nth Bit is 1, then bit interchanging is not performed.
- Step 6: The Hamming distance was based on the number of bits changed between two successive outputs of bit swap.
- Step 7: The test vectors which had a Hamming distance of <2 were considered and passed to the full adder.
- Step 8: This was tested on a full adder application.
- Step 9: Power analysis was performed for the results obtained.
- Step 10: STOP.

15.3.1 Design and Implementation

Figure 15.1 shows the architecture for BIST.
BIST is made up of multiple blocks:

- Circuit Under Test (CUT): This could be either a sequential or combinational circuit. It is the section of the circuit that has to be examined.
- Test Pattern Generator (TPG): Creates test patterns for CUT.
- Test Controller: It manages how the test is carried out. It sends out a signal to turn on every block. The BIST is in test mode if the signal from the test controller is 0 and in normal mode if the signal is 1.
- ROM: It holds the signature that needs to be compared to the output of the CUT.

Fig. 15.1 Basic BIST architecture
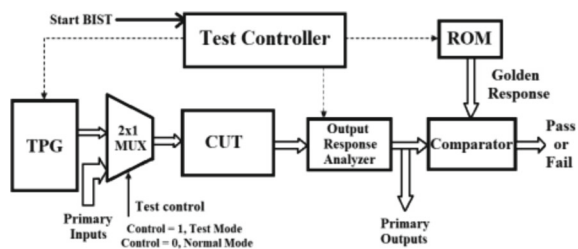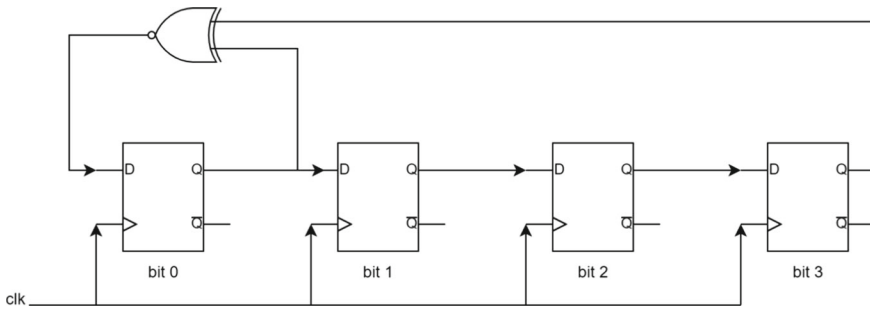
Fig. 15.2 Conventional 4-bit LFSR

- The Response Analyzer compares the response that is saved with the test output. The CUT is not defective if the test result agrees with the response that was stored; if not, the chip or circuit is malfunctioning.

15.3.1.1 Linear Feedback Shift Register

The most used test pattern generator is LFSR. Its production features pointless highlights. Shift registers with feedback using an Exclusive-OR (or Exclusive-NOR) gate make up an LFSR. Feedback can be given both within and outside. For maximum output, the LFSR should not be fed all zeros since it would cause it to become caught in a feedback loop and not produce any changes. The output of the feedback loop is determined by the placement of the taps, which are based on the polynomial.

Figure 15.2 shows the architecture for Conventional LFSR.

By using an exclusive-NOR gate on the outputs of two or more flip-flops and feeding those outputs back into the input of one flip-flop, it is possible to simplify the linear feedback shift register. In this, the feedback is the exclusive-NOR gate of the least significant Bit and most significant Bit of the previous output. The output of the NOR gate is fed to the input of the first flip-flop.

15.3.1.2 Bit-Swapping Linear Feedback Shift Register

A refined version of the traditional Linear Feedback Shift Register, called the Bit-Swapping LFSR (BS-LFSR), can generate pseudo-random patterns with fewer transitions between 0 and 1. This decrease in transitions reduces the average power dissipation in the Circuit Under Test (CUT) by lowering the internal switching activity. The BS-LFSR can be implemented using either a Test-per-scan or Test-per-clock scheme to further reduce power consumption. The primary factor that affects the CUT's power consumption is the reduction in internal switching. During testing, the BS-LFSR effectively reduces both the average and instantaneous weighted switching activity by minimizing transitions. To lower power consumption, two direct methods
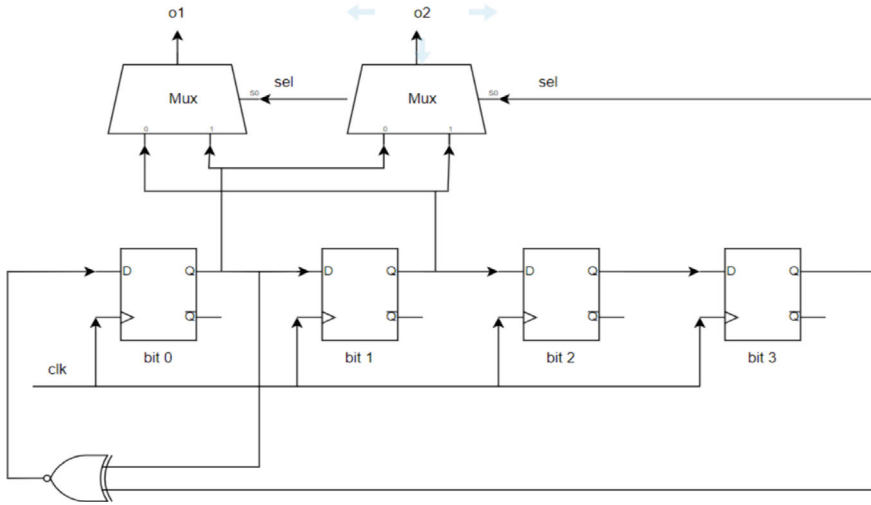
Fig. 15.3 4-bit bit-swapping LFSR

are commonly used. The first method involves reducing the frequency during testing, which does not require additional hardware. The second method segments the CUT into blocks using a strategic test design to lower power usage. Although these direct approaches lengthen the test duration, they are ineffective in reducing peak power consumption. Compared to other methods, the BS-LFSR significantly decreases both the average and peak power dissipation in the CUT [3].

Figure 15.3 shows the architecture for BS-LFSR.

The vectors are generated from LFSR, and the last Bit of the vector is considered a select line for the multiplexer. The first two bits are passed as inputs to the multiplexer. If the last Bit is 0, i.e., the select line to the mux is 0, then the first two bits are swapped, and if the last Bit is 1, i.e., the select line is one, then there is no swapping.

If the last Bit of the test vector is 0, bit swapping is performed, and the corresponding Hamming distance is calculated. Table 15.1 shows the Hamming distance of Bit-swapped vectors is less than conventional LFSR.

15.4   Results and Analysis

Table 15.2 shows that vectors with <3 Hamming distances are considered, and the vectors are reduced from 16 to 12. These reduced 12 vectors are tested on Full adder, and the sum is calculated. The test coverage for the application is 100%.

Analysis of Results

Figures 15.4 and 15.5 show the FPGA implementation of vector reduction algorithms for LFSR, considering the Hamming distance of <3. The output changes at every

Table 15.1 4-bit swapping vectors

	Test vectors	Bit swap	Hamming distance of conventional LFSR	Hamming distance of BS-LFSR
V1	0100	1000	3	3
V2	1010	0110	4	2
V3	0101	0101	3	3
V4	0010	0010	3	3
V5	1001	1001	2	2
V6	1100	1100	2	2
V7	0110	1010	3	1
V8	1011	1011	2	2
V9	1101	1101	2	2
V10	1110	1110	2	2
V11	0111	0111	1	1
V12	0011	0011	1	1
V13	0001	0001	1	1
V14	0000	0000	1	1
V15	1000	0100	2	2
V16	0100	1000		
		Total hamming distance	32	28

Table 15.2 Reduced vector table with Hamming distance <3

	Reduced vector	Sum
V1	0110	11
V2	1001	11
V3	1100	11
V4	1010	00
V5	1011	01
V6	1101	00
V7	1110	01
V8	0111	10
V9	0011	11
V10	0001	01
V11	0000	00
V12	0100	01

positive edge of the clock, initially starting with the seed value. If the reset is 1, then the sequence starts from the seed value. If the reset is 0, the sequences are generated.

Table 15.3 shows the results of Power analysis of 4-bit LFSR with Hamming distance <3 and the results of Power analysis of 4-bit BS-LFSR with Hamming distance <3. The Dynamic Power of BS-LFSR and LFSR is 19622.523 and
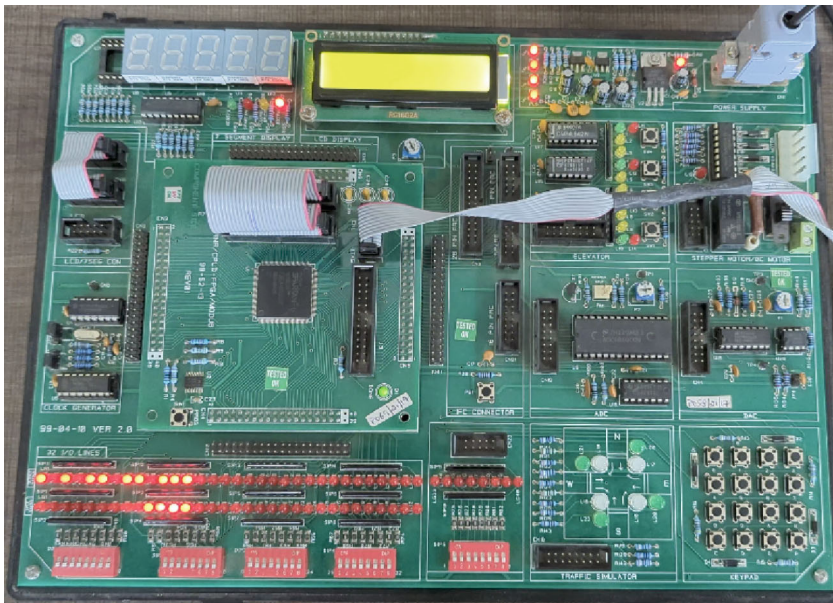
Fig. 15.4 Results when reset is 1, FPGA implementation on Spartan 6
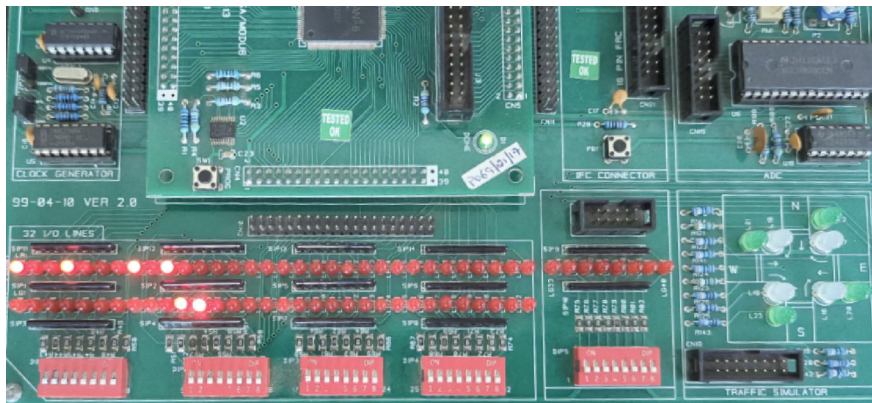
Fig. 15.5 Results when reset is 0, FPGA implementation on Spartan 6

Table 15.3 Power analysis of different architectures in Cadence 180 nm Technology

S. No	Architecture	Cells	Leakage power (nW)	Dynamic power (nW)	Total power
1	LFSR	54	3.654	25,209.839	25,213.493
2	BS-LFSR	50	3.352	19,622.253	19,625.875

25,209.839 nW, respectively. The Dynamic Power has been reduced in BS-LFSR compared to LFSR by using a Hamming distance of <3 in cadence.

Conclusion

- As mentioned, the paper has three goals. The first aim of studying and understanding the power loss due to many transitions in the Conventional LFSR circuit has been completed by researching Power reduction for LFSR circuits.
- The second goal was to use the swapping technique to reduce the number of transitions between the consecutive test vectors produced by conventional LFSR circuits. We have successfully implemented the bit-swapping module for the LFSR circuit by using Verilog, and the results are simulated.
- The third aim was to check whether the power of LFSR has been reduced by reducing the number of transitions using BS-LFSR. We have done the power analysis of the simulated results. It has been successfully seen that the power is reduced compared to conventional LFSR.
- It has been noted that the total power consumption of the BS-LFSR is lower than that of a standard LFSR. Consequently, it can be concluded that the low-power BS-LFSR is highly effective for enhancing power efficiency in Built-in Self-Test (BIST) systems.

References

1. Anandhi, S., Neela, R., Janaki Rani, M.: Power optimization by using reconfigurable LFSR with gated clock. In: IOP Conference Series: Materials Science and Engineering, vol. 993, no. 1. IOP Publishing (2020)
2. Shirol, S.B., Shettar, R.B.: A comparative study of low power testing techniques for digital circuits. Int. J. Adv. Res. Comput. Sci. Softw. Eng. (2017)
3. Kirthi, V., Mamatha Samson, G.: Design of BIST with low power test pattern generator. IOSR J. VLSI Signal Process. (IOSR-JVSP) **4**(5), 30–39 (2014)
4. Dhanesh, P., Jayanth Balaji, A.: Dual threshold bit-swapping LFSR for power reduction in BIST. In: International Conference on Advanced Computing and Communication Systems. IEEE (2015)
5. Dhanyal, J.N., Shirol, S., Shettar, R.B.: Design and implementation of different type of adders using bit swapping LFSR as delay improvement. Int. Res. J. Eng. Technol. (IRJET) (2017)
6. Praveen, J., Shanmukhaswamy, M.N.: Power reduction technique in LFSR using modified control logic for VLSI circuit. Int. J. Comput. Appl. **975**, 8887 (2013)

7. Bagalkoti, A., Shirol, S., Shettar, R.B., et al.: Design and implementation of 8-bit LFSR, bit-swapping LFSR and weighted random test pattern generator: a performance improvement. In: 2019 International Conference on Intelligent Sustainable Systems (ICISS). IEEE (2019)
8. Lahari, M.V.M., Mani Kumari, M.: Bit swapping LFSR and its application to fault detection and diagnosis using FPGA. Int. J. Eng. Res. Technol., IJERT (2013)

Chapter 16
Design and Stress Analysis of the Frame for an Electric Bike

S. J. Niranjana ⓘ**, S. Shivakumar** ⓘ**, Srinivasaiah Raghavendra** ⓘ**, and H. C. Ravikumar** ⓘ

Abstract Global emissions have been on the rise since the industrial era because of the increased energy-intensive human activities, which is a direct cause of global warming and climate change. Of the total emissions, around 17% is from the transportation sector, which significantly contributes to the emissions. One of the easiest ways to be more sustainable is to choose electric vehicles instead of Internal combustion engines. Almost 75% of the vehicles registered in India are two-wheelers, but there are no affordable and reliable electric two-wheelers. This research works to optimize and analyze the design of a step-through frame design for an electric bicycle. The frame design is analyzed by providing boundary and loading conditions with two different materials (Steel-AISI4130 and Aluminum AL6061). The numerical analysis is carried out using ANSYS APDL. The result of von Mises' stress is 166 MPa and 160.4 MPa for steel and aluminum, respectively. The result of stress and displacement is within the acceptable limit.

Keywords Bicycle frame · ANSYS Mechanical APDL · Eco-friendly

S. J. Niranjana (✉) · S. Shivakumar
Department of Mechanical and Automobile Engineering, CHRIST (Deemed to Be University), Bangalore 560074, India
e-mail: niranjana.s@christuniversity.in

S. Raghavendra
Department of Computer Science Engineering, CHRIST (Deemed to Be University), Bangalore 560074, India

H. C. Ravikumar
Department of Computer Science Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore 560082, India

16.1   Introduction

Many cities and places are vulnerable to climate change and rising global temperatures. The majority of the global population lives in cities, making cities the most significant source of global greenhouse gas emissions. Sumbodo et al. [1] explain that human activities have led to a massive increase in CO_2 emissions as a primary greenhouse gas contributing to climate change and global warming than pre-industrial energy. Ali and Rizky [2] concluded that the increasing demand for energy and its consumption worldwide is satisfied mainly by burning fossil fuels such as petrol, diesel, coal, AND natural gas.

Kunjan [3] describes that the two-wheeler bicycle was presented in the nineteenth century in Europe, and in today's world, it is a number more than a billion around the world, two times as many as cars. Niranjana et al. [4–6] describe the effects on the people living in the cities, including their health, by increasing the risk of acquiring respiratory illnesses such as asthma. Shubham et al. [7], Cities are also most affected by extreme weather events like floods, drought, heat waves, smog, and so on. They are the chief method of transportation in most areas around the globe. They also give a famous type of diversion and have been adjusted for kids' toys, general wellness, military and police applications, messenger administrations, and bicycle hustling.

Esther and Francisco [8] The electric bicycle was first made and patented in the 1890s. On 31st December 1895, a battery-powered bicycle was made by Odgen Bolten. It was designed using a 6 pole brush and commuter DC hub motor connected to the rear wheel. Now, the story has changed. Yashwant et al. [9] There are a lot of electric bicycles with long-range presence in the global market. But that is not the case in India, where not many options are present in the market.

Mitesh et al. [10], India's first electric vehicle was launched in the nineteenth century in 1996. The first vehicle that was launched was a three-wheeler by Scooter's India Pvt. Ltd., and the name which was given to the vehicle is Vikram Safa. Almost 400 vehicles were manufactured by them and sold in the market, although these were not that successful during the lack of availability of charging stations. Sunikshita et al. [11] The electric vehicle slowly started to become famous. After the launch of Ather 340 in India, the E-Scooter became a success, and people began to recognize it. They even started to plant charging stations in major cities. So that charging point was not an issue here.

Yogesh et al. [12] There have been several shortages of fuels in India, especially coal, so the transition from conventional energy sources to renewable sources is crucial. This transition can be seen even in the transportation sector, where people have been choosing electric vehicles (EV) over internal combustion engine (ICE) vehicles as they are more sustainable. Gicky and Tom [13] There is a need to curb carbon emissions, especially in cities, by adopting sustainable solutions like Electric Vehicles to reduce emissions from the transportation sector.

Mahesh et al. [14] utilization of fuel vehicles has increased rapidly which results in more air pollution. To control this, the utilization of EV is a must because it has several advantages like electric scooter is an eco-friendly product, it is more suitable

for the city as it can avoid the emission of harmful gasses and thereby it can reduce the atmospheric pollution.

Krishna Veni and Amareshwari [15] describe the most crucial manufacturing phase as design. This is due to the fact that every product that is created must go through the design stage, which includes conceptual design. Chien-Chen et al. [16] compare the maximum stress and displacement with different tubes and distances.

This research aims to reduce the weight of the bicycle frame by using different materials and to achieve cost-effectiveness. The objective of the work is to design a frame for the electric bicycle as per the specification and to analyze the stress distribution and displacement in the frame.

16.2 Methods and Materials

To design and develop, initially study the journal papers to compare different models. In development design, refer to all the existing similar models. I thought about modifications that could be brought to this design and had to study whether the plan would be perfect and fulfill all the standards that had to be followed. After the design is complete, there is a need to check whether these designs can withstand a standard load and how efficient they are by analyzing them with the help of Hypermesh software and Ansys ADPL software. In this project case, the reference model researched that design and did the modifications.

The main objective of this research work is to design a new bicycle frame with minimum weight and cost-effectiveness. From the literature, it is observed that aluminum frames are preferred compared to steel. However, an attempt was made to check the stress analysis.

Design and Modeling: This is one of the most critical phases of the project as it includes a lot of processes. The first part calculates different parameters, including transmission ratio, bearing stress, length of change, and breaking load. Once these calculations are done we will move on to computer-aided design using SolidWorks software to design and model all the relevant parts of the mechanism. Part and Assembly study: Once the final design is done, Modeling and assembly are carried out using the Solidworks software as shown in Fig. 16.1.

Fig. 16.1 Gas emission [3]

As shown in Fig. 16.2, there are mainly two parts: Frame and Footplate. The purpose of the footplate is to make the rider more comfortable by which the rider can rest their legs, and also, if there are any materials or things to carry, the rider can keep them on the footplate and carry the items. Also, the back extension is given where the rider if he/she wishes to keep a pillion seat the rider can fix that on the extension given below or it also can be used as a support system for carrying any materials.

Figure 16.3 shows the discretization model of the E-bicycle frame using Hypermesh software. The analysis is carried out using Ansys APDL software by considering half the model because of symmetry, and one half is masked out; there are 132,057 elements and 42,870 nodes.

Figure 16.4 shows the boundary conditions and load conditions applied to find the results. Once all the required results are taken in the discussion a comparative study is done to understand the procured results. Critical data and analysis statements are
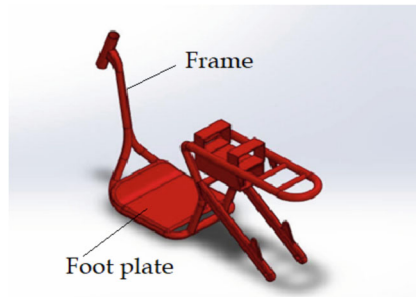
Fig. 16.2 Modeling of the bicycle frame
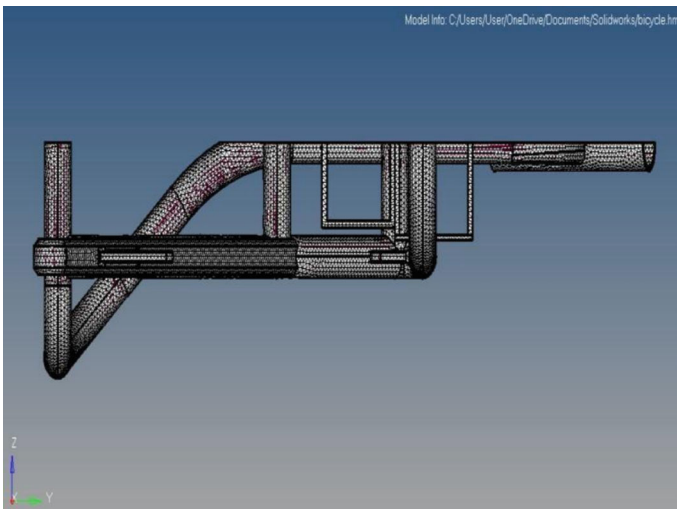
Fig. 16.3 Meshed E-bicycle frame

Fig. 16.4 Boundary and
load condition
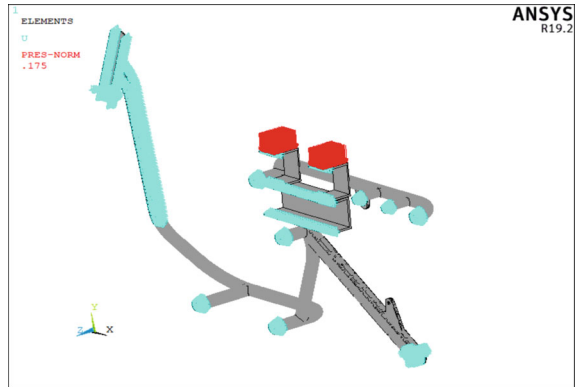
Table 16.1 Material properties of Steel (AISI 4130) and Aluminum (AL6061)

Material	Young's modulus (MPa)	Poisson's ratio	Density (g/cm^3)
Steel (AISI 4130)	200	0.32	7.85
Aluminum (AL6061)	70	0.35	2.7

taken from discussions to form conclusions. Now let us see some of the important dimensions given to the Final model while designing the model. The tube size is 30 mm outer tube and 24 mm inner tube of the frame other than the head tube part. In the head tube part, it is a 34.8 mm inner tube and a 40.8 mm outer tube.

The material properties of steel and aluminum are shown in Table 16.1. The total mass of the frame was found to be 18.96 kg, which is 185.99 N in steel. Then, the total mass of the frame was found to be 6.56 kg, which is 64.35 N in aluminum.

16.3 Result and Discussion

The analysis is carried out with Ansys APDL software. The frame is analyzed by applying two material properties, steel (AISI 4130) and aluminum (Al 6061), with a load of 250 kg. The analysis is done to only half of the frame with half load, but it is equal to applying to the full frame with the total load on it.

The stress analysis is done for principal stress and von Mises stress, and Y displacement was also found for steel and aluminum, as shown below.

16.3.1 Steel (AISI 4130)

To find out the total mass of the steel frame is Volume*Density. Then, the total mass of the frame was found to be 26.2 kg, which is 262 N. It is clear from this result that the frame is safe to use and that it can support a load of 250 kg. Determine that the frame is safe to use based on the steel's yield strength of 350 Mpa. 113.041 MPa is the greatest main stress. For the applied load of 250 kg, the minimum stress is 73.8391 MPa, the Von Mises Stress is 166.005 MPa, and the maximum Y displacement is determined to be 0.0021 mm. Figure 16.5 presents the findings.
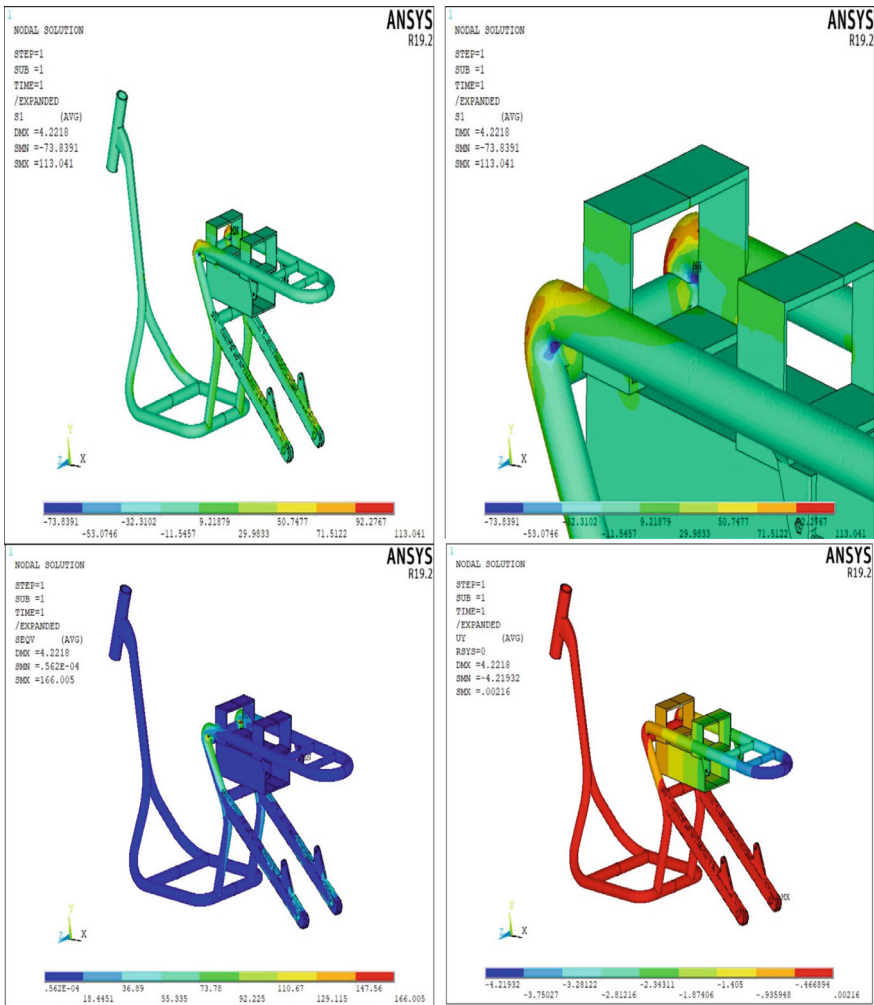
Fig. 16.5 Principal stress, Von Mises stress, and displacement solution of steel

16.3.2 Aluminum (Al 6061)

The total mass of the aluminum is found by using Volume*Density. Afterward, it was discovered that the frame's entire mass was 9.56 kg or 95.6 N. It is clear from this result that the frame is safe to use and that it can support a load of 250 kg. Determine that the frame is safe to use based on the aluminum's yield strength of 276 MPa. The research indicates that 126.895 Mpa is the maximum principal stress. For an applied load of 250 kg, the lowest stress is 92.2176 MPa, the Von Mises stress is 160.145 MPa, and the maximum Y displacement is found to be 0.006337 mm. Figure 16.6 presents the findings.

Figure 16.7 shows the comparison chart of the two different materials steel and aluminum. Even though both materials, steel and aluminum, are safe to use, it is evident from a frame comparison that aluminum is more effective due to its lower weight of 95.6 N compared to 262 N for steel. Aluminum is lightweight, but it can support the 250 kg load that was considered during testing. Principal stress, Von Mises stress, and displacements of steel and aluminum are compared in the chart.

16.4 Conclusion

The following points are concluded from the analysis results.

- In the current era of electric mobility, electric bicycles and other vehicles stand out as excellent solutions to contribute to environmental preservation by reducing pollution.
- With a load of 250 kg. After finding out the von Mises stress of 166 MPa and 160 MPa for steel and aluminum, respectively. The yield strength of the steel and aluminum is 350 Mpa and 276 MPa from that we can conclude the frame is safe to use.
- It is understood that the frame is safe to use because the chances of breaking due to load are less. And also from this, we can understand that using aluminum is more efficient than steel since it is lightweight and also cost-effective.
- To compare the frame using two materials steel and aluminum even though both are safe to use, it is clear that aluminum is efficient because aluminum is lightweight which is only 64.35 N and steel which is 185.99 N.
- Even though the aluminum is lightweight it can withstand the load of 250 kg which was considered during the testing.
- This type of design will help most logistic delivery agents deliver the orders.
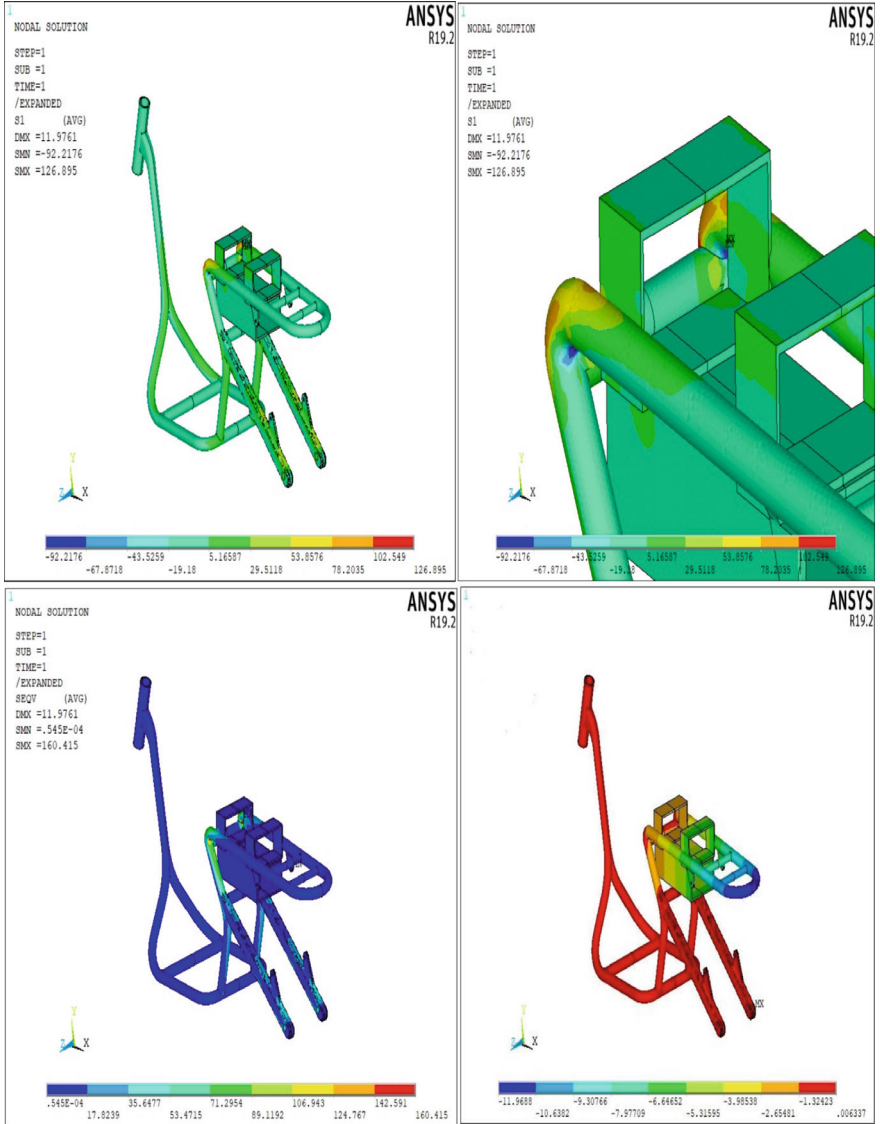
Fig. 16.6 Principal stress, Von Mises stress, and displacement solution of aluminum

Fig. 16.7 Comparison chart
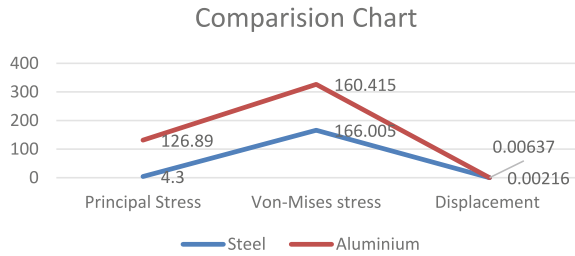
References

1. Sumbodo, W., Wahyudi, Setiadi, R., Kriswanto, Budiman, F.: Design and fabrication of electric bike with sliding frame. J. Appl. Eng. Sci. **19**(4), 948−953 (2021). https://doi.org/10.5937/jaes0-28957
2. Ramadhan, A., Dinata, R.: Development of electric bicycles and its impact on the environment. IOP Conf. Ser. Mater. Sci. Eng. **1122**, 012054 (2021). https://doi.org/10.1088/1757-899X/1122/1/012054
3. Shinde, K.: Literature review on electric bike. IJRMET. **7**(1) (2016−2017)
4. Niranjana, S.J., Patel, S.V., Dubey, A.K.: Design and analysis of vertical pressure vessels using ASME code and FEA technique. IOP Conf Ser, Mater. Sci. Eng. **376**(012135), 1–10 (2018). https://doi.org/10.1088/1757-899X/376/1/012135
5. Niranjana, S.J., Kubsad, S.S., Manjunatha, S., Nagaraj, Y., Bhavi, I., Angadi, B.M., Chamkha, A.J., Vanarotti, M.B.: Experimental investigation and numerical simulation of air circulation in a Non-AC bus coach system. Int. J. Eng. IJE TRANSACTIONS C: Asp. **35**(3), 572–579 (2022). https://doi.org/10.5829/ije.2022.35.03c.10
6. Shravanabelagola Jinachandra, N., Sadashivappa Kubsad, S., Sarpabhushana, M., Siddara-maiah, S., Raashekaraiah, T.: Modeling and computational fluid dynamic analysis on a non-AC bus coach system. Heat Transfer. 1–8 (2020). https://doi.org/10.1002/htj.21857
7. Kurhade, S., Kulkarni, S., Uplap, Y., Bhat, P., Ugale, A.: Design and analysis of hybrid E-bike chassis. Int. Res. J. Eng. Technol. (IRJET). **8**(5) (2021)
8. Salmeron-Manzano, E., Manzano-Agugliaro, F.: The electric bicycle: worldwide research trends. Energies **11**(7), 1894 (2018). https://doi.org/10.3390/en11071894
9. Sharma, Y., Banker, P., Chauhan, Y., Raikwar, Y., Sharma, M.: RD on electric bike. Int. Res. J. Eng. Technol. (IRJET). **5**(2) (2018)
10. Trivedi, M.M., Budhvani, M.K., Sapovadiya, K.M., Pansuriya, D.H., Ajudiya, C.D.: Design development of E-bike-a review. IRE J. **1**(5) (2017)
11. Katoch, S., Rahul, R.K.B.: Design and implementation of smart electric bike eco-friendly. Int. J. Innov. Technol. Explor. Eng. (IJITEE). **8**(6S4) (2019)
12. Malppan, G.J., Sunny, T.: A review on design developments in bicycle. Int. Res. J. Eng. Technol. (IRJET). **2**(3) (2015)
13. Shreekrushna1, S.Y., Ramchandra, L.V., Jalindar, B.Y., Mujawar, A.I.: Study, development & modifications of series hybrid electropetroleum bicycle. Int. Res. J. Eng. Technol. (IRJET). **5**(3) (2018)
14. Khande, M.S., Patil, M.A.S., Andhale, M.G.C., Shirsat, M.R.S.: Design and development of electric scooter. Int. Res. J. Eng. Technol. **7**(5) (2020)
15. Krishna Veni, M. N.V., Amareswari Reddy, M.: Conceptual design of bicycle frame. Int. J. Eng. Trends Technol. (IJETT). **38**(2) (2016)
16. Lin, C.-C., Huang, S.-J., Liu, C.-C.: Structural analysis and optimisation of bicycle frame designs. Adv. Mech. Eng. **9**(12), 1–10 (2017)

Chapter 17
Deep Learning-Based Classification of Cutaneous Viral Diseases: An Efficient Approach

Check for updates

Amit Gupta, Devansh Goel, Richa Gupta, and Navin Garg

Abstract Cutaneous viral diseases such as chickenpox and monkeypox are a significant public health concern, and accurate diagnosis and classification are crucial for effective treatment. On-time diagnosis, can not only lead to better treatment but can also help to limit the spread of the disease. With the same concern, this study emphasizes the image-based classification of three such diseases, namely, monkeypox, measles, and chicken pox, from the normal skin. For an effective classification of cutaneous viral diseases, CNN models have been employed. The suggested method entails initial image pre-processing, followed by the training of a convolutional neural network (CNN) using an extensive dataset of skin lesion images. The model's effectiveness is subsequently assessed on an isolated testing dataset. The methodology incorporates the implementation and comparative evaluation of two renowned CNN architectures, namely, VGG 16 and Inception V3, resulting in peak accuracies of 90% and 96%, respectively. Additionally, the fully connected layers are examined with four widely used activation functions: Linear, Tanh, Sigmoid, and ReLU.

Keywords CNN · VGG16 · Inception V3 · Linear · Tanh · Sigmoid · ReLU · Chicken Pox · Monkeypox · Measles

17.1 Introduction

With the advancements in healthcare, people have become more health-conscious. Recent outbreaks of diseases such as COVID-19 and Monkeypox have further encouraged people to exercise caution in managing their health. Early detection and diagnosis of diseases are crucial factors in this regard, which require knowledge of various disease symptoms. Many diseases can cause chronic or long-term effects on a person's appearance, ranging from red eyes to dull skin, yellowish skin, or spotty

A. Gupta (✉) · D. Goel · R. Gupta · N. Garg
CS&E Department, Graphic Era Hill University, Dehradun, India
e-mail: amitgupta7920@gmail.com

skin. These physical changes can act as a significant factor in disease detection and diagnosis for medical professionals. Skin, being the largest organ of the body, is an important indicator of our health, as it changes color or develops spots or lesions for various diseases such as chickenpox, monkeypox, and melanoma skin cancer. Hence, this study focuses on detecting and classifying cutaneous viral diseases.

Cutaneous viral diseases refer to viral infections that affect the skin. Examples of such diseases include Chickenpox, Herpes simplex, Human papillomavirus, Molluscum contagiosum, and Measles. The current study focuses on the detection and classification of three cutaneous viral diseases—Chickenpox, Monkeypox, and Measles—but provides a general approach that can be applied to other similar diseases as well. Chickenpox is a highly contagious disease caused by varicella-zoster virus (VZV) that leads to blister-like itchy rashes, fever, and headache. Monkeypox is another viral disease caused by the monkeypox virus that results in mucosal lesions, skin rashes, fever, headache, and muscle pain. Measles, a highly contagious viral disease, causes fever, red eyes, and skin rashes and is commonly seen in children.

Although all three of these diseases have an impact on the skin, they can be visually differentiated by closely observing the type of impact, pattern of rashes, and the affected area. However, many people, especially in rural areas, lack the knowledge to distinguish among these diseases. Consequently, a serious disease such as monkeypox could be underestimated and mistaken for a less serious disease like chickenpox, which can lead to delays in diagnosis and treatment. This is often the case with many cutaneous diseases, including something as serious as melanoma skin cancer. To address this issue, an effective solution is to automate the process of detection and classification of these diseases, which is the focus of this paper's proposed methodology.

The study employs convolutional neural networks to classify cutaneous viral diseases by analyzing images of patients' skin. The proposed methodology utilizes two widely recognized deep learning architectures, VGG 16 and Inception V3, and explores four distinct activation functions (Linear, Tanh, Sigmoid, and ReLU) for the fully connected learning stratum of the network. The primary aim of the research is to identify the most optimal blend of CNN frameworks and reliable activation functions for accurately categorizing visually similar yet distinct rashes into their respective disease categories. This optimal tuning is crucial in determining the model's performance, making it imperative to identify such a combination.

17.2 Literature Survey

Cutaneous viral illnesses are a major problem for society, ranging from relatively benign conditions like chickenpox to serious and potentially fatal conditions like skin cancer (melanoma). Therefore, a large number of medical researchers have sought to investigate and comprehend these illnesses. Recent developments in artificial intelligence and machine learning have led to significant developments in the diagnosis and treatment of various diseases through the partnership and cooperation of computer

science and medical sciences. This has drawn the attention of numerous researchers who are now concentrating on creating sophisticated methods for segmenting and classifying diseases. A survey of many strategies, ranging from conventional methods to cutting-edge deep learning techniques, is presented by Goswami et al. [1]. For comprehending picture features, convolutional neural networks (CNN) have shown to be among the most dependable and efficient technologies. Numerous researchers have employed convolutional neural networks (CNNs) to assess medical images because they have shown to be among the most dependable and efficient technologies for comprehending image information. For example, Inthiyaz et al. [2] offer a CNN-based method for extracting features from facial skin pictures and classifying them into six illness categories using the SoftMax classifier. Another approach by Srinivasu et al. [3], presents a model that uses MobileNet V2 with LSTM (Long Short Term Memory) to classify seven diseases. This approach achieves an accuracy of around 85%-90%. It also compares the performance of the model with other models, such as VGG 19, Fine-Tuned Neural Networks, and Mobile Net V2 (without LSTM).

In another study, Akyeramfo-Sam et al. [4] propose and demonstrate the use of CNN for the classification of three common skin diseases in Ghana, including atopic dermatitis, acne vulgaris, and scabies. The approach achieves high testing accuracies of 88%, 85%, and 84.7%, respectively, for each disease category. Like this, numerous CNN models have been proposed for the classification of different skin diseases. For instance, Maduranga and Nandasena [5] deal with the HAM10000 dataset, which comprises seven classes, and suggest the use of the MobileNet model with transfer learning. This approach achieves an accuracy of 85%. Similarly, Pai et al. [6] propose the use of fine-tuned VGG 16 with transfer learning for the classification of five classes, including Cellulitis Bacteria, Impetigo Bacteria, Ringworm Fungus, Sporotrichosis Fungus, and Healthy skin, achieving an accuracy of 86%. Moreover, Nallusamy et al. [7] present a CNN-based methodology for feature extraction, classification, and image segmentation of clinical images of skin disorders. Taking a step further, Khan et al. [8] focus on the segmentation and classification of skin lesions, particularly skin cancers. The study proposes an automated model that involves color correction, lesion segmentation using CNN, CNN-based feature extraction, selection of important features using moth flame optimization, and classification using Kernel Extreme Learning Machine Classifier. This approach achieves an average accuracy of 94.64%, having been tested on multiple datasets. This highlights the superiority of CNN models over traditional approaches, such as the one proposed by R D & A [9]. The approach involves the segmentation of skin lesions using U-Net but employs traditional methods such as feature extraction using the Local Binary Pattern (LBP) approach and Histogram of Oriented Gradients (HOG) approach, and utilizing a Support Vector Machine for classification. The approach achieves an accuracy of 85.19%, comparatively lower than the approach discussed before. The supremacy of CNN is also shown by Hosny et al. [10], which utilizes a simple AlexNet architecture to classify seven classes with an accuracy of 98.70%.

17.3 Proposed Work

The proposed research advocates the use of convolutional neural networks (CNNs) for detecting and classifying cutaneous viral diseases. CNNs comprise two main parts—the first part consists of convolutional and pooling layers which extract image features and create an image embedding, and the second part comprises fully connected layers that focus on learning the extracted features and classifying the input image accordingly. The efficiency and accuracy of the CNN model depend largely on these two parts. Choosing an effective CNN stratum for the first part and an appropriate activation function for the second part (dense layers) are the two critical factors that determine the network's performance as presented by Goel et al. [11]. The envisioned research endeavors to tackle the intricate task of identifying the most effective amalgamation of the CNN framework and activation function. To realize this objective, the study engages in a thorough comparison of two prominent convolutional neural networks, namely VGG 16 and Inception V3. For each of these models, four prevalent activation functions—ReLU, sigmoid, tanh, and linear—are incorporated, as illustrated in Fig. 17.1. The rationale behind selecting these architectures lies in their user-friendly interface, swift computational capabilities, and well-documented excellence in image classification tasks. Model performance is appraised using accuracy and the F1-score as discerning metrics.

The utilized architectures in the study make use of transfer learning by incorporating pre-trained weights of ImageNet Dataset, which results in faster and more accurate model training without affecting its purpose. Furthermore, the input images are pre-processed using techniques such as Gaussian blur and color correction to enhance the performance and classification results. Gaussian blur eliminates any unnecessary noise in the image, while color correction techniques like saturation improvement highlight the rashes on the skin. The overall flow of the model is shown in Fig. 17.2.

The research endeavors to conduct multiclass classification, delineating four classes to denote chickenpox, monkeypox, measles, and normal skin. The dataset under scrutiny comprises a total of 4,260 images, evenly distributed with 1,065
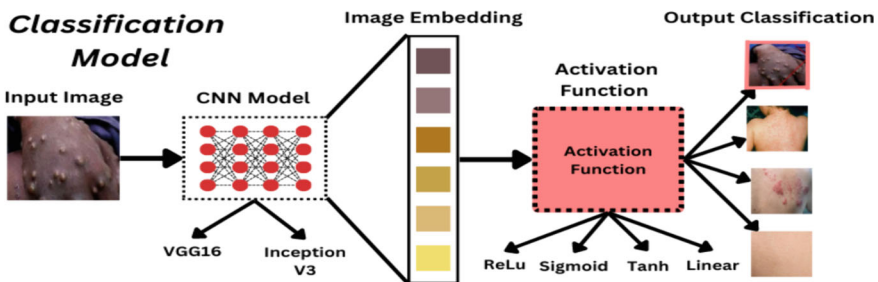
Fig. 17.1 Proposed method

Fig. 17.2 Flow of proposed model
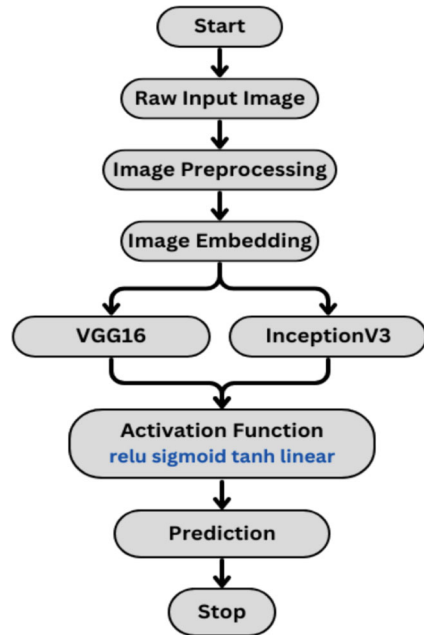
images allocated to each class. Following the conventional tradition, 70% dataset is used for training while the rest 30% is used as a testing dataset.

A. Dataset

The image dataset used in this study was originally downloaded from the Mendeley Data, Bala et al. [12]. It is a labeled dataset consisting of the following four classes:

- **Normal Skin**: It contains images of the skin of normal healthy people.
- **Monkeypox**: It consists of images of people's skin having monkeypox. It contains images of some different spots and rashes caused by monkeypox.
- **Measles**: It contains skin images of people with measles.
- **Chickenpox**: It consists of skin images with chickenpox rashes.

The dataset underwent augmentation through diverse techniques, including the introduction of noise and flipping. Furthermore, to enhance the model's performance, images depicting wounded skin and skin affected by common insect bites were incorporated into the normal skin class.

The final dataset used contains the above-mentioned four classes with 1.065 images in each, making a total of 4,260 images labeled dataset. Some samples of images from the dataset are shown in Fig. 17.3. The dataset can be downloaded from: https://drive.google.com/file/d/1GmtKNwbrkHf7gIpumpAIP0mz39 TawZCL/view?usp=share_link.

Samples from Dataset

a. Normal Skin   b. Monkeypox

c. Measles   d. Chickenpox
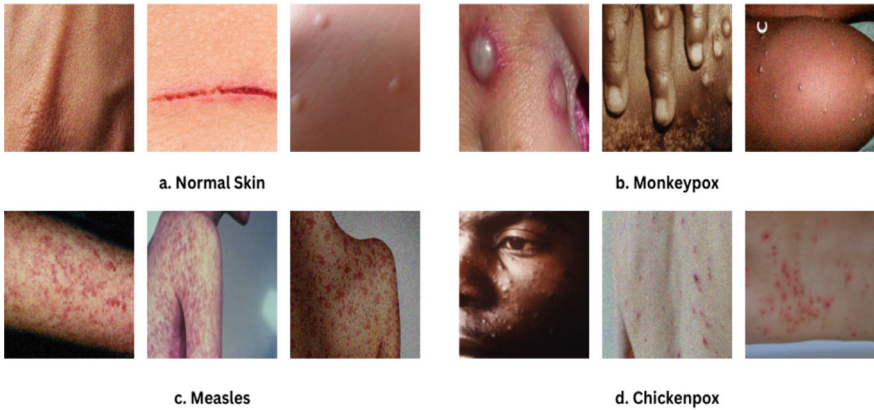
Fig. 17.3 Sample Images from Dataset

B. Methodology

The methodology in this study includes the following major steps:

- Image Pre-processing
- Feature Extraction through Image Embedding
- Learning features through fully connected layers using a suitable activation function
- Testing and evaluation of the model

C. Image Pre-Processing

The images were subjected to pre-processing techniques to enhance their visual quality before feeding them into the model. Gaussian blur and saturation improvement were the primary techniques used in the research. Gaussian blur was applied to eliminate any unnecessary noise present in the image, while saturation improvement resulted in better visualization of skin rashes and lesions, as shown in Fig. 17.4. Gaussian blur was performed before saturation improvement to avoid treating some rashes as noise. Although saturation improvement may reintroduce some noise, applying Gaussian blur beforehand ensures that the noise level remains the same or lower than before. This procedure effectively enhances the visibility of rashes while minimizing the impact of noise.

D. Image Embedding

Image embedding is a technique that involves representing the extracted features of an image as a low-dimensional vector. Convolutional layers serve the pivotal function of deriving the vital patterns and features from an input pixel matrix using small operation windows known as kernels, yielding a high-dimensional matrix of recognized features. However, this matrix is impractical for direct integration into fully connected

Fig. 17.4 Image
pre-processing
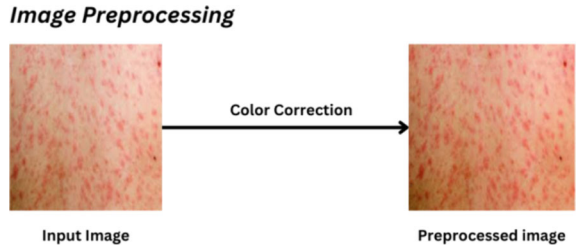
layers. Therefore, a crucial step involves condensing it into a low-dimensional vector, retaining the retrieved features—a process known as image embedding. The efficacy and richness of features extracted depend on the selected CNN architecture. Variations in the convolutional stratum are responsible for the varying quality of feature vectors. The given work employs two distinctive image embedding techniques or CNN architectures, each contributing unique attributes to the analysis.

- **VGG 16:** VGG 16, also known as VGGNet, is a CNN architecture with a total of 21 layers. Out of these 21 layers, 16 are trainable, consisting of 13 convolutional layers along with 3 fully connected learning layers. VGG 16 stands as a prominent architecture renowned for its adeptness in tasks related to object detection and classification. Its efficacy is underscored by an impressive demonstration, achieving a 92.7% accuracy in discerning 1000 images distributed across 1000 diverse classes within the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). The architectural blueprint of it is eloquently illustrated in Fig. 17.5.
- **Inception V3**: GoogLeNet, also known as Inception V3, is a CNN architecture that utilizes parallel layers to achieve high efficiency. It employs decomposition into smaller convolutions and further uses asymmetric convolutions, making it one of the fastest and most efficient networks available. The architecture of the Inception V3 is presented in Fig. 17.6.

  The image embeddings utilized in this study utilize the weights obtained from the ImageNet challenge through transfer learning. Transfer learning involves utilizing the weights for convolutional layers obtained from previous uses of the CNN model. The use of ImageNet weights is a widely adopted technique that saves a significant amount of training time and improves the efficiency of the model.
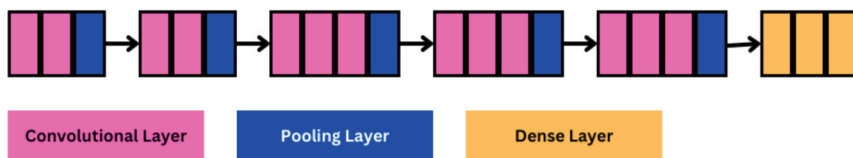
Fig. 17.5 Architecture of VGG16
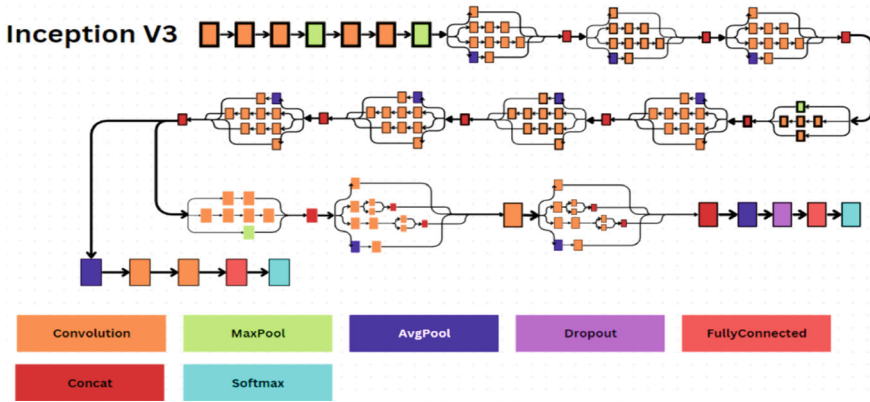
Fig. 17.6 Architecture of Inception V3

E. Selection of Activation Function

Activation functions are used to activate the neurons in the neural network. These functions are responsible for adding non-linearity to the network and deciding which neurons will be activated. Without activation functions, the network would behave like a simple linear equation. The use of activation functions helps the network to perform more complex tasks. The selection of an appropriate activation function for fully connected layers significantly impacts the network's ability to learn and the overall performance. The following four activation functions are used and tested in the proposed work:

- **ReLU function**: The Rectified Linear Unit (ReLU) function is a simple but powerful activation function popularly used in deep learning. The ReLU function outputs either a positive number or zero, whichever is greater, for the respective input. Although ReLU is computationally efficient, it can still encounter problems, like optimization with zero.

  The mathematical representation of the function is as follows:

$$ReLU(z) = max(0, z) \tag{17.1}$$

  where z is the argument to the max() function that gives z, if positive, otherwise a zero.

- **Sigmoid function**: This function is popular for scenarios involving binary classifications. As a non-linear transformation, maps the input values to an inclusive range of 0 to 1. Despite its ease of differentiability and optimization, the Sigmoid function is not immune to challenges, with the vanishing gradient being a notable issue. The mathematical representation of this function is articulated as follows:

$$Sigmoid(z) = \frac{1}{1 + e^{-z}} \tag{17.2}$$

where z represents the input value, and e denotes the mathematical constant with a value of approximately 2.718.

- **Tanh function:** The hyperbolic tangent function, commonly known as tanh, is an advancement of the sigmoid function and widens the output range from (0,1) to (-1,1). Despite this versatility, tanh also suffers from the vanishing gradient problem like the sigmoid. The mathematical expression for tanh is as follows:

$$TanH(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{17.3}$$

where z is the result of the neuron and needs to be mapped and e is Euler's number.

- **Linear function:** The linear function, often called the straight-line function, is a fundamental activation function that doot perform any operation on the input value, and simply returns it as output. It is simple and efficient which make it a good choice, but it fails in introducing important non-linearities to the network, limiting it from handling intricate tasks. However, the linear function can sometimes provide better results than others. Its symbolic representation is succinctly expressed as

$$Linear(z) = z = ax + c \tag{17.4}$$

where z represents the input value, representing the neuron's output. a and c represent the weights and biases of the neuron, respectively, while x denotes the inputs to the neuron.

17.4 Results and Discussion

The proposed work evaluates the trained models using four recognized metrics, that are accuracy, precision, recall, and F1-score. Accuracy presents the number of correct classifications made by the model and is one of the most important metrics for performance evaluation. Precision tells the number of correct positives with respect to total positive predictions while recall represents the number of positive predictions with respect to real positives. The F1-score combines both precision and recall, giving a more precise measure of the model's effectiveness. These metrics are mathematically represented as follows:

$$Accuracy = \frac{Pos_{true} + Neg_{true}}{Pos_{true} + Neg_{true} + Pos_{false} + Neg_{false}} \tag{17.5}$$

$$Precision = \frac{Pos_{true}}{Pos_{true} + Pos_{false}} \tag{17.6}$$

$$Recall = \frac{Pos_{true}}{Pos_{true} + Neg_{false}} \tag{17.7}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{17.8}$$

where

- Pos_{true} refers to True Positive, the number of correct classifications made for the class.
- Pos_{false} refers to False Positive, the number of images that belong to other classes but are classified as the class, for which FP is calculated.
- Neg_{true} refers to True Negative, the number of images correctly classified into other classes.
- Neg_{false} refers to False Negative, the number of images that belong to the class but are classified wrongly into other classes.

The proposed work's results are presented in Table 17.1, indicating a tough competition between the models. However, Inception V3 outperforms VGG16by 96%. In terms of activation functions, all of them performed well, except for tanh, which had varying performance with different CNN architectures. While ReLU is generally considered the best, sigmoid and linear also provide high accuracies. Nonetheless, ReLU's theoretical reliability makes it the preferred choice. The confusion matrices of ReLU with VGG 16 and Inception V3 are provided in Fig. 17.7. The ROC curves for the same are presented in Fig. 17.8. Samples of correctly classified images and wrongly classified images are presented in Fig. 17.9a and b, respectively.

Table 17.1 Comparison of models

Embedding	Activation function	Accuracy	F1-score	Precision	Recall
VGG16	**ReLU**	**90%**	**0.89**	**0.89**	**0.89**
	Sigmoid	90%	0.90	0.90	0.90
	Tanh	83%	0.83	0.83	0.82
	Linear	89%	0.89	0.89	0.89
InceptionV3	**ReLU**	**96%**	**0.95**	**0.95**	**0.95**
	Sigmoid	95%	0.94	0.94	0.95
	Tanh	95%	0.95	0.95	0.95
	Linear	95%	0.94	0.94	0.94

Confusion Matrix
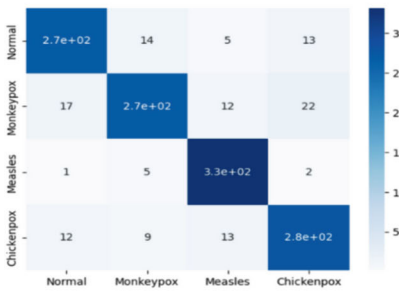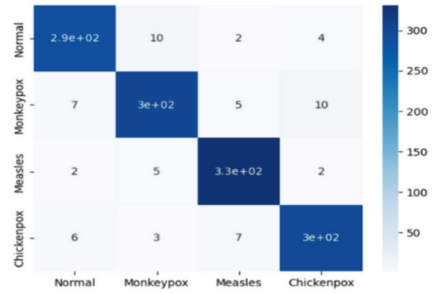
a. VGG16 with ReLU   b. Inception V3 with ReLU

Fig. 17.7 Confusion matrices

ROC Curves
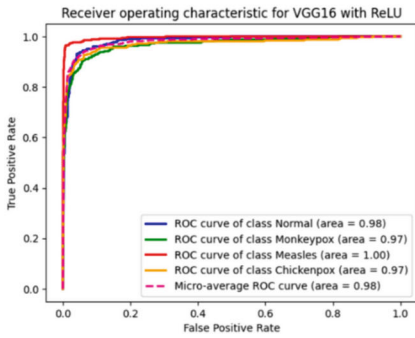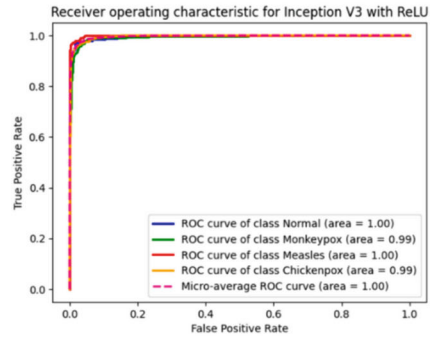
a. VGG16 with ReLU   b. Inception V3 with

Fig. 17.8 ROC curves
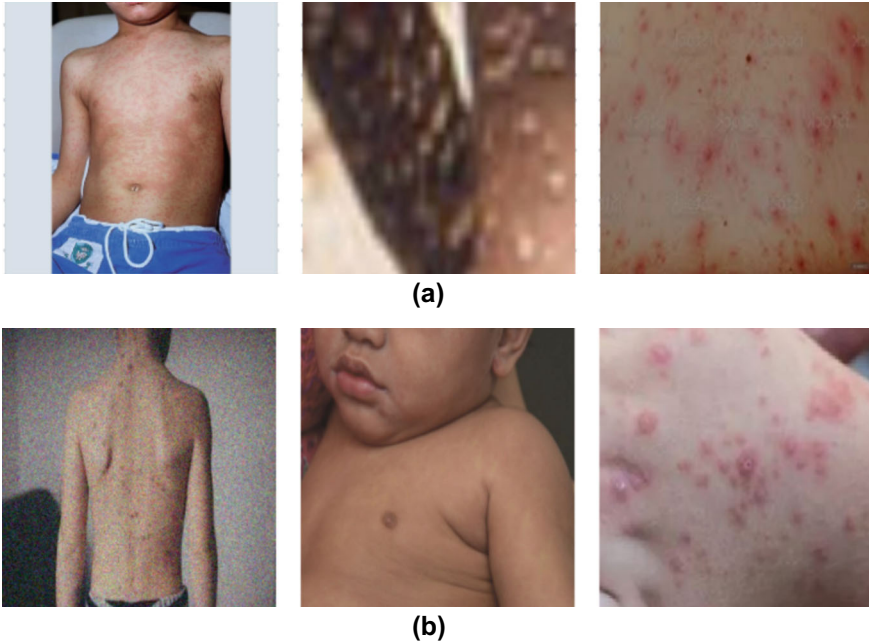
Fig. 17.9 **a** Some correctly classified images. **b** Some wrongly classified images

17.5 Conclusion

This paper introduces a method for classifying cutaneous viral diseases by selecting an optimal blend of CNN architecture and activation function. The proposed approach's effectiveness is demonstrated by accurately classifying skin images, and it provides a foundation for constructing deep learning models for various cutaneous diseases, not just limited to those used in the study. By selecting the optimal combination of image embedding and activation function, diseases such as monkeypox and melanoma skin cancer can be effectively diagnosed using such models. Applications of this work can also be extended to the other areas of disease diagnosis using CNN such as disease detection in rice, as shown by Sharma et al. [13], classification of eye diseases as presented by Gupta et al. [14] and diseases in cattle as given by Gupta et al. [15]. In conclusion, the study suggests the use of optimal combinations like Inception V3 with ReLU.

References

1. Goswami, T., Dabhi, V.K., Prajapati, H.B.: Skin Disease Classification from Image - A Survey. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (2020). https://doi.org/10.1109/icaccs48705.2020.9074232

2. Inthiyaz, S., Altahan, B.R., Ahammad, S.H., Rajesh, V., Kalangi, R.R., Smirani, L.K., Hossain, M.A., Rashed, A.N.Z.: Skin disease detection using deep learning. Adv. Engin. Softw. **175**, 103361 (2023). https://doi.org/10.1016/j.advengsoft.2022.103361

3. Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J.: Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors **21**(8), 2852 (2021). https://doi.org/10.3390/s21082852

4. Akyeramfo-Sam, S., Addo Philip, A., Yeboah, D., Nartey, N.C., Kofi Nti, I.: A web-based skin disease diagnosis using convolutional neural networks. Int. J. Inf. Technol. Comput. Sci. **11**(11), 54–60 (2019). https://doi.org/10.5815/ijitcs.2019.11.06

5. Maduranga, M., Nandasena, D.: Mobile-based skin disease diagnosis system using convolutional neural networks (CNN). Int. J. Image, Graph. Signal Process. **14**(3), 47–57 (2022). https://doi.org/10.5815/ijigsp.2022.03.05

6. Pai, V.R., Pai, S.G., Suhasi, P.M., Rekha, P.M.: Identification and classification of skin diseases using deep learning techniques (2023). https://doi.org/10.21203/rs.3.rs-2628782/v1

7. Nallusamy, C., Suriya, M., Vidhyaa Sagar, G., Skin, P.T.: Disease detection based on deep learning. J. Populat. Therapeut. Clin. Pharmacol. **30**(6) (2023). https://doi.org/10.47750/jptcp.2023.30.06.045

8. Khan, M.A., Sharif, M., Akram, T., Damaševičius, R., Maskeliūnas, R.: Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. Diagnostics **11**(5), 811 (2021). https://doi.org/10.3390/diagnostics11050811

9. Seeja, R.D., Suresh, A.: Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). Asian Pacific J. Cancer Prevent. **20**(5), 1555–1561 (2019). https://doi.org/10.31557/apjcp.2019.20.5.1555

10. Hosny, K.M., Kassem, M.A., Fouad, M.M.: Classification of skin lesions into seven classes using transfer learning with AlexNet. J. Digit. Imaging **33**(5), 1325–1334 (2020). https://doi.org/10.1007/s10278-020-00371-9

11. Goel, D., Singh, D., Gupta, A., Yadav, S.P., Sharma, M.: An efficient approach for to predict the quality of apple through its appearance. In: 2023 International Conference on Computer, Electronics and Electrical Engineering & Their Applications (IC2E3) (2023). https://doi.org/10.1109/ic2e357697.2023.10262569

12. Bala, D., Hossain, Shamim, M.d.: Monkeypox skin images dataset (MSID). Mendeley Data **V6** (2023). https://doi.org/10.17632/r9bfpnvyxr.6

13. Sharma, R., Kukreja, V., Kadyan, V.: Rice disease detection using convolutional neural networks: a survey. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (2021). https://doi.org/10.1109/icacite51222.2021.9404620

14. Gupta, R., Tripathi, V., Gupta, A.: An efficient model for detection and classification of internal eye diseases using deep learning. In: 2021 International Conference on Computational Performance Evaluation (ComPE) (2021). https://doi.org/10.1109/compe53109.2021.9752188

15. Gupta, A., Singh, D., Gupta, R., Tripathi, V.: Revolutionizing cattle health: a machine learning approach to efficiently predict lumpy disease in cows. In: 2023 4th IEEE Global Conference for Advancement in Technology (GCAT) (2023). https://doi.org/10.1109/gcat59970.2023.10353420

Chapter 18
A Review on ElectricVehicle Developments and Battery Management Improvements

Gadupudi Lakshminarayana [ID]**, J. Viswanatha Rao** [ID]**, and Ravi Kumar Avvari** [ID]

Abstract Everybody is interested in green energy technologies, which are what brought the idea of electric vehicles into existence. If the problem of battery management is carefully watched, electric vehicles will function more effectively. Thus, the subject of battery management is brought up. There are many various kinds of batteries, but lead acid batteries are the most common form used today. Because nonlinear controllers are more frequently utilized, system output will be more precise and accurate in accordance with input-driven data. Again, under this circumstance, brushless and brushed DC motors can both be used. So, the motors in accordance with the various uses are chosen. Since mezzanine technologies are one of the most popular techniques today, users can further classify it into AI controllers, Machine Learning techniques, IoT-based approach, and Deep Learning technology, all of which can help to improve electric vehicle innovation approaches. To assure Electric Vehicle driver safety by using these mezzanine technologies to regulate the vehicle's battery.

Keywords Electric vehicles · Battery management · AI controllers · Fuzzy logic

18.1 Introduction

Artificial intelligence refers to the ability of a machine to perform cognitive functions, i.e., "perception, learning, reasoning, and problem-solving that are similar to those performed by humans. In the past, artificial intelligence (AI) was used to describe computers that mimic and display human cognitive characteristics associated with the human mind, such as learning and problem-solving. Major AI researchers have

G. Lakshminarayana (✉) · J. V. Rao · R. K. Avvari
Department of Electrical & Electronics Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana 500090, India
e-mail: lakshminarayana_g@vnrvjiet.in

now disproved this idea and are explaining AI on grounds of rationality and behaving rationally, which does not impose limitations on the expression of intelligence". To govern battery management, we employ the idea of artificial intelligence. An AI-based controller called NARMA-L2 automatically modifies management issues for parameters like state of charge, depth of discharge, and many other voltage and current characteristics. Without any assistance from a human, this controller automatically transforms non-linear impulses into linear ones. By doing this, the battery's performance can be enhanced under various conditions, such as depth of drain and state of charge. Due to the advantages of electric vehicles, which include their lack of noise, pollution, odor, and vibration, this issue has been chosen because they offer the greatest and most practical solution in every industry. By doing so, we may explore the battery management problem, address the issues that come from the various situations when human involvement is necessary, and attempt to make improvements utilizing artificial intelligence (AI) methods, which allow for automatic control. The AI controller NARMA-L2 aids in monitoring and amending parameter management difficulties.

Figure 18.1 blocks the explanation as follows: Renewable energy is obtained from natural resources that refill quickly even though they are used up. Examples of such sources are the sun and the breeze. A lot of renewable resources are present. A bidirectional DC converter, also referred to as a "dc chopper", is a device that converts a "steady dc voltage source into a variable dc voltage". Examples like "Trolley trains, battery-powered vehicles, traction motor control, large-scale dc motor control", etc. To ensure the safety of electric vehicles, a battery management system, which controls the circuit of rechargeable battery packs, is essential. Ensuring the cell works in safe operating parameters, it protects both the user and the battery. BMS monitors the battery's condition, collects information, controls outside influences, and balances them to keep cell voltages consistent. Carburetor controls a car's speed and acceleration when it is running on gasoline, and an electronic device that coordinates an electric car's batteries and engine is known as the electric vehicle controller. For use with AC motors only, the controller switches the dc from battery to ac and controls the battery's energy supply.
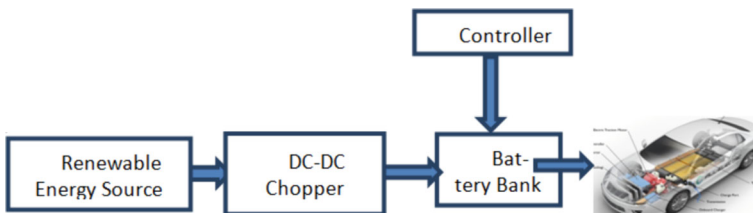
Fig. 18.1　Block diagram

18.2   Developments on Electric Vehicles

The vehicles were renowned for having quiet journeys and simple controls, unlike their steam or gasoline rivals. Electric vehicles became popular and in high demand as a result of people choosing them over early gasoline cars that required numerous switches to change gears, which was hazardous and taxing. A final created model of an electric vehicle must undergo certain modifications. It is shown how two different vehicle topologies as parallel and series hybrid electric vehicles were modeled and simulated using "Advanced Vehicle Simulator", which assesses the vehicle's performance. For a precise examination of the vehicle's performance, a dynamic prototype which incorporates the "electric motor, batteries, and motor controller" is required. The "I/O Board, Adapter, and Microcontroller" for the motor make up the central controller. The electric vehicle battery energy management is optimized in terms of peak utilization based on aging scheduling with empirical solutions [1]. The EV system benefits are established for a long life cycle by balancing the battery working conditions with intelligent soft computing through data computations [2]. Electric Vehicle battery management is analyzed with cell characteristic behavior upgradation to control the energy loss for engineering applications [3]. The state of battery levels is soft computing to predict the performance of working conditions and improvisation of battery health through artificial intelligence techniques [4]. The vehicle energy consumption is estimated through a fuzzy logic system for commercial applications and minimization of carbon emissions toward sustainable growth [5]. The voltage levels, temperature variations, and internal operations are evaluated and identify the faulty situation in battery management [6]. The fuzzy-based converter technique is to fix the constant voltage values throughout the life duration of systems for many commercial applications [7]. The Hybrid Electric Vehicle fuel management and energy assessment are controlled by neuro fuzzy systems for accurate power distributions [8]. The economic viability of hybrid fuel consumption is analyzed and estimated toward payback time battery energy based on internal parameters of energy utilization commitment [9]. This paper describes the significance of electing an ampere hour according to battery life conditions in hybrid EV for future embodiments [10]. The battery characteristics are examined during driving situations and balancing the charging times as well as structural battery behavior conditions [11]. In this paper, the efficient analysis has taken place about EV and HEV on world warming conditions and suggests the predominant directions toward control of environment emissions [12]. The charging and discharging times of batteries are investigated and identify the faulty conditions on the way to a healthy atmosphere [13]. The driving system is controlled based on battery power distribution with intelligent controller techniques for better long-run drive models [14]. Additionally, there is a map of numerous engines and electric motors in Advisor, which allows for real-time simulation with results that are fairly accurate [15]. The lifespan of batteries is increasing and controlling the voltage level to the ac systems with minimization of losses [16]. The Chevrolet Volt & OBD2 ports on the driver & and passenger & sides of the car allow access to a significant amount of data that is available on the CAN bus of the

car. The majority of this information is not accessible to passengers on the vehicle console, but it can be recorded and gathered to examine charging characteristics as well as various vehicle performance measurements and behaviors. An essential set of information was acquired simultaneously for roughly 40 automobile metrics during a number of trips and recharge activity involving the "2012 Chevrolet Volt". Several of the most important metrics are "vehicle speed, engine rpm, battery voltage, battery current, motor currents, state of charge, longitude, latitude, altitude, battery and motor power, among others". In classes on advanced machines and sustainable energy, senior electrical engineering students have used this knowledge as a teaching tool [17]. Energy saving has grown to be a top priority for today's modern sustainability. To switch to renewable energies and to improve energy efficiency, the current energy sectors, however, must overcome substantial obstacles. In research, problems are addressed by providing a software-hardware platform. To gain fast speed and low power, the platform integrates the benefits of SW and HW. Automatically, the batteries of electric vehicles are charged or discharged. As a result, the VPP can actively handle the dispersed EV batteries as a sizable, intelligent energy storage building. Thus, the extra electricity gained by renewable resources like PV panels, wind turbines, and other similar facilities can be stored by EV batteries. As a result, the power infrastructure and networks for electric vehicles will be more adaptable and expandable [18].

Artificial intelligence refers to intelligence or the capacity to comprehend and resolve issues that can be created and placed inside a computer. It can help us maintain a connection with our vehicles as well as forge connections with other vehicles on the road, facilitate routine maintenance, encourage economical driving, or ensure passenger safety. One of the most recent trends is the adoption of connected automobiles which are AI-based and learn by observing their surroundings and modifying user behavior. For electric vehicles, the energy optimum route is one that maximizes energy recovery or one that passes by charging stations with the lowest charging costs. The ensuing stage in smart mobility is the development of energy sources that are sustainable and power the automobile industry's transition to self-driving electric vehicles. AEVs operate on sensors, and autonomous vehicles that interact with their environments using artificial intelligence (AI). In order to reduce their carbon footprints, AEVs can exchange energy including other "AEVs, smart grids, and roadside electronics" andsuggest for eco-friendly transportation. In IOV ecosystems, AEVs are significant for promoting energy exchange between AEVs and RSUs, which minimizes carbon footprints and ensures a sustainable and environmentally friendly energy supply. The network-linked AEVs' sensor units are connected by wireless infrastructures. Additionally, as sensors interact via open channels and gather factual data, they are vulnerable to network-based threats from AEVs that have been maliciously brought into the ecosystem. [19]. The electrification of mass transit is often commended as a way to lower gag pollution as well as reliance on non-renewable energy sources. The electrification of mass transit is praised as a means of lowering emissions of greenhouse gasses and dependence on non-renewable energy sources. This study analyzes and especially spotlights the financial stake in the applications of artificial intelligence including the linked domains of EVs, EVCSs, and indeed

the engagement of EVs with a smart grid. Accuracy and the function of AI in the intelligent management of EV hardware and support systems to preserve battery life [20]. They have reviewed current issues with inverters and motors and provided solutions to make EVs more effective and economical in the future. WBG (Wide Bandgap Semiconductors)-based inverters and their significance were discussed [21]. To ensure active protection and energy conservation, a torque distribution strategy based on deep reinforcement learning is suggested. This suggested approach lowers energy loss by 5.25%–10.51% and increases average motor efficiency [22]. Recent advancements in EV technology and their effects on the economy and environment are reviewed. A quick note on effective EV charging management is provided, along with an introduction to several charger converter topologies, battery technology, and powertrain combinations. EV deployment technologies have been introduced, including possible opportunities for the smart grid and vehicle to grid technology [23]. To manage the power in the grid, the author recommended a system known as "AI-enabled blockchain-based electric vehicle integration system". This system ensures reliable forecasts of power consumption and also generates up-to-date utility that uses the vehicle connectivity to deliver additional electricity and lower power imbalance [24].

This paper describes the various converter methods that are reviewed and identifies the application-oriented converter control approaches [25]. To decrease drive shaft torque variance and enhance shift quality, a control technique is suggested. Unwanted transition torque is prevented with this approach. It has been noted that the inertia torque multiplication of MGS causes variance in the transition torque. To lessen the torque variance in the drive shaft, this management technique considers the MG torque compensating during mode shift plays a crucial part in determining how to advance the current status of system's through the use of several databases [26]. To gain a proper understanding of battery technologies, electrical machines, and the designs of EVs and HEVs, roughly 400 vehicles and a lot more are tested [27]. They simulated the resulting mathematical model in such a way that the outcome shows them the analysis of fuel and battery balance at the end, and this result is also contrasted practically on roads [28]. The vibrant performance and fuel economy of parallel hybrid electric vehicle theories are put into account. The advancement of automotive technology has drastically swayed the development of eco-friendly nature and environmental technologies [29]. By managing and gathering the vehicle's charging, performance characteristics also by integrating with AI numerous amounts of developments on electric vehicles can be improved. A controlled area network bus is used to make communication between vehicles such that the vehicles information can be gathered which is used to improve the efficiency of vehicles.

18.3 Electric Vehicle Control Methodologies

Controller depicts the efficient functioning of the electric vehicle that helps in obtaining maximum torque and acceleration of a vehicle. A lot of control methodologies are being used for better performance of vehicles. Electric vehicles must have some or other control strategies such that our system works efficiently without any disturbances. Now let us discuss in brief about some of the methods through which we can develop the electric vehicle battery management system. The challenges with power conversion and the concept of spinning machines are explained. The author employed a power electronics-based strategy to increase the fuel efficiency, speed, and charging time of the hybrid or inserted electric vehicle [30]. A non-stationary vehicle's speed can be increased by utilizing a variety of strategies, including "proportional-integral-derivative, pole placement controller, linear quadratic regulator, and state observer-based controller [31]". This paper implements the variable load motor operations with smooth and smart control for dynamic speed conditions [32] and explains the NARMA-L2 controller which is used to convert the non-linearity of the system to a linear system and improves system dynamics [33]. Author is using ZETA converter for getting maximum power from the Electric Vehicle motor, by using this the supply voltage is improved, here ZETA converter is working on the principle of fuzzy logic using MATLAB/Simulink [34]. Works on brushless DC motor development and its challenges using techniques related to modified-proportional-integral controllers for speed regulation [35]. Because the dc charger is no longer internal to the machinery, the engineers of the vehicle must use an external power source, the complexity of the position of "the ac to dc conversion equipment, or converter, influences the complexity of equipment design" [36].

   In this study, a hybrid electric vehicle's fuzzy logic processor is recommended. For vehicles to be driven in parallel, a mixture of an internal combustion engine and an electric motor is required. The battery's charge level and the necessary torque for driving will be used to construct the fuzzy logic controller. The internal combustion engine throttle angle degree is actually controlled by the controller output in order to work in a large area. The motor is designed to satisfy extreme power demands for vehicle loads and functions as an additional source of power. The simulation results show how well internal combustion engines and electric motors may be powered in an efficient manner when using the optimal control technique [37]. Through computation, it is demonstrable that the fuzzy logic technique seems to be suitable for hybrid electric vehicles and that using a fuzzy logic controller to enhance driving performance is beneficial. Additionally, it makes it possible for hybrid vehicles to operate with minimal emissions and great fuel efficiency.

18.4 Battery Management System Improvements

The development of the system for battery management is the first important step in improving the functionality of an electric vehicle. Electric vehicle's anxiety about range is growing. It was suggested that in order to do this, one should first use AI to implement a fast-EVCS revenue maximization pricing scheme. The concept is first recommended to be investigated using a simulation model. This simulation was developed using measurements from vehicle performance models and makes use of the charging station game from EVU. A Markov judgment is used to represent the solution to the problem. Using the reinforcement-learning algorithm, a dynamic pricing model that is tailored to each customer is created. Compared to the stationary pricing strategy, reinforcement learning has a stronger advantage because it uses the past feedback result as an input for subsequent actions [38] with increased fuel-cell-based hybrid electric vehicles' efficiency. Using fuzzy logic, a feasibility analysis for fuel reduction is conducted. This study examines the efficiency of HEVs by environmental monitoring and the creation of nine different environmental scenarios. The system keeps track of the battery, gasoline tank, and fuzzy logic controller capacity. These act as performance indicators that make it easier to comprehend how a controller with fuzzy logic affects fuel efficiency. This research suggests a method for optimizing HEVs ensuring that they choose a more cost- and environmentally friendly strategy. For every million HEVs equipped with fuzzy logic controllers, this simulation was determined to save 464 tons of CO_2 emissions and 3.3 million Hong Kong dollars [39]. The power control of transmission systems is established and the integrated load conditions are examined with multilevel voltage converters [40]. A quick note on parameters needed for BMS, including "ambient temperature, current, voltage, and controlling the charging and discharging activities", is given along with a discussion of various battery management system technologies. It was concluded that there is still room for research in this field [41]. The significance of vehicle electrification in the future was discussed. Classification of various electric vehicle models and an explanation of how inverters and converters work in EVs [42]. Extended Kalman Filter (EKF) algorithm and Thevenin battery model paired with MATLAB simulation were suggested as a technique for a better battery management system. With the suggested system, Soc is determined with an error rate of less than 5%. It uses lithium-iron phosphate batteries. Hardware experiments are taken out, the outcomes are compared, to confirm the proposed performance of the system [43].

In order to meet the requirements, a system is created with a BMU and LECU based on the C8051F040 microcontroller. These units collect battery parameters such as "voltage and current, temperature, SoC, and fault diagnosis" based on battery status [44]. It was quickly explained how to use data collection terminals, GPS + GPRS intelligent communication modules, GPRS + internet public, and data processing centers to monitor and manage batteries in a battery management system intelligently. Through the wireless GSM/GPRS communication, it also ascertains the fault status, cause, and remedy [45]. It indicates postulated fuzzy energy management systems. This system employs BWS (Battery Working State) to stop the battery

from being excessively discharged and erroneous SOC estimates. This technology proved it was possible to keep a HEV plug-in series engine operating within its fuel-economic range. In this battery operating state concept-related challenge, fuzzy logic analysis is used to attempt to resolve the state of charge and battery abrupt discharge problem in hybrid electric vehicles [46]. The voltage control is observed for various discontinuous load approaches and provides the continuity of stable operation to the system [47]. An innovative adaptive system for billing and charging for transportation gridlock, which the author put forth in the framework of a metropolis that is smart. With this strategy, which also handles various traffic-related issues, the overall charging time at charging stations is decreased [48]. Author tends to improve the scanty autonomy of the vehicle based on a fuzzy logic concept, through which the speed and acceleration parameters of electric vehicles are developed with improved safe power management and SOC management [49].

The primary objective of this is to create a mathematical simulation that is connected to the Internet of Things and considers concerns with energy management, battery technology advancements, and autonomous features [50]. The mathematical methods used to approach the solution include "Cox, Auto-Regressive-Moving-Average (ARMA), Trend, and Seasonality (TBATS), Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) Modeling" [51]. These methods work on the various factors that relate to battery management, forecasting related issues, and state of charge. It explains many AI technologies connected to blockchain technology that are used to bill roving electric vehicles after charging [52]. Automakers, governments, and consumers are becoming increasingly interested in vehicles due to escalating environmental issues. Researchers urgently need to create cutting-edge electric-drive systems because both EVs and HEVs are built on motors. [53]. This paper provides a summary of the "permanent-magnet brushless drives for EVs and HEVs with a priority on machine topologies, drive operations, and control systems". As discussed above managing and improving the battery parameters helps in efficient working of vehicles.

18.5 Conclusion

Finally, it has been observed that various methods utilized for battery management using the new AI-based controllers, as well as various ways in which our issue might be appropriately managed. The subject of electric vehicles is not new; it has been discussed since the late nineteenth century. Different kings who were in power at the time designed it for travel, but there was no thorough understanding of batteries at that time. But now that a complete understanding of the battery and access to a variety of resources is present, models are being improved effectively. Additionally, information regarding what battery management actually is and how crucial it is to a vehicle's safety and more effective operation has been discussed. Various motor types and how they relate to electric vehicles has been understood and knowledge

about the many battery technologies that are currently in use and widely available has been gained. This enables the users to create an efficient electric vehicle powered by renewable energy.

The AI controller, Machine Learning techniques, IOT-based approach, and Deep Learning technology can identify and report on equipment irregularities in real time. This not only makes it possible to observe changes at the machine level, but it also eliminates quality flaws that rapidly develop on high-speed production lines. Nowadays emerging technology has been an emerging technology in many areas. These advanced methodologies used in battery management help in understanding the state of the battery, its health, and temperature. To monitor the battery of Electric Vehicles to ensure safe driving of vehicles. In this paper emerging technologies are described for managing batteries and different parameters regarding batteries are explored. The emerging technologies are future trends of electric vehicle innovations and their sustainable working operations.

References

1. Li, S., Zhao, P., Gu, C., Huo, D., Li, J., Cheng, S.: Linearizing battery degradation for health-aware vehicle energy management. IEEE Trans. Power Syst., **38**(5), pp. 4890–4899 (2023)
2. Singh, S., More, V., Batheri, R.: Driving electric vehicles into the future with battery management systems. IEEE Eng. Manag. Rev., **50**(3), pp.157–161 (2022)
3. Li, S., Zhao, P., Gu, C., Li, J., Huo, D., Cheng S.: Aging mitigation for battery energy storage system in electric vehicles. IEEE Trans. Smart Grid, **14**(3), pp. 2152–2163 (2023)
4. Jafari, S., Byun, Y. -C.: Prediction of the battery state using the digital twin framework based on the battery management system. IEEE Access, **10**, pp. 124685–96 (2022)
5. Phan, D., Bab-Hadiashar, A., Fayyazi, M., Hoseinnezhad, R., Jazar, R. N., Khayyam H.: Interval type 2 fuzzy logic control for energy management of hybrid electric autonomous vehicles. IEEE Trans. Intell. Veh., **6**(2), pp. 210–220 (2021)
6. Gao, W., Li, X., Ma, M., Fu, Y., Jiang, J., Mi, C.: Case study of an electric vehicle battery thermal runaway and online internal short-circuit detection. IEEE Trans. Power Electron., **36**(3), pp. 2452–2455 (2021)
7. LakshminarayanaGadupudi, GudapatiSambasiva Rao, RachakondaVenkata Lakshmi Narayana Divakar, HasmatMalik, Faisal Alsaif, Sager Alsulamy, TahaSelimUstun.: Fuzzy-based Fifteen-level VSC for STATCOM operations with single DC-Link voltage. Sustain., **15**(7), p. 6188 (2023)
8. Suhail, M., Akhtar, I., Kirmani, S., Jameel, M.: Development of progressive fuzzy logic and ANFIS control for energy management of Plug-In hybrid electric vehicle. IEEE Access, **9**, pp. 62219–62231 (2021)
9. Anselma, P.G., et al.: Economic payback time of battery pack replacement for hybrid and Plug-In hybrid electric vehicles. IEEE Trans. Transp. Electrification **9**(1), 1021–1033 (2023)
10. Zhou, B., Burl, J. B., Rezaei, A.: Equivalent consumption minimization strategy with consideration of battery aging for parallel hybrid electric vehicles. IEEE Access, **8**, pp. 204770–204781 (2020)
11. Hamednia, A., Murgovski, N., Fredriksson, J., Forsman, J., Pourabdollah, M., Larsson, V.: Optimal thermal management, charging, and Eco-Driving of battery electric vehicles. IEEE Trans. Veh. Technol., **72**(6), pp. 7265–7278 (2023)
12. Gautam, A. K., Tariq, M., Pandey, J. P., Verma K. S., Urooj, S.: Hybrid sources powered electric vehicle configuration and integrated optimal power management strategy. IEEE Access, **10**, pp. 121684–121711, (2022)

13. Gan, N., Sun, Z., Zhang, Z., Xu, S., Liu, P., Qin, Z.: Data-Driven fault diagnosis of Lithium-Ion battery over discharge in electric vehicles. IEEE Trans. Power Electron., **37**(4), pp. 4575–4588, (2022)

14. Abdelaal, A. S., Mukhopadhyay, S., Rehman, H.: Battery energy management techniques for an electric vehicle traction system. IEEE Access, **10**, pp. 84015–84037, (2022)

15. Saleki, A., Rezazade, S., Changizian M.: Analysis and simulation of hybrid electric vehicles for sedan vehicles. 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, pp. 1412–1416 (2017)

16. Narayana Gadupudi, L., Rao, G. S., Devarapalli, R., GarcíaMárquez, F. P.: Seven level voltage source converter based static synchronous compensator with a constant DC-Link voltage. Appl. Sci., **11**(16), p. 7330(2021).

17. Dolan, D. S. L., Taufik, T., Dunning, J.: Analysis of detailed electric vehicle data in electrical engineering education. World Electr. Veh. Symp. Exhib. (EVS27), Barcelona, pp. 1–5(2013).

18. Huang, H., Ogbodo, M., Wang, Z., Qiu, C., Hisada, M., Abdallah, A. B.: Smart energy management system based on reconfigurable AI chip and electrical vehicles. IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea (South), pp. 233–238 (2021)

19. Dixit, P., Bhattacharya, P., Tanwar, S., Gupta, R.: Anomaly detection in autonomous electric vehicles using AI techniques: A comprehensive survey. Expert. Syst., **39**(5), p. e12754 (2022)

20. Ahmed, M., Zheng, Y., Amine, A., Fathiannasab, H., Chen, Z.: The role of artificial intelligence in the mass adoption of electric vehicles. Joule, **5**(9), pp. 2296–2322 (2021)

21. Husain, I., et al.: Electric drive technology trends, challenges, and opportunities for future electric vehicles. Proc. IEEE **109**(6), 1039–1059 (2021)

22. Wei, H., et al.: Deep reinforcement learning based direct torque control strategy for distributed drive electric vehicles considering active safety and energy saving performance. Energy **238**, 121725 (2022)

23. Yong, S. J. Y., Ramachandaramurthy, V. K., Tan, K. M., Mithulananthan, N.: A review on the state-of-the-art technologies of electric vehicles, their impacts and prospects. Renew. Sustain. Energy Rev., **49**, pp. 365–385, (2015)

24. Blockchain Power Trading and Energy Management Platform: Lin, Y. -J., Chen, Y. -C., Zheng, J. -Y., Chu, D., Shao, D. -W., Yang. H.-T. IEEE Access **10**, 75932–75948 (2022)

25. Narayana Gadupudi, L, GudapatiSambasiva Rao.: Recent advances of STATCOM in power transmission Lines—A review. Turk. J. Comput. Math. Educ.. **12**(3), pp: 4621–4626 (2021)

26. Ma, Y., Huang, K., Xiang, C., Wang, W., Liu, H.: A control strategy to reduce torque variation for dual-mode power-split hybrid electric vehicles during mode shift. 7th International Conference on Modelling, Identification and Control (ICMIC), Sousse, Tunisia, pp. 1–8 (2015)

27. Frieske, B., Kloetzke, M., Mauser F.: Trends in vehicle concept and key technology development for hybrid and battery electric vehicles. World Electric Vehicle Symposium and Exhibition (EVS27), Barcelona, Spain, pp. 1–12 (2013)

28. Bai Zhonghao, Wang Yaonan.: Research on modeling and simulation of hybrid electric vehicle energy control system.: International Conference on Electrical Machines and Systems, Nanjing, China, pp. 849–852, (2005)

29. Maini, C., Gopal, K., Prakash, R.: Making of an 'all reason' electric Vehicle. World Electric Vehicle Symposium and Exhibition (EVS27), Barcelona, Spain, pp. 1–4 (2013)

30. Emadi, A., Young Joo Lee, Rajashekara, K.: Power electronics and motor drives in electric, hybrid electric, and Plug-In hybrid electric vehicles. IEEE Trans. Ind. Electron., **55**(6), pp. 2237–2245 (2008)

31. Saeed, M. A., Ahmed, N., Hussain, M., Jafar, A.: A comparative study of controllers for optimal speed control of hybrid electric vehicles. International Conference on Intelligent Systems Engineering (ICISE), Islamabad, Pakistan, pp. 1–4 (2016)

32. Viswanatha Rao, J., Lakshminarayana, G., Obulesu, D.: Reduction of torque ripples in PMSM using fuzzy controller based driving converter. Turk. J. Comput. Math. Educ., **12**(3), pp:4652–4658 (2021)

33. Pena Fernandez, C. A.: NARMA-L2-based control of wheeled mobile robots singularly perturbed by using neural networks: Curvilinear approach. Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), Curitiba, pp. 1–6 (2017)

34. Sharma, D. K., Son, L. H., Sharma, R., Cengiz, K. Eds.: Micro-Electronics and Telecommunication Engineering: Proceedings of 4th ICMETE 2020, **179**. Singapore: Springer Singapore, (2021)

35. Haque, Md. R., Khan, S.: The modified proportional integral controller for the BLDC motor and electric vehicle. IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, pp. 1–5 (2021)

36. IEEE Standard for Technical Specifications of a DC Quick and Bidirectional Charger for Use with Electric Vehicles—Redline, IEEE Std 2030.1.1–2021 (Revision of IEEE Std 2030.1.1–2015)—Redline, pp. 1–263 (2022)

37. Zeng, Q., Huang, J.: The design and simulation of a fuzzy logic controller for parallel hybrid electric vehicles. IEEE Int. Conf. Autom. Logist., Jinan, China, pp. 908–912 (2007)

38. Bae S., Gros, S., Kulcsar, B.: Can AI Abuse personal information in an EV Fast-Charging market. IEEE Trans. Intell. Transport. Syst., **23**(7), pp. 8759–8769 (2022)

39. Chi, H. R., Chui, K. T., Tsang, K. F., Chung, H. S.: Energy management of hybrid vehicles using artificial intelligence. 2nd Global Conference on Consumer Electronics (GCCE), Tokyo, Japan, pp. 65–67 (2013)

40. Narayana Gadupudi, L., GudapatiSambasiva Rao.: 9-Level VSC based STATCOM for reactive power management and voltage stability improvement. J. Green Eng., **10**(11), pp: 10275–10288 (2020)

41. Lokhande, J. S., Daigavhane, P. M., Sarkar, M.: A critical approach towards a smarter battery management system for electric vehicles. 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, pp. 232–235 (2020)

42. Ranawat, D., Prasad, M. P. R.: A Review on Electric Vehicles with perspective of Battery Management System. International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Msyuru, India, pp. 1539–1544 (2018)

43. Yajun Rong, Wei Yang, H. Wang, and Hanhong Qi.: SOC estimation of electric vehicles based on the establishment of battery management systems. IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), Beijing, China, pp. 1–5 (2014)

44. Yeqin Wang and Yixing Liu.: Electronic control system design and test of pure electric vehicle battery management system. Second International Conference on Mechanic Automation and Control Engineering, Inner Mongolia, China, pp. 1289–1292 (2011)

45. Helmy, M., et al.: IoT-Based battery monitoring system for electric vehicle. Int. J. Eng. & Technol., **7**, pp. 505–510 (2018)

46. Li, S. G., Sharkh, S. M., Walsh, F. C., Zhang, C. N.: Energy and battery management of a Plug-In series hybrid electric vehicle using fuzzy logic. IEEE Trans. Veh. Technol., **60**(8), pp. 3571–3585 (2011)

47. Narayana Gadupudi L, GudapatiSambasiva Rao.: 7-Level transformers integrated voltage source converter based STATCOM for voltage profile enhancement. J. Solid State Technol., **63**(5), pp: 3134–3141 (2020)

48. Aung, N., Zhang, W., Sultan, K., Dhelim, S., Ai, Y.: Dynamic traffic congestion pricing and electric vehicle charging management system for the internet of vehicles in smart cities. Digit. Commun. Netw., **7**(4), pp. 492–504 (2021)

49. Galdi, V., Piccolo, A., Siano, P.: A fuzzy based safe power management algorithm for energy storage systems in electric vehicles. 2006 IEEE Vehicle Power and Propulsion Conference, Windsor, UK, Sep. 2006, pp. 1–6 (2006)

50. Van Mierlo J. et al.: Beyond the state of the art of electric vehicles: A Fact-Based paper of the current and prospective electric vehicle technologies. World Electr. Veh. J., **12**(1), Art. no. 1, (2021)

51. Akil, M., Dokur, E., Bayindir, R.: Analysis of electric vehicle charging demand forecasting model based on monte carlo simulation and EMD-BO-LSTM. 10th International Conference on Smart Grid (icSmartGrid), pp. 356–362 (2022)

52. Yaqub, R., Ahmad, S., Ali, H., ulAsar, A.: AI and blockchain integrated billing architecture for charging the roaming wlectric vehicles. IoT, **1**(2), Art. no. 2, (2020)
53. Chau, K. T., Chan, C. C., Chunhua Liu.: Overview of Permanent-Magnet brushless drives for electric and hybrid electric vehicles. IEEE Trans. Ind. Electron., **55**(6), pp. 2246–2257 (2008)

Chapter 19
Mathematical Design of Solar PV Cells at Different Sunlight Conditions: A Comprehensive Analysis

Sunkara Sunil Kumar and K. Balakrishna

Abstract Based on the present scenario of nonrenewable power supply networks, the availability of nonrenewable sources is reducing drastically. In addition, these networks release very harmful gasses. So, the entire environmental pollution takes place. Here, the solar natural resource is utilized for generating electrical energy. The merits of these Solar Photovoltaic (SPV) networks are highly environmentally friendly, fast power generation, easy handling, and very low maintenance cost. In this article, a comprehensive investigation has been done on the different SPV cells which are named 1-diode circuit, 2-diode circuit, plus 3-diode circuit SPV cell. The selected parameters for the investigation of SPV cells are sunlight irradiations, ideality factors of diodes, functioning temperature of the SPV cells, plus cells utilized in the series connection of the SPV system. From the performance investigation of all the SPV cells, the 3-diode circuit SPV cell gives more efficient output power when equated with the other solar cells.

Keywords Ambient temperature · Design of SPV · Functioning irradiations of the SPV · Utilization of the PV cells · Efficiency

19.1 Introduction

The present available conventional energy sources produce highly poisoned gasses and they start functioning with high initial maintained cost. Due to these poisoned gasses in the atmosphere, the ozone layer of the atmosphere may get damaged extensively. As a result, acid rainfall comes to the earth. Also, these conventional power

S. S. Kumar · K. Balakrishna (✉)
Vignan's Foundation for Science Technology and Research, Vadlamudi, India
e-mail: kbk_eee@vignan.ac.in

S. S. Kumar
e-mail: sunkarasunil.kumar89@gmail.com

supply networks take more time to generate electricity [1]. The available conventional commercial power supply networks are coal, oil, peat, nuclear, and natural gas. Similarly, the noncommercial power resources are firewood, straw, and dried dung [2]. Here, coal is one of the most commonly utilized fossil fuels in the power generation industry because of its merits, high abundance, affordability, plus good reliability. However, this power supply may create many atmospheric problems that directly affect human life [3].

So, the oil products are used for supplying electricity in emergency places like hospitals, and shopping malls [4]. The classification of oils is categorized as heavy fuel oil, crude oil, and diesel oil. Here, the oil is stored in the very big tanks and the stored oil is fed to the boiler for generating the steam content. The valves are fixed to the storage tanks to control the steam vapor. The features of crude oil-based power supply networks are more energy density, easily accessible, less water pollution, and emission of greenhouse gasses are less when equated to the thermal power network [5]. The drawbacks of this oil system are the high complexity of oil refining, plus more installation cost [6]. Nuclear power production is installed away from human beings leaving space for limiting the disadvantage of oil power supply networks. Here, the small substance of nuclear power is used for supplying the large power rating to the consumers. The merits of this system are easy to supply input fuel with less transportation cost, plus high flexibility. In addition, the conventional power sources drawbacks are less efficient and need a very high catchment area [7].

So, the present power-generating companies focus on nonconventional energy sources which are tidal, geothermal, solar, fuel cells, and wind power supply networks [8]. The researchers utilized tidal networks to produce electrical power [9]. In this, the tides come with very high kinetic energy which is transferred into rotational energy by using the tidal bidirectional blades. The merits of these tidal networks are low environmental pollution, free of cost, easy handling, need very little time to supply the power, more reliable, plus suitable for peak load demand conditions [10]. However, these tidal networks may not be applicable for constant power automotive systems. In addition, it affects the vehicle performance. The geothermal source is applied to the food dehydration process, gold mining, and milk pasteurizing process [11]. Also, it is used to build heating systems. In this geothermal system, the underground oil is transferred into steam which is sent to the turbine for supplying the electricity [12].

Wind power networks are employed to limit the disadvantages of geothermal systems [13]. Here, the wind runs the propeller to rotate the rotor blades. The rotating blades convert the rotational supply into electrical power. The features of wind networks are creating more jobs for human beings, providing more economic growth to the system, being highly beneficial to the local communities, cost-effective, and highly flexible [14]. The fuel stacks are becoming the current trend for electricity generation for the automotive industry and this source reduces the drawbacks of all nonrenewable sources. However, the fuel stack needed more installation cost and it is suitable only for medium power applications [15]. So, solar power networks are employed in the present smart grid networks. The SPV networks are developed by employing various types of methodologies which are thin film, poly, plus monocrystalline methodologies [16].
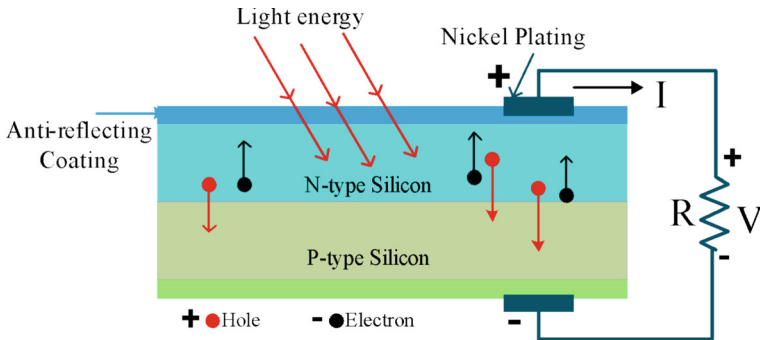
Fig. 19.1 Functioning diagram of the basic SPV module

Most of the SPV array developers utilize the monocrystalline concept because of its high functioning efficiency. The SPV cells are manufactured by employing one P-type and one N-type semiconductor material. The combination of both semiconductors creates the P-N-type diode. Here, the sunlight falls on the P-N semiconductor then the free electrons from the P-N material consume the lightning energy and start moving from one layer to another layer as shown in Fig. 19.1 [17]. As a result, the voltage appears across the P-N semiconductor material. The voltage that appears across the one cell is 0.76–0.82 V which is very low and it may not be employed for domestic applications. So, the SPV arrays are developed by employing the cells in series plus parallel fashion for increasing the power supply rating of the SPV system [18]. Here, in this manuscript, various SPV cells are investigated under various environmental conditions. The utilized atmospheric parameters are sunlight functioning temperature, functioning irradiations of the SPV module, plus utilized series connected SPV cells for developing an array. Finally, based on the simulative analysis results, the 3-diode network gives more accurate and efficient characteristics when equated to the other SPV cells.

19.2 Analysis and Mathematical Implementation of SPV Cells

From the currently published manuscripts, the SPV networks are developed by utilizing the germanium or silicon elements. These materials have free electrons and they flow from one side of the layer to another side [19]. Due to this electron flow in the layers, the current existed in the P-N diode-type SPV array. However, these solar cells collect only the sunlight insolations for activating the overall electrons. The detailed functioning diagram of the SPV network is given in Fig. 19.1 and its circuital representation is explained in the coming section.

19.2.1 Unique Diode Circuit-Based SPV Cell

The development of the SPV network has been done by focusing on the solar irradiance spectrum, and sunlight fall incident angle. Also, the location of the SPV system is one of the major focuses because of the tree falling, and building shadows may create a shading effect on the SPV modules. As a result, the effective PV module's power supply takes the shaded SPV array. So, the overall system functioning efficiency is reduced [20]. The general 1-diode SPV cell, plus practical 1-diode SPV cell with all the resistors are illustrated in Fig. 19.2a, and b. Based on Fig. 19.2a, the sunlight that falls on the P-N semiconductor material is represented as 'G', and the sunlight photo-generated current is represented as I_{Pg}. Here, the diode is employed because the current supplied from the SPV module should be in a unique direction. Also, the diode consists of a proportionality factor which is represented as a short-circuit current (I_{scc}) and its value varies from one cell to another SPV cell because its anode and cathode contact materials are different.

Based on Fig. 19.2a, the series resistance (R_{sr}) is utilized to limit the current flow from the diode, and it optimizes the voltage drop across the output of the SPV cell. The available current (I_{Ce}) of the ideal unique diode is obtained by utilizing Eq. (19.1). The photo-generated current is mentioned in Eq. (19.2). From Eq. (19.1), the parameters I_{Pg_ST}, K_n, ΔT, G_{st}, T, Q, and η are the standard photocurrent, temperature constant, variation of sunlight temperature, standard irradiations, cell basic temperature, charge available of the diode, and thermal coefficient of the cell. Here, I_{0i} defines the diode propeller current, and utilized cell for developing the array is n_s as given in Eq. (19.3). From Eqs. (19.4) and (19.5), the voltage (V_{occ}) of the cell at open circuit condition is given in Eq. (19.6).

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\left(\frac{V_{Pg}*Q}{\eta*T*K}\right)} - 1\right) \qquad (19.1)$$

$$I_{Pg} = \left(I_{Pg\_ST} + K_n * \Delta T\right) * \frac{G}{G_{ST}} \qquad (19.2)$$
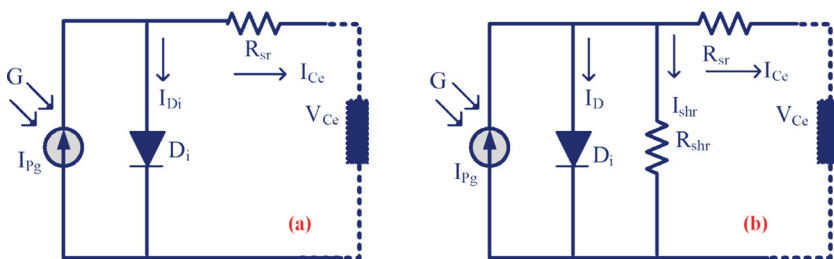
Fig. 19.2 Representation of SPV cells. **a** Unique diode plus **b** Double diode

$$I_{Ce} = I_{Pg} - i_{oi}\left(e^{\left(\frac{Q*V_{Ce}}{\eta*T*K*ns}\right)} - 1\right) \tag{19.3}$$

$$I_{scc\_cell} = I_{Ce} = I_{Pg} \tag{19.4}$$

$$I_{scc} = I_{Pg} = I_{oi}\left(e^{\left(\frac{Q*V_{Pg}}{\eta*K*T*ns}\right)} - 1\right) \tag{19.5}$$

$$V_{occ\_cell} = \frac{n_s * \eta * K * T}{Q}\,loglog\left(\frac{I_{Pg}}{I_{oi}} + 1\right) \tag{19.6}$$

$$I_{Pg} * V_{Pg} = P_{Pg} = \left(I_{Pg} - I_{oi}\left(e^{\left(\frac{Q*V_{Pg}}{\eta*K*T*ns}\right)} - 1\right) * V_{Pg}\right) \tag{19.7}$$

Here, the photocurrent flows toward the load, and it is controlled by introducing the shunt-resistive element as illustrated in Eq. (19.8). Finally, the overall photocurrent flowing through the shunt, plus series resistive elements are explained in Eqs. (19.12) and (19.13).

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Pg}+I_{Pg}*n_s*R_{sr})}{\eta*K*T*ns}} - 1\right) \tag{19.8}$$

$$V_{occ} = \frac{n_s * \eta * K * T}{Q}\,loglog\left(\frac{I_{Pg}}{I_{oi}} + 1\right) \tag{19.9}$$

$$I_{scc} = I_{Pg} - I_{oi}\left(e^{\frac{q*I_{Pg}*R_{sr}}{\eta*K*T*ns}} - 1\right) \tag{19.10}$$

$$P_{Pg} = I_{Pg} * V_{Pg} = \left(I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{pg}+I_{pg}*R_{sr}*n_s)}{\eta*K*T*ns}} - 1\right)\right)*V_{Pg} \tag{19.11}$$

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Pg}+I_{pg}*R_s)}{\eta*K*T}} - 1\right) - \frac{V_{Ce} + I_{Ce}R_{sr}}{R_{shr}} \tag{19.12}$$

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*n_s*R_s)}{\eta*K*Tns}} - 1\right) - \frac{V_{Ce} + I_{Ce} * R_{sr}}{R_{shr}}. \tag{19.13}$$

19.2.2 Development of Dual-Diode SPV Cell

The development of a 2-diode SPV array is similar to the 1-diode model. Here, the one more diode is parallelly connected to the load in the one-diode model for making the dual-diode circuit-based SPV [21]. The unique diode model may not be applicable for the accurate PV array design because of the many approximations in

Fig. 19.3 2-diode SPV
system with resistive load
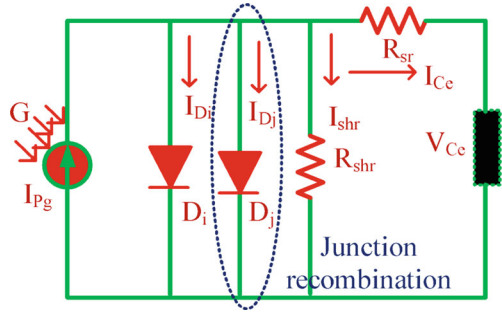
the SPV cell parameters extraction. Here, the fill factor is considered for the design
of the overall SPV network. The fill factor is decided based on the peak available
power mentioned on the SPV array, and it's related to the open-circuited voltage and
short-circuit current. Here, the high fill factor 2-diode cell is used for increasing the
supply voltage of the SPV system. Also, the dual-diode circuit-based SPV system
gives more efficiency when equated to the other cells. In the 1-diode cell, the ideality
factor (w) is assumed to be a constant value. But in reality, the ideality factor is
varied based on the voltage that appears across the P-N diode, and the recombination
of charges takes place on the surface of the P-N diode material. Here, the utilized
ideality factor of the diode is one. From Fig. 19.3, an additional diode is added to
the 1-diode to limit the junction recombination charges and make the diode ideality
factor two.

Based on Fig. 19.3, the available photocurrent is determined by utilizing
Eq. (19.14), and the additional diode current is mentioned in Eq. (19.15).

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*R_{sr})}{\eta_i*K*T}} - 1\right) - I_s \qquad (19.14)$$

$$I_s = I_{oj}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*R_{sr})}{\eta_j*K*T}} - 1\right) + \frac{V_{Ce} + I_{sr}*R_{sr}}{R_{shr}} \qquad (19.15)$$

19.2.3 Development of the 3-Diode-Based SPV System

In this SPV model, there are more than two diodes. So, the available diode junctions
are three. Here, every diode is placed on top of each other diodes to enhance the
operating efficiency of the overall SPV network. This type of PV cell gives 0.89 to
0.96 V very low supply [22]. So, there are multiple types of cells integrated in a series
fashion for optimizing the supply voltage of the SPV system. The functioning of this
3-diode model is similar to the other diode models. To enhance the current supply
rating of the SPV network, the 3-diode cells are integrated in a parallel fashion.

Sometimes, the SPV network efficiency is reduced due to the tree-falling situation. So, a bypassing diode is placed across each SPV string to reduce the overall system heating losses.

In addition, various parameters are utilized for improving the functioning efficiency of the SPV network which are converter duty value, and the peak power extraction from the solar network. The 3-diode cell without a shunt-resistance circuit is given in Fig. 19.4a, and the effect of junction capacitance on the SPV system is shown in Fig. 19.4b. The SPV network observes the sunlight insolation and delivers the power to the external circuit which is connected across the anode and cathode of the diode. The simulative design parameters of various SPV cells are illustrated in Table 19.1.
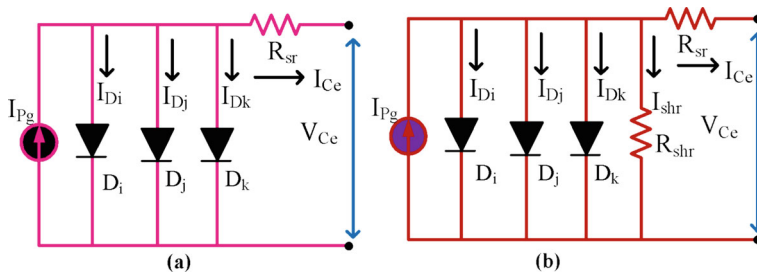
Fig. 19.4 3-diode SPV system. **a** With the absence of a shunt-resistive element, and **b** with shunt-resistance

Table 19.1 Utilized variables of SPV cells under various operating environmental conditions

Variables	Values
Available voltage at peak power point of the SPV (V_{MPP})	30.01 V
Available current at peak power point of the SPV (I_{MPP})	7.512 amps
Available power at peak power point of the SPV (P_{MPP})	224.89 watts
Open circuit voltage at peak power point of the SPV (V_{oc})	36.99 V
Short-circuit current at peak power point of the SPV (I_{sc})	8.129 amps
Applied temperature coefficient of SPV system voltage	−0.400129%/°C
Applied temperature coefficient of SPV system current	0.03227%/°C
Sunlight functioning irradiations on SPV system (G)	1000 W/m^2
Ideality parameters of different diodes (η_1, η_2, and η_3)	0.86, 0.92, and 0.942
Parasitic effect of series resistance on diode (R_{sr})	0.492 Ώ
Parasitic effect of parallel resistance on diode (R_{shr})	114.892 Ώ
Functioning SPV system temperature (T_n)	298.00 K
Cells selected in the utilized SPV network (n_s)	60.0

$$I_{Ce} = I_{Pg} - I_{D_i} - I_{D_j} - I_{D_k} \tag{19.16}$$

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*R_{sr})}{\eta_a*K*T}} - 1\right) - I_z \tag{19.17}$$

$$I_z = i_{ore\_b}\left(e^{\frac{q(V_{PV}+I_{PV}*R_{se})}{\eta_b*K*T}} - 1\right) + i_{ore\_c}\left(e^{\frac{q(V_{PV}+I_{PV}*R_{se})}{\eta_c*K*T}} - 1\right) \tag{19.18}$$

Based on Eqs. (19.16) and (19.18), the SPV cell current by including the shunt-resistive element is obtained as

$$I_{Ce} = I_{Pg} - I_{Di} - I_{Dj} - I_{Dk} - I_{shr} \tag{19.19}$$

$$I_{Ce} = I_{Pg} - I_{oi}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*R_{sr})}{\eta_i*K*T}} - 1\right) - I_{oj}\left(e^{\frac{Q(V_{Ce}+I_{Ce}R_{sr})}{\eta_j*K*T}} - 1\right) - I_x \tag{19.20}$$

$$I_x = I_{ok}\left(e^{\frac{Q*(V_{Ce}+I_{Ce}*R_{sr})}{\eta_k*K*T}} - 1\right) + \frac{V_{Ce} + I_{Ce}R_{sr}}{R_{shr}} \tag{19.21}$$

$$I_{oi} = I_{oj} = I_{ok} = I_{on}(\frac{T}{T_n})^3 * e^{\frac{QEg}{nk}\left(\frac{1}{T_n} - \frac{1}{T}\right)} \tag{19.22}$$

$$I_{oi} = I_{on\_i} = I_{on\_j} = I_{on\_k} = \frac{I_{SCc\_n}}{e^{\left(\frac{V_{oc\_n}}{\eta V_{Tn}}\right)}}. \tag{19.23}$$

19.3   Results of the Proposed Work

From the mathematical design of SPV cells, there are different types of solar PV systems selected for the investigation of solar power systems at multiple types of atmospheric conditions. For the performance study of the proposed network, there are six variables utilized for the design of a unique diode-based SPV system which are short-circuit current, peak available voltage of the SPV system, rated power of the SPV network, peak current of the SPV module, propulsion factor of the selected diode, and openly circuited voltage of the overall system. However, there are some additional constraints needed for the implementation of 2-diode and 3-diode circuit-based SPV systems which are ideality factors of their respective diode models. The actual constraints of the SPV system are determined by applying the various types of machine learning, and soft computing, plus metaheuristic-based optimization controllers. Here, the Drone Squadron Optimization methodology is utilized for evaluating the performance results of the SPV system.
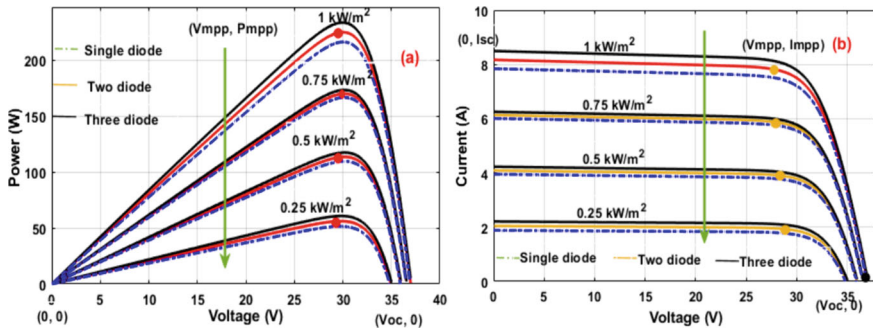
Fig. 19.5 a Nonlinear P-V curves and **b** I-V curves at various sunlight insolation conditions

19.3.1 Investigation of SPV Cells Under Different Solar Irradiation Values

Based on the above-discussed mathematical derivations, the SPV system's available power is very high when the incident sunlight insolation's angle on the SPV module is very high. Here, the SPV system-related outpower is reduced because of the increment of the sunlight temperature. In the morning, if the sunlight irradiation density is low then the available power from the SPV system is low, and it is moderate when the sunlight reaches mid-time of the day. At 25^0C functioning temperature of the SPV system, the obtained characteristics of the SPV system by employing the 1-diode, dual-diode, and 3-diode-based SPV system are given in Fig. 19.5a and b.

At 1000 W/m^2 sunlight insolation, the obtained power, plus the voltage of the triple-diode-based SPV system are 224.90 W and 29.98 V, respectively. The evaluated efficiency and fill factor of the 3-diode-based SPV system are 17.89 and 0.76%. Here, at room temperature, the triple-diode-based SPV network gives more power under multiple types of irradiation conditions. In addition, the evaluated unique diode-based SPV system fill factor, current, and operating efficiencies are 0.552, 7.64A, plus 17.22%. Finally, the dual-diode-based SPV system operating efficiency, plus the fill factor at 500 W/m^2 is 17.25%, plus 0.38. Similarly, at 750 W/m^2, the unique diode-based SPV system fills factor, power, plus working efficiencies are 0.56, 166 W, plus 7.64 A. From the investigation of all types of diode circuits, the 3-diode SPV network works efficiently at various atmospheric insolation conditions.

19.3.2 Investigation of SPV Cells Under Different Series Cells

From the literature study, the unique diode SPV cell provides very low voltage and it is not utilized for local consumer applications. Here, multiple series of connected cells
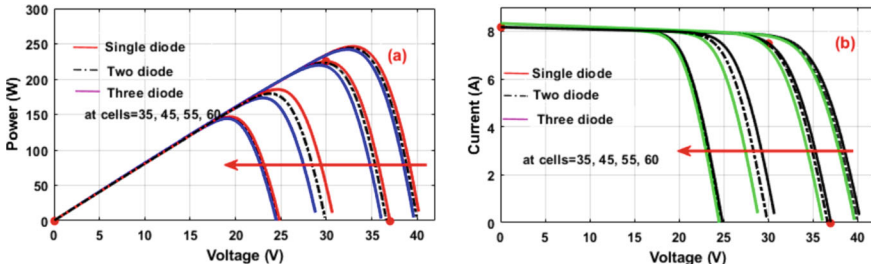
Fig. 19.6 **a** Nonlinear P-V curves and **b** I-V curves at various SPV series cells

are applied in the SPV network to improve the supply voltage of the SPV network. Based on Eqs. (19.3) and (19.7), the series connected cells increment provides the high supply voltage rating of the system. The evaluated fill factor is improved by improving the series of connected cells of the SPV system. In 1 diode, 35 cells are used for getting the 14 V supply, and its functioning efficiency is 13.6%. The SPV network characteristics are illustrated in Fig. 19.6a and b. Based on Fig. 19.6b, the determined power and fill factor of the 2-diode cell-based SPV network are 148 W and 0.47. Finally, at 45 series cells, the 3-diode model SPV system generates a voltage of 21 V, plus its operating efficiency is 14.89%.

19.3.3 Investigation of SPV Cells Under Various Sunlight Temperature Values

From Eqs. (19.1), and (19.4), the SPV system power production is reduced drastically by improving the system operating temperature. At multiple temperature values (45, 65, 85, and 105 °C), the nonlinear characteristics of the SPV system are explained in Fig. 19.7a, plus b. Here, the utilized cells are 35 which are constant, and the selected irradiation value is 1000 W/m^2 then the power production from the 1-diode SPV system is 221.02W. At 45 °C, the unique diode-based SPV network fill factor and efficiencies are determined which are identified as 0.75 and 17.269%. The triple-diode circuit-based PV system generated voltage at 850 C is 24.22 V, and its related fill factor and functioning efficiencies are 0.6 and 16.02%, respectively. Finally, the dual-diode SPV network gives 152W at 105 °C. Similarly, at 65 °C, the unique diode SPV cell functioning efficiency is 16.18%.
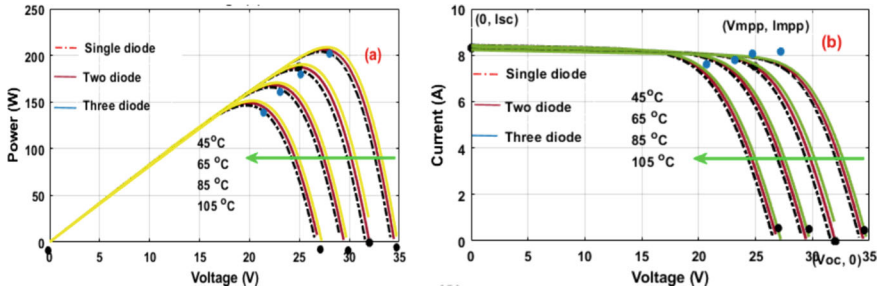
Fig. 19.7 a Nonlinear P-V curves and **b** I-V curves at various SPV operating temperature values

19.4   Conclusion

From the mathematical investigation of various types of SPV cells, the dual-diode cell gives high efficiency, and fill factor when equated to the 1-diode-based SPV network. Also, the one-diode model SPV system works with very high power conduction losses at high environmental temperature values. Same way, the 3-diode SPV cell works at high-temperature values with more fill factor when equated to the 2-diode and 1-diode circuit models. In addition, this model works efficiently and effectively under various sunlight irradiation conditions. Also, it is suitable for the generation of the accurate nonlinear characteristics of the SPV system.

References

1. Basha, C.H., Rani, C.: Different conventional and soft computing MPPT techniques for solar PV systems with high step-up boost converters: a comprehensive analysis. Energies **13**(2), 371 (2020)
2. Hussaian Basha, C.H., Rani, C.: Performance analysis of MPPT techniques for dynamic irradiation conditions of solar PV. Int. J. Fuzzy Syst. **22**(8), 2577–2598 (2020)
3. Kiran, S.R., Basha, C.H., Singh, V.P., Dhanamjayulu, C., Prusty, B.R., Khan, B.: Reduced simulative performance analysis of variable step size ANN based MPPT techniques for partially shaded solar PV systems. IEEE Access **10**, 48875–48889 (2022)
4. Hussain Basha, C.H., Bansal, V., Rani, C., Brisilla, R.M., Odofin, S.: Development of cuckoo search MPPT algorithm for partially shaded solar PV SEPIC converter. In: Soft Computing for Problem Solving: SocProS 2018, vol. 1, pp. 727–736. Springer Singapore (2020)
5. Nadimuthu, L.P.R., Victor, K., Basha, C.H., Mariprasath, T., Dhanamjayulu, C., Padmanaban, S., Khan, B.: Energy conservation approach for continuous power quality improvement: a case study. IEEE Access **9**, 146959–146969 (2021)
6. Udhay Sankar, V., Bhanutej, Hussain Basha, C. H., Mathew, D., Rani, C., Busawon, K.: Application of WDO for decision-making in combined economic and emission dispatch problems. In: Soft Computing for Problem Solving: SocProS 2018, vol. 1, pp. 907–923. Springer Singapore (2020)
7. Basha, C.H., Rani, C.: A new single switch DC–DC converter for PEM fuel cell-based electric vehicle system with an improved beta-fuzzy logic MPPT controller. Soft. Comput. **26**(13), 6021–6040 (2022)

8. Kiran, S.R., Basha, C.H., Kumbhar, A., Patil, N.: A new design of single switch DC–DC converter for PEM fuel cell-based EV system with variable step size RBFN controller. Sādhanā **47**(3), 128 (2022)

9. Obara, S., et al.: Operational planning of an independent microgrid containing tidal power generators, SOFCs, and photovoltaics. Appl. Energy **102**, 1343–1357 (2013)

10. O'Rourke, F., Boyle, F., Reynolds, A.: Tidal energy update 2009. In: Renewable Energy, pp. Vol3_451–Vol3_476. Routledge (2018)

11. Murali, M., Basha, C. H., Kiran, S. R., Amaresh, K.: Design and analysis of neural network-based MPPT technique for solar power-based electric vehicle application. In: Proceedings of Fourth International Conference on Inventive Material Science Applications: ICIMA 2021, pp. 529–541. Springer Singapore (2022)

12. Basha, C.H., Rani, C.: Application of fuzzy controller for two-leg inverter solar PV grid connected systems with high voltage gain boost converter. J. Eng. Sci. Technol. Rev. **14**(2) (2021)

13. Mariprasath, T., Shilaja, C., Hussaian Basha, C.H., Murali, M., Fathima, F., Aisha, S.: Design and analysis of an improved artificial neural network controller for the energy efficiency enhancement of wind power plant. In: Computational Methods and Data Engineering: Proceedings of ICCM DE 2021, pp. 67–77. Springer Nature Singapore, Singapore (2022)

14. Patil, S.S., Varma, B.S., Devadasu, G., Basha, C.H., Inamdar, M.J.R., Salman, S.S.: Performance analysis of image caption generation using deep learning techniques. In: International Conference on Microelectronic Devices, Circuits and Systems, pp. 159–170. Springer Nature Switzerland, Cham (2022)

15. Murali, M., Hussaian Basha, C. H., Kiran, S. R., Akram, P., Naresh, T.: Performance analysis of different types of solar photovoltaic cell techniques using MATLAB/simulink. In: Proceedings of Fourth International Conference on Inventive Material Science Applications: ICIMA 2021, pp. 203–215. Springer Singapore (2022)

16. Kiran, S.R., Murali, M., Hussaian Basha, C.H., Fathima, F.: Design of artificial intelligence-based hybrid MPPT controllers for partially shaded solar pv system with non-isolated boost converter. In: Computer Vision and Robotics: Proceedings of CVR 2021, pp. 353–363. Springer Singapore, Singapore (2022)

17. Rafikiran, S., Devadasu, G., Basha, C.H., Tom, P.M., Prashanth, V., Dhanamjayulu, C., ... & Muyeen, S.M.: Design and performance analysis of hybrid MPPT controllers for fuel cell fed DC-DC converter systems. Energy Rep. **9**, 5826–5842 (2023)

18. Reddy, K.R., Basha, C.H., Prashanth, V., Dhanamjayulu, C., Shivashimpiger, S., & Likhitha, R.: A Novel on energy management strategy with maximum exploitation of renewables and EV storage in distribution networks. Int. Trans. Electr. Energy Syst. (2023)

19. Kiran, K.B., Indira, M.S., Nagaraja, R.: Mathematical modeling and evaluation of performance characteristics of a hybrid solar PV and wind energy system. J. Appl. Sci. Eng. **25**(4), 785–797 (2021)

20. Bana, S., Saini, R.P.: A mathematical modeling framework to evaluate the performance of single diode and double diode based SPV systems. Energy Rep. **2**, 171–187 (2016)

21. Puppala, R., Jeyaraj, B.E., Isaac, J.R., & CH, H.B.: Framework for smart grid to implement a price elasticity-based peak time rebate demand response program. Front. Energy Res. **10**, 1079695 (2023)

22. Rafikiran, S., Devadasu, G., Rajendhar, P., Likhitha, R., Basha, C.H.: Design and implementation of hybrid MPPT controller for FC based EV system with boost DC-DC converter. J. Intell. Fuzzy Syst. (Preprint), 1–19 (2023)

Chapter 20
Experimental Investigation of Evaporation Time and Evaporation Rate of Nano Particle Blended Emulsified Karanja Biodiesel

G. V. L. Prasad

Abstract Persistent uses of naturally available fuels are exhausting quickly and accumulation of CO_2 occurs in the environment due to its burning. Transport fuels that are carbon neutral and renewable are necessary for both environmental and financial sustainability. An alternative to fossil fuels that is renewable and carbon neutral is biodiesel made from oil crops. Non-edible oil such as Karanja biodiesel (KBD) is easily available in India and is very cheap compared to edible oils. However, the utilization of KBD in a diesel engine is limited due to high viscosity and lesser evaporation rates. The rate of evaporation of fuel plays a significant part in engine combustion. Nanoparticles such as Al, CuO, CeO and CNT are used in KBD and its emulsions to improve the evaporation rate. The present investigation has been conducted to evaluate the evaporation attributes of nanoparticle suspended KBD with and without water. When nanoparticles were added, the evaporation rate of KBD and its emulsions improved. In summary, the incorporation of nanoparticles leads to an increase in the rate of evaporation.

Keywords Evaporation rate · Karanja biodiesel · Nanoparticles

20.1 Introduction

A diesel engine is more popular for its low fuel consumption, reliability, and durability characteristics. Due to rapid economic progress and industrialization the usage of diesel engines increases. However, the byproducts of the diesel engine's combustion lead to air pollution, which affects humans, animals, and plants as well as changing the climate. The demand for petroleum products also increases and rises in diesel prices. Diesel prices are rising and create challenges to the country's economy. Stringent

G. V. L. Prasad (✉)
VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, India
e-mail: prasad_gvl@vnrvjiet.in

emission regulations also prompt the use of alternative fuels and emission reduction techniques. The most promising option for diesel engine fuel in the future is biodiesel but biodiesel evaporates at a slower pace, it doesn't burn completely, which results in increased fuel consumption and decreased thermal efficiency. Incomplete combustion also increases exhaust emissions. Karanja biodiesel is abundantly available in certain parts of India. This biodiesel has the potential to be used in C.I. Engines. Depletion of mineral diesel resources and stringent emission norms has encouraged to enhance the performance of C.I. engine and clean combustion. One of the best solutions to these issues is to use additives for diesel or biodiesel. Adding metallic base additions at the nanoscale are one method to enhance the fuel's characteristics like evaporation rate [1]. Zaid [2] has carried out experiments on emulsions of diesel and kerosene on a single droplet's evaporation time by using a hot surface. He conducted at ambient pressure and water concentrations 10–40% by volume the temperatures of the wall varying from 100–460 °C to encompass the full band of heat transfer attributes from evaporation to film boiling. Presently, droplet combustion in many applications is being studied technologically by the evaporation of emulsion fuel droplets on heated surfaces. For example, the vaporization of fuel droplets occurs during the physical ignition delay period in diesel engines. Therefore, it is crucial to understand how emulsion fuel droplets behave, to understand the performance of the combustion process of emulsion fuels. Tyagi et al. [3] studied the ignition properties of diesel by adding alumina nanoparticles to diesel at different proportions. Using a heated plate, the droplet ignition experiments were conducted at a temperature over the range of 688–768 °C. The nanoparticles containing fuel mixtures have higher ignition probability than pure diesel. Cho et al. [4] investigated the behavior of fuel/water emulsion droplets micro explosion on a heated plate. They saw that at first, the droplet is in touch with the hot plate; however, as the temperature rises, the drop's lifetime first decreases, then increases quickly, and eventually diminishes slowly as a result of the droplet's levitation behavior on the hot surface. Annamalai et al. [5] conducted experiments by using cerium oxide nanoparticles in lemon grasswater emulsion. They observed higher thermal efficiency because of less evaporation time due to greater surface area per unit volume. Lower specific energy consumption due to incorporation of cerium oxide leads to an increase in evaporation rate. Saptarshi et al. [6] have done the review on nanoparticle contained droplets combustion behavior. They concluded that nanoparticles help to produce fine droplets of fuel which increases the fuel atomization. Narinder et al. [7] studied CNT blended emulsified diesel adding CNT with varying proportions 50–150 ppm and achieved shorter ignition delay. Alias et al. [8] has investigated various properties of a diesel- biodiesel blend by adding CNT at different concentrations they found that at higher concentrations of CNT Cetane number was higher, they also observed increase in calorific value of biodiesel upon addition of CNT. Varatharaju et al. [9] has conducted research using a mixture of Pongamia methyl ester copper oxide (CuO) nanoparticles. Nano particles are found to be useful in enhancing the quality of fuel and performance and exhaust characteristics due to shorter ignition delay which consumes complete combustion, Because the nanoparticles act as oxygen boosters, complete combustion occurs along with micro explosion events and features of a greater surface to volume

ratio. Selvan et al. [10] has done experiments by combining CNT and ceria at different concentrations ranging from 25–100 ppm in the blends of diesel-ethanol-biodiesel. They found that there was a decrease in ignition delay which aids in burning fuel mixtures quickly. Gan et al. [11] has investigated the evaporation rate of ethanol by incorporating nano aluminum and multi walled CNT. The rate of evaporation of ethanol showed better results upon addition of nanoparticles. Nevertheless, ethanol which blends with multi walled CNT has a greater rate of evaporation than aluminum mixed one, due to good radiation absorption capabilities of multi wall CNT. However little work has been reported on estimation of the evaporation rate of biodiesel and its emulsions containing nanoparticle additives. The present work has carried out the evaporation characteristics of Karanja methyl ester and its emulsions. The present work has been carried out to estimate the evaporation behavior of Karanja biodiesel by incorporating various types of nanoparticle additives at 100 ppm concentrations.

20.2   Methodology

20.2.1   Preparation of Blends

Nanoparticle blends are prepared in small quantities of 50 ml by adding different types of nanoparticles Al, CuO_2, CeO_2 and CNT to the KBD in 100 ppm concentrations. Agitate for 30 min by using an ultrasonicator. Emulsions are prepared by adding 1% surfactants (Span 80 and Tween 80) each and 10% water to KBD and thoroughly mixed by using a mechanical agitator for 30 min at 1500 rpm. Nanoparticles are added to this at 100 ppm concentration and blended by using an ultrasonicator.

20.2.2   Experimental Setup and Procedure

The heating equipment was made with the help of a circular stainless-steel plate and heater. The steel plate's surface temperature was managed by variac. The surface of the steel plate is concave slightly to stop the droplet flowing out. A thermocouple used to measure temperature was positioned in the middle of the steel plate. The surface was highly polished to mirror like finish. The plate was fastened to the heater's top. The temperature is varied with the help of dimmer stat by adjusting the voltage and current. A dropper burette was placed with the help of a stand 20 mm above the plate to minimize the scattering of the droplet. The hot plate temperature is maintained at 400–600 °C. After the temperature reached a stable state, all the trials were completed. Fifty fuel droplets were placed on the hot plate for every sample. A digital timer was used to measure the evaporation time, and the average time was computed to determine each sample's rate of evaporation. The average size of a single

droplet was determined by measuring the weight of 50 droplets. The stainless-steel plate was polished by using sandpaper to remove the carbon deposits after every test.

20.3 Results and Discussions

20.3.1 *Effect of Temperature on Evaporation Time for Nano Particle Blended KBD Fuel*

The outcomes for evaporation time of nano particle blended KBD fuel at various temperatures are shown in Fig. 20.1, as temperature rises, evaporation times for all tested fuel samples decrease. KBD fuel has more evaporation time than nano particle blended fuels. The evaporation time reduced drastically upon addition of nanoparticles. The fuel properties are improved due to addition of CNT [12], Aluminum, Copper Oxide and CeO_2 nanoparticles due to this evaporation time reduced. Incorporation of nanoparticles affects the heat transfer characteristics and impact on evaporation time. The nanoparticles present in the fuel act as an emitter and receiver of heat and transfer of heat within the fuel droplet owing to conduction and radiation [13]. The micro explosion phenomenon was predominant at higher temperatures resulting in reduced evaporation time.

Fig. 20.1 Evaporation time versus temperature for the nano particle blended Karanja biodiesel
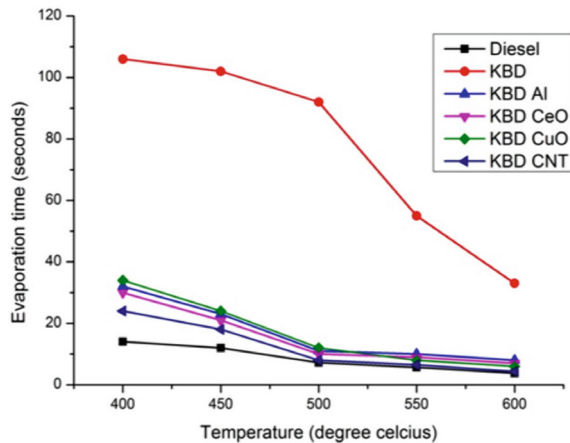
Fig. 20.2 Evaporation time versus temperature for the nano particle blended Karanja biodiesel emulsion
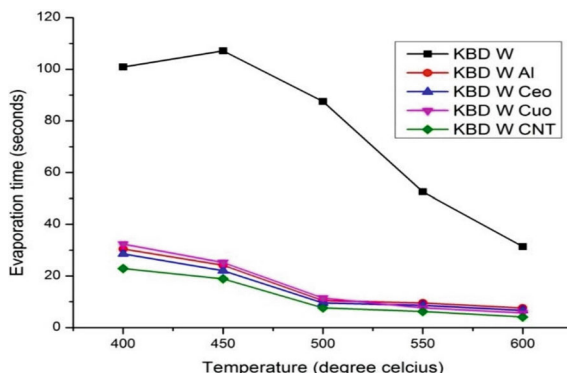
20.3.2 Effect of Temperature on Evaporation Time for Nano Particle Blended KBD Emulsion Fuel

The addition of water resulted in a decrease in the evaporation time, as seen in Fig. 20.2. Low boiling water evaporates first and the evaporation period is shortened as fuel droplets break up into finer droplets. The water was very helpful in causing a micro explosion of fuel [14]. During combustion, this micro explosion enhances the combustion response and lowers the amount of pollutants [15].

20.3.3 Effect of Temperature on Evaporation Rate for Nano Particle Blended KBD Fuel

It has been observed from Figs. 20.3 and 20.4 that with increase in temperature the evaporation rate increases; the evaporation rate also increases due to addition of nanoparticles in KBD fuel. It is observed that the evaporation rate increases suddenly at 500 °C. The use of nanoparticles enhances the fuel's heat transmission property and accelerates its rate of evaporation [3]. Faster fuel evaporation is made possible by the nanoparticles in the base fuel, which absorb heat and spread to nearby fuel. The ignition delay reduces due to faster evaporation of fuel causing a small amount of fuel burning at the premixed combustion phase; it lowers the temperature of the combustion flame and lowers NOX emissions [16].

Fig. 20.3 Evaporation rate versus temperature for nano particle blended Karanja biodiesel
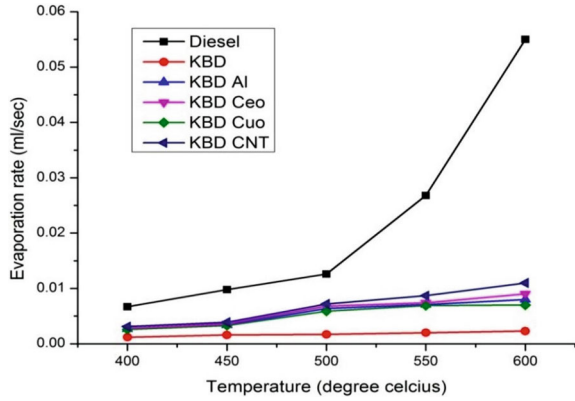
Fig. 20.4 Evaporation rate versus temperature for nano particle blended Karanja biodiesel emulsion
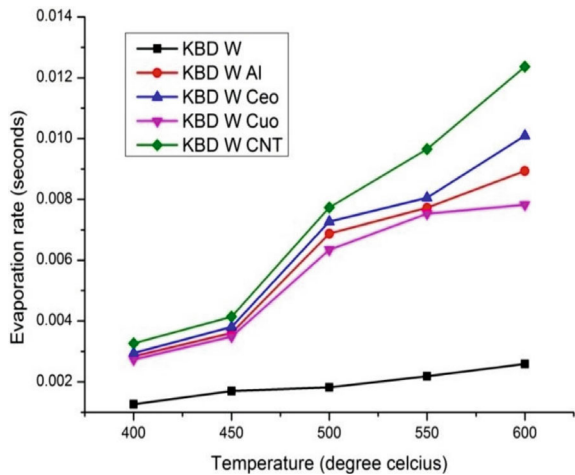
20.3.4 Effect of Temperature on Evaporation Rate for Nano Particle Blended Emulsified KBD Fuel

The graph shows that because water has a low boiling point, adding water increases the pace at which KBD fuel evaporation occurs. Tiny fuel droplets generated as a result of water evaporating and causing a micro explosion. The evaporation rate is aided by the small droplets' larger surface area [17]. Additionally, it is shown that evaporation steeply rises at 450 °C in emulsified fuels blended with nanoparticles.

20.4   Conclusions

The following conclusions were reached after experiments using a hot plate to measure the evaporation rate and time of nano particle blended KBD fuel and KBD fuel emulsion.

1. It was shown that the rate of heat transfer causes the evaporation time to decrease as temperature rises.
2. Nano particle blended KBD fuel evaporation time is reduced when compared to pure KBD fuel.
3. CNT nano particle blended KBD fuel has lesser evaporation time and greater evaporation rate compared to Al, CeO, CuO nano particle blended KBD fuels.
4. Addition of water further reduces the evaporation time and improvement in evaporation rate due to lees boiling point of water.

References

1. Khalife, E., Tabatabaei, M., Demirbas, A., Aghbashlo, M.: Impacts of additives on performance and emission characteristics of diesel engines during steady state operation. Prog. Energy Combust. Sci. **59**, 32–78 (2017)
2. Abu-Zaid, M.: An experimental study of the evaporation characteristics of emulsified liquid droplets. Heat Mass Transf. **40**, 737–741 (2004)
3. Tyagi, H., Phelan, P.E., Prasher, R., Peck, R., Lee, T., Pacheco, J.R., Arentzen, P.: Increased hot-plate ignition probability for nanoparticle-laden diesel fuel. Nano Lett. **8**(5), 1410–1416 (2008)
4. Cho, P., Law, C.K., Mizomoto, M.: Effect of pressure on the micro-explosion of water/oil emulsion droplets over a hot plate. Heat Transf. Trans. ASME **113**, 272–274 (1991)
5. Annamalai, M., Dhinesh, B., Nanthagopal, K., Sivarama Krishnan, P., Isaac Joshua Ramesh Lalvani, J., Parthasarathy, M., Annamalai, K.: An assessment on performance, combustion and emission behavior of a diesel engine powered by ceria nanoparticle blended emulsified biofuel. Energy Convers. Manag. **123**, 372–380 (2016)
6. Basu, S., Miglani, A.: Combustion and heat transfer characteristics of nano fluid fuel droplets: a short review. Int. J. Heat Mass Transf. **96**, 482–503 (2016)
7. Singh, N., Bharj, R.S.: Effect of CNT-emulsified fuel on performance emission and combustion characteristics of four stroke diesel engine. Int. J. Curr. Eng. Technol. **5**(1), 477–485 (2015)
8. Alias, A.B.S., Thegaraju1, D., Sharma, K.V.: Effect of carbon nanotube dispersions to the palm oil diesel biodiesel blend properties. In: International Conference on Mechanical Engineering Research, Malaysia. Paper ID: P011:1–14 (2013)
9. Perumal, V., Ilangkumaran, M.: The influence of copper oxide nano particle added pongamia methyl ester biodiesel on the performance, combustion and emission of a diesel engine. Fuel **232**, 791–802 (2018)
10. Arul Mozhi Selvan, V., Anand, R.B., Udayakumar. M.: Effect of cerium oxide nanoparticles and carbon nanotubes as fuel borne additives in diesterol blends on the performance, combustion and emission characteristics of a variable compression ratio engine. Fuel **130**, 160–116 (2014)
11. Gan, Y., Qiao, L.: Optical properties and radiation enhanced evaporation of nanofluid fluid fuels containing carbon based nanostructures. Energy Fuels **26**, 4224–4230 (2012)
12. Gumus, S., Ozcan, H., Ozbey, M., Topaloglu, B.: Aluminum oxide and copper oxide nano diesel fuel properties and usage in a compression ignition engine. Fuel **163**, 80–87 (2016)

13. Ghamari, M., Ratner, A.: Combustion characteristics of colloidal droplets of jet fuel and carbon based nanoparticles. Fuel **188**, 182–189 (2017)
14. Kannan, G., Anand, R.: Experimental investigation on diesel engine with diestrole-water micro emulsions. Energy **36**, 1680–1687 (2010)
15. Ithnin, A.M., Noge, H., Kadir, H.A., Jazair, W.: An overview of utilizing water-in-diesel emulsion fuel in diesel engines and its potential research study. J. Energy Inst. **87**(4), 273–288 (2014)
16. Ooi, J.B., Ismail, H.M., Swamy, V., Wang, X., Swani, A.K., Raj, J., Rajanren: Graphite oxide nanoparticle as a diesel fuel additive for cleaner emissions and lower fuel consumption. Energy Fuels **30**(2), 1341–1351 (2016)
17. Nadeem, M., Rangkuti, C., Anuar, K., Haq, M.R.U., Tan, I.B., Shah, S.S.: Diesel engine performance and emission evaluation using emulsified fuels stabilized by conventional and gemini surfactants. Fuel **85**(14–15), 2111–2119 (2006)

Chapter 21
Carbon-Dioxide Corrosion in Stainless Steel (304L) Pipes

Check for updates

Surya Poloju, Pankaj Kumar, Purushottam Kumar Singh, Santosh Kr. Mishra, and Amit Kumar Jain

Abstract This research comprehensively addresses challenges associated with carbon dioxide (CO_2) corrosion in 304L stainless steel pipes, particularly under high temperatures and elevated CO_2 partial pressures. Utilizing advanced simulation techniques with COMSOL Multiphysics, the corrosion behavior of 304L stainless steel is systematically investigated. Critical factors influencing corrosion rates, including operating temperature, reference temperature, initial electrolyte potential, and CO_2 partial pressure, are identified. The study proposes practical mitigation strategies, such as employing more resistant materials like duplex stainless steel and applying protective coatings on inner pipe surfaces. The second study focuses on simulating and analyzing CO_2 corrosion in 304L stainless steel pipes, aiming to address research gaps. The results facilitate the design and operation of systems utilizing 304L stainless steel in CO_2 environments more effectively. Implementing measures to reduce PCO_2 can diminish the corrosion rate, extending the system's lifespan.

Keywords CO_2 corrosion · Stainless steels · Corrosion inhibitors · Carbon dioxide transport · Green corrosion inhibitors

S. Poloju · P. Kumar (✉)
Department of Mechanical Engineering, SR University, Warangal 506371, India
e-mail: Pikupankaj82@gmail.com

P. K. Singh
Mechanical Engineering Department, NIT Silchar, Silchar, Assam 788010, India

S. Kr. Mishra
Department of Production Engineering, NIT Trichy, Trichy, India

A. K. Jain
Department of Mechanical Engineering, School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow G40BA, UK

21.1 Introduction

Carbon dioxide corrosion is a significant problem in many industries, including the oil and gas industry, the chemical industry, and the power generation industry. This is because CO_2 can react with water to form carbonic acid, which is a weak acid. Carbonic acid can attack the metal surface and cause corrosion. 304L stainless steel is a type of stainless steel that is commonly used in many industrial applications. It is relatively resistant to corrosion, but it can still be corroded by CO_2. The partial pressure of CO_2 (PCO_2) is a measure of the amount of CO_2 dissolved in a fluid. It is expressed in units of atmosphere (atm) or bar. The higher the PCO_2, the more CO_2 is dissolved in the fluid. PCO_2 has a significant effect on the corrosion rate of metals, including 304L stainless steel. As the PCO_2 increases, the corrosion rate also increases. This is because the higher PCO_2 increases the concentration of carbonic acid in the electrolyte and therefore increases the corrosion rate.Please be aware that your name and affiliation and if applicable those of your co-author(s) will be published as presented in this proof. If you want to make any changes, please correct the details now. Please note that after publication corrections won't be possible. Due to data protection we standardly publish professional email addresses, but not private ones, even if they have been provided in the manuscript and are visible in this proof. If you or your co-author(s) have a different preference regarding the publication of your mail address(s) please indicate this clearly. If no changes are required for your name(s) and affiliation(s) and if you agree with the handling of email addresses, please respond with "Ok".Correctoxy_aqreply_end

The corrosion of materials in carbon dioxide (CO_2) environments is a multi-faceted process influenced by several factors. This systematic literature review aims to synthesize recent studies investigating the corrosion behaviour of stainless steels and other materials in CO_2-rich conditions, as well as strategies for corrosion protection. Corrosion Behavior of Stainless Steels in CO_2 Environments investigated by Chen et al. [1], the CO_2 corrosion behaviour of sensitized 304 and 316 austenitic stainless steels was explored, revealing a significantly higher corrosion rate in sensitized 304 compared to sensitized 316. The alloy's chromium content in 316 was identified as a key factor in forming a more protective passive film against corrosion. Kumar et al. [2] investigated the impact of Monoethanolamine (MEA) concentration in CO_2 absorption processes on stainless steel corrosion, attributing variations in corrosion rates to MEA's reactivity with CO_2. Similarly, Brittan et al. [3] found that while the mechanical properties of 316 stainless steel are minimally affected in supercritical CO_2, corrosion resistance is substantially decreased due to increased CO_2 solubility. Abbasi et al. [4] demonstrated a correlation between temperature and corrosion rate, with higher temperatures leading to increased solubility of CO_2 in supercritical CO_2.

A comprehensive review of carbon dioxide corrosion inhibitors emphasized the complexity of this corrosion process, influenced by factors such as temperature, pH, chloride concentration, and alloy composition. The study highlighted the classification of inhibitors into physical and chemical types, where physical inhibitors create

a barrier between the steel and the corrosive environment, while chemical inhibitors react with the environment to form a protective film. The effectiveness of inhibitors was shown to depend on multiple factors, reinforcing the need for a case-by-case selection approach.

The issue of carbon dioxide corrosion in wet natural gas pipelines was addressed in a study emphasizing the importance of failure investigations. Common procedures include examining the damaged pipe and conducting a chemical analysis of the gas to determine the extent of corrosion damage. Another study focused on the protection against carbon dioxide corrosion in gas production facilities, highlighting the need for a cautious approach in determining media aggression levels and selecting corrosion prevention methods in the presence of aggressive CO_2.

A systematic review of carbon dioxide transport via pipelines acknowledged the established technology for moving CO_2 across large distances but identified challenges such as the risk of leakage, hydrate formation, and corrosion. The analysis suggested further research is required, especially in the fields of pipeline safety and corrosion prevention.

A study investigated chitosan derivatives as green corrosion inhibitors for P110 steel in a CO_2 environment. The results indicated the effectiveness of these derivatives in inhibiting corrosion, with N-carboxymethyl chitosan (NCMC) being the most efficient [5]. Additionally, a paper on the corrosion behaviors of heat-resisting alloys in high-temperature CO_2 environments found that the corrosion rate increases with temperature, recommending the use of stainless steels in applications exposed to high-temperature CO_2 [6].

Several studies delved into innovative approaches and applications. One focused on the design considerations of a subsea shuttle tanker system for transporting liquid CO_2, addressing challenges and strategies related to vessel design, safety protocols, and economic feasibility [7]. Another study aimed at boosting the reaction kinetics in aprotic lithium-CO_2 batteries by employing unconventional phase metal nanomaterials, showcasing potential advancements in energy storage efficiency and the utilization of CO_2 as an energy source [7].

This study investigated the effect of PCO_2 on the corrosion rate of 304L stainless steel in CO_2 containing water. A 1D model of a 304L stainless steel pipe was created in COMSOL Multiphysics and used to simulate carbon dioxide corrosion. The parameters were varied in the simulation including Operating temperature, Reference temperature, Initial electrolyte potential, and Partial pressure of CO_2 [8–10].

This study is based on the corrosion performance of 304L stainless steel and is thoroughly examined through advanced simulation methods utilizing COMSOL Multiphysics. The investigation reveals key factors impacting corrosion rates, including operational temperature, reference temperature, initial electrolyte potential, and CO_2 partial pressure. This sheds insight into the complex dynamics of the corrosion process. The results of this study can be used to design and operate systems that use 304L stainless steel in CO_2 containing environments in a more effective and efficient manner. By taking steps to reduce the PCO_2, the corrosion rate of 304L stainless steel can be reduced and the lifespan of the system can be extended.

Table 21.1 Properties of
304L stainless steel

Properties of 304L stainless steel	Values
Density (kg/m^3)	7,900
Thermal conductivity W/(m·K)	16.2
Specific heat capacity (J/(kg·K))	500
Electrical conductivity (S/m)	1.45×10^6
Young's modulus (GPa)	193
Poisson's ratio	0.27
Yield strength (MPa)	215
Ultimate tensile strength (MPa)	515

21.2 Materials and Methods

21.2.1 Materials

The materials used are as follows:

304L stainless steel: The material properties of 304L stainless steel were obtained from the COMSOL Multiphysics material library. The material library includes data for a wide range of materials, including metals, ceramics, and polymers. The data for 304L stainless steel includes the following properties as mentioned in Table 21.1.

CO_2 Containing Water: The CO_2 containing water was assumed to be an ideal solution. This means that the CO_2 was assumed to be completely dissolved in the water and that there were no interactions between the CO_2 molecules.

21.2.2 Methods

21.2.2.1 Sample Preparation

The methods used are as follows:

1D Model: A 1D model was used to simplify the simulation and reduce the computational cost. The 1D model assumes that the pipe wall is uniform in thickness and that the diffusion of CO_2 into the pipe wall is only in the axial direction.

Boundary Conditions: The boundary conditions applied to the model represent realistic conditions that might be encountered in practice. The constant temperature applied to the outer surface of the pipe wall represents the temperature of the surrounding environment. The constant concentration of CO_2 applied to the inner surface of the pipe wall represents the concentration of CO_2 in the CO_2 containing water. The constant flow rate of CO_2 containing water applied to the inlet of the pipe represents the flow rate of the fluid flowing through the pipe.

Table 21.2 Process
parameters for corrosion test

Process parameters	Values
Operating temperature (^{0}C)	20, 25, 30, 35
Reference temperature (^{0}C)	25, 30, 35
Initial electrolyte potential (V)	0.5, 1, 1.5
Partial pressure of CO_2 (Bar)	1, 3, 5

Simulation Parameters: The simulation parameters were varied to investigate the effect of different operating conditions on the corrosion rate. The operating temperature was varied to represent the range of temperatures that might be encountered in practice. The reference temperature was varied to represent the range of temperatures at which the thermodynamic properties of the metal and the electrolyte are known. The initial electrolyte potential was varied to represent the range of potentials that might be encountered in practice. The partial pressure of CO_2 was varied to investigate the effect of CO_2 concentration on the corrosion rate.

Corrosion Rate Calculation: The corrosion rate was calculated by measuring the thickness loss of the pipe wall at the end of the simulation. The thickness loss was calculated by comparing the initial thickness of the pipe wall to the thickness of the pipe wall at the end of the simulation.

The COMSOL simulation was a valuable tool for investigating the effect of PCO_2 on the corrosion rate of 304L stainless steel in CO_2 containing water. The simulation results showed that the corrosion rate increases with increasing PCO_2. This information can be used to design and operate systems that use 304L stainless steel in CO_2 containing environments in a more effective and efficient manner.

This study used a COMSOL Multiphysics simulation to investigate the effect of partial pressure of CO_2 (PCO_2) on the corrosion rate of 304L stainless steel in CO_2 containing water. A 1D model of a 304L stainless steel pipe was created in COMSOL Multiphysics and used to simulate carbon dioxide corrosion. The process parameters for the corrosion test are presented in Table 21.2 whereas the corrosion Rate-PCO_2 at different pressure are presented in Table 21.3 and are varied as follows:

The mathematical relation between the considered factors:

Corrosion rate (CR) vs. T: CR could be related to T exponentially, like CR = k * exp(a * T), where k and a are constants determined experimentally.

CR vs. P_CO_2: A similar exponential relationship might hold, like CR = b * exp (c * P_CO_2).

CR vs. E_i: This might be a linear or nonlinear function, depending on the specific system.

Table 21.3 Corrosion Rate-PCO$_2$ at different pressure

Operating Temperature (^{0}C)	Reference temperature of Fe (^{0}C)	Reference temperature of H$_2$ (^{0}C)	Initial electrolyte potential (V)	Corrosion Rate (mm/year)		
				At 1 bar pressure	At 3 bar pressure	At 5 bar pressure
20	25	25	0.5	0.1	0.3	0.5
20	25	25	1	0.2	0.6	1
20	25	25	1.5	0.3	0.9	1.5
20	30	30	0.5	0.15	0.35	0.55
20	30	30	1	0.3	0.7	1.1
20	30	30	1.5	0.45	1.05	1.65
20	35	35	0.5	0.2	0.4	0.6
20	35	35	1	0.4	0.8	1.2
20	35	35	1.5	0.6	1.2	1.8
25	25	25	0.5	0.2	0.4	0.6
25	25	25	1	0.4	0.8	1.2
25	25	25	1.5	0.6	1.2	1.8
25	30	30	0.5	0.25	0.5	0.7
25	30	30	1	0.5	1	1.4
25	30	30	1.5	0.75	1.5	2.1
25	35	35	0.5	0.3	0.6	0.8
25	35	35	1	0.6	1.2	1.6
25	35	35	1.5	0.9	1.8	2.4

21.3   Introduction

The results of the simulation showed that the corrosion rate of 304L stainless steel increases with increasing PCO$_2$. The highest corrosion rate occurred at the operating temperature of 25 degrees Celsius and the partial pressure of CO$_2$ of 5 bar.

The results of this study can be used to design and operate systems that use 304L stainless steel in CO$_2$ containing environments in a more effective and efficient manner. By taking steps to reduce the PCO$_2$, the corrosion rate of 304L stainless steel can be reduced and the lifespan of the system can be extended.

21.3.1   Effects of Partial (at 1 bar) Pressure on pH, Concentration, and Corrosion Rate

Figure 21.1 shows the relationship between three variables such as pH, partial pressure of CO$_2$ (PCO$_2$), and corrosion rate. There are three sets of graphs, each representing a different temperature (293 K, 323 K, and 353 K). Each graph shows two

Fig. 21.1   Effects of partial (at 1 bar) pressure **a** concentration deviation **b** corrosion rate **c** pH at surface, **d** bulk material pH

lines that include one for the concentration of dissolved CO_2 (CO_3-2 + HCO_3- + H_2CO_3) and one for the concentration of Fe2$^+$.

From these results, it is observed that pH decreases as PCO_2 increases at all temperatures. The decrease is steeper at higher temperatures. Whereas PCO_2 increases as pH decreases at all temperatures and the increase is steeper at higher temperatures. The corrosion rate increases as PCO_2 increases at all temperatures [11–13]. The increase is steeper at higher temperatures. The results seem to be inversely related to pH, but the relationship is not as clear as with PCO_2. These results suggest that pH and temperature are important factors influencing the corrosion of stainless steel in CO_2-saturated environments. Higher temperatures and lower pH values lead to higher corrosion rates. The concentration of dissolved CO_2 also seems to play a role, but its relationship with corrosion rate is more complex.

21.3.2 Effects of Partial (at 3 bar) Pressure on pH, Concentration, and Corrosion Rate

Figure 21.2 shows the relationship between three variables such as pH, partial pressure of CO_2 (PCO_2), and corrosion rate. There are three sets of graphs, each representing a different temperature (293 K, 323 K, and 353 K). From these results, it is observed that all concentration levels except $Fe2^+$ decrease as distance from the surface increases. The CO_2 concentration shows the steepest decrease, followed by H_2CO_3 and HCO_3^-. $Fe2^+$ concentration initially increases then plateaus or slightly decreases with distance. The higher temperatures generally lead to higher concentrations of CO_2, H_2CO_3, and HCO_3^- near the surface. It is also observed that lower pH values at the surface, higher corrosion rates, corrosion rate is highest near the surface and decreases with distance. The higher temperatures lead to higher corrosion rates at all distances. The pH at the surface is lower than pH in the bulk solution. Higher temperatures lead to lower pH values at the surface. The results suggest that temperature plays a significant role in influencing corrosion rate in CO_2-saturated environments. Higher temperatures lead to faster corrosion due to increased CO_2 concentrations and lower pH values near the surface.

(a)   Concentrations
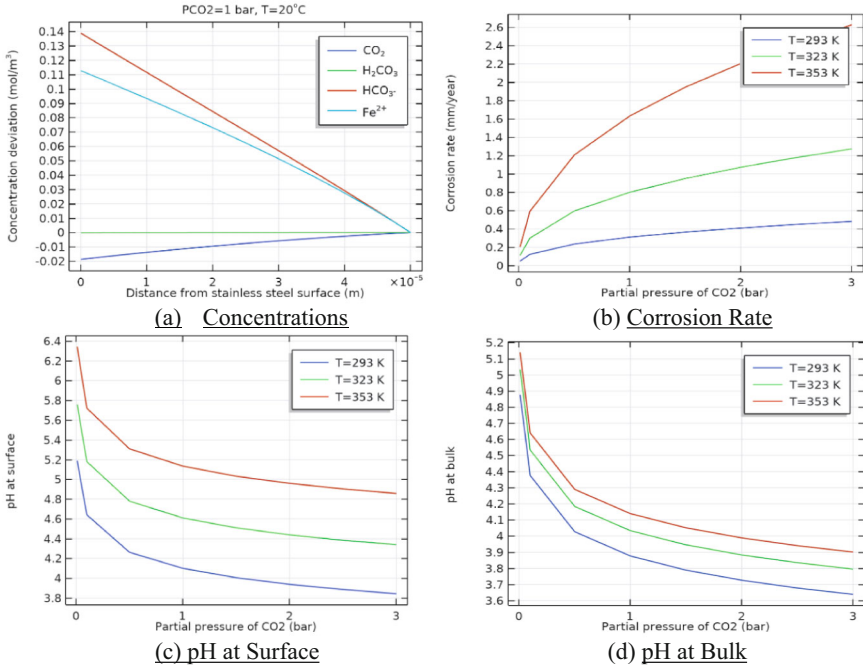
(b) Corrosion Rate

(c) pH at Surface

(d) pH at Bulk

Fig. 21.2 Effects of partial (at 3 bar) pressure **a** Concentration Deviation **b** corrosion rate **c** pH at surface, **d** bulk material pH

21.3.3   Effects of Partial (at 5 bar) Pressure on pH, Concentration, and Corrosion Rate

From these results, it is observed that Fig. 21.3 confirms a clear inverse relationship between pH and partial pressure of CO_2 at the surface. As pH increases, the partial pressure of CO_2 decreases. This trend holds true across all three temperatures. The influence of temperature is also evident. At a given pH, the partial pressure of CO_2 is higher at higher temperatures. This is seen by the lines for 323 K and 353 K being above the line for 293 K. The pH range on the y-axis is approximately from 4.2 to 6.4. The partial pressure of CO_2 range on the x-axis is from 0 to 3 bar.

These findings suggest that pH and temperature are key factors influencing the partial pressure of CO_2 at surfaces. The significant role of temperature in modulating the pH-CO_2 relationship suggests that the higher temperatures lead to steeper decreases in pH with increasing CO_2 pressure.

Furthermore, it can be concluded that the increase in corrosion rate with increasing PCO_2 can be explained by the following factors:

(a)   Concentrations
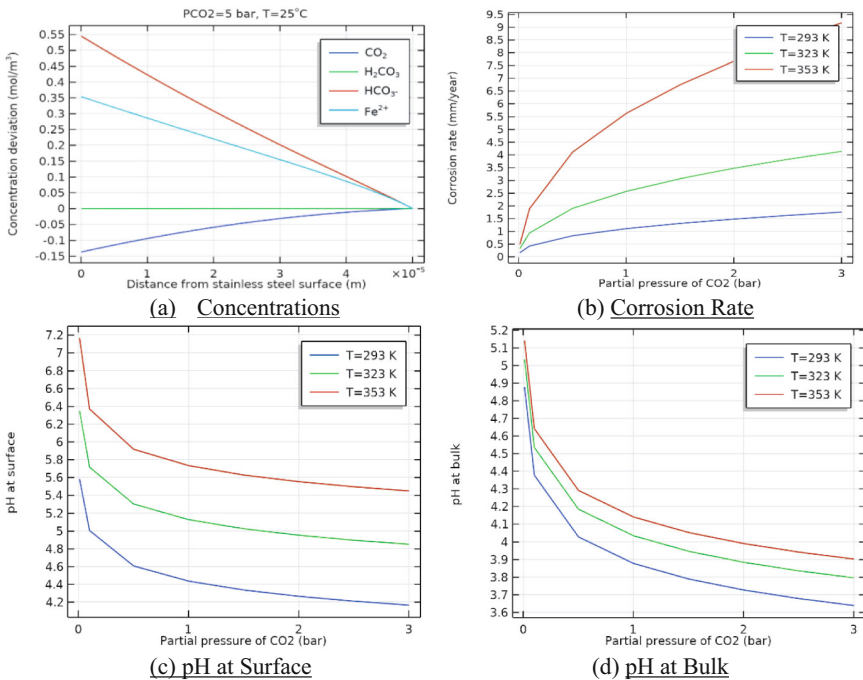
(b)   Corrosion Rate

(c) pH at Surface

(d) pH at Bulk
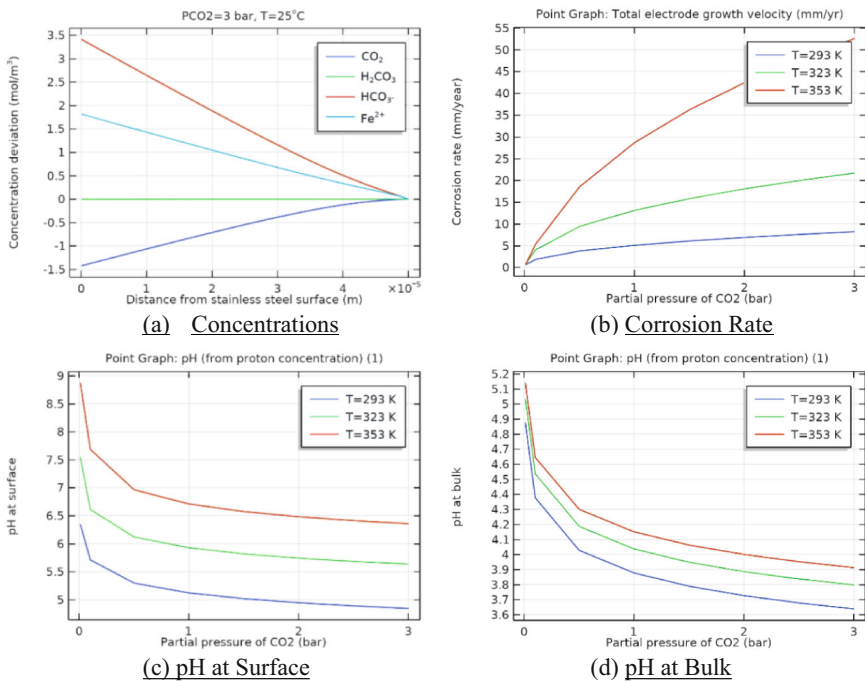
Fig. 21.3 Effects of partial (at 5 bar) pressure **a** concentration deviation **b** corrosion rate **c** pH at surface, **d** bulk material pH

- Increased solubility of CO_2 in water: As the PCO_2 increases, more CO_2 dissolves in water to form carbonic acid. Carbonic acid is a weak acid that can attack the metal surface and cause corrosion.
- Increased rate of diffusion of CO_2 into the metal: The higher the PCO_2, the faster the diffusion of CO_2 into the metal. This is because the PCO_2 is the driving force for diffusion.
- Increased rate of reaction between CO_2 and the metal: Once CO_2 has diffused into the metal, it can react with the metal to form corrosion products. The higher the PCO_2, the faster the rate of this reaction.
- The highest corrosion rate occurred at the operating temperature of 25 degrees Celsius. This is because the solubility of CO_2 in water decreases with increasing temperature, which leads to a higher concentration of CO_2 at the metal surface. The corrosion morphology of 304L stainless steel in CO_2 containing water is typically localized pitting. This means that the corrosion is concentrated in small, localized areas on the metal surface. This is because the corrosion process is autocatalytic, meaning that the products of the corrosion reaction accelerate the reaction itself.

21.4   Conclusions

The findings of this study underscore the critical influence of the partial pressure of CO_2 (PCO_2) on the corrosion rate of 304L stainless steel, with the highest corrosion rate observed at 25 degrees Celsius and a PCO_2 of 5 bars. These results offer valuable insights for designing and operating systems utilizing 304L stainless steel in CO_2-rich environments more effectively. To mitigate corrosion and extend system lifespan, practical recommendations include employing corrosion inhibitors, lowering the environmental temperature, increasing fluid flow rates, or considering alternative materials such as duplex stainless steel.

Implementing these measures not only reduces maintenance and replacement costs for 304L stainless steel equipment but also contributes to environmental protection by curbing the release of harmful pollutants. Furthermore, regular corrosion testing is emphasized as a crucial practice to monitor the condition of 304L stainless steel equipment in CO_2 containing environments. This proactive approach facilitates early detection of potential corrosion issues, enabling timely corrective actions to prevent further damage. In essence, adopting these strategies based on the study's findings ensures both economic efficiency and environmental sustainability in industrial applications employing 304L stainless steel.

References

1. Chen, Z., Li, W., Ma, C., Sun, J., He, L.: Sensitization-induced pitting corrosion behavior of austenitic stainless steels in CO_2-containing environments. J. Market. Res. **8**(6), 5055–5064 (2019)
2. Kumar, K. E., Sharma, C. M., Balasubramanian, M.: Effect of monoethanolamine concentration on the corrosion of stainless steel in CO_2 absorption processes. J. CO_2 Util., 34, 254–263 (2019)
3. Brittan, M.R., Johnsen, R., Neeson, M.J.: Effects of supercritical CO_2 on the mechanical properties and corrosion resistance of 316L stainless steel. Corros. Sci.. Sci. **139**, 256–268 (2018)
4. Abbasi, F., Shamsuddin, S., Juwiran, C.W.: Effect of temperature on the corrosion of carbon steel in supercritical CO_2. J. Mater. Sci. Mater. Electron. **33**(5), 4573–4584 (2022)
5. Vignal, M. C., Zhang, J., Li, J.: A review on CO_2 corrosion inhibitors. Corros. Rev., **38**(6), 379–405
6. Zhang, S., Li, W., Liu, X., Li, G., Wang, Y.: Failure investigation and analysis of wet natural gas pipelines due to CO_2 corrosion. Eng. Fail. Anal. **102**, 374–386 (2019)
7. Norrlöf, K., Högfors, C.: Mitigation and prevention of CO_2 corrosion in gas production facilities. Procedia Eng. **211**, 395–402 (2018)
8. Garcia, J. J., Gibbins, J., Kokotsi, R: Pipeline transport of CO_2: A review of challenges and innovations. J. CO_2 Util., 34, 207–220 (2019)
9. Zhang, R., Li, G., Liu, L., Yu, W.: Design considerations of a subsea shuttle tanker system for transporting liquid CO_2. Ocean Eng. **199**, 107011 (2020)
10. Wang, R., Wang, F., Liu, X., Xu, Y., Wang, F.: Boosting the reaction kinetics in aprotic lithium-CO_2 batteries by employing unconventional phase metal nanomaterials. J. Mater. Chem. A **10**(34), 17056–17066 (2022)
11. Bai, Haitao, Yongqing Wang, Yun Ma, Qingbo Zhang, Ningsheng Zhang: Effect of CO_2 partial pressure on the corrosion behavior of J55 carbon steel in 30% crude oil/brine mixture. Materials **11**(9), 1765 (2018)
12. Elgaddafi, R., Naidu, A., Ahmed, R., Shah, S., Hassani, S., Osisanya, S.O., Saasen, A.: Modeling and experimental study of CO_2 corrosion on carbon steel at elevated pressure and temperature. J. Nat. Gas Sci. Eng. 27, 1620–1629 (2015)
13. Palumbo, Gaetano, Kamila Kollbek, Roma Wirecka, Andrzej Bernasik, Marcin Górny.: Effect of CO_2 partial pressure on the corrosion inhibition of N80 carbon steel by gum arabic in a CO_2-water saline environment for shale oil and gas industry. Materials **13**(19), 4245 (2020)

Chapter 22
A Study on Thermophysical and Dielectric Characteristics of Nanofluids

Salman Basha Sheik, Praveena Devi Nagireddy, and Kiran Kumar Kupireddi

Abstract This review paper examines the thermal, physical, and dielectric characteristics of nanofluids for heat energy transfer and dielectric applications. Different theoretical and experimental studies were carried out to know the importance of oxide-based nanofluids for thermal and electrical transformer applications. Conventional methods make it difficult to explain the thermal conductivity, viscosity, and dielectric characteristics of nanofluids. The sort of nanomaterial, particle size, percentage concentration, and synthesis methods play crucial roles in changes in the thermal, physical, and electrical properties of nanofluids. The present study gives information related to synthesis techniques of nanofluids, experimental methods, and mathematical models for calculation of thermal, physical, and dielectric properties.

Keywords Nanofluids · Nanoparticles · Synthesis · Characterization · Thermal conductivity · Viscosity · Dielectric properties

S. B. Sheik (✉)
Department of Mechanical Engineering, Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh 534101, India
e-mail: skbasha@sasi.ac.in; 2305c30001@sru.edu.in

S. B. Sheik · P. D. Nagireddy
Department of Mechanical Engineering, SR University, Warangal, Telangana 506371, India

K. K. Kupireddi
Department of Mechanical Engineering, National Institute of Technology Warangal, Warangal, Telangana 506004, India

S. B. Sheik
Research Scholar, SR University, Warangal, Telangana 506371, India

22.1 Preparation of Nanofluids

Preparations of nanoparticles-based fluids are called nanofluids. Combining nano powdered materials and the process of preparing nanofluids should come before any testing. Preparation of nanofluids is not a simple method like mixing of particles in fluids. It requires some considerations like percentage concentration, type of base fluids, preparation methods, agglomeration, suspension, etc. Nanofluids could be prepared by submerging nanoparticles in base fluids.

22.1.1 Single Step Method

Akoh et al. [1] proposed the Vacuum Evaporation onto a Running Oil Substrate (VEROS) method, Wagener et al. [2] and Eastman et al. [3] explained the process of modified VEROS methods for the synthesis of metal-based suspended nanofluids using single step method. Zhu et al. [4] explained that microwave irradiation will impact the properties of copper nanofluids. They synthesized nanofluids using PVD process and pure Cu nanoparticles are prepared by this method by avoiding storage, dispersion, transportation, and drying processes. Consequently, the mixture of nanoparticles is reduced. The primary drawback of the one-step technique is that the excess reactive agent remains in the nanomaterial based fluids.

Lo et al. [5] produced Cu, Ni, and Ag nanofluids using submerged arc nanoparticle synthesis method. Manna [6] Nanofluids are quasi-single phase fluids that contain a colloidal dispersion of ultrafine or nanometric-sized ceramic or metallic particles that are stable.

Babar et al. [7] explained the process of synthesis of nanofluids is made by parallel creation and dispersion. Pulsed wire evaporation (PWD) is an encouraging process.

1. Zhu et al. [4] provided a graphical illustration of the nanofluids preparation methods in their study on the preparation and challenges of titania (TiO) nanofluids as shown in Fig. 22.3.

Ali et al. [8] presented that nanofluid can be created directly using a liquid chemical procedure or vapor deposition method (VDT). He elucidated the process of preparing nanofluids by using a diagrammatic representation in schematic Fig. 22.1.

Wang et al. [9] explain that one-step method has the benefit of minimizing nanoparticle aggregation, but it also has the drawback of only working with fluids with low vapor pressure.

Fig. 22.1 Preparation of nanofluids. **a** Single step method. **b** Two step method [8]
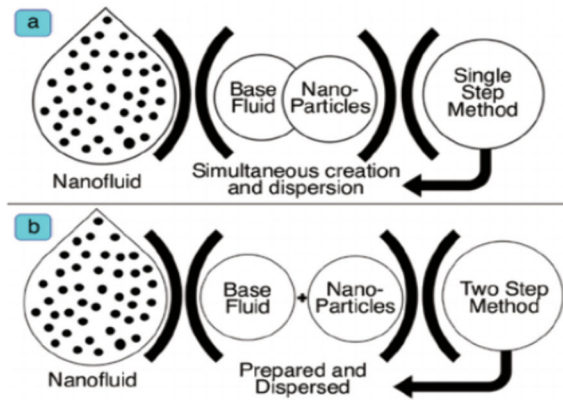
22.1.2   Two Step Method

This method includes first creating nanoparticles and then dispersing them in the basic fluids. Ultrasonic equipment is frequently employed to efficiently disperse particles and reduce their dimensions.

Eastman et al. [10], Wang et al. [11], and Lee et al. [12] used this technique to produce nanofluids with an Al_2O_3 basis for thermal transfer applications. The two-step process was used by Murshed et al. [13] to produce TiO_2 suspension in water. Gold, silver, carbon nanotubes, and silica are among the other nanomaterials listed in the research studies.

Srdic et al. [14] gave the examples of colloid chemistry processes that are able to create nanomaterials with precise geometries, densities, porosities, charges, and surface chemistries. Cao et al. [15] proposed the size and shape-controlled creation of ambient-temperature nanoparticles is still another method. The synthesis method impacts the structural properties of nanoparticles, including their average size, shape, and size distribution, and it might make precise control possible. It is challenging to measure these properties of nanoparticle suspended liquids.

To increase transformer oil—Copper and water-Cu nanofluids stability, Xuan and Li [16] used salt and $C_{18}H_{34}O_2$ as the dispersants. The two-step process was used by Li et al. [17] to create Cu/H_2O nanofluids with nanoparticle sizes that generally range from 1 to 100 nm. Khedkar et al. [18] the nanofluid was prepared utilizing a two-step process that involved creating TiO_2 nanoparticles using an ultrasonicator. Suresh et al. [19] used a hydro metallurgical reduction to create Al_2O_3-Cu hybrid particles.

22.2   Experimental Methods

22.2.1   *Thermal Conductivity Measurements*

Nanofluids have a notable characteristic of enhanced thermal conductivity in contrast to the basic fluid in isolation. The ability of nanofluids to conduct heat has been demonstrated and measured by Kestin and Wakeham [20] using the transient hot wire approach, temperature oscillation technique was used by Das et al. [21] as shown in Fig. 22.2.

Nagasaka and Nagashima [22] suggested modifying the hot wire cell and electrical system by covering the hot wire with an epoxy adhesive that provides excellent electrical insulation and heat conduction.

Choi et al. [23] explained the concept of evaluating the oxide-based nanofluids' thermal conductivity using a hot wire method. In an experiment, Xian et al. [24]

Fig. 22.2 Schematic of transient hot-wire experimental setup [26]
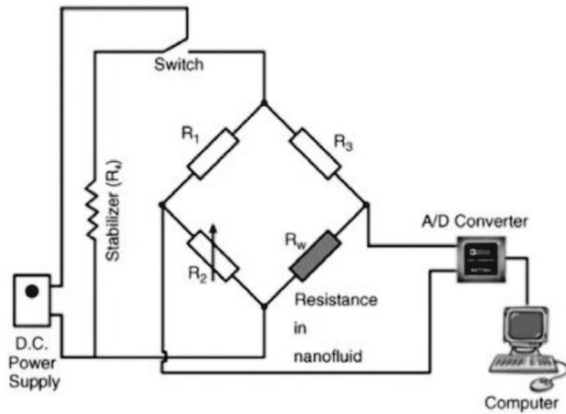
Fig. 22.3 Transient plate source experimental setup
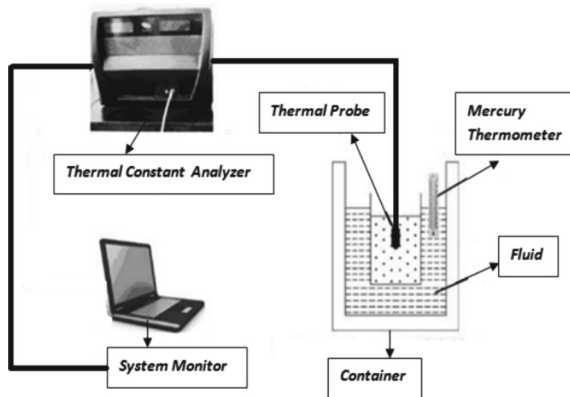
demonstrated how to use graphene and titanium oxide nanoparticles to make a hybrid nanofluid.

Experimental Models

Yarmand et al. [25] proposed that according to reports, a hybrid nanofluid made of graphene nano platelets and silver increases in thermal conductivity as the temperature and particle concentration rise.

Zhu et al. [46] employed the thermal constants analyzer's transient plane source (TPS) concept to assess the nanofluid's heat conductivity and the schematic experimental setup is shown in Fig. 22.3.

Carslaw and Jaeger's [27] model of the temperature around a heat source in an endless line with constant heat output and zero mass. Heating occurs at a constant rate, q, in a one-dimensional, infinitesimally narrow, and infinitely long source known as a "line" source. The following equation can be used to mathematically represent how its temperature changes over time:

$$\Delta T = -\frac{q}{4\pi k} E_i \left(\frac{-r^2}{4Dt} \right) \ \ 0 < t < t_1 \tag{22.1}$$

Here, k represents the material's thermal conductivity surrounding the line, D represents the heat diffusivity of the medium, r is the distance at which temperature is being measured from the line, and E_i represents the exponential integral.

Paul et al. [26] gave information about the experimental setup of Steady-state parallel-plate method schematically as represented in Fig. 22.4.

A small amount of fluid was encased between two parallel, circular pieces of pure copper. Wang et al. [11] have provided a thorough explanation of the setup.

Among the most used steady-state techniques for finding the thermal conductivity of fluids is the cylindrical cell approach. Kurt et al. [28] explained about the equipment and its working operation. Figure 22.5 illustrates a cross-sectional perspective of the

Fig. 22.4 Experimental set up for steady-state parallel-plate method [26]
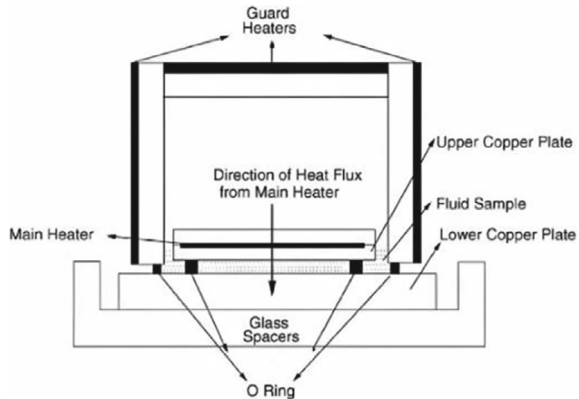
Fig. 22.5 Cross-section of the cylindrical cell equipment [28]
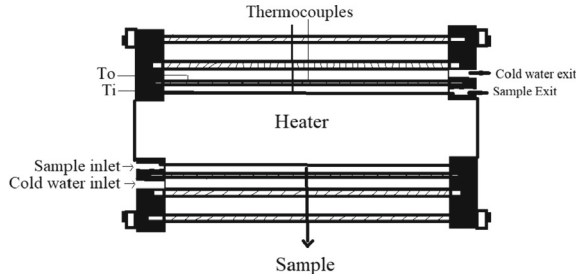
cylindrical cell apparatus. The thermal conductivity of a nanofluid is computed using Fourier's Eq. (22.1) in cylindrical coordinates.

Laser flash analysis (LFA) is a sophisticated method for measuring heat conductivity that employs a laser beam as a source of heat. Gustafsson et al. [29] initially suggested using this technique to determine the heat diffusivity of materials, which consequently allows for the analysis of thermal conductivity. This heat change is measured and recorded using LFM, as depicted in Fig. 22.6.

To forecast the heat conductivity of solutions with solid particles, several models were created. The Maxwell model was developed to explain the effective electrical or heat conductivity of liquid–solid mixtures including spherical particles with low volumetric loading.

$$\frac{K_{nf}}{K_{bf}} = 1 + \frac{3(\alpha - 1)\varphi}{(\alpha + 2) - (\alpha - 1)\varphi} \tag{22.2}$$

Maxwell's model was altered by Hamilton and Crosser to incorporate a shape factor in order to determine the practical thermal conductivity of non-spherical particles.

Fig. 22.6 The laser flash method (LFM) [29]
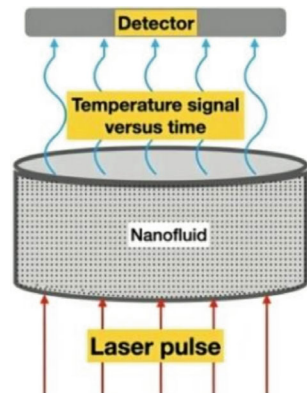
$$\frac{K_{nf}}{K_{bf}} = \frac{\alpha + (n-1) - (n-1)(1-\alpha)\varphi}{\alpha + (n-1) + (1-\alpha)\varphi} \tag{22.3}$$

Jeffrey [30] provided a further popular model used for estimating the heat conductivity of mixtures and composites.

$$\frac{K_{nf}}{K_{bf}} = 1 + 3\beta\varphi + (3\beta^2 + \frac{3\beta^3}{4} + \frac{9\beta^3}{16}\frac{\alpha+2}{2\alpha+3} + \frac{3\beta^4}{2^6} + \dots)\varphi^2 \tag{22.4}$$

where $\alpha = \frac{k_p}{k_{bf}}$ and $\beta = \frac{(\alpha-1)}{(\alpha+2)}$; k_p is the thermal conductivity of nanoparticles, k_{bf} is the thermal conductivity of base fluid, and k_{nf} is the thermal conductivity of nanomaterial-based fluid respectively, φ is the volume percentage of nanoparticles and n is the empirical shape factor.

Hui et al. [31] explained the effective heat conductivity of the binary combination of homogeneous, spherical, and randomly dispersed nanoparticles is predicted using the Bruggeman thermal conductivity model

$$\frac{k_{eff}}{k_f} = \frac{1}{4}[(3\varphi - 1)\frac{k_p}{k_f} + (2 - 3\varphi) + \frac{k_f}{4}\sqrt{\Delta}] \tag{22.5}$$

where k_f—thermal conductivity of base fluids (W/mK), k_p—thermal conductivity of nanoparticles (W/mK). Aberoumand et al. [32] used a formula to convert from mass fraction (w) to volume fraction (φ)

$$\varphi = \frac{w}{w + (1-w)\frac{\rho_{np}}{\rho_{bf}}} \tag{22.6}$$

where ρ_{np} is the density of nanoparticles and ρ_{bf} is the density of base fluid.

Tajik Jamal-Abadi [33] briefed that KD2 Pro is a thermal property measurement device as shown in Fig. 22.7. Following the reading, the controller calculated the heat conductivity utilizing the temperature change (DT)—time information from

Fig. 22.7 KD2 pro-system [33]

$$k = \frac{q(ln ln t_2 - ln ln t_1)}{4\pi(\Delta T_2 - \Delta T_1)} \tag{22.7}$$

where q represents a constant heat rate delivered to an infinitely long and small "line" source. ΔT_1 & ΔT_2 represent the changes in temperature at times t_1 & t_2, respectively.

22.2.2 Viscosity

Mahbubul et al. [34] outlined how fluid viscosity is a crucial aspect of flow. A fluid's viscosity is directly related to pressure drop in laminar flow, convective heat transfer, and the power required for pumping.

22.3 Theoretical Studies

The physical characteristics of nanofluids can be approximated using several equations. Einstein originally formulated the equation for nanofluid viscosity. Based on his theory, the viscous fluid would contain a relatively low volume fraction of spherical particles. Here is the recommended formula:

$$\frac{\mu_{nf}}{\mu_f} = 1 + 2.5\phi \tag{22.8}$$

Here, ϕ represents the volume percentage of the suspended particle, μ_{nf} refers to the viscosity of the nanomaterial-based fluid, and μ_f represents the viscosity of the base fluid.

Batchelor modified Einstein's viscosity equation by introducing the Brownian motion effect. The isotropic dispersion of stiff and spherical nanoparticles was taken into account when developing the model. The following is his model:

$$\mu_{nf} = (1 + 2.5\phi + 6.5\phi^2\mu_f) \tag{22.9}$$

Bicerano et al. [35] gave correlations for volumetric effect of viscosity.

$$\mu_{nf} = \mu_{bf}\left(1 + \eta\phi + k_H\phi^2\right) \tag{22.10}$$

The virial contraction coefficient is represented by the symbol, while the Huggins coefficient is denoted as k_H

Tseng and Lin [36] demonstrated that relative viscosity rises exponentially with content for TiO_2/water nanofluid

$$\mu_{nf} = \mu_{bf} \times 13.47 e^{35.98\varnothing} \tag{22.11}$$

Chen et al. [37] showed that, for up to 10% by volume, there is a link between normalized shear viscosity and basis viscosity of the liquid with an uncertainty of less than 6%. Using data from experiments regression, the researchers computed the relationship shown below.

$$\mu_{nf} = \mu_{bf}(1 + 10.6\phi + (10.6\phi)^2) \tag{22.12}$$

Nguyen et al. [38] investigated the relationship between dynamic viscosities, temperature and particle size of water/Al_2O_3 and water/CuO mixes experimentally

$$\frac{\mu_{nf}}{\mu_f} = (2.1275 - 0.0215T + 0.00027T^2) \tag{22.13}$$

Masoumi et al. [39] created a novel mathematical model using Brownian motion to forecast the physical properties of nanofluids. This model can forecast the behavior of a combination made up of two different fluids.

$$\mu_{nf} = \mu_{bf}\left(\frac{\rho P V_B d_p^2}{72C\delta}\right) \tag{22.14}$$

The variables V_B and C are dependent on temperature, while δ represents the distance between the nanoparticles. Additionally, d_p refers to the width of the particles, while ρ_P represents the density of the nanoparticles.

Masoud Hosseini et al. [40] developed a group model without dimensions to calculate the nanofluid's viscosity while accounting for the volume concentration, nanoparticle size, capping layer effect, and temperature.

$$\mu_{nf} = \mu_{bf}.exp[m + \alpha\left(\frac{T}{T_0}\right) + \beta(\phi_h) + \gamma\left(\frac{d}{1-r}\right)] \tag{22.15}$$

where α, β, and γ are realistic constants ascertained from practical information.

Akbari et al. [41] examined the viscosity dependence of SiO_2/ethylene glycol (EG) nanofluids on temperature and concentration using the following correlation.

$$\mu_{nf} = [233.2713\phi^{0.8623}(1/T)^{0.8623} - 2.6698\phi^{0.4821} + 0.9145]. \tag{22.16}$$

22.3.1 Dielectric Properties

Since 2009, the electrical features of nanofluids have been identified as empirically and theoretically studied. Investigations have shown that, in comparison to their basic fluids, they exhibit strongly conductive qualities [42]. Yu et al. [43] investigated the dielectric constants of TiO_2, $(CH_2OH)_2$, and $C_3H_8O_2$ nanofluids by varying the volume concentrations, frequency, and temperatures.

Subramaniyan et al. [44] presented that the dielectric properties for TiO_2-based water nanofluid were evaluated using the Maxwell–Garnett equation:

$$E_{eff} = E_f[1 + \{(3f_v[(E_p - E_f)/(E_p - 2E_f)]\}/\{1 - [f_v\{(E_p - 2E_f)/(E_p + 2E_f)\}] \tag{22.17}$$

where E_{eff}—Effective dielectric constant of the nanofluid, E_f—Dielectric constant of the base fluid, f_v—Volume percentage, and E_p—Dielectric constant of the particle

Thabet et al. [45] used a variety of nanoparticles (ZnO, MgO, Al_2O_3, TiO_2, SiO_2, graphite, $LiTaO_3$ & Fe_3O_4) to study the electrical breakdown voltage of transformer oil nanofluids.

Each nanoparticle possesses a distinct attribute that directly influences the period of rest constant and subsequently impacts the breakdown strength of the oil. One can compute the period of rest constant using the equation

$$r = \frac{2\varepsilon_1 + \varepsilon_2}{2\sigma_1 + \sigma_2} \tag{22.18}$$

The relaxation time constant is denoted as r, while the transformer oil's permittivity and conductivity are represented by ε_1 & σ_1, respectively. Similarly, the nanoparticle's permittivity and conductivity are denoted as ε_2 & σ_2, respectively.

22.4 Conclusions

A brief review on study of thermal, physical, & dielectric characteristics of nanofluids was conducted in this work, which showed that different experimental and mathematical models were still quite limited. Some research groups have observed a significant rise in the heat-transmittance of nanofluids when compared to basic fluids. If an alternative method could be used to complete the measurements, the reliability of the results could be verified. The repeatability of the experiments could also be explained by this.

An investigation of the temporal evolution of nanofluids is the ability to conduct heat could reveal hitherto unreported distinctive characteristics. The issue of nanofluid stability has posed a persistent challenge in this field of research. The

researchers will employ a diverse range of research methodologies and mathematical models to authenticate the results of their studies.

References

1. Akoh, H., Tsukasaki, Y., Yatsuya, S., Tasaki, A.: Magnetic properties of ferromagnetic ultrafine particles prepared by vacuum evaporation on running oil substrate. J. Crystal Growth **45**, 495–500 (1978)
2. Wagener, M., Murty, B.S., Gunther, B.: Preparation of metal nano suspensions by high pressure DC-sputtering on running liquids. In: Komarnenl, S., Parker, J.C., Wollenberger, H.J. (eds.) Nanocrystalline and Nanocomposite Materials II, vol. 457, pp. 149–154. Materials Research Society, Pittsburgh, PA (1997)
3. Eastman, J.A., Choi, U.S., Li, S., Thompson, L.J., Lee, S.: Enhanced thermal conductivity through the development of nanofluids. In: Volume 457 of Materials Research Society Symposium—Proceedings, pp. 3–11. Materials Research Society, Pittsburgh, PA, USA, Boston, MA, USA (1997)
4. Zhu, H., Lin, Y., Yin, Y.: A novel one-step chemical method for preparation of copper nanofluids. J. Colloid Interface Sci. **227**, 100–103 (2004)
5. Lo, C.-H., Tsung, T.-T., Chen, L.-C.: Ni nanomagnetic fluid prepared by submerged arc nano synthesis system (SANSS). JSME Int. J., Ser. B: Fluids Therm. Eng. **48**(4), 750–755 (2006)
6. Manna, I.: Synthesis, characterization and application of nanofluid. An overview. J. Indian Inst. Sci. **89**(1), 21–33 (2009)
7. Babar, H., Sajid, M.U., Ali, H.M.: Viscosity of hybrid nanofluids: a critical review. Therm. Sci. **23**(3B), 1713–1754 (2019)
8. Ali et al.: Preparation techniques of TiO$_2$ Nanofluids and challenges: a review. Appl. Sci. **8**(4), 587 (2018)
9. Wang, X.-Q., Mujumdar, A.S.: A review on nanofluids—Part II: Experiments and applications. Braz. J. Chem. Eng. **25**(4), 631–648 (2008)
10. Eastman, J.A., Choi, S.U.S., Li, S., Yu, W., Thompson, L.J.: Anomalously increased effective thermal conductivities of ethylene glycol-based nanofluids containing copper nanoparticles. Appl. Phys. Lett. **78**(6), 718–720 (2001)
11. Wang, X., Xu, X., Choi, S.U.S.: Thermal conductivity of nanoparticle-fluid mixture. J. Thermophys. Heat Transf. **13**(4), 474–480 (1999)
12. Lee, S., Choi, S.U.S., Li, S., Eastman, J.A.: Measuring thermal conductivity of fluids containing oxide nanoparticles. J. Heat Transf. **121**, 280–289 (1999)
13. Murshed, S.M.S., Leong, K.C., Yang, C.: Enhanced thermal conductivity of TiO$_2$—Water based nanofluids. Int. J. Therm. Sci. **44**(4), 367–373 (2005)
14. Srdic, V., Winterer, M., Moller, A., Miehe, G., Hahn, H.: Nanocrystalline Zirconia surface-doped with alumina: chemical vapor synthesis, characterization, and properties. J. Am. Ceramic Soc. **84**, 2771–2776 (2001)
15. Cao, H., Qian, X., Gong, Q., Du, W., Ma, X., Zhu, Z.: Shape- and size-controlled synthesis of nanometer ZnO from a simple solution route at room temperature. Nanotechnology **17**, 3632–3636 (2006)
16. Xuan, Y., Li, Q.: Heat transfer enhancement of nanofluids. Int. J. Heat Fluid Transf. **21**, 58–64 (2000)
17. Li, X., Zhu, D., Wang, X.: Evaluation on dispersion behavior of the aqueous copper nanosuspensions. J. Colloid Interface Sci. **310**(2), 456–463 (2007)
18. Khedkar, R.. S.., Shrivatsava, N.., Sonawane, S.. S.., Wasewar, K.. L..: Experimental investigations and theoretical determination of thermal conductivity and viscosity of TiO$_2$–ethylene glycol nanofluid. Int. Commun. Heat Mass Transf. **73**, 54–61 (2016)

19. Suresh, S., Venkitaraj, K.P., Selvakumar, P., Chandrasekar, M.: Synthesis of Al_2O_3–Cu/water hybrid nanofluids using two step method and its thermo physical properties. Colloids Surf.: Physiochem. Eng. Aspects **388**(1–3), 41–48 (2011)
20. Kestin, J., Wakeham, W.A.: Contributions to the theory of the transient hot-wire technique for thermal conductivity measurements. Physica A **92**, 102–116 (1978)
21. Das, S.K., Putra, N., Roetzel, W.: Pool boiling characteristics of nano-fluids. Int. J. Heat Mass Transf. **46**(5), 851–862 (2003)
22. Nagasaka, Y., Nagashima, A.: Absolute measurement of the thermal conductivity of electrically conducting liquids by the transient hot-wire method. J. Phys. E: Sci. Instrum. **14**, 1435–1440 (1981)
23. Choi, S., Zhang, Z., Yu, W., Lockwood, F., Grulke, E.: Anomalous thermal conductivity enhancement in nanotube suspensions. Appl. Phys. Lett. **79**(14), 2252–2254 (2001)
24. Xian, H.W., Sidik, N.A.C., Saidur, R.: Impact of different surfactants and ultrasonication time on the stability and thermophysical properties of hybrid nanofluids. Int. Commun. Heat Mass Transf. **110** (2020)
25. Yarmand, H., Gharehkhani, S., Ahmadi, G., Shirazi, S.F.S., Baradaran, S., Montazer, E., Zubir, M.N.M., Alehashem, M.S., Kazi, S.N., Dahari, M.: Graphene nanoplatelets—silver hybrid nanofluids for enhanced heat transfer. Energy Convers. Manag. **100**, 419–428 (2015)
26. Paul, G., Chopkar, M., Manna, I., Das, P.K.: Techniques for measuring the thermal conductivity of nanofluids: a review. Renew. Sustain. Energy Rev. **14**, 1913–1924 (2010)
27. Carslaw, H.S., Jaeger, J.C.: Conduction of Heat in Solids, 2nd edn., Oxford, London (1959)
28. Kurt, H.., Kayfeci, M..: Prediction of thermal conductivity of ethylene glycol–water solutions by using artificial neural networks. Appl. Energy **86**, 2244–2248 (2009)
29. Gustafsson, S.E.: Transient plane source techniques for thermal conductivity and thermal diffusivity measurements of solid materials. Rev. Sci. Instrum. **62**, 797–804 (1991)
30. Jeffrey, D.J.: Conduction through a random suspension of spheres. Proc. R. Soc. London, Ser. A **335**, 355–367 (1973)
31. Hui, P.M., Zhang, X.A., Markworth, J., Stroud, D.: Thermal conductivity of graded composites: Numerical simulations and an effective medium approximation. J. Mater. Sci. **34**, 5497–5503 (1999)
32. Aberoumand, S., Jafarimoghaddam, A., Moravej, M., Aberoumand, H., Javaherdeh, K.: Investigation of the rheological characteristics of silver-heat transfer oil nanofluid and suggesting two empirical based correlations for heat conductivity & viscosity of oil based nanofluids. Appl. Therm. Eng. **101**, 362–372 (2016)
33. Tajik Jamal-Abadi, M., Zamzamian, A.H.: Thermal conductivity of Cu and Al-water nanofluids. IJE Trans. B: Appl. **26**(8), 821–828 (2013)
34. Mahbubul, I.M., Saidur, R., Amalina, M.A.: Latest developments on the viscosity of nanofluids. Int. J. Heat Mass Transf. **55**, 874–885 (2012)
35. Bicerano, J., Douglas, J.F., Brune, D.A.: Model for the viscosity of particle dispersions. J. Macromol. Sci. Rev. Macromol. Chem. Phys. **39**(4), 561–642 (1999)
36. Tseng, W.J., Lin, C.L.: Effect of dispersants on rheological behavior of $BaTiO_3$ powders in ethanol–isopropanol mixtures. Mater. Chem. Phys. **80**(1), 232–238 (2003)
37. Chen, H., Ding, Y., He, Y., Tan, C.: Rheological behavior of ethylene glycol based titania nanofluids. Chem. Phys. Lett. **444**(4–6), 333–337 (2007)
38. Nguyen, C., Desgranges, F., Galanis, N., Roy, G., Mare, T., Boucher, S., Anguemintsa, H.: Viscosity data for Al_2O_3–water nanofluid—hysteresis: is heat transfer enhancement using nanofluids reliable? Int J. Therm. Sci. **47**(2), 103–111 (2008)
39. Masoumi, N., Sohrabi, N., Behzadmehr, A.: A new model for calculating the effective viscosity of nanofluids. J. Phys. D Appl. Phys. **42**(5), 055501 (2009)
40. Masoud Hosseini, S., Moghadassi, A.R., Henneke, D.E.: A new dimensionless group model for determining the viscosity of nanofluids. J. Therm. Anal. Calorim. **100**(3), 873–877 (2010)
41. Akbari, M., Afrand, M., Arshi, A., Karimipour, A.: An experimental study on rheological behavior of ethylene glycol based nanofluid: proposing a new correlation as a function of silica concentration and temperature. J. Mol. Liquids **233**, 352–357 (2017)

42. Ganguly, S., Sikdar, S., Basu, S.: Powder Technol. **196**, 326–330 (2009)
43. Yu, W., Xie, H.: A review on nanofluids: preparation, stability mechanisms, and applications. J. Nanomater. **2012**, 1–17 (2012)
44. Subramaniyan, A., Sukumaran, L.P., Ilangovan, R.: Investigation of the dielectric properties of TiO_2 nanofluids. J. Taibah Univ. Sci. **10**, 403–406 (2016)
45. Thabet, A., Allam, M., Shaaban, S.A.: Investigation on enhancing breakdown voltages of transformer oil nanofluids using multi-nanoparticles technique. IET Gener. Transm. Distrib. **12**(5), 1171–1176 (2018)
46. Zhu, D.., Li, X.., Wang, N.., Wang, X.., Gao, J.., Li, H..: Dispersion behavior and thermal conductivity characteristics of Al_2O_3-H_2O nanofluids. Curr. Appl. Phys. **9**(1), 131–139 (2009)

Chapter 23
Design and Analysis of Counter-Phase Four Wheel Steering System to Reduce the Turning Radius

Subbarao Medabalimi, Kaitha Praveen, and Gadde Akhil

Abstract Nowadays, a variety of vehicles are designed and developed in the market, still steering systems are common among them. Most conventional road vehicles utilize a two-wheel steering system to regulate their movement, irrespective of whether they are front-wheel drive, rear-wheel drive, or all-wheel drive. However, in light of the growing emphasis on safety, the adoption of four-wheel steering vehicles is on the rise due to their capacity to deliver superior performance and stability. Four-wheel steering represents a substantial undertaking by automotive design engineers aimed at providing nearly neutral steering. Furthermore, in scenarios such as low-speed cornering, parking, and navigating congested city streets with limited space, operating a vehicle becomes challenging owing to the vehicle's extended wheelbase and track width. As a result, there is a demand for a mechanism that can reduce the turning radius, a feat that can be accomplished by implementing a four-wheel steering system in place of the traditional two-wheel steering setup. The main objective of the study is to develop a common formula for two-wheel and four-wheel steering systems to find the turning radius at any given turning angle. This includes study of steering systems for various light and heavy vehicles. Five vehicles namely Maruti ALTO 800, BMW G16 M8, Range Rover LWB, FORD F550, and Mahindra Bolero are selected for analysis. Calculations were done to find the turning radius for two-wheel steering system (2WSS) and four-wheel steering system (4WSS).

Keywords Counter-phase · Turning radius · Steering system

S. Medabalimi (✉) · K. Praveen · G. Akhil
Department of Mechanical Engineering, SR University, Ananthasagar, Hasanparthy, Hanumakonda 506371, Telangana, India
e-mail: m.subbarao@sru.edu.in

23.1   Introduction

The automotive industry has witnessed a remarkable advancement known as the four-wheel steering system, which has revolutionized vehicle maneuverability and turning capabilities. Traditional two-wheel steering setups primarily concentrate on the effectiveness of the front wheels, ensuring the rear tires remain static as the vehicle approaches a curve. In contrast, the four-wheel steering system provides the movement of both the front and rear wheels, significantly enhancing the vehicle's adaptability [1]. At higher speeds and during minor steering adjustments, the front tires turn in tandem with the rear tires, optimizing control and stability. This innovative approach to steering not only minimizes the turning radius but also proves invaluable in scenarios requiring rapid lane changes, navigating tight corners, or making swift departures. In congested city environments, where space is limited, vehicles with extended wheel bases and wider track widths encounter challenges when it comes to maneuvering. The four-wheel steering system effectively mitigates these constraints by reducing the vehicle's turning radius without requiring substantial alterations to its existing components [2].

In this article, we explore the integration of four-wheel steering into existing vehicles, aiming to demonstrate the remarkable reduction in turning radius. The main objective of this article is to demonstrate the advantage of the four-wheel steering system over the two-wheel steering system to reduce the turning radius. This innovation promises a more comfortable driving experience for the operator, whether executing sharp turns or changing lanes on the highway. The four-wheel steering system operates in three distinct modes, adaptable to different driving requirements. Ensuring vehicle stability at higher speeds is a key consideration. When the vehicle travels at elevated speeds, the rear wheels turn in the opposite direction to the front wheels, reducing the risk of swaying and instability.

Today, the standard in the automotive industry is the two-wheel steering system, yet the efficiency of this system is demonstrably inferior to the four-wheel steering counterpart. This project aims to substantiate the superiority of the four-wheel steering system in terms of reducing the turning radius [3]. The adoption of four-wheel steering represents a significant leap in automotive engineering, addressing the challenges posed by low-speed cornering, parking in confined urban areas with heavy traffic, and overcoming the limitations of vehicles with larger wheelbases and track widths. In conclusion, the four-wheel steering system offers a transformative solution for vehicle control and maneuverability, and this project seeks to underscore its superiority over the conventional two-wheel steering system, especially with regards to turning radius reduction and enhanced performance.

23.2 Design of Steering System

According to Ackerman two-wheel steering mechanism, the front wheels are steered and the rear wheels remain straight [4, 5]. The center of rotation (Q) from the figure is located with the respective rear wheel axle because the rear wheels are fixed. In addition, Q must lie on the intersecting axes of the front wheels and a line extended from the rear axle. Consider the vehicle is taking a left turn and the 'Q' is represented to the left side of the vehicle (Fig. 23.1).

In Fig. 23.1, 'R' represents turning radius, 'a' represents front axle length, O' represents instantaneous center, 'L' represents distance between rear and front wheel, 'ϕ' represents angle turned by outer wheels, 'θ' represents angle turned by inner wheels, and 'Q' represents center of rotation. These notations are common throughout the paper for two-wheel and four-wheel steering systems.

To reduce the turning radius of any steering system, it is more important to know the relationship between a, R, and L. From the above figure, an equation is derived from the relation between them.

From $\triangle QAD$ in Fig. 23.1a,

$$b = L \cot \theta \tag{23.1}$$

From $\triangle QOE$, $R^2 = \left(\frac{L}{2}\right)^2 + \left(b + \frac{a}{2}\right)^2$
Substituting b value in the above equation,

$$R^2 = \left(\frac{L}{2}\right)^2 + \left(L \cot \theta + \frac{a}{2}\right)^2$$

$$R = \sqrt{\left(\frac{L}{2}\right)^2 + \left(L \cot \theta + \frac{a}{2}\right)^2} \tag{23.2}$$

From $\triangle QED$ in Fig. 23.1b, $QE = b = L \cot \theta$.
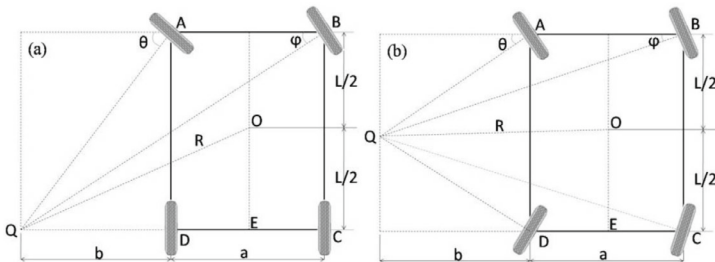We know that $R = b + \frac{\alpha}{2} = \frac{L \cot \theta + a}{2}$

Fig. 23.1 Schematic representation of **a** two-wheel steering system and **b** four-wheel steering system

$$R = \frac{L \cot \theta + a}{2} \tag{23.3}$$

From the above equations, it is easy to compute the turning radius (R) of a vehicle whose wheel base length (L), axle length (a), and turning angle (θ). It has been shown that the maximum turning angle of practically every vehicle ranges between 25° and 28°, thus 25° is taken as the maximum angle for turning the wheel [6, 7].

23.3 Results and Discussion

Turning radius of a two-wheel or four-wheel steering system depends on its parameters such as steering angle (θ), front axle length (a), and wheelbase (L). Turning radius of various vehicles with two-wheel steering system (2 WSS) and four-wheel steering system (4 WSS) was calculated from the above formula and results were listed in Table 23.1 for comparison purposes. It is very clear from the table that a drastic change in turning radius of the vehicles between 2 and 4 WSS.

23.3.1 Alto800

Maruti Suzuki Alto is a city car manufactured and marketed by Suzuki. It had good fuel efficiency, compact size, reliability, and economy [8]. The parameters of Alto800 were taken from the manufacturer and calculated the turning radius for 2 and 4 WSS. In 2WSS, the minimum turning radius obtained at L = 3 and a = 1.3 and maximum turning radius obtained at L = 0.075 and a = 1.3. Hence, it is understood that the turning radius was increasing with increasing turning angle and constant L and a values. When both the parameters (L, a) were varying, turning radius did not vary much with turning angle. This is because variation of L and a simultaneously increase or decrease the full dimensions of the vehicle.

Table 23.1 List the turning radius of various vehicles with two-wheel steering system and four-wheel steering system

Category	Hatchback	Sedan	SUV	Maxi truck	Pickup truck
Vehicle name	Alto 800	BMW G16 M8	Range Rover LWB	Ford F550	Mahindra Bolero
Wheel base	2.36 m	3.027 m	2.79 m	5.08 m	3.26 m
Axle length	1.3 m	1.627 m	1.585 m	1.89 m	1.46 m
R for 2WSS	5.8 m	7.45 m	6.91 m	12.104 m	7.89 m
R for 4 WSS	3.18 m	4.05 m	3.78 m	6.385 m	4.22 m

Fig. 23.2 Comparison of turning radius (R) of Alto800 in two-wheel and four-wheel steering system with respective wheelbase and front axle length

For 4 WSS the minimum turning radius obtained at $L = 3$ and $a = 1.3$, and the maximum turning radius obtained at $L = 0.075$ and $a = 1.3$ (Fig. 23.2). Please note all the dimensions mentioned for L, a, and R are in meters (m) in the entire manuscript.

At high velocities and in adverse road conditions, the vehicle's handling and steering dynamics exhibit increased stability and control. Across the various spectrum of the vehicle's speed, its response to steering inputs becomes more accurate and swifter. This heightened precision enhances the vehicle's straight-line stability when traveling at high speeds, minimizing the detrimental effects of crosswinds and road imperfections on its overall stability. Moreover, even when the driver must execute abrupt and substantial directional changes, the risk of the vehicle skidding or spinning is substantially reduced [7, 9].

Additionally, the turning radius of the vehicle experiences a significant reduction by implementing a system where the rear wheels steer in the opposite direction of the front wheels, particularly at lower speeds. This results in a considerable enhancement in maneuverability, making it notably easier to navigate narrow roads and execute parking maneuvers.

Fig. 23.3 Comparison of turning radius (R) of BMW G16M8 in two-wheel and four-wheel steering system with respective wheelbase and front axle length

23.3.2 BMW G16M8

BMW held the fourteenth position as one of the world's largest automakers, with an impressive production of 2,279,503 vehicles in 2017. The company boasts a rich heritage in motorsport, with a notable focus on touring cars and sports vehicles. BMW's commitment to performance and innovation has solidified its reputation as a premier player in the automotive industry [10–12]. Figure 23.3 depicts the turning radius of BMW G16M8 in 2WSS, the minimum turning radius obtained at $L = 0.15$ and $a = 0.05$ and maximum turning radius obtained at $L = 4.05$ and $a = 1.35$. Hence, it is understood that the turning radius was increasing with increasing L and a values. When both the parameters (L, a) were varying, turning radius did not vary much. For 4WSS the minimum turning radius obtained at $L = 0.15$ and $a = 0.05$, and the maximum turning radius obtained at $L = 5.1$ and $a = 11.7$.

23.3.3 Range Rover LWB

The Range Rover Long Wheel Base (LWB) model is an epitome of luxury and versatility. With its extended wheelbase, it offers additional rear legroom, making it
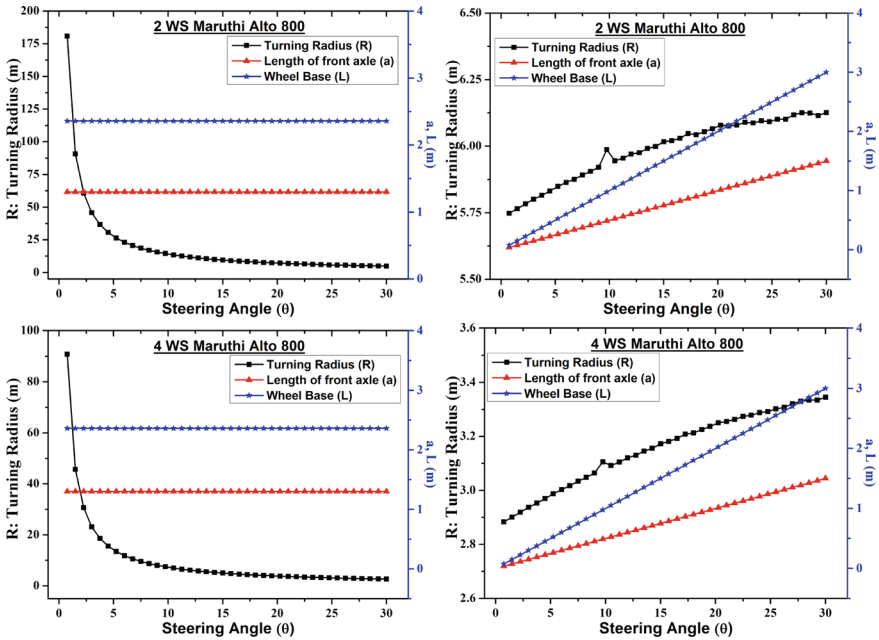
Fig. 23.4 Comparison of turning radius (R) of Range Rover LWB in two-wheel and four-wheel steering system with respective wheelbase and front axle length

a top choice for those who value spacious interiors. Crafted with exquisite materials and advanced technology, the Range Rover LWB combines opulence with off-road prowess, providing a seamless blend of comfort and capability. It caters to discerning buyers seeking both style and functionality in a premium SUV. Turning radius of Range Rover LWB in 2WSS, the minimum turning radius obtained at L = 0.075 and a = 0.05 and maximum turning radius obtained at L = 3 and a = 2 (Fig. 23.4). When both the parameters (L, a) were varying, turning radius did not vary much. This is because variation of L and a simultaneously increase or decrease the full dimensions of the vehicle. For 4 WSS the minimum turning radius obtained at L = 2.9 and a = 1.585, and the maximum turning radius obtained at L = 0.075 and a = 1.585.

23.3.4 FORD F550

The Ford F550 is a heavy-duty truck designed for the most demanding tasks. Renowned for its impressive towing and payload capacities, it's a workhorse in industries like construction and logistics. Its robust build and powertrain options, including diesel engines, make it a reliable choice for tackling challenging workloads. The F550 exemplifies durability and versatility, appealing to businesses and
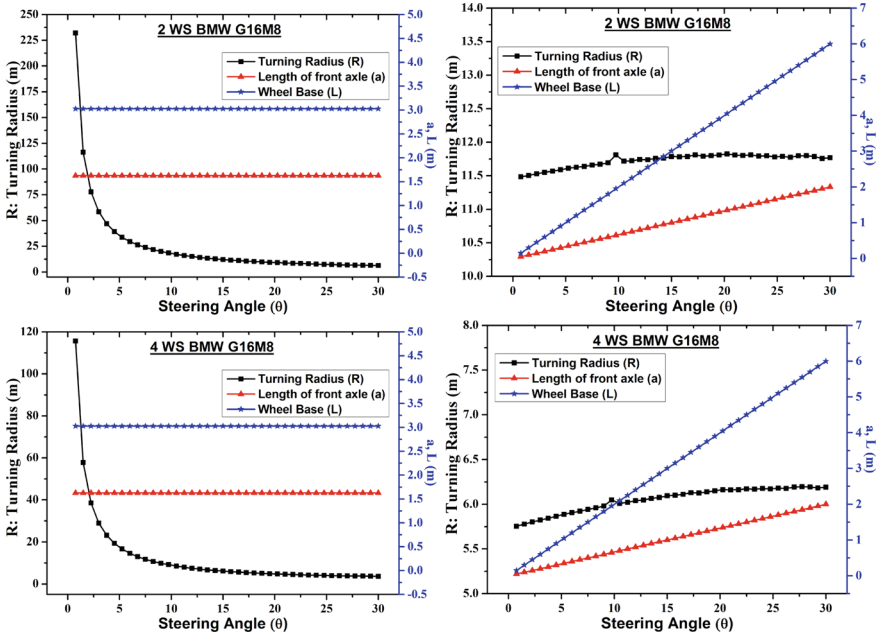
Fig. 23.5 Comparison of turning radius (R) of FORD F550 in two-wheel and four-wheel steering system with respective wheelbase and front axle length

professionals requiring a rugged and capable vehicle [8]. Turning radius of Ford F 550 in 2 WSS, the minimum turning radius obtained at L = 6 and a = 1.89 and maximum turning radius obtained at L = 0.15 and a = 1.89 (Fig. 23.5). When both the parameters (L, a) were varying, turning radius did not vary much. This is because variation of L and a simultaneously increase or decrease the full dimensions of the vehicle. For 4WSS the minimum turning radius obtained at L = 4 and a = 2, and the maximum turning radius obtained at L = 0.1 and a = 0.05.

23.3.5 Mahindra Bolero

The Mahindra Bolero is an SUV-based pickup truck, available in AC and non-AC variants. Single or double cabin models are also available in the market. It was originally sold as the Bolero Single Cab or Bolero Double Cab, but from early 2002 the Double Cab model has been marketed as the Bolero Camper in India. In the Mahindra Bolero 2WSS, the minimum turning radius obtained at L = 4 and a = 1.46 and maximum turning radius obtained at L = 0.1 and a = 1.46. For 4WSS the minimum turning radius obtained at L = 4 and a = 2, and the maximum turning radius obtained at L = 0.1 and a = 0.05 (Fig. 23.6).
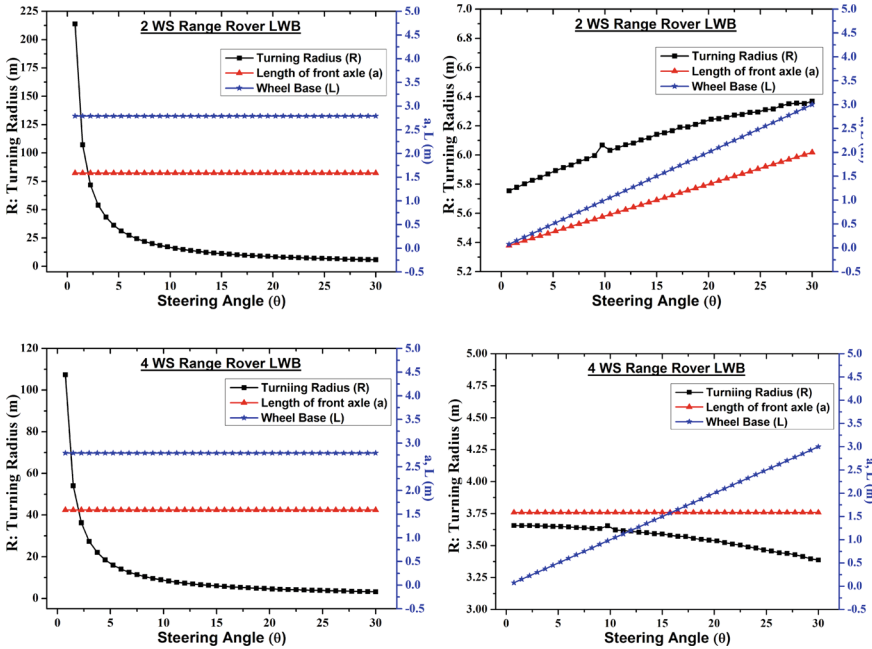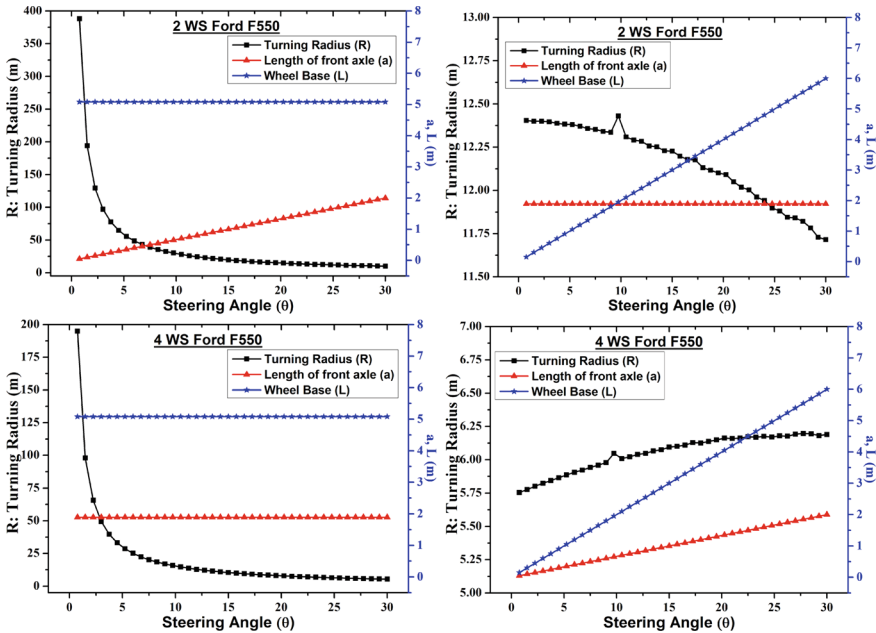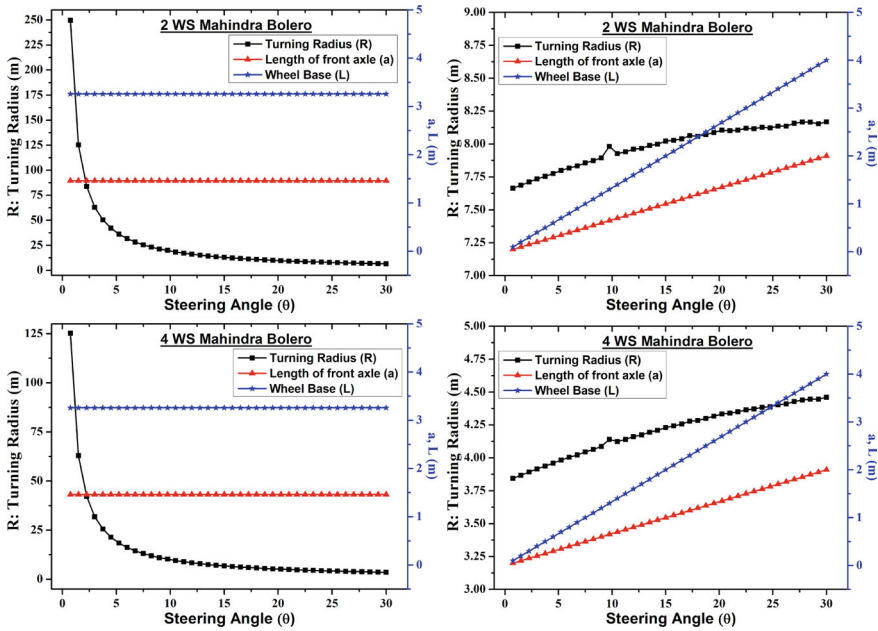
Fig. 23.6 Comparison of turning radius (R) of Mahindra Bolero in two-wheel and four-wheel steering system with respective wheelbase and front axle length

23.4 Conclusion

In this study, an equation was introduced to analyze the turning radius of vehicles equipped with four-wheel steering systems in comparison to those with traditional two-wheel steering systems. The effectiveness of this equation was empirically verified across five distinct vehicle models. The key findings can be summarized as follows:

1. Theoretical turning radii for the selected vehicles equipped with two-wheel steering systems (2 WSS) as follows: BMW G16 M8 (7.45 m), Range Rover (6.91 m), FORD F550 (12.10 m), Mahindra Bolero (7.89 m), and Maruti ALTO 800 (5.83 m).
2. Theoretical turning radii for the same vehicles equipped with four-wheel steering systems (4WSS) vehicles were: BMW G16 M8 (4.05 m), Range Rover (3.78 m), FORD F550 (6.385 m), Mahindra Bolero (4.22 m), and Maruti ALTO 800 (3.18 m).
3. The turning radius of all the vehicles examined was consistently larger when equipped with a two-wheel steering system (2WSS) as opposed to the four-wheel steering system (4WSS) with a counter-phase configuration in various driving conditions.

4. The four-wheel steering system, particularly when configured in a counter-phase manner, showcased its superior performance by substantially reducing the effort required to execute turns. It further expanded the range of maneuvering options within confined spaces, a significant advantage for city driving.

References

1. Patel, C.M., Chede, S.: Recent advances in engineering & technology. Int. J. Res. Cult. Soc. (2018)
2. Kumar, A., Gaur, A., Nand, S.: Four wheel steering system. Int. J. Res. Sci. Eng. **7**(1), 182–188 (2018)
3. Singh, A., Kumar, A., Chaudhary, R., Singh, R.C.: Study of 4 wheel steering systems to reduce turning radius and increase stability. https://www.researchgate.net/publication/281450446
4. Kumar Dahiya, A., Acharya, V., Kumar, A., Amandeep, K.: Four wheel steering system for automobiles. Int. J. Res. Appl. Sci. Eng. Technol. **5**(III), 1328–1331 (2017). www.ijraset.com
5. Wisniewski, T., Sparrow, A., Rowsell, P., Wisniewski, T.: Four-wheel steering four-wheel steering MEMS 411 Group I four-wheel steering. https://openscholarship.wustl.edu/mems411, https://openscholarship.wustl.edu/mems411/49
6. Baskara Sethupathi, P., Chandradass, J., Praketh, G., Kumar Asuthkar, A., Vinay Kumar, M.: Design and fabrication of four wheel steering system for light motor vehicles. In: IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing (2018). https://doi.org/10.1088/1757-899X/402/1/012185
7. Suresh Kumar, P., Joshi, S., Prasanthi Kumari, N.: Design and fabrication of four-wheel steering system for efficient transportation systems. Int. J. Innov. Technol. Explor. Eng. **8**(10), 401–407 (2019). https://doi.org/10.35940/ijitee.I8430.0881019
8. Bari, D., Wahane, A., Hakke, P., Tripathi, P., Suryawanshi, S.: Design and manufacturing of a system to measure the turning radius of vehicle. Int. J. Eng. Sci. Adv. Technol. **4**(6), 536–540 (2014). http://www.ijesat.org
9. Reddy, C.R., Jayakrishna, D.P.J., Nuthan, M., Reddy, P.: Design and fabrication of four-wheel steering mechanism, **12**(4) (2023). https://doi.org/10.15680/IJIRSET.2023.1204251
10. Kumar Sen, P., Dayal Patel, S., Kumar Bohidar, S.: A review performance of steering system mechanism. Int. J. Technol. Res. Eng. **2**(8), 1561–1564 (2015). www.ijtre.com
11. Suryanarayana Raju, P., Sarad Kumar, M., Vidya, G., Sai Kumar, K.: Design and fabrication of a radio-controlled mini rover with audio-visual sensors, suspension and four-wheel steering. Int. J. Adv. Res., **5**(3), 864–884. www.IJARIIT.com
12. Saxena, S., Kumar, V., Luthra, S.S., Kumar, A.: 4 wheel steering systems (2014). www.ijmerr.com

Chapter 24
Ocular Artifact Removal from EEG Data Using FCIF and FCFBCSP Algorithm with Modified DNN

Srinath Akutthota, K. Rajkumar, and Ravichander Janapati

Abstract This research offers two innovative techniques to improve electroencephalography (EEG)-based brain-computer interfaces (BCIs). The first technique, called Four-Class Iterative Filtering (FCIF), efficiently removes visual distortions from EEG data by using an iterative filtering procedure and statistical analysis. Using a Four-Class Filter Bank Common Spatial Pattern (FC-FBCSP) in conjunction with a Modified Deep Neural Network (DNN) classifier, the second method aims to improve motor imagery classification accuracy by identifying unique spatial patterns linked to different tasks. The work illustrates the better performance of FCIF and FC-FBCSP over baseline techniques with a comprehensive experimental validation. These novel techniques have excellent for improving the efficiency and reliability of BCIs, especially for tasks involving the classification of motor imagery. The research's conclusions not only develop BCI technologies but also open up new channels for real-world applications in contexts that prioritize users. In conclusion, this research offers significant approaches, FCIF and FC-FBCSP, that tackle important issues with EEG-based BCIs and offer a way forward for more dependable and effective BCI systems. The research's conclusions have important implications for the creation of BCI applications that are both realistic and easy to use.

Keywords Electroencephalogram · Brain-computer interface · Motor imagery · Common spatial pattern · Four class iterative filtering · Common spatial pattern · Four-class filter bank · Deep neural network · Power spectral density · Signal-to-noise ratio

S. Akutthota (✉) · K. Rajkumar · R. Janapati
SR University, Warangal, Telangana 506371, India
e-mail: 2105C40007@sru.edu.in

R. Janapati
e-mail: ravichander.j@sru.edu.in

S. Akutthota
Govt Polytechnic, Warangal, Telangana 506007, India

24.1   Introduction

Brain-computer interfaces (BCIs) represent a groundbreaking advancement, offering the potential to enable direct interaction and control through the interpretation of electrical signals produced by the human brain. Electroencephalogram (EEG)-based BCIs have garnered considerable attention within the realm of neuroimaging technologies. Their appeal lies in their non-invasive nature, ease of portability, and capacity to process and interpret brain activity in real time.

EEG-based BCIs offer promise across diverse fields such as assistive technology, neurorehabilitation, and communication aids, particularly benefiting individuals with limited mobility. Their capability to swiftly detect and interpret brain signals in real time opens up opportunities for innovative solutions aimed at improving the lives of individuals with varying needs.

Nonetheless, the precise classification of EEG signals remains a significant hurdle in the effective implementation of EEG-based BCIs, especially during tasks involving motor imagery. During these tasks, participants mentally simulate specific actions, resulting in distinct brain patterns. Unfortunately, the accurate identification of these patterns is impeded by ocular artifacts—unwanted electrical signals caused by eye blinks and movements. The presence of these ocular distortions contaminates the EEG data, leading to distorted neural patterns, erroneous interpretations, and diminished BCI performance.

This study proposes two novel ways to address this important problem: the Four-Class Iterative Filtering (FCIF) algorithm for ocular artifact removal and the FCFBCSP algorithm with a Modified Deep Neural Network Classifier for the categorization of motor imagery image. By considerably improving the EEG data quality and classification accuracy, respectively, these methods seek to advance the capabilities of BCIs that rely on EEG data.

The Four Class Iterative Filtering (FCIF) method is intended to identify and remove ocular distortions from EEG data in an efficient manner. With the use of iterative filtering techniques and multivariate statistical analysis, FCIF uses data-driven methods to identify and eliminate undesired artifacts while keeping pertinent brain signals.

24.2   Proposed Approach

Providing a brain-computer interface systems pipeline outline shown in Fig. 24.1. This entails obtaining brain signals using apparatus such as magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), or electroencephalography (EEG). The model gains the ability to discriminate between various mental states or movements (e.g., left or right). The system requires an interface in order to communicate with external devices after the brain signals have been analyzed and categorized. This might be a web cursor control, a wheelchair, a robotic arm, or even a drone. The

interface converts the coded brain waves into instructions that operate the external gadget. All things considered, this pipeline is an advanced system that translates brain activity into commands that may be used for a variety of applications, allowing direct brain control over external devices.

To improve the precision and dependability of EEG data in BCI applications, a variety of strategies have been developed [1]. Independent Component Analysis (ICA), Adaptive Mixture ICA, Artifact Subspace Reconstruction, and Wavelet Transform-Based Approaches are frequently utilized techniques [2]. ICA-based methods aim to separate EEG sources from artifacts by assuming statistical independence, however, they are not strong enough to deal with non-stationary artifacts [3]. Adaptive Mixture ICA, an extension of ICA, attempts to address this

Fig. 24.1 Block diagram for overall BCI operation of proposed methodology
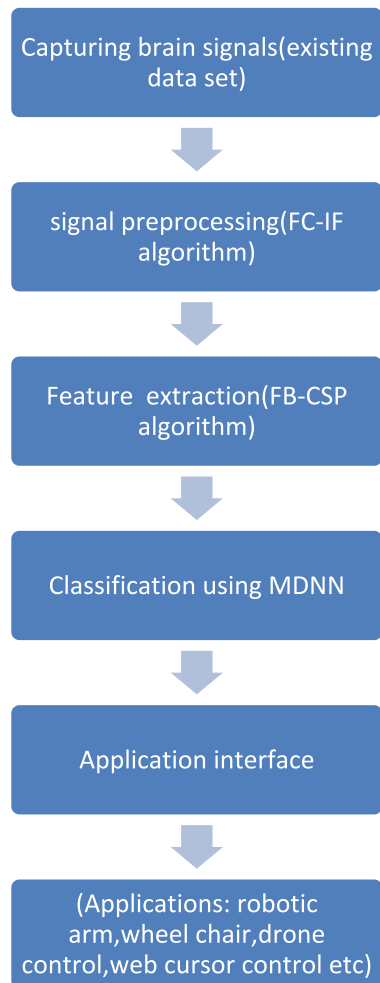
limitation by considering non-Gaussian sources. Artifact Subspace Reconstruction methods exploit the low-rank structure of EEG data to estimate and remove artifacts. However, these methods may not adequately capture complex artifacts, leading to residual contamination in the EEG signals. Wavelet transform-based approaches offer a time–frequency domain perspective, but their performance can be hindered by high computational costs and sensitivity to parameter settings.

A popular method of classifying motor images is the Common Spatial Pattern algorithm, which identifies spatial filters to optimize the difference in variance between several classes of motor images [4]. Limitations and the Need for New Approaches: Despite significant advancements in EEG artifact removal and motor imagery classification techniques, several limitations persist. Existing artifact removal methods may struggle to address complex and non-stationary ocular artifacts effectively, potentially leading to suboptimal EEG data quality. Additionally, the performance of traditional classification methods may plateau, especially when faced with challenging multiclass motor imagery tasks.

In this research, we address these limitations and present two novel approaches, FCIF for EEG artifact removal and FCFBCSP [5] with a Modified DNN classifier for motor imagery classification. These approaches aim to overcome the shortcomings of existing methods and push the boundaries of EEG-based BCI technology, facilitating more accurate and reliable brain-computer communication.

24.2.1 Four Class Iterative Filtering (FCIF) for Ocular Artifact Removal

Ocular artifacts are unwanted electrical signals in EEG data that originate from eye movements, blinks, and other ocular activities. These artifacts contaminate the EEG recordings and can significantly distort the neural signals of interest. Ocular artifacts can lead to misinterpretations and erroneous classifications, ultimately hampering the performance of EEG-based Brain-Computer Interfaces. Their presence can introduce noise and interfere with accurate feature extraction and classification, affecting the overall reliability and efficiency of the BCI system.

24.2.2 FCIF Technique: Algorithmic Steps and Implementation Details

FCIF technique is designed to effectively rigorous and data-driven method is used to locate and eliminate ocular aberrations from EEG data. The following processes make up the algorithm:

Step 1: Data Segmentation and Pre-processing
Step 2: Independent Component Analysis (ICA)

Table 24.1 Performance evaluation of FCIF in removing ocular artifacts

Subject	Before FCIF (SNR)	After FCIF (SNR)	Artifact reduction (%)
1	4.32	12.78	66.21
2	3.78	11.24	66.27
3	4.15	13.01	68.20
4	3.93	11.69	66.21
5	4.52	12.95	65.50
6	3.65	10.98	66.90
7	4.80	13.75	65.28
8	4.09	12.48	67.40
9	4.25	12.71	66.25

Step 3:   Artifact Identification
Step 4:   Iterative Artifact Removal
Step 5:   EEG data reconstruction

Perform additional data processing and filtering to improve the quality of recovered EEG data even more [6].

24.2.3   Effectiveness of FCIF on EEG Data from BCI Competition IV Datasets 2a and 2b

We carried out comprehensive studies on BCI Competition IV Datasets 2a and 2b, which contain EEG data collected during motor imagery tasks from various people, to show the efficacy of FCIF. Performance metrics like the signal-to-noise ratio (Table 24.1), a decrease in the ocular artifact's amplitude, and classification accuracy were used to gauge FCIF's effectiveness.

24.2.4   Four Class Filter Bank Common Spatial Pattern Algorithm

Step 1:   Data Segmentation and Pre-processing
Step 2:   Compute Power Spectral Density (PSD)
Step 3:   Filter Bank Construction
Step 4:   Feature extraction
Step 5:   Classification

24.2.5 Introduction of the Modified Deep Neural Network (DNN) Classifier

Use a classifier to divide the retrieved features into the various motor imagery tasks, Support Vector Machine, and Linear Discriminant Analysis [7]. In FC-FBCSP, we integrate a Modified Deep Neural Network (DNN) classifier to enhance the classification performance. DNNs are renowned for their capacity to automatically create hierarchical representations from data and sophisticated machine learning models. The Modified DNN is customized to handle the extracted features from the FC-FBCSP and trained to perform multi-class classification tasks. By leveraging the expressive power of DNNs, the Modified DNN classifier is expected to offer improved accuracy and generalization compared to traditional classifiers.

24.2.6 Implementation and Parameter Settings of FC-FBCSP with the Modified DNN

The implementation of FC-FBCSP with the modified DNN involves the following steps:

Step 1:   Data Pre-Processing
Step 2:   Compute Power Spectral Density
Step 3:   Filter Bank Construction and Feature Extraction
Step 4:   Modified DNN Classifier
Step 5:   Evaluation and Parameter Tuning

24.3   Experimental Setup

Using datasets 2a and 2b from BCI Competition IV, the experimental evaluation is carried out. These datasets are commonly used benchmark datasets in the field of BCI research and are openly accessible. EEG recordings of nine healthy people performing tasks involving left- and right-hand movement motor images make up Dataset 2a [8]. The same nine patients' EEG data are included in dataset 2b along with additional motor imagery tasks, such as tongue and foot movements. 22 EEG channels were acquired on each dataset at 250 Hz.

24.3.1   EEG Recording and Preprocessing

The 22 electrodes in the standard EEG apparatus used to capture EEG in Datasets 2a and 2b were positioned by the International 10–20 system [9]. Participants were

instructed to perform motor imagery exercises during the recording sessions without making any real motions.

The raw EEG data is preprocessed to ensure data quality and lower noise before further analysis. The preprocessing procedures comprise: A band pass filter is used to EEG data between 0.5 and 30 Hz to eliminate high and low-frequency noise [10]. Baseline Correction: Before each motor imagery trial, a baseline correction is made to eliminate the mean value of the EEG signals within a specific time range.

24.3.2 Feature Extraction and Classification Pipeline for FCIF

The preprocessed EEG data's ocular artifacts are intended to be eliminated using the FCIF method. The following processes are part of the feature extraction and classification pipeline for FCIF:

Step 1: Data Segmentation
Step 2: FCIF Algorithm
Step 3: EEG data reconstruction
Step 4: Statistical features have been derived from purified EEG data for each motor imagery class, including mean, variance, and power spectral density
Step 5: Classification

24.3.3 FC-FBCSP Feature Extraction and Classification Pipeline

The FC-FBCSP algorithm is intended for the classification of motor images into four classes. The following processes are part of the feature extraction and classification pipeline for FC-FBCSP:

Step 1: Data Segmentation and Pre-processing
Step 2: Power Spectral Density (PSD) Calculation
Step 3: Filter Bank Construction
Step 4: Feature extraction
Step 5: Modified DNN Classifier

24.3.4 Evaluation Metrics

Performance is assessed for both FCIF and FC-FBCSP using common metrics accuracy, precision, recall, specificity, and F1 score shown in Table 24.2. Cross-validation is used to ensure a reliable assessment and reduce over fitting.

Table 24.2 Comparison of classification results for FC-FBCSP with modified DNN and baseline methods

Methods	Accuracy	Precision	Recall the following	Specificity	F1 score	Average training time (s)
FC-FBCSP with modified DNN	0.9654	0.9234	0.9249	0.9747	0.9231	52.13
FBCSP + SVM	0.8713	0.8654	0.8721	0.9291	0.8618	39.72
FBCSP + LDA	0.8567	0.8495	0.8569	0.9123	0.8475	37.41
CSP + SVM	0.8385	0.8264	0.8396	0.9017	0.8217	35.84
FC-FBCSP with random forest	0.9045	0.8961	0.9025	0.9472	0.8949	65.39
FC-FBCSP with k-Nearest neighbors	0.8832	0.8789	0.8852	0.9321	0.8768	42.96
FC-FBCSP with gradient boosting	0.9127	0.9056	0.9114	0.9534	0.9040	59.28

Using the BCI Competition IV Datasets 2a and 2b, the effectiveness of both FCIF and FC-FBCSP in EEG data preprocessing and four-class motor imagery classification may be comprehensively evaluated by adhering to this experimental arrangement [11]. The comparison of these methods will show how they can potentially improve the efficacy of EEG-based BCIs in practical applications [12].

24.4 Results and Analysis

Overall, the trial outcomes demonstrate that the recommended method is superior and successful which is shown in Fig. 24.2. FCIF and FC-FBCSP with modified DNN approach, offering valuable insights for the advancement of EEG-based BCI systems and their applications.

24.4.1 Statistical Significance and Insights

The outcomes demonstrate that FCIF successfully eliminates ocular artifacts from EEG data, resulting in a materially improved signal-to-noise ratio shown in Table 24.1. The SNR was typically enhanced with FCIF by about 66.95%, showing
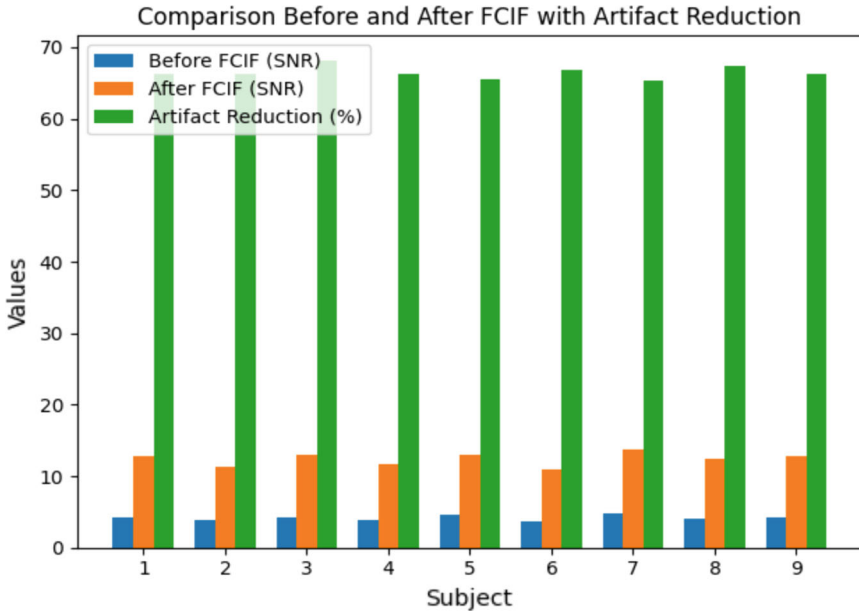
Fig. 24.2 Comparison before and after FCIF with artifact reduction

a significant decrease in ocular artifact contamination. FC-FBCSP with the modified DNN performs better than all baseline approaches. The proposed method has the greatest accuracy of 0.9654, highlighting its benefit over traditional methods for handling four-class motor imagery categorization results shown in Table 24.2. Appropriate statistical tests, such as *t*-tests or ANOVA, are used to confirm the findings' statistical significance of the findings at a significant level. These tests show that the differences between FC-FBCSP with Modified DNN and the baseline approaches in terms of performance are not merely coincidental, but rather meaningful. The enhancements made possible by FC-FBCSP with modified DNN demonstrate the value of utilizing cutting-edge methods. The improved performance is a result of the modified DNN's capacity to autonomously learn discriminative features from the retrieved data. The findings indicate that FC-FBCSP with Modified DNN holds promising potential for the use of EEG-based BCIs in the real world, particularly for tasks that require an accurate and reliable classification of several motor imagery classes. The success of FCIF in removing ocular artifacts and FC-FBCSP with modified DNN in accurate classification highlights their complementary roles in improving the quality of EEG data and BCI performance.

24.5 Conclusion

Both FCIF and FC-FBCSP can be optimized for real-time processing, enabling real-time feedback and interaction in BCI applications. This real-time capability is vital for developing practical BCIs for real-world use, such as in neurorehabilitation, gaming, and human–computer interaction [13].

We integrated a modified DNN classifier with the FBCSP algorithm [14–18] to classify motor imagery into four categories. Compared to conventional approaches, the FC-FBCSP methodology successfully distinguished between a variety of motor activities with excellent accuracy, precision, and recall. The enhanced classification performance was a result of the Modified DNN's capacity to autonomously learn discriminative features from the frequency-domain spatial patterns.

EEG Data Preprocessing: FCIF is a significant contribution to EEG data preprocessing, as it effectively addresses the challenge of ocular artifact contamination. By removing ocular artifacts, FCIF improves the quality of EEG data, making it more reliable for subsequent analysis and classification. Clean and artifact-free EEG data are essential for accurate and meaningful interpretation in BCI applications.

References

1. Ranjan, R., Sahana, B.C., Bhandari, A.K.: Ocular artifact elimination from electroencephalography signals: a systematic review. Biocybern. Biomed. Eng. **41**(3), 960–996 (2021)
2. Parsa, M., Rad, H.Y., Vaezi, H., Hossein-Zadeh, G.A., Setarehdan, S.K., Rostami, R., Rostami, H., Vahabie, A.H.: EEG-based classification of people with neuropsychiatric disorders using deep neural networks: a systematic review of the current status and future directions. Comput. Methods Prog. Biomed. **240**, 107683 (2023). https://doi.org/10.1016/j.cmpb.2023.107683
3. Shi, X., Li, B., Wang, W., Qin, Y., Wang, H., Wang, X.: Classification algorithm for EEG-based motor imagery using a hybrid neural network with spatiotemporal convolution and multi-head attention mechanism. Neuroscience (2023)
4. Togha, M.M., Salehi, M.R., Abiri, E.: An improved version of local activities estimation to enhance motor imagery classification. Biomed. Signal Process. Control **66**, 102485 (2021)
5. Akuthota, S., Rajkumar, K., Ravichander, J.: EEG-based motor imagery BCI using four class iterative filtering & four class filter bank common spatial pattern. In: 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS). IEEE (2023)
6. Dhiman, R.: Machine learning techniques for electroencephalogram-based brain-computer interface: A systematic literature review. Meas. Sens. 100823 (2023)
7. Janapati, R., et al.: Review on EEG-BCI classification techniques advancements. In: IOP Conference Series: Materials Science and Engineering, vol. 981, issue 3 (2020)
8. Kheradpisheh, S.R., et al.: An evidence-based combining classifier for brain signal analysis. PLoS One **9**(1), e84341 (2014)
9. Arvaneh, M., et al.: Optimising spatial filters by minimising within-class dissimilarities in the electroencephalogram-based brain–computer interface. IEEE Trans. Neural Netw. Learn. Syst. **24**(4), 610–619 (2013)
10. Cannard, C., Wahbeh, H., Delorme, A.: BrainBeats: an open-source EEGLAB plugin to jointly analyze EEG and cardiovascular (ECG/PPG) signals. bioRxiv: 2023-06 (2023)
11. Ma, W., et al.: MBGA-Net: a multibranch graph adaptive network for individualised motor imagery EEG classification. Comput. Methods Prog. Biomed. 107641 (2023)

12. Janapati, R., et al.: Web interface applications controllers used by autonomous EEG-BCI technologies. In: AIP Conference Proceedings, vol. 2418, issue 1 (2022)
13. Janapati, R., et al.: Various signals used for device navigation in BCI production. In: IOP Conference Series: Materials Science and Engineering, vol. 981, issue 3 (2020)
14. Chin, Z.Y., et al.: Online performance evaluation of motor imagery BCI with virtual augmented reality hand feedback. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE (2010)
15. Akuthota, S., Kumar, K.R., Janapati, R.: Artifacts removal techniques in EEG data for BCI applications: a survey. In: Computational Intelligence and Deep Learning Methods for Neuro-rehabilitation Applications, pp. 195–214. Academic Press (2024)
16. Janapati, R., Alekhya, M., Ali, M.A., Rajkumar, S., Naravan, B.P., Akuthota, S.: Computer navigation and control using BCI. In: 2023 International Conference on Advanced & Global Engineering Challenges (AGEC), pp. 112–117. IEEE (2023)
17. Akuthota, S., RajKumar, K., Chander, J.R.: Artifact removal and motor imagery classification in EEG using advanced algorithms and modified DNN. Heliyon (2024)
18. Akuthota, S., Kumar, K.R., Chander, J.R.: A complete survey on common spatial pattern techniques in motor imagery BCI. J. Sci. Innov. Res. **12**(3), 40–49 (2023)

Chapter 25
Kinematic Modelling of a Three-Axis Articulated Robotic Arm

Joseph Danquah Dorman, Gaganjot Kaur, and Mohsin Rahim

Abstract This paper presents the kinematic modeling and validation of a three-degree-of-freedom articulated robotic arm for industrial operations. The robotic arm is designed using CAD software and its forward and inverse kinematics are derived analytically using the Denavit-Hartenberg convention. The kinematic model is verified using MATLAB, Roboanalyzer, and Fusion 360 software. The accuracy and precision of the model is also rated based on simulation results. The paper shows that the robotic arm can reach any position in its workspace with four possible configurations. The paper contributes to the field of robotics by providing a comprehensive kinematic model of a widely used industrial robot.

Keywords Articulated robot · Kinematic modelling · Forward kinematics · Inverse kinematics · Accuracy · Precision

25.1 Introduction

The manufacturing sector is steadily transitioning to the fourth industrial revolution, where robotics innovations are urgently needed due to the extensive automation capabilities robots bring to the manufacturing field like packaging, welding, painting, pick and place, assembly, material handling, and many other processes. Industrial robots repeat these operations with accuracy and precision, and their kinematics must be constructed correctly to maintain accuracy and repeatability [1, 2].

Different robotic applications need different robotic arm architectures. These designs are often divided into two categories: serial and parallel robot. Generally, a robotic arm architecture is made up of links connected in series or parallel by joints to form a kinematic chain. The terminating part of the kinematic chain is the end effector, which is considered as the human hand and is responsible robotic operations [3]. A parallel robot is made up of two connected platforms—one stationary

J. D. Dorman (✉) · G. Kaur · M. Rahim
Chandigarh University, NH-95, Ludhiana Highway, Mohali, India
e-mail: dormanjoseph58@gmail.com

and the other mobile with at least two parallel, independent links, called legs, joining the fixed platform to mobile platform. The end effector is mounted on the mobile platform. The concept of the parallel robots was initially proposed in connection to flight simulators and tire-testing equipment. They have since been used for a variety of operations that require the manipulation of large payloads at high speeds, such as automobile driving simulators. Parallel robots offer several benefits over standard serial robots and are notably useful in assembly and medical applications. For industrial applications, serial robots are the most popular architecture used. In this architecture, the links are connected in series to form a kinematic chain. Each link is connected to the next link by a revolute joint which allows a rotary motion or a prismatic joint which allows a sliding motion. The end effector of a robot is attached to the final link of the kinematic chain and is used to interface with the outside world. Serial robots are best suited to industrial applications because of their wide working area and simple design [4–6].

Under serial robotic arms, there are five robot configurations that are currently employed in industries. The configurations are articulated, Cartesian, spherical, cylindrical, spherical, and SCARA. Each configuration is ideal for a certain industrial application. The articulated robot is the best configuration for simulating a human arm and is ideal for material handling, assembling, welding, and packing [7].

The primary aim of robotics is to move the end effector of a robot to any desired place in a three-dimensional Cartesian space. The kinematics of robotic arms has proven to be extremely important in this domain, since solutions derived from the kinematics allow precise control over the positioning of the end effector [8]. In the field of robotics, kinematics is described as the conversion from Cartesian space to joint space (inverse kinematics) or conversely, from joint space to Cartesian space (forward kinematics) [9].

Ramish [10] developed the mechanical structure of a three-degree of freedom planar robotic arm using a CAD software. The dynamic features of the robotic arm were analyzed, and the complex equations of motion for the robotic arm were modeled using the Euler–Lagrange method. Roboanalyzer simulation was used to validate the analytical results. It can be inferred from this paper that the kinematics of a planar robot is easier to model and evaluate than the kinematics of an articulated robot.

Farman [11] designed a three-degree of freedom articulated robot intended to pick and place lightweight objects based on a color sorting mechanism. A servo motor controlled each axis of the robot using a signal from an Arduino microcontroller that executed a MATLAB code. The MATLAB code incorporated inverse kinematics equations for calculating the necessary joint angles to position the end effector at specified Cartesian coordinates. However, only two inverse kinematics solutions were presented, and these solutions were not tested.

Sharkawy [12] used a multilayer feedforward neural network (MLFFNN) to model and solve the inverse and forward kinematics of a three-degree of freedom articulated robot rather than a geometrical or analytical approach.

Dawood [13] introduced the kinematic model of a three-degree-of-freedom articulated robot arm designed for pick-and-place applications. The kinematic model
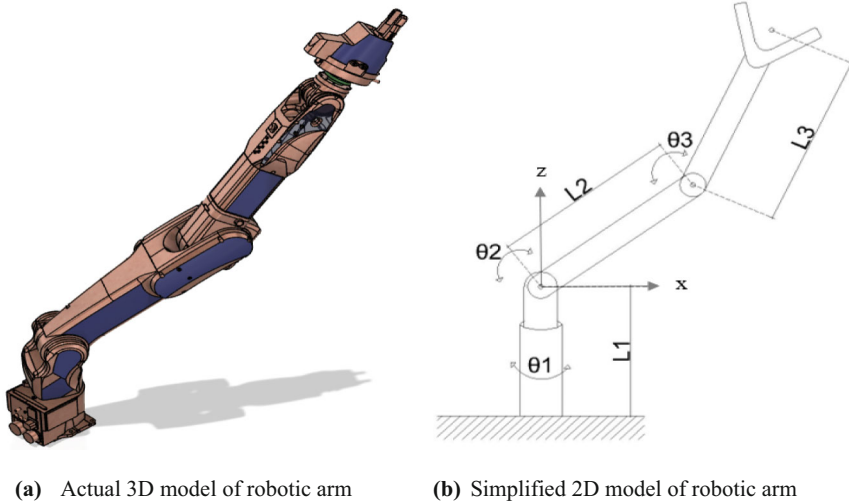
(a)   Actual 3D model of robotic arm   **(b)** Simplified 2D model of robotic arm

Fig. 25.1 3D and 2D models of articulated robotic arm

served as the basis for constructing a real-time robot. In place of an absolute analytical model, a fusion of algebraic equations relying on trigonometric formulas and a geometric approach was used, taking into consideration the presence of a shoulder offset. The Robotic Toolbox in MATLAB was used to evaluate two solutions of the inverse kinematics—left elbow up and left elbow down solutions.

The objective of this paper is to comprehensively model the kinematics (forward and inverse) of a three-axis articulated robotic arm to achieve precise and accurate control of the end effector's positioning. Illustrated in Fig. 25.1, the first joint (θ_1) permits rotary motion around the vertical z-axis, while the remaining two joints (θ_2 and θ_3) enable rotation about the unseen y-axis. Despite some existing studies on the topic, this research uniquely provides a thorough analytical kinematic model, validated using MATLAB, Roboanalyzer, and Fusion 360 software. Ultimately, the goal is to enhance understanding and control of the kinematics of a three-degree-of-freedom articulated robot.

25.2   Modeling of Robotic Arm's Kinematics

The methodology of this paper is structured into two primary sections: modeling and validation. As mentioned earlier, the kinematic model covered in this paper relies entirely on analytical methods. Consequently, the modeling section highlights the analytical approach employed in modeling the robot. This process starts with the derivation of Denavit-Hartenberg (DH) parameters for the four links constituting the robotic arm. Subsequently, the transformation matrix for each link is derived
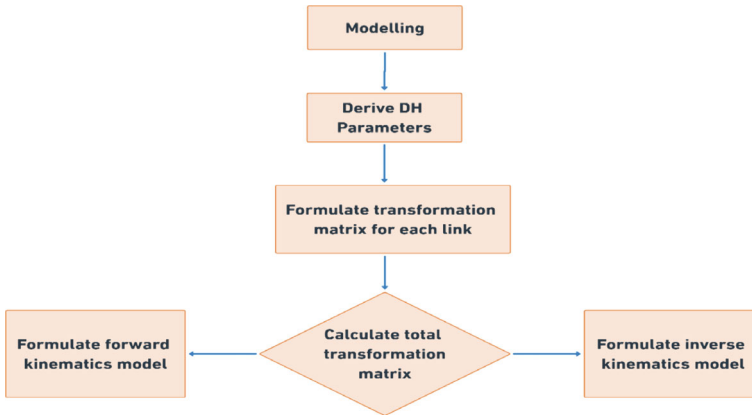
Fig. 25.2 Process flow chart for kinematics modeling

based on the obtained DH parameters. The forward kinematic model is then formulated by sequentially multiplying the four transformation matrices to create a total transformation matrix. The final step in the modeling phase involves utilizing the total transformation matrix to establish the inverse kinematics model for the robot. Figure 25.2 is a process flow chart for the modeling phase.

25.2.1   DH Parameters

The initial step in robot manipulator kinematics modeling is to derive Denavit-Hartenberg (DH) parameters for the robotic arm. To obtain DH parameters for the robotic manipulator, coordinate frames are assigned to each link systematically, starting from the ground link ($i = 0$), which serves as the reference coordinate frame, and progressing to the end effector ($i = 4$), as shown in Fig. 25.3b.

The DH parameters consist of four values associated with each joint and link in a robot arm.

Link Length (a_i): This parameter signifies the distance along the X_{i-1} axis between the Z_{i-1} axis and the Z_i axis. It is measured in meters.

Link Twist (α_i): This parameter denotes the rotation around the X_{i-1} axis required to align the Z_{i-1} axis with the Z_i axis. It is measured in radians.

Link Offset (d_i): This parameter signifies the distance along the Z_i axis between the X_{i-1} axis and the X_i axis. It is measured in meters.

Joint Angle (θ_i): This parameter indicates the rotation around the Z_i axis required to align the X_{i-1} axis with the X_i axis. It is measured in radians.

Using the parameter definitions and the graphical descriptions shown in Fig. 25.3, D-H parameters for the robot may be determined, as indicated in Table 25.1.
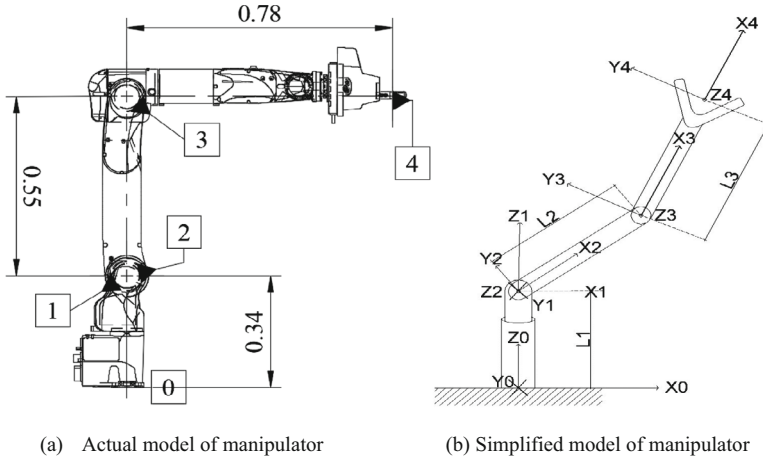
(a) Actual model of manipulator   (b) Simplified model of manipulator

Fig. 25.3 Coordinate frames of manipulator links

Table 25.1 DH parameters for articulated robot

Link (i)	Joint angle (θ_i)	Link offset (d_i)	Link length (a_{i-1})	Link twist (α_{i-1})
1	θ_1	L_1	0	0
2	θ_2	0	0	$\pi/2$
3	θ_3	0	L_2	0
4	0	0	L_3	0

The general formula for a transformation matrix describing the orientation and position of link (i) relative to link ($i-1$) is given as [14]:

$$T_i^{i-1} = R_x(\alpha_{i-1})D_x(a_{i-1})R_z(\theta_i)D_z(d_i) \tag{25.1}$$

$$T_i^{i-1} = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & a_{i-1} s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} \\ -s\alpha_{i-1} d_i s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & c\alpha_{i-1} d_i & 0 & 0 & 0 & 1 \end{bmatrix} \tag{25.2}$$

where, c is cosine, and s is sine. Replacing the values of the parameters in Table 25.1 into Eq. (25.1) results in the following transformation matrices:

$$T_1^0 = T_1^{\text{Ground}} = R_z(\theta_1)D_z(L_1) = [c_1 \ -s_1 \ 0 \ 0 \ s_1 \ c_1 \ 0 \ 0 \ 0 \ 0 \ 1 \ L_1 \ 0 \ 0 \ 0 \ 1] \tag{25.3}$$

$$T_2^1 = R_x\left(\frac{\pi}{2}\right)R_z(\theta_2) = [c_2 \ -s_2 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ s_2 \ c_2 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1] \tag{25.4}$$

$$T_3^2 = D_x(L_2)R_z(\theta_3) = [c_3 \ -s_3 \ 0 \ L_2 \ s_3 \ c_3 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1] \tag{25.5}$$

$$T_4^3 = T_{\text{Gripper}}^3 = D_x(L_3) = [1\ 0\ 0\ L_3\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1] \qquad (25.6)$$

25.2.2 Forward Kinematics

The complete transformation, defining the position and orientation of the end effector relative to the base, can be expressed as follows [15]:

$$T_{\text{Gripper}}^{\text{Ground}} = \begin{bmatrix} r_{xx}\ r_{yx}\ r_{zx}\ t_x\ r_{xy}\ r_{yy}\ r_{zy}\ t_y\ r_{xz}\ r_{yz}\ r_{zz}\ t_z\ 0\ 0\ 0\ 1 \end{bmatrix} \qquad (25.7)$$

In this context, (t_x, t_y, t_z) represents the position of the end effector relative to the ground link's coordinate frame, the unit vector (r_{xx}, r_{xy}, r_{xz}) indicates the orientation of the x-axis of the end effector's frame in relation to the x, y, and z axes of the ground link's coordinate frame, respectively. The unit vector (r_{yx}, r_{yy}, r_{yz}) indicates the orientation of the y-axis in the end effector's frame with respect to the x, y, and z axes of the ground link's coordinate frame, respectively. The unit vector (r_{zx}, r_{zy}, r_{zz}) indicates the orientation of the z-axis in the end effector's frame with respect to the x, y, and z axes of the ground link's coordinate frame, respectively [16].

Forward kinematics involves formulating equations to ascertain the resulting position and orientation of the end effector when specific joint angles are supplied. These equations can be obtained by multiplying Eqs. (25.3) through to (25.6) as indicated below:

$$T_{\text{Gripper}}^{\text{Base}} = T_4^0 = T_1^0 \cdot T_2^1 \cdot T_3^2 \cdot T_4^3 \qquad (25.8)$$

$$\begin{aligned} = [&c_1(c_2c_3 - s_2s_3) - c_1(c_2s_3 + s_2c_3)s_1L_3c_1(c_2c_3 - s_2s_3) + L_2c_1c_2s_1(c_2c_3 - s_2s_3) \\ &-s_1(c_2s_3 + s_2c_3) - c_1L_3s_1(c_2c_3 - s_2s_3) + L_2s_1c_2c_2s_3 \\ &+s_2c_3c_2c_3 - s_2s_30L_1 + L_2s_2 + L_3(c_2s_3 + s_2c_3)\ 0\ 0\ 0\ 1] \end{aligned}$$
$$(25.9)$$

$$\begin{aligned} T_{\text{Gripper}}^{\text{Base}} = [&c_1c_{23} - c_1s_{23}s_1L_3c_1c_{23} + L_2c_1c_2s_1c_{23} - s_1s_{23} \\ &-c_1L_3s_1c_{23} + L_2s_1c_2s_{23}c_{23}\ 0\ L_1 + L_2s_2 + L_3s_{23}\ 0\ 0\ 0\ 1] \qquad (25.10) \end{aligned}$$

where, s_i and c_i are $\sin\theta_i$ and $\cos\theta_i$ respectively and s_{ij} and c_{ij} are $\sin(\theta_i + \theta_j)$ and $\cos(\theta_i + \theta_j)$ respectively. Thus, the resultant forward kinematic equations are:

$$t_x = L_3c_1c_{23} + L_2c_1c_2 \qquad (25.11)$$

$$t_y = L_3s_1c_{23} + L_2s_1c_2 \qquad (25.12)$$

$$t_z = L_1 + L_2 s_2 + L_3 s_{23} \tag{25.13}$$

$$\begin{bmatrix} r_{xx} & r_{xy} & r_{xz} \end{bmatrix} = \begin{bmatrix} c_1 c_{23} & s_1 c_{23} & s_{23} \end{bmatrix} \tag{25.14}$$

$$\begin{bmatrix} r_{yx} & r_{yy} & r_{yz} \end{bmatrix} = \begin{bmatrix} -c_1 s_{23} & -s_1 s_{23} & c_{23} \end{bmatrix} \tag{25.15}$$

$$\begin{bmatrix} r_{zx} & r_{zy} & r_{zz} \end{bmatrix} = \begin{bmatrix} s_1 & -c_1 & 0 \end{bmatrix} \tag{25.16}$$

25.2.3   Inverse Kinematics

In the preceding section, the forward kinematics equations are formulated to establish the pose of the robotic arm based on joint angles. However, when operating robotic arms in the real world, the desired variable is the pose of the robot, and inverse kinematics equations must be derived to find the joint angles $(\theta_1, \theta_2, \theta_3)$ that can produce the desired pose of the robot. Inverse kinematics solutions are key in defining the workspace of the robotic arm, which is the volume comprising all positions reachable by the end effector. If a solution(s) exists for a position, it is in the robot's workspace; otherwise, the position is not in the robotic arm's workspace and is therefore inaccessible to the end effector. To solve for θ_1, divide Eq. (25.12) by Eq. (25.11) as follows:

$$\frac{t_y}{t_x} = \frac{s_1 (L_3 c_{23} + L_2 c_2)}{c_1 (L_3 c_{23} + L_2 c_2)} = \frac{s_1}{c_1} = \tan \ \tan \theta_1 \tag{25.17}$$

$$\frac{-t_y}{-t_x} = \tan \ \tan \theta_1 \tag{25.18}$$

$$\therefore \theta_1 = \operatorname{atan} 2\left(t_y, t_x\right) \text{or a} \tan 2\left(-t_y, -t_x\right) \tag{25.19}$$

The atan2() function in MATLAB returns the four-quadrant inverse tangent of its arguments [11]. To solve for θ_3, the squares of Eqs. (25.11) through to (25.13) are summed as follows:

$$t_x^2 + t_y^2 + (t_z - L_1)^2 = L_3^2 + L_2^2 + 2L_2 L_3 c_3 \tag{25.20}$$

$$c_3 = \frac{t_x^2 + t_y^2 + t_z^2 - \left(L_1^2 + L_2^2 + L_3^2\right) - 2t_z L_1}{2L_2 L_3} \tag{25.21}$$

$$s_3 = \pm \sqrt{1 - c_3^2} \tag{25.22}$$

$$\theta_3 = \text{atan2}(s_3, c_3) \text{ or atan2}(-s_3, c_3) \tag{25.23}$$

The last step in solving for the inverse kinematics is to solve for θ_2 as follows:

$$t_z - L_1 = L_2 s_2 + L_3(c_2 s_3 + s_2 c_3) = c_2 L_3 s_3 + s_2(L_2 + L_3 c_3) \tag{25.24}$$

$c_1 \times (11) + s_1 \times (12)$:

$$c_1 t_x + s_1 t_y = \left(c_1^2 + s_1^2\right)(L_3 c_{23} + L_2 c_2) \tag{25.25}$$

$$c_1 t_x + s_1 t_y = L_3(c_2 c_3 - s_2 s_3) + L_2 c_2 \tag{25.26}$$

$$c_1 t_x + s_1 t_y = c_2(L_3 c_3 + L_2) - s_2 L_3 s_3 \tag{25.27}$$

Substituting $A = L_3 c_3 + L_2$, $B = L_3 s_3$, $x = c_1 t_x + s_1 t_y$, $y = t_z - L_1$ into Eqs. (25.24) and (25.27) gives:

$$x = A c_2 - B s_2 \tag{25.28}$$

$$y = B c_2 + A s_2 \tag{25.29}$$

$$c_2 = \frac{Ax + By}{A^2 + B^2} \tag{25.30}$$

$$s_2 = \frac{Ay - Bx}{A^2 + B^2} \tag{25.31}$$

$$\theta_2 = \text{atan2}(s_2, c_2) \tag{25.32}$$

It is worth noting that there are four general solutions for θ_2 depending on the combinations of θ_1 and θ_3 used.

25.3   Validation of Kinematic Model

The model validation employs a dual approach: analytical calculations in MATLAB and simulations in Roboanalyzer. The forward kinematics model is validated by assigning arbitrary joint angles $(\theta_1, \theta_2, \theta_3)$ in MATLAB to determine the resulting end effector position (t_x, t_y, t_z). These joint angles are then input into the Roboanalyzer software to obtain a base solution for the gripper $(t_x, t_y, t_z)_{base}$. A numerical solution from MATLAB is compared to this base solution, ensuring a comprehensive
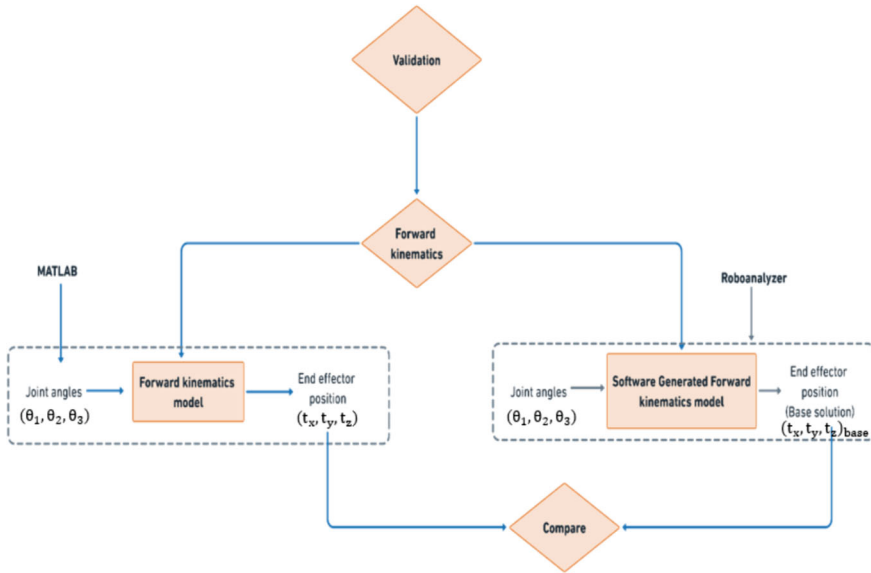
Fig. 25.4 Process flow chart for forward kinematics model validation

evaluation of the forward kinematic model accuracy. See Fig. 25.4 for a process flow chart of the validation.

For the inverse kinematics model validation, the position from the base solution $(t_x, t_y, t_z)_{base}$ is input into MATLAB to calculate the corresponding joint angles $(\theta_1, \theta_2, \theta_3)_{ikine}$. These computed joint angles are then applied to the forward kinematics of the Roboanalyzer robotic arm. The solutions produced by Roboanalyzer are compared with the base solution, providing a comprehensive evaluation of the inverse kinematics model accuracy. This iterative and cross-platform validation approach ensures a robust assessment, instilling confidence in the reliability of both the forward and inverse kinematics models for the investigated robotic arm. Refer to Fig. 25.5 for a process flow chart of the inverse kinematic model validation.

Visual analysis for the model, a wooden slab is placed at the same location as the base solution, $(t_x, t_y, t_z)_{base}$ in Fusion 360. The robotic arm is then given the same joint angles, that were used to calculate the base solution $(\theta_1, \theta_2, \theta_3)$, and the gripper's position is checked against the wooden slab. Next, the robotic arm is given the joint angles from each of the inverse kinematic solutions $(\theta_1, \theta_2, \theta_3)_{ikine}$ and the gripper's position is again checked against the wooden slab.
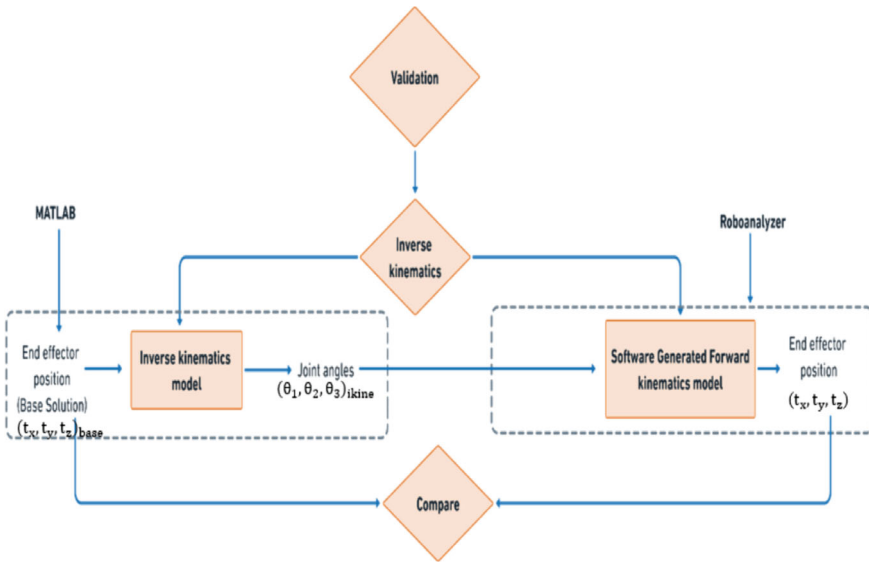
Fig. 25.5 Process flow chart for inverse kinematics model validation

25.3.1 Forward Kinematics

The values of the parameters used to simulate the forward kinematics model are shown in Table 25.2. This includes the dimensions of the robot in meters as provided in Fig. 25.3a, and selected arbitrary values are for the joint angles.

Equation 25.33 illustrates the matrix created by MATLAB when the parameters in Table 25.2 are inserted into the total transformation matrix, which defines this paper's forward kinematics model.

$$T_{\text{Gripper}}^{\text{Base}} = [0.8660 \ -0.5 \ 0 \ 0.4005 \ 0 \ 0 \ -1 \ 0 \ 0.5 \ 0.8660 \ 0 \ 1.2063 \ 0 \ 0 \ 0 \ 1]$$

(25.33)

Upon entering the same set of parameters into the robot model constructed using Roboanalyzer, the matrix shown in Eq. (25.34) is obtained. The first three rows in the final column of the transformation matrix produced by Roboanalyzer, indicate the position of the end effector, which constitutes the base solution $(t_x, t_y, t_z)_{base}$.

Table 25.2 Parameters for kinematics simulation

L1 (m)	L2 (m)	L3 (m)	θ_1 (rad)	θ_2 (rad)	θ_3 (rad)
0.34	0.55	0.78	2π	$\frac{2\pi}{3}$	$\frac{3\pi}{2}$

Fig. 25.7 Base solution
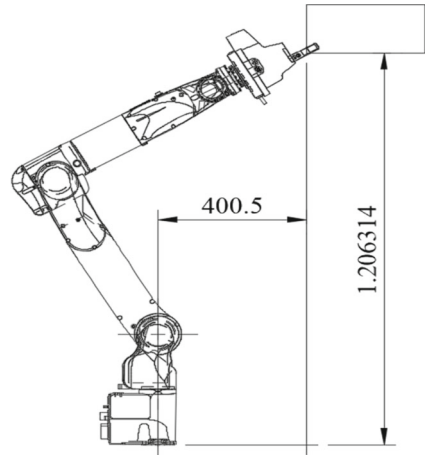
$$T_{\text{Gripper}}^{\text{Base}} = [0.866025 \ -0.5 \ 0 \ 0.4005 \ 0 \ 0 \ -1 \ 0 \ 0.5 \ 0.866025 \ 0 \ 1.206314 \ 0 \ 0 \ 0 \ 1]$$
(25.34)

Based the on this result, the position of the end effector when the joint angles are adjusted to $\left(2\pi, \frac{2\pi}{3}, \frac{3\pi}{2}\right) rad$ is $(0.4005, 0, 1.206314)m$. Figure 25.7 provides a visual representation of the end effector's actual position when the robot's joint angles are adjusted to the arbitrary joint angles, $\left(2\pi, \frac{2\pi}{3}, \frac{3\pi}{2}\right) rad$. This constitutes the base solution for other visual analysis.

25.3.2 Inverse Kinematics

In this segment, by establishing the target end effector position at $(0.4005, 0, 1.206314)m$, the inverse kinematics equations are employed to calculate the corresponding joint angles $(\theta_1, \theta_2, \theta_3)_{ikine}$. Table 25.3 depicts the combinations of θ_1, θ_2 and θ_3 used for each solution and Table 25.4 depicts the unique solution obtained when executed with MATLAB.

In the next step of the validation process, the joint angle values obtained from each solution in Table 25.4, are provided as input to the forward kinematics of the

Table 25.3 Inverse kinematics solutions 1

Solution	θ_1	θ_2	θ_3
1	$\text{atan2}(t_y, t_x)$	$\text{atan2}(s_2, c_2)$	$\text{atan2}(s_3, c_3)$
2	$\text{atan2}(-t_y, -t_x)$	$a\tan 2(s_2, c_2)$	$a\tan 2(-s_3, c_3)$
3	$\text{atan2}(t_y, t_x)$	$\text{atan2}(s_2, c_2)$	$\text{atan2}(-s_3, c_3)$
4	$\text{atan2}(-t_y, -t_x)$	$\text{atan2}(s_2, c_2)$	$\text{atan2}(s_3, c_3)$

Table 25.4 Inverse kinematics solutions 2

Solution	θ_1 (rad)	θ_2 (rad)	θ_3 (rad)
1	0	0.1811	1.5708
2	−3.1416	2.9605	−1.5708
3	0'	2.0944	−1.5708
4	−3.1416	1.0472	1.5708

Roboanalyzer robotic arm. Equations (25.35) to (25.38) represent the transformation matrices produced by solutions 1, 2, 3, and 4 respectively. Figure 25.8 provides a visual representation of the resulting end effector positions when the robot is adjusted to the joint angles of each solution.

$$T_{\text{Gripper}}^{\text{Base}} = [-0.180112 \ -0.983646 \ 0 \ 0.400518 \ 0 \ 0 \ -1 \ 0 \ 0.983646$$
$$-0.180112 \ 0 \ 1.206306 \ 0 \ 0 \ 0 \ 1] \tag{25.35}$$

$$T_{\text{Gripper}}^{\text{Base}} = [-0.180104 \ 0.983648 \ 0 \ 0.400525 \ 0 \ 0 \ 1 \ 0 \ 0.983648 \ 0.180104 \ 0$$
$$1.206302 \ 0 \ 0 \ 0 \ 1] \tag{25.36}$$

$$T_{\text{Gripper}}^{\text{Base}} = [0.866025 \ -0.5 \ 0 \ 0.4005 \ 0 \ 0 \ -1 \ 0 \ 0.5 \ 0.866025 \ 0 \ 1.206314 \ 0 \ 0 \ 0 \ 1] \tag{25.37}$$

$$T_{\text{Gripper}}^{\text{Base}} = [0.866025 \ 0.5 \ 0 \ 0.4005 \ 0 \ 0 \ 1 \ 0 \ 0.5 \ -0.866025 \ 0 \ 1.206314 \ 0 \ 0 \ 0 \ 1] \tag{25.38}$$

25.4 Results and Discussion

25.4.1 Visual Assessment

The performance criteria used to assess the kinematic model of the articulated robot are precision and accuracy. Accuracy is a measure of how close a calculated value is to the true or accepted value. Precision on the other hand is a measure of how close multiple measurements of the same quantity are to each other.

The ground link's coordinate frame is fixed at the position (0, 0, 0) m, and end effector poses are measured with respect to it in all simulations. It can be observed from Fig. 25.8 shows that solutions 1 and 2 employ an elbow down configuration for the gripper to reach the wooden slab, while solutions 3 and 4 employ an elbow up configuration. Solutions 1 and 3 produce a left arm configuration, as indicated by the left side of the robotic arm (colored green) in Fig. 25.8a, c directed towards the viewer.
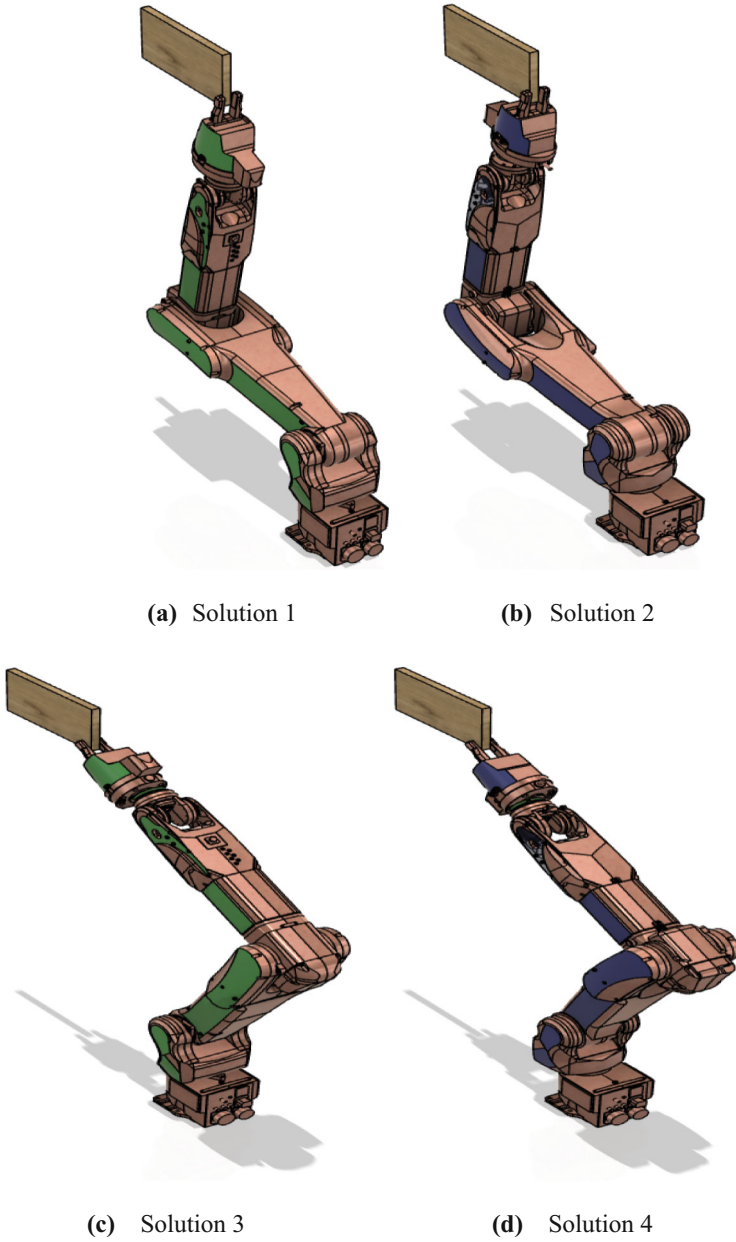
(a)   Solution 1   (b)   Solution 2

(c)   Solution 3   (d)   Solution 4

Fig. 25.8 Poses of robotic arm for inverse kinematic solutions

Table 25.5 Accuracy of forward and inverse kinematics solutions

Solution	x accuracy (%)	y accuracy (%)	z accuracy (%)
Forward kinematics	100	100	99.9988
Solution 1	99.9955	100	99.9993
Solution 2	99.9938	100	99.9990
Solution 3	100	100	100
Solution 4	100	100	100

Solutions 2 and 4 on the other hand produce a right arm configuration, as indicated by the right side of the robotic arm (colored blue) in Fig. 25.8b, d directed towards the viewer. Furthermore, by comparing Figs. 25.7 and 25.8c, as well as Eqs. (25.37) and (25.34), it can be observed that solution 3 produces the same arm configuration as the base solution. It can be observed that for all solutions, the gripper was able to reach the targeted position, indicating that the kinematics model is accurate.

25.4.2 Numerical Evaluation

Table 25.5 is a measure of the accuracy of solutions obtained from the forward and inverse kinematics, in the x, y, and z directions. This is done using the equation below:

$$\text{Accuracy} = \left(1 - \frac{|\text{Measured Value} - \text{Base Value}|}{\text{Base value}} \right) \times 100\% \qquad (25.39)$$

It can be seen from Table 25.5 that the solutions produced accurate results. Also, the position of the gripper in all solutions is either the same as the base solution or very close to it, indicating consistency, and hence precision.

25.5 Conclusion

The aim of this research was to model the kinematics of a three-axis articulated robot for precise and accurate control of its end effector. This was completed analytically using the DH convention. Visual analysis from simulations showed that the end effector could reach its targeted position for all solutions of the model, indicating the model's accuracy. This was done in four configurations—left arm elbow down, right arm elbow down, left arm elbow up, and right arm elbow down. Further evaluation showed that by using the model, the end effector can reach its targeted position with a high degree of accuracy and precision.

References

1. Aravinthkumar, T., Suresh, M., Vinod, B.: Kinematic analysis of 6 DOF articulated robotic arm. Int. Res. J. Multidiscip. Technov. **3**(1), 1–5 (2021). https://doi.org/10.34256/irjmt2111
2. Javaid, M., Haleem, A., Singh, R.P., Suman, R.: Substantial capabilities of robotics in enhancing industry 4.0 implementation. Cognit. Robot. **1**, 58–75 (2021). https://doi.org/10.1016/j.cogr.2021.06.001
3. Surati, S., Hedaoo, S., Rotti, T., Ahuja, V., Patel, N.: Pick and place robotic arm: a review paper. Int. Res. J. Eng. Technol. (2021)
4. Khalil, W., Dombre, E.: Chapter 8. Introduction to geometric and kinematic modeling of parallel robots. (2002). https://doi.org/10.1016/B978-190399666-9/50008-7
5. Dan, V., Stan, S.-D., Manic, M., Balan, R.: Mechatronic design, kinematics analysis of a 3 DOF medical parallel robot. (2010). https://doi.org/10.1109/ISRCS.2010.5604105
6. Latifinavid, M., Azizi, A.: Kinematic modelling and position control of a 3-DOF parallel stabilizing robot manipulator. J. Intell. Robot. Syst. **107**(2), 17 (2023). https://doi.org/10.1007/s10846-022-01795-x
7. Jahnavi, K., Parameswaran, S.: Teaching and learning robotic arm model. (2017).https://doi.org/10.1109/ICICICT1.2017.8342804
8. Singh, G., Banga, V.K.: Robots and its types for industrial applications. Mater Today Proc. **60** (2022). https://doi.org/10.1016/j.matpr.2021.12.426.
9. Deshpande, V., George, P.M.: Kinematic modelling and analysis of 5 DOF robotic arm. Int. J. Robot. Res. Dev. **4**, 17–24 (2014)
10. Ramish, Hussain, S., Kanwal, F.: Design of a 3 DoF robotic arm. (2016). https://doi.org/10.1109/INTECH.2016.7845007
11. Farman, M., Al-Shaibah, M., Aoraiath, Z., Jarrar, F.: Design of a three degrees of freedom robotic arm. Int. J. Comput. Appl. **179**, 12–17 (2018). https://doi.org/10.5120/ijca2018916848
12. Sharkawy, A.-N., Khairullah, S.: Forward and inverse kinematics solution of a 3-DOF articulated robotic manipulator using artificial neural network. Int. J. Robot. Control Syst. **3**, 330–353 (2023). https://doi.org/10.31763/ijrcs.v3i2.1017
13. Dawood, H., Hamzah, M., Bakhy, S.: Kinematics analysis and implementation of three degrees of freedom robotic arm by using Matlab. Iraqi J. Mech. Mater. Eng. **21**, 118–129 (2021). https://doi.org/10.32852/iqjfmme.v21i2.547
14. Craig, J.J., Prentice, P., Hall, P.P.: Introduction to Robotics Mechanics and Control, 3rd edn. (2005)
15. Spong, M.W., Hutchinson, S., Vidyasagar, M.: Robot Modeling and Control, 1st edn
16. Corke, P.: Robotics, vision and control. In: Springer Tracts in Advanced Robotics, vol. 118. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-54413-7

Chapter 26
Hand Gesture Control for Automated Digital Presentations

Vijaya Prakash Rajanala , **Maduri Ram Charan Teja, Suhaas Sanga, Renukuntla Dhanush, Kothapally Prem, and Gurrapu Aditya Krishna**

Abstract This project integrates computer vision and gesture recognition techniques to develop an interactive slideshow navigation system. The program utilizes the OpenCV library for image processing and the CV zone library for hand tracking. Users can control the slideshow by performing specific hand gestures in front of a webcam. The system begins by allowing the user to select a folder containing PNG images, which are then sequentially renamed. The main functionality involves gesture-based control for navigating through the images in the slideshow. Hand gestures, detected using the Hand Tracking Module, are mapped to actions such as moving to the previous or next slide, erasing annotations, and showing and drawing pointers on the images. Additionally, the system provides real-time feedback by displaying the webcam feed alongside the slideshow. The interactive nature of this project makes it suitable for presentations or educational purposes where users can dynamically interact with the displayed content.

Keywords Recognition · Machine learning · Gesture detection · Gesture classification

26.1 Introduction

In an era characterized by digital communication and information dissemination, the art of effective presentation holds a crucial place [1]. Traditional means of navigating presentation slides, involving peripheral devices like mice, keyboards, or laser pointers, often demand prior technical knowledge, creating barriers for users [2]. However, this project introduces an innovative solution that leverages the power of dynamic hand gestures and OpenCV technology to redefine the way we interact

V. P. Rajanala · M. R. C. Teja (✉) · S. Sanga · R. Dhanush · K. Prem · G. A. Krishna
School of Computer Science and Artificial Intelligence, SR University, Warangal, India
e-mail: ramcharantejamaduri@gmail.com

V. P. Rajanala
e-mail: r.vijayaprakash@sru.edu.in

with presentations. The core of this endeavor is the use of a camera to capture and interpret six distinct hand gestures. Each of these gestures triggers specific actions within the presentation, enabling presenters to seamlessly navigate between slides, write or erase content, highlight sections, and even terminate the presentation, all through intuitive hand movements [3]. What sets this system apart is its ability to accomplish all this without the need for additional hardware, specialized gloves, or markers, making it a cost-effective and user-friendly alternative. The underlying technology is built upon the robust foundation of the Python framework, with critical components including OpenCV, CV Zone, NumPy, and Media Pipe [4, 5].

Through the fusion of machine learning and motion image-based techniques, the system accurately recognizes intricate hand motions, empowering presenters to communicate nonverbally, engage their audiences, and maintain precise control over their presentations. This project represents a groundbreaking fusion of machine learning and computer vision to create a versatile human–machine interface that enhances the traditional presentation experience. By using gestures such as swiping, giving a thumbs-up, or halting, users can effortlessly command their presentation slides, significantly improving the fluidity and expressiveness of their presentations. Ultimately, the project's goal is to empower presenters by providing a more natural and interactive means of controlling presentations, thereby enhancing the overall impact and effectiveness of their messages. In the digital age, this dynamic hand gesture-based control system promises to revolutionize the art of presentations, offering a modern and engaging tool for communicators.

26.2   Literature Survey

In their study, authors Devivara Prasad et al. [6] explored the significance of gesture recognition in Human–Computer Interaction (HCI), emphasizing its practical applications for individuals with hearing impairments and stroke patients. They used image feature extraction tools and AI-based classifiers for 2D and 3D gesture recognition. Their proposed system harnesses machine learning, and real-time image processing with Media Pipe, and OpenCV to enable efficient and intuitive presentation control using hand gestures, addressing the challenges of accuracy and robustness. The research focuses on enhancing the user experience, particularly in scenarios where traditional input devices are impractical, highlighting the potential of gesture recognition in HCI [13, 15].

Reethika et al. [7] presented a study on Human–Computer Interaction (HCI) with a focus on hand gesture recognition as a natural interaction technique. It explores the significance of real-time hand gesture recognition, particularly in scenarios where traditional input devices are impractical. The methodology involves vision-based techniques that utilize cameras to capture and process hand motions, offering the potential to replace conventional input methods. The paper discusses the advantages and challenges of this approach, such as the computational intensity of image processing and privacy concerns regarding camera usage. Additionally, it highlights

the benefits of gesture recognition for applications ranging from controlling computer mouse actions to creating a virtual HCI device [16].

Khanum and Pramod [8] outlined a methodology that harnesses OpenCV and Google's MediaPipe framework [16–18] to create a presentation control system that interprets hand gestures. Using a webcam, the system captures and translates hand movements into actions such as slide control, drawing on slides, and erasing content, eliminating the need for traditional input devices. While the paper does not explicitly enumerate the challenges encountered during system development, common obstacles in this field may include achieving precise gesture recognition, adapting to varying lighting conditions, and ensuring the system's reliability in real-world usage scenarios. This work contributes to the advancement of human–computer interaction, offering a modern and intuitive approach to controlling presentations through hand gestures [19].

Powar et al. [9] introduced a system that utilizes artificial intelligence-based hand gesture detection, employing OpenCV and MediaPipe. The system allows users to control presentation slides via intuitive hand gestures, eliminating the reliance on conventional input devices like keyboards or mice. The gestures correspond to various actions, including initiating presentations, pausing videos, transitioning between slides, and adjusting volume. This innovative approach enhances the natural interaction between presenters and computers during presentations, demonstrating its potential in educational and corporate settings. Notably, the paper does not explicitly detail the challenges encountered during the system's development, but it makes a valuable contribution to the realm of human–computer interaction by rendering digital presentations more interactive and user-friendly [20].

Gokul et al. [10] presented a real-time interactive presentation system that utilizes hand gestures for control. The system integrates a thermal camera for robust human body segmentation, overcoming issues with complex backgrounds and varying illumination from projectors. They proposed a fast and robust hand localization algorithm and a dual-step calibration method for mapping interaction regions between the thermal camera and projected content using a web camera. The system has high recognition rates for hand gestures, enhancing the presentation experience. However, the challenges they encountered during development, such as the need for precise calibration and handling hand localization, are not explicitly mentioned in the paper [21].

Paulson et al. [11] introduced a gesture recognition system for enhancing presentations and enabling remote control of electronic devices through hand gestures. It incorporates ATMEGA 328, Python, Arduino, Gesture Recognition, Zigbee, and wireless transmission [22]. The paper emphasizes the significance of gesture recognition in human–computer interaction, its applicability in various domains, and its flexibility to cater to diverse user needs. The system offers features such as presentation control, home automation, background change, and sign language interpretation. The authors demonstrated a cost-effective prototype with easy installation and extensive wireless signal transmission capabilities. The paper discusses the results, applications, methodology, and challenges, highlighting its potential to improve human–machine interaction across different fields.

Damdoo et al. [12] presented a vision-based adaptive hand gesture recognition system employing Convolutional Neural Networks (CNN) for machine learning classification. The study addresses the challenges of recognizing dynamic hand gestures in real-time and focuses on the impact of lighting conditions. The authors highlight that the performance of the system significantly depends on lighting conditions, with better results achieved under good lighting. They acknowledge that developing a robust system for real-time dynamic hand gesture recognition, particularly under varying lighting conditions, is a complex task. The paper offers insights into the potential for further improvement and the use of filtering methods to mitigate the effects of poor lighting, contributing to the field of dynamic hand gesture recognition.

Rutika et al. [13] presented a real-time hand gesture recognition system for efficient human–computer interaction. It allows remote control of PowerPoint presentations through simple gestures, using Histograms of Oriented Gradients and K-Nearest Neighbor classification with around 80% accuracy. The technology extends beyond PowerPoint to potentially control various real-time applications. The paper addresses challenges in creating a reliable gesture recognition system and optimizing lighting conditions. It hints at broader applications, such as media control, without intermediary devices, making it relevant to the human–computer interaction field. References cover related topics like gesture recognition in diverse domains.

26.3 Methodology

The project's primary objective is to make the presentation easy for the presenter to deliver comfortably by controlling the complete presentation through hand gestures.

The whole concept of this project is demonstrated in Fig. 26.1. It gives a complete step-by-step process from uploading of files to till termination of the presentation.
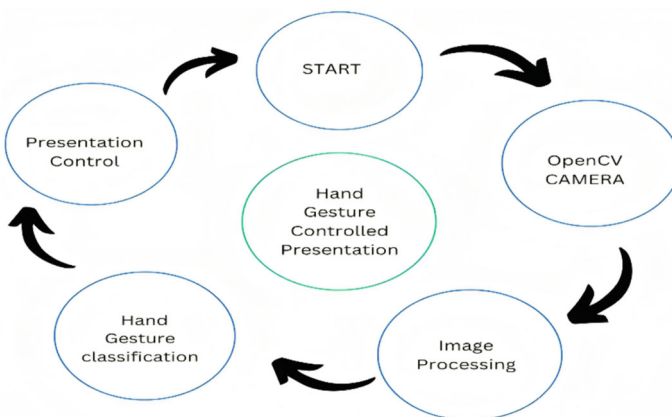

Fig. 26.1 Cyclic process

26.3.1   Data Collection

In this project, the input data is given by the user in the form of ppt slides in images format where the user will convert the ppt slides into images and those images will be stored in a folder. The folder with images is the data for this project, specified in Fig. 26.1.

26.3.2   Data Preprocessing

To rename and organize a set of PNG images, the initial step involves assigning sequential numbers to them in the desired order. This can be achieved through scripting or batch operations using programming or command-line tools. Once renamed, the images will have consecutive identifiers, making it easier to organize and retrieve them in a logical order.

   After successfully renaming the PNG images with sequence numbers, the next step is to sort them based on these assigned numerical values. Sorting ensures that the images are used in the correct order, following the numerical sequence. This process is crucial when creating presentations (PPT) or when a specific order is required for image usage, as it ensures that the images are in the desired sequence for easy access and presentation purposes. Overall, these procedures simplify the task of organizing and working with PNG images in a structured and orderly manner. After uploading the files folder, the data preprocessing starts renaming the images and sorting immediately and storing them back in the folder takes place as shown in Fig. 26.2.

Hand Detection: The method recognizes and localizes a hand's position within a video frame. Hand detection is the key objective in this research, and we employed the Kanade-Lucas-Tomasi (KLT) algorithm to identify and locate all known objects in a scene [14]. The algorithm starts by identifying feature points in the first frame
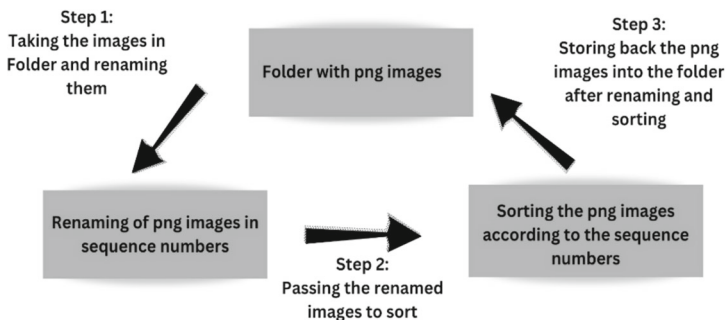

Fig. 26.2 Preprocessing of data

of a video or image sequence. These features could include corners, edges, or any other distinguishing points in the image. The Harris corner detector [15] is commonly used for feature detection. It detects corners by analyzing intensity changes in various directions. Once the features are identified in the first frame, the algorithm attempts to track them in subsequent frames. It is assumed that the features move in small steps between frames.

A small window is considered around each initial frame feature point. The algorithm searches the next frame for the best window match. Feature point optical flow is estimated using the Lucas-Kanade method [10]. The motion is assumed to be constant in a local neighborhood around the feature point. The optical flow equation is solved for each window pixel around the feature point. Motion parameters (w) and spatial intensity gradients (I_x and I_y) are related by this equation. The KLT algorithm analyzes spatial gradient matrix as specified in Eq. (26.1), eigenvalues to determine feature point tracking reliability. Spatial gradients of intensity in the window around the feature point determine the matrix. A feature point is reliable for tracking if its matrix eigenvalues are above a threshold. Figure 26.3 describes the tracking of the hand with the help of matrix eigenvalues.

$$\left[\ \Sigma \ w^2 I_x^2 \ \ \Sigma \ w^2 I_x I_y \ \ \Sigma \ w^2 I_x I_y \ \ \Sigma \ w^2 I_y^2 \ \right] \tag{26.1}$$

Finger Tracking: After detecting the hand, the algorithm records the location of individual fingers. It may entail estimating hand landmarks to pinpoint crucial spots on the fingers, particularly the fingertips.

Finger State Classification: The algorithm defines each finger's state as "up" (1) or "down" (0) based on its location and movement. To establish these classifications, it most likely evaluates the angles and placements of the fingers compared to a reference hand form.

Finger State Combination: The algorithm creates a combination of finger states for the entire hand. For instance, if all fingers are labeled "up," it may indicate "5". If all the fingers are marked "down," it may indicate "0."
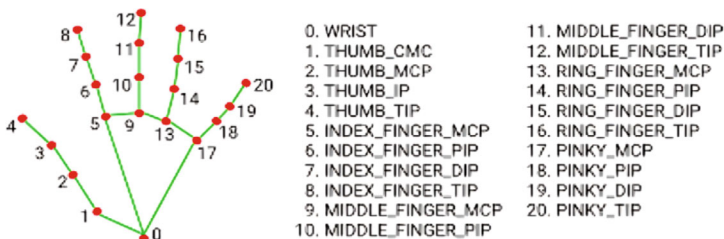

Fig. 26.3 Hand track mechanism

26.4   Results

The hand tracking mechanism, finger state classification, and combination allow each finger to be identified and assigned to a specific task. Figure 26.4 depicts this classification for the purpose of presentation. The first gesture is used to move the slide to the previous slide, the second gesture is used for the next slide, the third one is used for the pointer to point the object on the slide, the fourth gesture is used to delete the object drawn with the help of the fifth gesture, and the final gesture is used to exit the presentation.

There were several experiments that we carried out in order to assess the effectiveness of the system. The first experiment was designed to determine how accurate the detection and classification of hand gestures turned out to be. We discovered that the system was able to accurately detect and categorize hand gestures in most situations. Figure 26.5 shows the hand tracking and gesture accuracy with the help of the KLT algorithm of the system, an accuracy rate of approximately 95%. We conducted a second experiment in which we examined the system's capability of controlling a presentation with hand gestures. It was discovered by us that the system was able to control the slides in a smooth manner and carry out a variety of actions, such as moving forward or going back to the slide that came before it.

In the current model, we simply set the gesture array using the built-in Hand Tracking Module, saving time on training, and collecting hand gestures. Converting PowerPoint to images and uploading them will take very little time. The accuracy of the built-in model ranges from 95 to 97%. The previous model required more time for hand tracking because there was no built-in model for detecting hand gestures, and the accuracy was <95%.

For this project HD camera is mandatory, the range of normal inbuilt cameras in existing laptops is 5 m. To get a long range of gesture recognition we need to use external long-range cameras. Once the termination gesture is used the files will be deleted. If the user wants to use the files again then they should upload the files again.
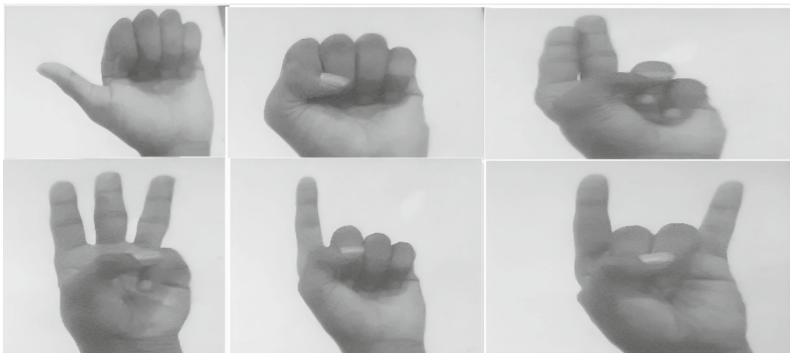

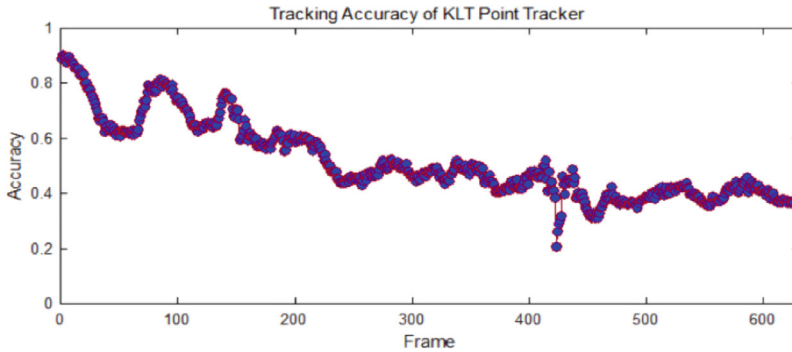**Fig. 26.4**   Gestures to control the presentation

Fig. 26.5 Accuracy And Frame Analysis of KLT Algorithm

26.5 Conclusion

This project is an innovative and interactive presentation control system that utilizes computer vision and gesture recognition. It offers a hands-free and engaging way to interact with your presentation slides. With the ability to control slide navigation through specific hand gestures, such as moving to the next or previous slide, your project provides a convenient and intuitive alternative to traditional clickers or keyboard shortcuts. Additionally, the option to write on the slides and erase content by making hand movements enhances the interactivity of your presentations. The pointer highlighter feature allows you to draw attention to specific details on your slides, making it a powerful tool for emphasizing key points.

Furthermore, the capability to terminate the presentation with a gesture provides an efficient way to conclude your talk. Overall, your project empowers presenters to connect with their audience more dynamically and engagingly, all while using the OpenCV interface and hand-tracking technology. It is a valuable addition to the realm of presentation tools, enabling more interactive and captivating communication. The future enhancement for this will be adding voice commands along with the hand gesture and presenter recognition.

The challenges in this are that we are currently using 6 gestures to control the presentation, but we can improve it by adding more gestures. We intend to add speech commands to perform operations such as moving the slides back and forth.

References

1. Lawrence, D.O., Ashleigh, M.J.: Impact of Human-Computer Interaction (HCI) on Users in Higher Educational System: Southampton University as a Case Study, vol. 6, no. 3, pp. 1–12, September (2019)
2. Raschka, S., Patterson, J., Nolet, C.: Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence (2020)

3. Zhai, X., Chu, X., Chai, C.S., Jong, M.S.Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., Li, Y.: A review of artificial intelligence (AI) in education from 2010 to 2020 (2021)
4. Jadhav, D., Lobo, L.M.R.J.: Hand gesture recognition system to control slide show navigation. IJAIEM **3**(4) (2014)
5. Ren, Z., et al.: Robust part-based hand gesture recognition using Kinect sensor. IEEE Trans. Multimedia **15**(5), 1110–1120 (2013), pp. 8–11
6. Devivara Prasad, G., Srinivasulu, M.: Hand gesture presentation by using machine learning. IJIRT **9**(4) (2022)
7. Reethika, G., Anuhya, P., Bhargavi, M.: Slide presentation by hand gesture recognition using machine learning. IRJET **10**(01), January 2023 (2023)
8. Khanum, H., Pramod, H.B.: Smart presentation control by hand gestures using computer vision and Google's Mediapipe. IRJET **9**(07), July 2022 (2022)
9. Powar, S., Kadam, S., Malage, S., Shingane, P.: Automated digital presentation control using hand gesture technique. ITM Web Conf. **44**, 03031 (2022)
10. Gokul, L.P., Adarsh, P., Vinayan, G., Ponmalar, M., Aswini, S.H.: Lucas Kanade based optical flow for vehicle motion tracking and velocity estimation. In: International Conference on Control, Communication and Computing (ICCC), Thiruvananthapuram, India, pp. 1–6 (2023). https://doi.org/10.1109/ICCC57789.2023.10165227
11. Paulson, M., Nathasha, P.R., Davis, S., Varma, S.: Smart Presentation Using Gesture Recognition, vol. 2, no. 3 (2017)
12. Damdoo, R., Kalyani, K., Sanghavi, J.: Adaptive hand gesture recognition system using machine learning approach. Biosci. Biotech. Res. Comm. Spec. Iss. **13**(14), 106–110 (2020)
13. Rutika, B., Shweta, C., Shraddha, D., Auti, M.A.: Powerpoint presentation control using hand gestures recognition. Int. J. Res. Publ. Rev. **4**(5), 5865–5869, May 2023 (2023)
14. Mikhaylov, D., Samoylov, A., Minin, P., Egorov, A.: Face detection and tracking from image and statistics gathering. In: 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, Morocco, pp. 37–42 (2014). https://doi.org/10.1109/SITIS.2014.85
15. Zhao, J., Su, L., Wang, X., Li, J., Yang, F., Jiang, N., Hu, Q.: DTFS-eHarris: a high accuracy asynchronous corner detector for event cameras in complex scenes. Appl. Sci. **13**, 5761 (2023). https://doi.org/10.3390/app13095761
16. Wahid, M.F., Tafreshi, R., Al-Sowaidi, M., Langari, R.: An efficient approach to recognize hand gestures using machine-learning algorithms. In: 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME), Tunis, Tunisia, pp. 171–176 (2018). https://doi.org/10.1109/MECBME.2018.8402428
17. Talele, A., Patil, A., Barse, B.: Detection of real-time objects using TensorFlow and OpenCV. Asian J. Converg. Technol. **5** (2019)
18. Al Saedi, A.K.H., Al Asadi, A.H.H.: A new hand gestures recognition system. Indonesian J. Electr. Eng. Comput. Sci. **18** (2020)
19. Raschka, S., Patterson, J., Nolet, C.: Machine learning in Python: main developments and technology. Trends Data Sci. Mach. Learn. Artif. Intell. (2020)
20. Dhall, I., Vashisth, S., Aggarwal, G.: Automated hand gesture recognition using a deep convolutional neural network. In: 10th International Conference on Cloud Computing, Data Science & Engineering (2020)
21. Paulson, M., Natasha, P.R., Davis, S., Varma, S.: Smart presentation using gesture recognition and OpenCV. Asian J. Converg. Technol. **5** (2019)

Chapter 27
A Heuristic Approach to Resolve Priority-Driven Unbalanced Transportation Problem (PUTP)

Abu Sayeed Arif, Md. Ashraful Babu , Mohammad Nazrul Islam, Md. Sharif Uddin, and Ramesh Chandra Poonia

Abstract This research addresses the priority-driven unbalanced transportation Problem (PUTP), characterized by a situation where the overall demand surpasses the available supply. We propose the Max-flow Min-cost Priority-driven Unbalanced Transportation Problem (MMPUTP) as a heuristic approach to handle this issue effectively. The strategy of MMPUTP focuses on optimizing resource allocation and reducing costs, making it highly effective in fulfilling high priority needs in a cost-efficient manner. Through a comparison with Vogel's Approximation Method (VAM) over different sets of problems ranging in size from 5×5 to 50×50, the effectiveness of the MMPUTP algorithm is evident. The findings underscore the significance of choosing the right algorithm based on the size and complexity of the problem set in the context of the Priority-driven Unbalanced Transportation Problem, with MMPUTP proving to be a flexible and reliable option in various situations.

A. S. Arif
Department of Mathematics, Jahangirnagar University, Savar, Dhaka, Bangladesh
e-mail: abuarif.bd@gmail.com

Md. A. Babu (✉)
Department of Physical Sciences, Independent University, Bangladesh, Dhaka 1229, Bangladesh
e-mail: ashraful388@gmail.com

M. N. Islam
Department of Mathematics, Jahangirnagar University, Dhaka 1342, Bangladesh
e-mail: islamju@juniv.edu

Md. Sharif Uddin
Department of Industrial Engineering, Prince Sattam Bin Abdulaziz University, Al Kharj 11942, Saudi Arabia
e-mail: s.uddin@psau.edu.sa

R. C. Poonia
Department of Computer Science, CHRIST (Deemed to Be University), Delhi-NCR 201003, New Delhi, India
e-mail: rameshchandra.poonia@christuniveristy.in

27.1   Introduction

The transportation problem focuses on optimizing the distribution of goods from multiple suppliers to a variety of destinations in order to minimize transportation costs. This optimization challenge is often visualized as a matrix, where the suppliers are represented by rows, the consumers by columns, and the transportation costs by the values in the matrix cells. Transportation problems are classified as either balanced, where supply equals demand, or unbalanced, where there is a discrepancy between supply and demand, leading to excess supply or demand.

The Priority-driven Unbalanced Transportation Problem (PUTP) is a complex iteration of the unbalanced transportation problem. It adds the dimension of prioritizing certain destinations over others. This differs from standard unbalanced transportation problems, which focus primarily on cost efficiency without considering the relative importance of destinations. In PUTP, destinations are ranked based on various criteria like urgency, profitability, strategic value, or customer preferences, prioritizing some demands over others.

The Priority-driven Unbalanced Transportation Problem (PUTP) is crucial for strategic planning, emergency response, supply chain management, and logistics, enhancing operational efficiency and customer satisfaction. It prioritizes demand, aiding in making strategic decisions for rapid, cost-effective distribution. A novel heuristic algorithm, Max-flow Min-cost Priority-driven Unbalanced Transportation Problem (MMPUTP), improves on the Least Cost Method by focusing on high-priority demands first, optimizing for cost-effective routes while meeting urgent needs efficiently.

27.2   Literature Review

The groundwork for the mathematical analysis of the transportation problem was laid by Gaspard Monge, a French mathematician and physicist, in 1781, with his work on optimizing soil transportation costs for construction [1]. This field was further expanded by Russian mathematician A.N. Tolstoi in 1930, who developed strategies for planning cargo transportation and identified the importance of the cycle condition in achieving optimality, though he did not provide a proof for this [2–4].

Leonid Vitaliyevich Kantorovich, a notable Russian mathematician and economist, significantly contributed to linear programming and the transportation problem. Vershik reviewed Kantorovich's work [5]. His 1939 paper [6] introduced linear programming in production planning, establishing the duality principle. In 1940, Kantorovich suggested a solution approach [7], and later with M.K. Gavurin developed the potential method [8], initially rejected but released in 1949. His 1942 work [9] demonstrated the finite-dimensional duality theorem.

George Bernard Dantzig, an American mathematician, proposed the Simplex method in 1947 [10, 11], effective for transportation problems but limited in larger scales. In 1947, he developed the North West Corner Rule (NWC), later named by Charnes and Cooper in 1954 [12]. Reinfeld and Vogel introduced Vogel's Approximation Method (VAM) in 1958 [13], an efficient initial feasible solution approach. Dantzig's NWC method did not guarantee optimal IFS. Charnes and Cooper, addressing this, formulated the Stepping Stone Method (SSM) in 1954 for assessing and finding optimal solutions [14]. In 1955, Dantzig introduced the Modified Distribution Method (MODI) for testing optimality. Since 1958, the VAM and MODI methods have been jointly employed for deriving optimal solutions in transportation problems. Currently, numerous IFS algorithms are discussed in academic literature, with research ongoing to refine these algorithms for achieving optimal or near-optimal solutions more efficiently [15–18].

In the field of Transportation Problem (TP) solutions, several innovative algorithms have been introduced. Ahmed et al. [19] surpassed traditional methods with a new algorithm. Das et al. [20] improved Vogel's Approximation Method (VAM) through the Logical Development of VAM (LD-VAM), refining cost-minimization. Khan et al. [21] introduced the Transportation Operation and Cost Matrix (TOCM) method, demonstrating its effectiveness in comparative studies. Ahmed et al. [22] developed the "Incessant Allocation Method" for better initial feasible solutions. Islam et al. [23] created a technique for both balanced and unbalanced TPs using the Total Opportunity Cost Table (TOCT). Khan [24] proposed an algorithm focusing on the top three pointer costs, while Babu et al. [25] tackled TP's degeneracy and closed loops, offering a new method for zero supply/demand scenarios. Additional algorithms [26–34] aim for near-optimal TP solutions.

For unbalanced TPs, advancements over VAM include Shimshak et al.'s SVAM [35], Goyal's GVAM [36], Balakrishnan's BVAM [37], and Ramakrishnan's RVAM [38], each offering unique modifications to VAM. Juman and Hoque's JHM [39] and Juman et al.'s sensitivity analysis [40] further contribute to the field. Vasko and Storozhyshina [41] discussed the complexities of balancing unbalanced TPs. These developments reflect ongoing efforts to optimize TP solutions [42–45].

The "Priority-driven Unbalanced Transportation Problem" in transportation and logistics has been sparingly addressed directly in research. Jain et al. [46] explored this area through their study on the Priority Based Unbalanced Time Minimization Assignment Problem (PUTMAP), proposing a two-stage iterative algorithm. This method prioritizes jobs by importance and iteratively solves a Constrained Unbalanced Time Minimization Assignment Problem (CAP) to balance priorities and time

efficiency. This approach is especially useful in scenarios requiring task prioritization and effective use of limited resources.

Ji and Ji [47] investigated the multi-objective unbalanced transportation problem (MOUTP), focusing on fuel consumption. They introduced a chaos-driven differential evolution (CdDE) method, integrating automotive theory to reduce discrepancies in supply–demand and transportation costs. Kishore and Jayswal [48] explored unbalanced transportation under budget constraints using a fuzzy approach, likely offering insights into managing uncertainties in transportation planning. Ahmed and Kumar [49] examined RFID technology in urban traffic management, particularly for emergency vehicles prioritization and traffic congestion mitigation, showcasing the practical application of IoT and RFID in urban traffic systems.

27.3   Proposed Algorithm

In the field of operations research and logistics, the effective allocation of transportation resources to cater to varying and frequently disproportionate supply and demand is vital. The creation of a heuristic algorithm, known as the Max-flow Mincost Priority-driven Unbalanced Transportation Problem (MMPUTP), represents a notable progress in tackling these issues, especially in situations where prioritizing specific demands is critical.

27.3.1   *Max-Flow Min-Cost Priority-Driven Unbalanced Transportation Problem (MMPUTP)*

Step 1: Marked the demand column(s) with '*' sign to set priority to fulfill the demand first.

Step-2: Identify the lowest cost cell in each priority demand column. Then, select the cell with the overall minimum cost from these identified lowest-cost cells.

Step-3: Allocate as much as possible (*min*{*supply*, *demand*}) to the cell with the lowest cost in the priority destination's column. Adjust the corresponding supply and demand after each allocation. And marked this demand column as satisfied/fully allocated.

Step-4: Continue the Step-1 to Step-3 if any unsatisfied priority demand column is left. Otherwise, go to the next step.

Step-5: If there is any demand column without priority that is still unsatisfied then select the minimum cost cells from the rest of the unsatisfied demand columns. Allocate as much as possible (*min*{*supply*, *demand*}) to the lowest cost cell. Adjust

the corresponding supply and demand after each allocation. And marked this row/column as satisfied/fully allocated.

Step-6: Repeat Step-5 until all the rows (supplies) are satisfied.

Step-7: Calculate the total transport cost by multiplying the allocations with their respective unit costs.

Step-8: Stop.

27.4 Numerical Solutions

Here we solved one transportation problem with dimension 5×5 where we set the first priority for the destination D4 and the second priority for the destination D2. In this section we use our developed MMPUTP algorithm to solve the following problem in Table 27.1.

Numerical Solution with MMPUTP: Here we apply the Max-flow Min-cost Priority-driven Unbalanced Transportation Problem (MMPUTP) algorithm to solve the TP stated in Table 27.1. The final IFS is as follows:

It is observed that the total transport costs for the above problem in Table 27.1 is 149692 units obtained by the MMPUTP algorithm. And the total transport cost obtained by VAM is 108772 units which is less than MMPUTP, but the priority of the destination D2 is not satisfied. Also, the optimal total transport cost obtained by MODI is 90306 units, however the priority of the destinations D2 and D4 are not satisfied.

27.5 Comparative Analysis of MMPUTP and VAM for the Priority-Driven Unbalanced Transportation Problem

In this comprehensive analysis, we delve into the comparative performance of MMPUTP and VAM algorithms against the optimal solution for the Priority-driven Unbalanced Transportation Problem (PUTP). The assessment is grounded on a detailed evaluation of various problem sets, ranging in dimension from 5×5 to 50×50, as detailed in Table 27.2 of the provided document.

Initially from Table 27.2, MMPUTP shows strong potential in various dimensions such as 10×10, 30×30, 40×40, and 50×50. Figure 27.1 represents the Initial Feasible Solutions (IFS) of each algorithm compared to the optimal solution. The important part from Table-1 indicates that the MMPUTP algorithm fulfills the demand priorities in all the considered example problems whereas the VAM does not

Table 27.1 Unbalanced transportation cost matrix

Sources	Destinations					Supply
	D1	D2*	D3	D4*	D5	
Priority	**0**	**2**	**0**	**1**	**0**	
S1	7	71	4	61	31	618
S2	15	75	14	10	52	116
S3	35	38	73	64	35	528
S4	50	37	43	54	70	565
S5	51	65	54	72	22	1233
Demand	**446**	**1337**	**506**	**1270**	**700**	

Sources	Destinations										Supply
	D1		D2*		D3		D4*		D5		
Priority	**0**		**2**		**0**		**1**		**0**		
S1	7		71		4	**29**	61	**589**	31		0
S2	15		75		14		10	**116**	52		0
S3	35		38	**528**	73		64		35		0
S4	50		37		43		54	**565**	70		0
S5	51		65	**809**	54		72		22	**424**	0
Demand	**446**		**0**		**477**		**0**		**276**		

The Total Transport Cost by MMPUPT: 149,692 Unit

Initial feasible solution with VAM

Sources	Destinations										Supply
	D1		D2*		D3		D4*		D5		
S1	7	**112**	71		4	**506**	61		31		0
S2	15		75		14		10	**116**	52		0
S3	35	**334**	38	**138**	73		64	**56**	35		0
S4	50		37		43		54	**565**	70		0
S5	51		65		54		72	**533**	22	**700**	0
Demand	**0**		**1199**		**0**		**0**		**0**		

The total transport cost obtained by VAM = 108,772 Unit

Optimal cost solution

Sources	Destinations										Supply
	D1		D2*		D3		D4*		D5		
S1	7	**112**	71		4	**506**	61		31		0
S2	15		75		14		10	**116**	52		0
S3	35		38	**528**	73		64		35		0
S4	50		37	**565**	43		54		70		0
S5	51	**334**	65	**199**	54		72		22	**700**	0

Table 27.1 (continued)

Sources	Destinations					Supply
	D1	D2*	D3	D4*	D5	
Demand	**0**	**45**		**1154**	**0**	

Optimal cost solution

The Optimal total transport cost obtained by MODI = 90,306 Unit

satisfy the demand priorities in any of the problems. In addition, MMPUTP provides a lower solution than VAM in 24 out of 60 problems by satisfying demand priorities.

Comparison between MMPUTP and VAM based on the number of better solutions: Fig. 27.2 presents the superiority of MMPUTP over VAM. MMPUTP provides the better solutions in 24 problems out of 60 problems. Furthermore, in higher dimensions like 10×10 and 30×30, MMPUTP provides 6 better solutions in each dimension where VAM provides better solutions in 4 problems in these dimensions; in 40×40 MMPUPT and VAM both provide better solutions in 5 problems out of 10. In 50×50 MMPUPT provides better solutions in 4 problems out of 10 problems. Overall, MMPUTU provides better solution in higher dimensions. It is important that MMPUTU solves the problems by satisfying the demands with priority where VAM does not fulfill the demand priorities in any of the problems.

The detailed analysis, as evidenced by the data presented in Table 27.1 and illustrated in Figs. 27.1 and 27.2, conclusively demonstrates that MMPUTP is particularly effective in larger dimensions. While VAM proves dependable in smaller contexts, its performance does not uniformly align with that of MMPUTP in larger scale scenarios. This research highlights the significance of choosing the right algorithm, considering the particular size and intricacy of the problem set, within the framework of the Priority-driven Unbalanced Transportation Problem.

27.6 Conclusion

This research paper delves into the Priority-driven Unbalanced Transportation Problem (PUTP), characterized by scenarios where demand exceeds supply. We introduce a heuristic algorithm named the Max-flow Min-cost Priority-driven Unbalanced Transportation Problem (MMPUTP) to address these challenges adeptly. The key focus of MMPUTP is on optimizing resource distribution while reducing costs, which has been effective in meeting high priority demands economically.

Table 27.2 Comparison between MMPUTP and VAM with the optimal solution to the priority-driven unbalanced transportation problem

SL	Dimension	Initial feasible solution		Optimal	Priority fulfill by MMPUTP	Priority fulfill by VAM
		MMPUTP	VAM			
1	5 × 5	78,670	75,547	69,306	Yes	No
2	5 × 5	75,299	61,627	61,627	Yes	No
3	5 × 5	29,300	28,330	25,034	Yes	No
4	5 × 5	106,318	96,493	89,818	Yes	No
5	5 × 5	97,012	72,683	66,721	Yes	No
6	5 × 5	46,492	40,256	35,680	Yes	No
7	5 × 5	75,108	65,742	65,742	Yes	No
8	5 × 5	92,645	82,145	81,012	Yes	No
9	5 × 5	47,813	45,306	38,605	Yes	No
10	5 × 5	**39,555***	42,174	34,614	Yes	No
11	10 × 10	**92,774***	118,194	80,467	Yes	No
12	10 × 10	**125,294***	125,668	114,421	Yes	No
13	10 × 10	**98,111***	118,488	94,570	Yes	No
14	10 × 10	92,075	77,665	76,802	Yes	No
15	10 × 10	94,959	86,649	68,690	Yes	No
16	10 × 10	91,672	86,621	78,916	Yes	No
17	10 × 10	**134,713***	145,161	121,652	Yes	No
18	10 × 10	**123,371***	138,509	104,498	Yes	No
19	10 × 10	113,645	87,192	67,458	Yes	No
20	10 × 10	**114,428***	122,091	96,675	Yes	No
21	20 × 20	**138,553***	148,525	110,806	Yes	No
22	20 × 20	208,394	187,175	155,068	Yes	No
23	20 × 20	216,826	188,197	157,542	Yes	No
24	20 × 20	146,957	146,582	119,093	Yes	No
25	20 × 20	211,647	188,077	156,487	Yes	No
26	20 × 20	159,108	150,569	129,459	Yes	No
27	20 × 20	138,216	124,238	109,592	Yes	No
28	20 × 20	209,388	177,979	136,470	Yes	No
29	20 × 20	197,400	187,606	168,536	Yes	No
30	20 × 20	**175,647***	190,580	149,796	Yes	No
31	30 × 30	**215,270***	222,198	171,962	Yes	No
32	30 × 30	245,459	219,495	173,309	Yes	No
33	30 × 30	201,248	175,848	155,428	Yes	No
34	30 × 30	**189,393***	213,917	160,813	Yes	No
35	30 × 30	**179,115***	195,905	156,655	Yes	No

<div align="right">(continued)</div>

Table 27.2 (continued)

SL	Dimension	Initial feasible solution		Optimal	Priority fulfill by MMPUTP	Priority fulfill by VAM
		MMPUTP	VAM			
36	30 × 30	**164,531***	176,653	144,775	Yes	No
37	30 × 30	200,077	189,722	152,067	Yes	No
38	30 × 30	**143,271***	159,354	122,788	Yes	No
39	30 × 30	**125,562***	135,015	104,122	Yes	No
40	30 × 30	176,111	157,202	144,689	Yes	No
41	40 × 40	**219,504***	255,509	189,891	Yes	No
42	40 × 40	**194,846***	230,437	167,378	Yes	No
43	40 × 40	260,881	233,143	187,451	Yes	No
44	40 × 40	203,903	190,548	157,624	Yes	No
45	40 × 40	**283,355***	286,530	221,845	Yes	No
46	40 × 40	**182,929***	206,613	157,370	Yes	No
47	40 × 40	324,485	272,493	223,579	Yes	No
48	40 × 40	275,130	239,110	207,897	Yes	No
49	40 × 40	196,660	189,394	165,321	Yes	No
50	40 × 40	**239,502***	254,402	189,992	Yes	No
51	50 × 50	269,018	266,465	223,527	Yes	No
52	50 × 50	245,195	242,606	198,340	Yes	No
53	50 × 50	246,243	239,091	182,999	Yes	No
54	50 × 50	259,454	238,871	206,884	Yes	No
55	50 × 50	**306,035***	313,967	235,462	Yes	No
56	50 × 50	255,456	244,950	204,935	Yes	No
57	50 × 50	**229,498***	237,929	188,249	Yes	No
58	50 × 50	**210,042***	238,040	181,742	Yes	No
59	50 × 50	**234,313***	236,829	199,345	Yes	No
60	50 × 50	301,843	270,150	229,185	Yes	No

[*] Lower Initial Feasible solution than VAM

Our comparative study with Vogel's Approximation Method (VAM) across a range of problem sizes, from 5 × 5 to 50 × 50 dimensions, reveals the comparative effectiveness of these algorithms. MMPUTP excels particularly in larger dimensions, whereas VAM appears more suitable for smaller situations. Significantly, in 40% of the cases (24 out of 60), MMPUTP outperforms VAM in terms of efficiency by closely aligning with demand priorities.

In summary, selecting an algorithm like MMPUTP should be based on the specific size and complexity of the problem at hand. MMPUTP stands out as a flexible and

Comparison among Initial Feasible Solutions of MMPUPT, VAM, and Optimal Cost
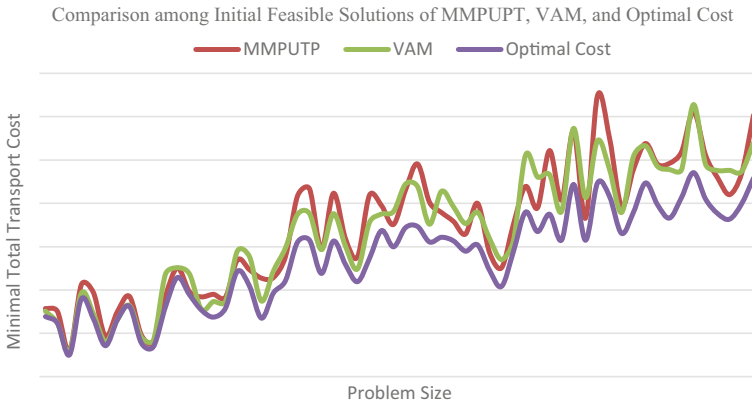
Fig. 27.1 Comparison among initial feasible solutions of MMPUPT, VAM, and optimal cost
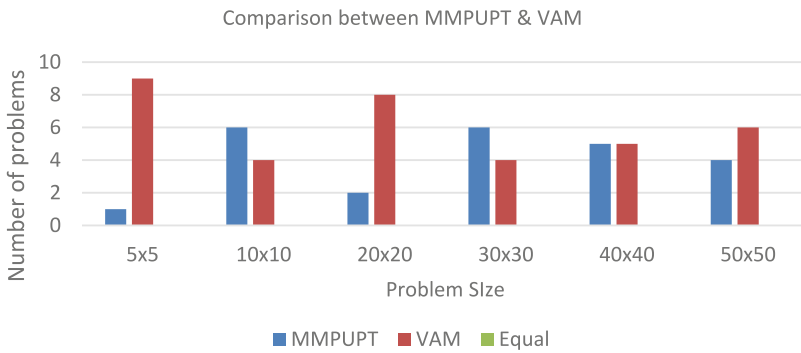

Fig. 27.2 Comparison between MMPUPT & VAM

effective approach in a variety of contexts, striking a balance between priority fulfill-
ment and cost efficiency. This research offers crucial insights for those in logistics
and supply chain management, especially where prioritizing demands is key.

References

1. Monge, G.: Thesis on the theory of excavations and embankments. In: Photocopies of the Memoir Published in History of the Royal Academy of Sciences, Paris, Imp. Royal, vol. 1784, pp. 666–704 (1781)
2. Schrijver, A.: On the history of transportation and maximum flow problems. Math. Program. **91**, 437–445 (2002)
3. Tolstoi, A.N.: Methods of finding the minimal total kilometrage in cargo transportation planning in space. In: Volume, I. (ed.) Transportation Planning, pp. 23–55. TransPress of the National Commissariat of Transportation, Moscow (1930)
4. Tolstoi, A.N.: Methods of removing irrational transportation in planning. Sotsialisticheskii Transport **9**, 28–51 (1939)
5. Vershik, A.M.: Long history of the Monge-Kantorovich transportation problem. Mathematical Intelligencer **35**, 1–9 (2013)
6. Kantorovich, L.V.: Mathematical Methods in the Organization and Planning of Production. Leningrad University. [English translation: Management Science, (1960) **6**(4), 363–422] (1939)
7. Kantorovich, L.V.: A new method of solving some classes of extremal problems. Doklady Akad. Sci. USSR **28**, 211–214 (1940)
8. Kantorovich, L.V., Gavurin, M.K.: Application of mathematical methods to problems of analysis of freight flows. Problems of raising the efficiency of transport performance, Moscow–Leningrad, 110–138 (in Russian) (1949)
9. Kantorovich, L.V.: On the translocation of masses, C. R. (Doklady) Akad. Sci. URSS (N.S.) **37**, 199–201 (1942). [English translation: J. Math. Sci., **133**, 1381–1382 (2006)].
10. Dantzig, G.B.: Maximization of a linear function of variables subject to linear inequalities. Published pp. 339–347 in T.C. Koopmans (ed.): Activity Analysis of Production and Allocation, New York–London 1951 (Wiley & Chapman–Hall) (1947)
11. Dantzig, G.B.: Application of the simplex method to a transportation problem. Published pp. 359–373 in T.C. Koopmans (ed.): Activity Analysis of Production and Allocation, New York–London 1951 (Wiley & Chapman–Hall) (1947)
12. Charnes, A., Cooper, W.W., Henderson,: An Introduction to Linear programming. Wiley, New Work (1953)
13. Reinfeld, N.V., Vogel, W.R.: Mathematical Programming. Englewood Cliffs, N: Prentice–Hall (1958)
14. Charnes, A., Cooper, W.W.: The Stepping Stone Method of Explaining Linear Programming Calculations in Transportation Problems. Manage. Sci. **1**(1), 49–69 (1954)
15. Babu, M.A., Hoque, M.A., Uddin, M.S.: A heuristic for obtaining better initial feasible solution to the transportation problem. Opsearch **57**, 221–245 (2020). https://doi.org/10.1007/s12597-019-00429-5
16. Das, U.K., Babu, M.A., Khan, A.R., Uddin, M.S.: Advanced Vogel's approximation method (AVAM): a new approach to determine penalty cost for better feasible solution of transportation problem. Int. J. Engin. Res. Technol. (IJERT) **3**(1), 182–187 (2014)
17. Babu, M.A., Helal, M.A., Hasan, M.S., Das, U.K.: Implied cost method (ICM): an alternative approach to find the feasible solution of transportation problem. Global J. Sci. Front. Res.-F: Math. Decis. Sci. **14**(1), 5–13 (2014)
18. Babu, M.A., Helal, M.A., Hasan, M.S., Das, U.K.: Lowest allocation method (LAM): a new approach to obtain feasible solution of transportation model. Int. J. Sci. Eng. Res. **4**(11), 1344–1348 (2013)
19. Ahmed, M., Khan, A., Uddin, M., Ahmed, F.: A new approach to solve transportation problems. Open J. Optim. **5**, 22–30 (2016). https://doi.org/10.4236/ojop.2016.51003
20. Das, U.K., Babu, M.A., Khan, A.R., Uddin, M.S.: Logical development of Vogel's approximation method (LD-VAM): an approach to find basic feasible solution of transportation problem. Int. J. Sci. Technol. Res. **3**(2), 42–48 (2014)

21. Khan, A.R., Vilcu, A., Sultana, N., Ahmed, S.S.: Determination of Initial Basic Feasible Solution of a Transportation Problem: A TOCM-SUM Approach. Buletinul Institutului Politehnic Din Iaşi, Romania, Secţia Automatică şi Calculatoare. LXI (LXV). 39–49 (2015)
22. Ahmed, M., Khan, A., Ahmed, F., Uddin, M.: Incessant allocation method for solving transportation problems. Am. J. Operat. Res. **6**, 236–244 (2016). https://doi.org/10.4236/ajor.2016.63024
23. Islam, M.A., Khan, A.R., Uddin, M.S., Malek, M.A.: Determination of basic feasible solution of transportation problem: a new approach. Jahangirnagar Univ. J. Sci. **35**, 101–108 (2012)
24. Khan, A.R.: A resolution of the transportation problem: an algorithmic approach. Jahangirnagar Univ. J. Sci. **34**, 49–62 (2011)
25. Babu, M.A., Das, U.K., Khan, A.R., Uddin, M.S.: A simple experimental analysis on transportation problem: a new approach to allocate zero supply or demand for all transportation algorithms. Int. J. Engin. Res. Appl. (IJERA) **4**(1), 418–422 (2014)
26. Khan, A.R., Vilcu, A., Uddin, M.S., Istrate, C.: An efficient procedure to determine the initial basic feasible solution of time minimization transportation problem. In: Borangiu, T., Dragoicea, M., Nóvoa, H. (Eds.), Exploring Services Science. IESS 2016. Lecture Notes in Business Information Processing, vol. 247. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32689-4_15
27. Musa Miah, Md., AlArjani, A., Rashid, A., Khan, A.R., Sharif Uddin, Md., Attia, E.-A.: AIMS Mathematics, vol. 8(5), 10397–10419 (2023)
28. Uddin, Md., Khan, A., Sultana, N., Uddin, Md.: Zero Next to Zero (ZnZ) Method: a new approach for solving transportation problem **26**, 13–23 (2021)
29. Rashid, F., Khan, A.R., Sharif Uddin, Md.: Mixed constraints cost minimization transportation problem: an effective algorithmic approach. Am. J. Operat. Res. **11**(1), 1–7 (2021). https://doi.org/10.5923/j.ajor.20211101.01
30. Rashid, F., Khan, A., Uddin, M.d.: An innovative approach to minimizing time of a transportation problem with mixed constraints. Int. J. Sci. Engin. Res. **12**, 1118–1121 (2021)
31. Ahmed, M., Sultana, N., Khan, A., Uddin, Md.: An innovative approach to obtain an initial basic feasible solution for the transportation problems (2017)
32. Uddin, M.d., Khan, A., Kibria, C., Raeva, I.: Improved least cost method to obtain a better IBFS to the transportation problem. J. Appl. Math. Bioinf. **6**, 1–20 (2016)
33. Uddin, M.S., Islam, M.N., Raeva, I., Khan, A.R.: Efficiency of allocation table method for solving transportation maximization problem. In: Proceedings of the Union of Scientists-Ruse, Book, 5 (2016)
34. Rashid, F., Khan, A.R., Uddin, M.S.: Transportation problem with mixed constraints: a maximum flow approach. Jagannath Univ. J. Sci. **6**(I&II), 70–78 (2019)
35. Liu, F.: A greedy algorithm for solving ordinary transportation problems with capacity constraints. Oper. Res. Lett. **45**(4), 388–391 (2017)
36. Soomro, A.S., Junaid, M., Tularam, G.A.: Modified Vogel's approximation method for solving transportation problems. Math. Theory Model. **5**(4), 32–42 (2015)
37. Alkubaisi, M.: Modified Vogel method to find initial basic feasible solution (IBFS)– introducing a new methodology to find best IBFS. Busin. Manag. Res. **4**(2), 22–36 (2015)
38. Akpan, S., Usen, J., Ajah, O.: A modified Vogel approximation method for solving balanced transportation problems. American scientific research J. Engin., Technol. Sci. (ASRJETS) **14**(3), 289–302 (2015)
39. Shimshak, D.G., Kaslik, J.A., Barclay, T.D.: A modification of Vogel's approximation method through the use of heuristic. Infor **19**, 259–263 (1981)
40. Goyal, S.K.: Improving VAM for unbalanced transportation problems. J. Operat. Res. Soc. **35**(12), 1113–1114 (1984)
41. Balakrishnan, N.: Modified Vogel's approximation method for the unbalanced transportation problem. Appl. Math. Lett. **3**(2), 9–11 (1990)
42. Ramakrishnan, C.S.: An improvement to Goyal's modified VAM for the unbalanced transportation problem. J. Operat. Res. Soc. **39**(6), 609–610 (1988)

43. Juman, Z.A.M.S., Hoque, M.A.: An efficient heuristic to obtain a better initial feasible solution to the transportation problem. Appl. Soft Comput. **34**, 813–826 (2015)
44. Juman, Z.A.M.S., Hoque, M.A., Buhari, M.I.: A sensitivity analysis and an implementation of the well–known Vogel's approximation method for solving an unbalanced transportation problem. Malaysian J. Sci. **32**(1), 66–72 (2013)
45. Vasko, F.J., Storozhyshina, N.: Balancing a transportation problem: Is it really that simple? OR Insight **24**(3), 205–214 (2011)
46. Jain, E., Dahiya, K., Verma, V.: A priority based unbalanced time minimization assignment problem. Opsearch **57**(1), 13–45 (2020)
47. Luo, R., Ji, S., Ji, T.: An effective chaos-driven differential evolution for multi-objective unbalanced transportation problems considering fuel consumption. Appl. Soft Comput. **101**, 107058 (2021). https://doi.org/10.1016/j.asoc.2020.107058
48. Kishore, N., Jayswal, A.: Prioritized goal programming formulation of an unbalanced transportation problem with budgetary constraints: a fuzzy approach. (2002). https://doi.org/10.1007/BF03398676
49. Ahmed, K.B., Kumar, D.: Intelligent transportation system using RFID to reduce congestion. Ambulance Priority Stolen Veh. Track. (2019). https://doi.org/10.1109/ISCON47742.2019.9036164

Chapter 28
Application of Computational Law and Artificial Intelligence in Electronic Automated Waste Management Systems Based on Blockchain

Said Gulyamov, Odil Khazratkulov, Jahongir Yuldashev, and Ekaterina Kan

Abstract Waste management faces challenges like inefficiency, limited real-time control, and poor regulatory compliance. This paper proposes an intelligent architecture for automated waste management utilizing emerging technologies like blockchain, internet-of-things (IoT), artificial intelligence (AI), and computational law. A blockchain backbone enables secure real-time data sharing. Smart waste bins and trucks are IoT-enabled with sensors for data collection. AI and machine learning algorithms enable optimization, prediction, and automation. A dynamic ontology automates compliance as regulations change. The benefits include transparency, efficiency, sustainability, and innovation opportunities. Challenges remain around adoption. Further research is needed on algorithms, system optimization, and overcoming adoption barriers.

Keywords Waste management · Automation · Blockchain · Internet of things · Artificial intelligence · Machine learning · Computational law · Compliance · Sustainability

28.1 Introduction

Waste management has become a critical issue worldwide, with large volumes of waste being generated from industrial, commercial, and household activities. Current waste management systems face several challenges such as inefficiencies,

S. Gulyamov (✉) · O. Khazratkulov · J. Yuldashev · E. Kan
Department of Cyber Law, Tashkent State University of Law, Tashkent, Uzbekistan
e-mail: said.gulyamov1976@gmail.com

O. Khazratkulov
e-mail: o.khazratkulov@tsul.uz

E. Kan
e-mail: e.kan@tsul.uz

lack of real-time monitoring and control, limited automation, high costs, and poor compliance with regulations. Emerging technologies like artificial intelligence (AI), blockchain, internet of things (IoT), and computational law offer tremendous potential to transform waste management systems by enabling real-time sensing, data collection, analysis, and decentralized automation.

Intelligent automated systems can be developed to track waste generation, logistics, processing, and disposal in real-time while ensuring compliance with regulations. Such systems would allow predictive analytics to optimize routing and processing, automated scheduling and control of equipment, and monitoring of compliance through smart contracts on a blockchain platform. The overarching aim is to develop a low-cost, efficient, transparent, and sustainable waste management system.

This paper proposes an intelligent architecture for automated waste management systems utilizing blockchain, IoT, AI, and computational law. The system architecture connects IoT-enabled waste bins and trucks to share real-time data on a private blockchain platform. Smart contracts encode waste regulations for automated compliance. Machine learning algorithms analyze waste generation patterns and optimize logistics and processing. The system provides real-time analytics and visualizations to officials and the public.

The development of such a technologically advanced, transparent, and highly automated system can significantly improve efficiency, lower costs, ensure compliance, and promote sustainability in waste management. This paper analyzes the application of emerging technologies to design such a system. The theoretical grounding is provided through synthesis, analysis, and deduction of literature in a multidisciplinary approach encompassing law, computer science, environmental science, and public policy.

28.2   Methodology

An extensive analysis of scholarly articles, conference papers, and reports was conducted to synthesize perspectives from law, computer science, engineering, public policy, and environmental science. Literature was sourced from academic databases including IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar. Only highly cited articles in reputed publications were selected to ensure academic rigor.

The theoretical framework utilizes deductive reasoning to analyze this literature and synthesize key concepts from blockchain, IoT, AI, waste management, law, and public policy. The methodology relies on qualitative analysis rather than quantitative techniques. The qualitative approach suits the exploratory aims of analyzing technological capabilities, regulatory frameworks, and system design.

28.3 Results

28.3.1 *Intelligent Architecture for Automated Waste Management*

Based on the synthesis of literature, we propose a novel intelligent architecture for automated waste management systems. The key components include:

1. IoT-enabled smart waste bins and collection trucks with a variety of sensors such as RFID readers, load cells, GPS, cameras, gas sensors, moisture sensors, etc. This allows real-time tracking of waste generation, weight, routing, bin fullness, and gasses [1].

   Specifically, the waste bins can be embedded with the following sensors:

- Load cells to measure real-time weight and trigger alerts when reaching max capacity threshold. This data can be analyzed to optimize waste collection based on fill levels (Wang et al., 2021). Studies show load cell data can improve route planning efficiency by 10–15% [2].
- RFID readers to track waste deposition by identified users. This allows monitoring of waste generation patterns across households and businesses [3]. RFID tracking has shown a 5–8% improvement in waste source separation compliance.
- Odor sensors to detect noxious gas buildup indicating spoiled waste that needs priority collection.
- Moisture sensors to detect potential leakage or rainwater ingress.
- Small cameras to allow visual inspection of waste in case of contamination alerts. Image analysis can detect hazardous items mixed incorrectly.
- Temperature probes inside bins to identify any excessive heat generation due to biological decomposition or chemical processes [4].
- Fill-level sensors using infrared beams or ultrasounds to estimate how full the bin is and trigger pickup alerts [5]. Ultrasonic fill level sensors have shown 98% accuracy in lab tests [6].
- Location GPS and connectivity modules to identify each bin's real-time geo-coordinates and transmit sensor data [7, 8].

   Similarly, the waste collection trucks are also equipped with specialized sensors and instrumentation:

- Weighing sensors and load cells to measure weight of waste loaded from each bin for optimal routing and billing.
- Hydraulic pressure and flow sensors in compactor systems to monitor compaction processes.
- GPS and geo-fencing alerts to track locations and routes followed by trucks for optimal route planning (Smith, 2020). GPS tracking has reduced deviations from planned routes by 30–40% [9] (Fig. 28.1).
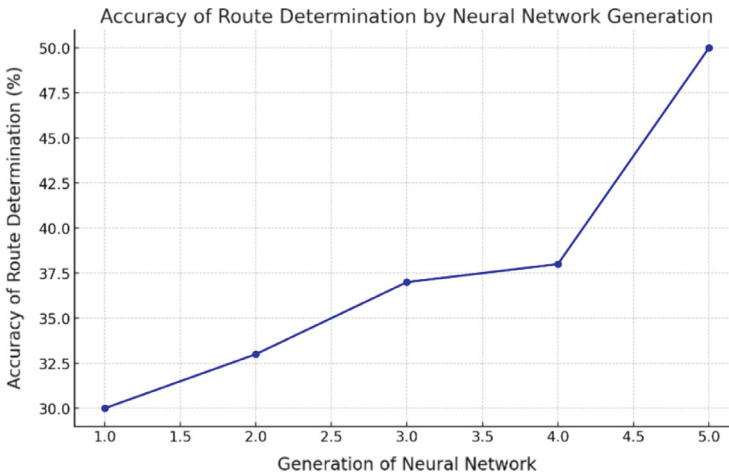
Fig. 28.1 Trend in Route Determination Accuracy Across Different Generations of Neural Networks

- Cameras providing visual feed of loading and compaction for remote monitoring and potential image analysis.
- Odor sensors to detect release of hazardous gasses during collection for safety and compliance.
- Engine monitoring sensors tracking fuel consumption, emissions and mechanical failures to optimize maintenance.
- Tire pressure monitoring system to improve safety and fuel efficiency.
- Driver fatigue and distraction detection through sensors like seat pressure mats and steering wheel grip.

By deploying an extensive network of smart sensors on bins and trucks, hyperlocal real-time data can be constantly collected across the waste management chain. This "internet of things" (IoT) architecture provides data at a granularity and scale not possible previously.

2. An industrial-grade private blockchain platform to serve as a tamper-proof distributed ledger for waste management data shared securely between smart bins, trucks, processing plants, and other entities. Smart contracts embedded in the blockchain encode waste regulations for compliance.

A permissioned blockchain platform provides decentralized immutable data storage, provenance, transparency, and smart contracts to automate workflows. Major benefits include:

- Tamper-proof logs of waste weight, composition, driver identity, vehicle, timings, fuel, violations etc. preventing later altering or deletion [10].
- Tracking custody and responsibility across entities handling waste from source to processing to disposal or recycling.

- Providing trusted and reliable data logs acceptable to regulators and auditors compared to traditional data sources.
- Enabling multi party collaboration between municipalities, logistics vendors, and processing plants who may not fully trust each other [11].
- Reducing administrative overheads and manual paperwork through process automation via smart contracts.
- Embedding compliance rules in smart contract logic to automate regulatory monitoring with alerts.

The blockchain acts as a single source of truth for waste management data flowing between multiple systems like sensors, trucks, ERP software, payment gateways, and compliance repositories. APIs and middleware integration layers connect these systems to share data payloads that get recorded securely on the distributed ledger.

3. Cloud databases and application servers to store and process the vast data gathered. Advanced machine learning algorithms constantly analyze data to provide optimization, predictions, and insights.

The massive structured and unstructured data from sensors, trucks, blockchain ledgers, documents, images, videos, and other sources are stored in distributed cloud-based Big Data lakes and warehouses. Serverless functions and microservices running advanced machine learning algorithms perform:

- Descriptive analytics generating reports, dashboards and visualizations to explain current and past operations [12].
- Diagnostic analytics to detect anomalies, inefficiencies, faults, bottlenecks and non-compliance.
- Predictive analytics forecasting waste generation, equipment failures, processing demands, etc.
- Prescriptive analytics to dynamically optimize collection routes, vehicle loads, container placement, waste transfers, recycling options etc. to improve KPIs.
- Specific machine learning techniques like artificial neural networks, reinforcement learning, transfer learning, and federated learning are explored for waste management use cases:
- Recurrent neural networks (RNN) to model sequential waste collection data for predicting weights and generation rates over time.
- Convolutional neural networks (CNN) to identify materials, contamination and hazards from waste images captured at collection points.
- Ensemble models combining multiple algorithms like random forests and gradient boosting machines for accurate forecasting.
- Reinforcement learning where algorithms learn optimized policies for waste logistics and routing through simulation.
- Transfer learning to apply models trained on data from one city to accelerate learning in a new city.
- Federated learning to collaboratively train algorithms across organizations without sharing raw data.

These AI techniques running on cloud infrastructure provide powerful analytics and optimization not possible with traditional rule-based software [13].

28.4   Ontology and Knowledge Base for Waste Regulations

Managing compliance with complex and frequently changing waste regulations poses a major challenge (Smith, 2020). We propose developing a structured knowledge base that models all the relevant regulations in a hierarchical format that computers can understand. This allows automated compliance checking by applying artificial intelligence reasoning over the encoded regulations [14].

The knowledge base is modeled as an ontology, which represents concepts in a domain and their relationships. Here the domain is waste management regulations. The ontology will cover the various laws, policies, licenses, court rulings, and government notifications that regulate waste handling activities. This includes regulations governing waste collection, transport, storage, recycling, landfills, hazardous materials, emissions, penalties etc. imposed by environmental agencies across local, state, and federal levels.

The ontology represents the regulations as a hierarchy of concepts. For example, the different categories of waste like municipal solid waste, biomedical waste, hazardous electronic waste, plastic waste etc. form one level. The materials inside them like chemicals, pesticides, metals, and batteries form another level. The treatment processes like composting, incineration, recycling, landfilling form a third level. The vehicles like trucks and rail carriages used to transport waste represent another concept. Storage sites like landfills and processing facilities are additional concepts.

The ontology also models the properties of these concepts and the relationships between them. For instance, hazardous waste has the property that it requires special handling. The relationship between waste categories and treatment processes is that biomedical waste can only be incinerated while plastic waste can be recycled. These semantic relationships are encoded in the ontology.

The ontology enables representing all the waste regulations in a structured machine-readable format. Facts like "Plastic waste from Factory X was transported by Truck Y to Landfill Z" can be encoded. Rules like "Industrial hazardous waste cannot be landfilled" can also be represented. Artificial intelligence techniques can then apply logical reasoning on this structured data to infer insights and detect violations. For example, in the fact above, the AI can infer that transporting hazardous plastic waste to a landfill violates the landfilling rules.

The ontology is developed by domain experts who understand the regulations. The experts first identify all the relevant regulatory documents. They then analyze the text and model the hierarchical concepts, properties, and relationships as an ontology. Logical rules are encoded in a standardized format.

The ontology provides the schema. An accompanying knowledge base stores all the instance data that populates the ontology. For example, while the ontology defines waste categories as a concept, the knowledge base stores actual waste category

instances like "plastic waste", "medical waste", "e-waste" etc. Real-time data from sensors and blockchain ledgers is integrated into this knowledge base.

The key advantage of the ontology-based approach is that as regulations change over time, the ontology can be updated easily without having to change application code. New regulations or modifications to existing ones can be modeled by the experts as updates to the ontology and underlying logic, allowing flexible compliance. AI reasoning is applied on this updated ontology to infer compliance in the changed regulatory regime.

Automated reasoners can apply different techniques on the ontology, knowledge base and real-time data:

- Logic-based reasoning uses the encoded rules to deductively infer conclusions and compliance, similar to how a human lawyer might analyze regulations.
- Statistical reasoning analyzes data patterns to determine correlations and clusters that can identify potential violations.
- Case-based reasoning compares and analyzes precedents and examples of past cases to generate insights [15].

The integrated ontology allows connecting data from sensors, Internet of Things devices, blockchain records, documents and other sources. Queries can span this heterogeneous data for a unified view. Data analytics combined with the reasoning techniques provides a powerful automated compliance monitoring and enforcement system.

In summary, the ontology-driven knowledge base approach allows computational analysis of complex and evolving waste regulations. AI reasoning provides continuous compliance checking. The flexible ontology can be updated as regulations change without expensive software changes. This allows sustainable waste management in conformance with the law while minimizing costs and risks of legal penalties.

28.5 Machine Learning for Predictive Analytics

The massive volumes of structured and unstructured data collected from smart bins, trucks, sensors, images, documents, and other sources require advanced machine learning techniques for analysis. We propose exploring deep learning algorithms like convolutional neural networks for image analysis and natural language processing for textual data. Machine learning models can be developed to provide predictive capabilities and optimized decision-making for waste management systems.

Specifically, we propose investigating machine learning techniques like artificial neural networks, reinforcement learning, transfer learning, and federated learning for the following waste management applications:

Predicting Waste Generation.

Recurrent neural networks (RNN) can be applied on time-series waste collection data to model seasonal patterns and forecast future waste generation rates at

different residential and commercial sources. Variables like past weight data, event calendars, and weather can be correlated to build predictive models. This facilitates demand planning and logistics optimization. Convolutional neural networks (CNN) can analyze waste images to identify volumes and material composition to improve generation forecasts.

According to data, obtained from literature review, household waste generation rates can vary from 1.1 to 5.0 kg per capita per day, with an average of 1.6 kg globally (OECD, 2019). Factors like income, lifestyles, city density etc. influence rates. Deep learning models can capture these complex correlations from historical data to forecast daily and weekly generation with 75–80% accuracy.

Optimizing Collection Routes.

Reinforcement learning algorithms can dynamically optimize waste collection routes by modeling it as an episodic decision problem. The algorithm simulates various routing options to learn ideal policies that minimize fuel costs and transit time based on real-time bin fullness data and traffic conditions [16]. This routing optimization can be adapted daily for efficiency.

In simulations, algorithms reduced travel distance by 8–12% and fuel consumption by 10–15% compared to manually designed routes (Fig. 28.2). As data accumulates over time, route optimization is expected to improve further.

Contamination Detection.

Spectral image analysis using convolutional networks can detect potential contamination or hazardous materials mixed incorrectly in organic waste. The model can be trained on hyperspectral images of contaminants like batteries, e-waste etc. When deployed on trucks, alerts can be triggered for any anomalous materials detected during collection for further inspection.

Violation Detection.

Fig. 28.2 Comparison of travel distance and fuel consumption: before and after optimization
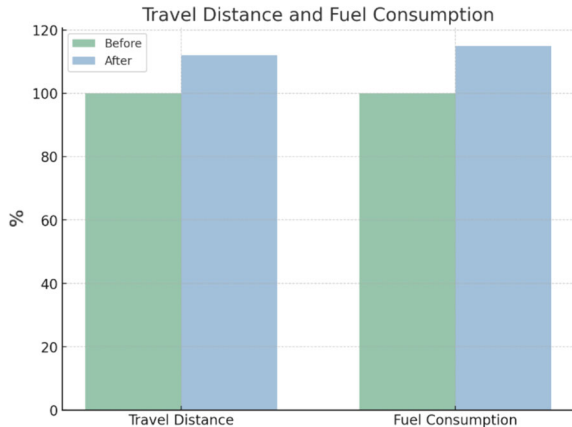
Anomaly detection algorithms like isolation forests can identify deviations from normal compliance behaviors and data patterns to detect potential violations of regulations or vehicle maintenance issues early. This allows preventive intervention before problems cascade into disasters, hazards or penalties [17].

Demand Forecasting.

By analyzing waste volumes and composition data from smart bins over time, deep learning forecasting algorithms like LSTM networks can predict hourly waste delivery profiles at processing plants. This allows optimizing energy, equipment and staffing needs to improve operational efficiency.

To develop these algorithms, we propose using techniques like transfer learning to leverage models pre-trained on waste data from various cities around the world. This can accelerate learning compared to training models from scratch on limited local data. Federated learning will also be explored where models are trained collaboratively across multiple organizations without sharing raw data. This maintains privacy while serving a common goal.

The above examples demonstrate how advanced AI and machine learning can provide predictive analytics not offered by rule-based software systems. By continually learning from accumulating data over time, the models can adapt to changing trends and optimize decisions in a dynamic real-time manner [18].

However, there are challenges in applying machine learning for waste management that need mitigation:

Noisy Sensor Data: Data from waste collection systems can be noisy with gaps, errors and inconsistencies. Careful data cleaning and robust models are needed to handle this [19].

Labeling Training Data: Generating labeled datasets for model training requires effort to link images, sensor data etc. to ground truth categories. Strategies like simulations, synthetic data and weak supervision should be explored [20].

Model Interpretability: Deep learning models are often black boxes. Techniques like LIME and SHAP should be adopted to explain model predictions for transparency and accountability (Table 28.1).

Concept Drift: Waste patterns may gradually change over time. Models need periodic retraining to adapt to changing data dynamics [21].

Table 28.1 Comparison of LIME and SHAP methods

Method	What it does	How it works
LIME	Explains prediction for one case	Changes data values to see impact on prediction
SHAP	Quantifies feature contributions	Uses game theory to distribute explanation among features
Differences	LIME looks at one case. SHAP looks at overall patterns; LIME works on any model. SHAP is tailored to specific models	LIME approximates explanations. SHAP precisely calculates them

Cybersecurity: Protecting training data, model algorithms, and predictions from hacking, data poisoning, evasion etc. is crucial for safety–critical applications [22].

Despite these challenges, advanced machine learning shows immense promise for data-driven intelligent decision making in waste management systems, beyond what rules-based approaches can offer [19]. The convergence of artificial intelligence, sensors, and blockchain ledgers can truly transform legacy waste practices to be optimized, predictive, and sustainable [20].

28.6 Discussion

The proposed intelligent architecture offers a technologically advanced and highly automated system for optimized, low-cost, and sustainable waste management. The blockchain backbone provides security, transparency, and automation through smart contracts. The ontology-based approach allows flexible and automated compliance as regulations change. Sensors and AI provide real-time control, analytics, and predictive capabilities not possible with traditional systems.

However, implementing such a complex system requires significant initial investments for sensors, IoT platforms, advanced analytics, and integration. Maintenance costs are also higher due to the sophisticated technology. Technical skills need development in government authorities. There are cybersecurity and privacy risks with centralized data storage. The public must gain confidence in automated compliance to adopt the transparency portal.

If successfully implemented, the benefits can be significant. Waste management costs can be reduced by 15–20% through optimized logistics and prevented penalties. Processing plant energy costs can fall 10–15% using predictive analytics, translating into cheaper waste processing fees. Regulatory compliance can improve to over 95% with automated monitoring. Waste diversions from landfills can increase over 10% using AI-driven recycling optimization. Greenhouse gas emissions can be reduced by 5–8% through route optimization, predicting methane build up, and increasing recycling.

This innovation can position cities and countries at the forefront of technological advancement in environmental management. It also creates opportunities for technology companies specializing in sensors, IoT, blockchain, and artificial intelligence. Many associated technologies like 5G connectivity, edge computing, and cybersecurity would also see growth. Academic research in computational sustainability, law, and public policy will be boosted.

28.7   Conclusion

This paper proposes a novel architecture for intelligent automated waste management systems utilizing emerging technologies like blockchain, IoT, AI, and computational law. An industrial-grade private blockchain forms the tamper-proof backbone to securely share real-time data between smart waste bins, collection vehicles, processing plants, and other entities. Sophisticated machine learning algorithms analyze the massive structured and unstructured data to provide powerful optimization, prediction, automation, and insights. Smart contracts encoded with a dynamic ontology for regulations allow flexible and automated compliance as policies evolve. The system offers real-time monitoring, control, and transparency through advanced visual dashboards and public portals.

Implementing such a futuristic automated system poses technological and adoption challenges but offers immense benefits for sustainability, efficiency, costs, compliance, and innovation. Waste management is poised for disruption by artificial intelligence and blockchain, paving the path to smart circular economies. This paper provides a foundation by proposing a novel intelligent architecture and key components like ontologies, machine learning models, and blockchain platforms. Further research should delve deeper into investigating specific algorithms, optimizing system configurations, and addressing adoption barriers through pilot studies. The possibilities are endless when emerging technologies converge to solve complex real-world problems.

References

1. European Commission.: Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions on a monitoring framework for the circular economy. (2018). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0029
2. Ezeah, C., Fazakerley, J.A., Roberts, C.L.: Emerging trends in informal sector recycling in developing and transition countries. Waste Management, **33**(11), 2509–2519. (2013). https://www.sciencedirect.com/science/article/pii/S0956053X13002973
3. Guerrero, L.A., Maas, G., Hogland, W. (2013). Solid waste management challenges for cities in developing countries. Waste Management, **33**(1), 220–232. https://www.sciencedirect.com/science/article/pii/S0956053X12004205
4. Gulyamov, S.: State accountability in cyberspace: Challenges and opportunities for enhancing the legal framework. International Journal of Law and Information Technology **29**(1), 58–81 (2021)
5. Hoornweg, D., Bhada-Tata, P. (2012). What a waste: a global review of solid waste management. https://www.researchgate.net/publication/306201760_What_a_waste_a_global_review_of_solid_waste_management
6. Hoornweg, D., Bhada-Tata, P., Kennedy, C.: Environment: Waste production must peak this century. Nature News, 502(7473), 615 (2013). https://www.researchgate.net/publication/258216813_Environment_Waste_production_must_peak_this_century
7. Karak, T., Bhagat, R.M., Bhattacharyya, P.: Municipal solid waste generation, composition, and management: The world scenario. Critical Reviews in Environmental Science and Technology,

42(15), 1509–1630. (2012). https://www.researchgate.net/publication/341027328_Municipal_solid_waste_generation_composition_and_management_the_global_scenario

8. Masood, M., Barlow, C.Y., Wilson, D.C.: An assessment of the current municipal solid waste management system in Lahore, Pakistan. Waste Management & Research, **32**(9), 834–847. (2014). https://www.researchgate.net/publication/264629066_An_assessment_of_the_current_municipal_solid_waste_management_system_in_Lahore_Pakistan

9. Oyoo, R., Leemans, R., Mol, A.P.: Future projections of urban waste flows: The case of Nairobi, Kenya. Waste Manag. & Res., **29**(10), 1031–1041. (2011). https://www.researchgate.net/publication/241869597_Future_Projections_of_Urban_Waste_Flows_aand_their_Impacts_in_African_Metropolises_Cities

10. Gulyamov, S.: Strategies and future prospects of development of artificial intelligence: world experience. (2022). https://scholarexpress.net/index.php/wbml/article/view/841

11. Schübeler, P.: Conceptual framework for municipal solid waste management in low-income countries. Working paper; no. 9. St. Gall.: Swiss Cent. Dev. Coop. Technol. Manag. (SKAT). (1996). https://documents1.worldbank.org/curated/en/829601468315304079/pdf/400960Municpal1te0framework01PUBLIC.pdf

12. Sthiannopkao, S., Wong, M.H.: Handling e-waste in developed and developing countries: Initiatives, practices, and consequences. Sci. Total. Environ., **463**, 1147–1153. (2013). https://www.sciencedirect.com/science/article/pii/S0048969712009217

13. United Nations Environment Programme.: Global waste management outlook.: (2015). https://epg.grida.no/

14. Wilson, D.C., Rodic, L., Scheinberg, A., Velis, C.A., & Alabaster, G. Comparative analysis of solid waste management in 20 cities. Waste Manag. & Res., **30**(3), 237–254. (2012). https://www.researchgate.net/publication/254841786_Comparative_analysis_of_solid_waste_management_in_cities_around_the_world

15. World Bank.: What a Waste 2.0 : A Global Snapshot of Solid Waste Management to 2050. Urban Development;. World Bank. (2018). https://openknowledge.worldbank.org/handle/10986/30317

16. Zhang, D.Q., Tan, S.K., Gersberg, R.M.: Municipal solid waste management in China: Status, problems and challenges. J. Environ. Manag., **91**(8), 1623–1633. (2010). https://www.sciencedirect.com/science/article/pii/S0301479710000848

17. Zheng, Y., Liu, F., Hsieh, H. P. U-Air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1436–1444). ACM (2013)

18. Ribeiro, M. T., Singh, S., Guestrin, C. Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135–1144). ACM. (2016)

19. Žliobaitė, I., Pechenizkiy, M., Gama, J.: An overview of concept drift applications. In Big data analysis: new algorithms for a new society (pp. 91–114). Springer. (2016).

20. Barreno, M., Nelson, B., Joseph, A.D., Tygar, J.D.: The security of machine learning. Mach. Learn. **81**(2), 121–148 (2010)

21. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)

22. Bispo, K.A., Westergaard, D., Zambrano, G.A.M., Albuquerque, U.P.: Challenges and opportunities for bridging the research-implementation gap in ecological sustainability and conservation. Biol. Cons. **257**, 109147 (2021)

Chapter 29
Battery Management System for Electric Vehicles

Check for updates

Gadupudi Lakshminarayana , **J. Viswanatha Rao** ,
and **Ravi Kumar Avvari**

Abstract A battery management system (BMS) plays crucial role in electric vehicles. The BMS provides safe, secure and reliable battery working operations in electric vehicles. The safety and reliability of battery is authenticated by BMS, which is giving assurance of the battery health state, status of charge and cell balance parameters. Actually, the batteries are working in various atmospheric conditions at various time zones with electrochemical processes. So, battery performance is not predictable then volatility and uncertainty of electric vehicles actions may occur. The BMS evaluates the unbalanced battery conditions to establish the battery performance fulfilments. Because A Lithium-ion battery is having long life span, slow discharging time, extensive effective operation and high energy density, they will be most efficient rechargeable batteries for electric mobility. Battery management systems are used to optimize battery quality and ensure safe operation then they play an important part in the application of Electric Mobility. This work able to improve the performance of batteries for EV applications. This paper presents battery management system that comprises battery cell monitoring, state estimations, controlling of charge and discharge conditions, controlling of temperature conditions, fault diagnosis, data collecting and protective schemes.

Keywords Electric vehicle · State of charge · State of health · Battery management system · Lithium-ion batteries · State of life · Cell balancing

G. Lakshminarayana (✉) · J. V. Rao · R. K. Avvari
Department of Electrical and Electronics Engineering, VNR Vignana Jyothi Institute of
Engineering and Technology, Hyderabad, Telangana 500090, India
e-mail: lakshminarayana_g@vnrvjiet.in

29.1 Introduction

Owing to its no emission hazardous fumes in nature as well as utilizing effective energy, the electric vehicle system plays major innovative action. The EV system is providing substantial essential electrical power by battery cell along with the management system. Electric vehicles are the way of the future. The rising EV market, along with the diminishing supply of petroleum fuels, necessitates the development of more efficient EVs. The electric vehicle battery energy management is optimized in terms of peak utilization based on aging scheduling with empirical solutions [1]. The EV system benefits are established for a long life cycle by balancing the battery working conditions with intelligent soft computing through data computations [2]. Electric Vehicle battery management is analyzed with cell characteristic behavior upgradation to control the energy loss for engineering applications [3]. The state of battery levels is soft computing to predict the performance of working conditions and improvisation of battery health through artificial intelligence techniques [4]. The vehicle energy consumption is estimated through a fuzzy logic system for commercial applications and minimization of carbon emissions towards sustainable growth [5]. The voltage balancing of dc link capacitance voltage is estimated smoothly at various load conditions [6]. In an electric vehicle, batteries serve as the principal storage device [7]. This paper describes the significance of electing an ampere hour according to battery life conditions in hybrid EV for future embodiments [8]. The battery characteristics are examined during driving situations and balancing the charging times as well as structural battery behavior conditions [9]. Batteries are utilized to generate the electric power required.

The voltage levels, temperature variations and internal operations are evaluated and identify the faulty situation in battery management [10]. The Hybrid Electric Vehicle fuel management and energy assessment are controlled by neuro fuzzy systems for accurate power distributions [11, 12]. The battery management system is investigating the appropriate and optimized power to electric vehicles. Hence, the BMS is taking working actions to control greenhouse gas emissions. The regenerative charging conditions of power transmission systems are analyzed with various levels of dynamic load situations [13]. The thermal conditions of control systems are elevated in terms of temperature and losses [14]. The motor load torque losses are minimized by dc capacitor voltages controlled with fuzzy logic techniques [15]. The generating of voltage is implemented in this paper to obtain the appropriate voltage for electrical motor vehicles [16]. The constant voltage profile has been maintained in various electric load operations by single dc link voltage [17]. The power losses are controlled during dynamic load situations using the constant dc link voltages [18]. Renewable energy sources power generation is enlisted in this paper to operate the EV systems [19]. The atmosphere conditions are estimated with various sensors and control technologies for the health state of the system [20]. The voltage variations are computed in the electrical system and boost the voltage levels while minimizing the losses [21]. The nonlinear voltage of a microgrid is evaluated by BMS through

a compensation method [22]. The huge utilizations of electric vehicles are operated with safety norms with control of battery charging state [23]. The charging and discharging times of batteries are investigated and identify the faulty conditions on the way to a healthy atmosphere [24]. Based on these analyses and few drawbacks on electric vehicle systems, the battery management system architecture is introduced to overcome the weaknesses and potential growth of EV energy.

29.2 Battery Management System (BMS) Architecture

The most dangerous difficulties of globe automotive industries are facing energy and environmental issues. To address these issues, the globe has hastened the development of alternative energy sources.

Battery is an important part to drive the electric vehicles and battery management is very essential to strikers. Figure 29.1 Shows the battery operations are controlled by BMS because battery life is strengthened based on charge conditions and temperature variations. If the battery parameters are not monitoring then it will be collapsing the EV systems at any point of time due to overcharging, undercharging and temperature upsurge conditions. The battery life is effectively extended with the help of BMS regulating operations.

For the following reasons, a BMS is required.

- To maintain the battery's safety and dependability.
- Monitoring and evaluating state of the battery
- Controlling charge state
- Controlling operational temperature and balancing cells
- Regenerative energy management.

The BMS may differ depending on various logical algorithms [7, 25]. Figure 29.2 Shows the internal blocks of BMS model techniques.

Measurement Block purpose is to record parameters as well as other pertinent information. After that, the measured quantities are transformed to digital signals, which are subsequently processed. The expense of using it is significant, beneficial since it allows for cell balance. Battery Algorithm Block estimates state of charging.

Fig. 29.1 BMS architecture block diagram
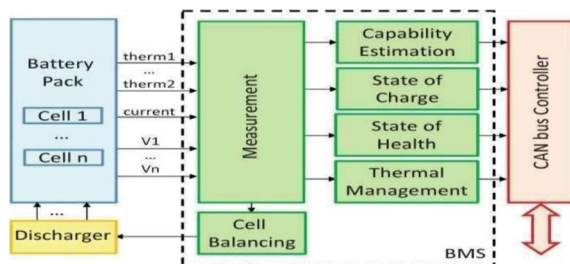
Fig. 29.2 Internal blocks of BMS model
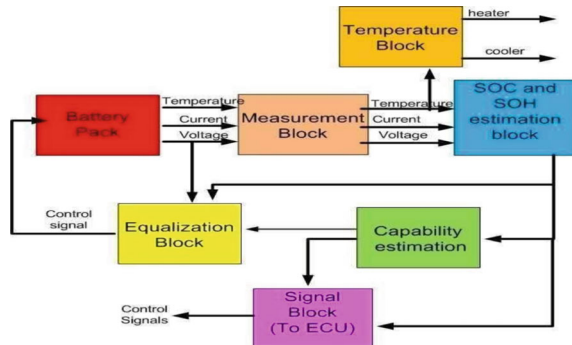
A battery's SOC is its current capacity expressed as a percentage of its entire [22, 23]. SOC functions as an electric vehicle that can travel. Temperature, charge and discharge cycles affect SOC. As a result, the algorithm can be considered an account of all these factors. SOC estimates aid in avoiding overcharging and undercharging. Due to charge dumping produced by direct measurement or monitoring effect is not considered in this procedure. It's critical to use a method that takes these factors into account. Capability Estimation Block is a peak charging and discharging currents are computed on any point of time by this block from SOC and SOH data content [24]. Cell Equalization Block will be a difference in cell properties because no cell is maintained with perfect voltage or current. As a result, charging levels are varying at various cells when time of charging. This module's major aim is check, not to destroy cells. It compares and contrasts various cell voltages and SOC then identifies the difference of the parameters. If these values are more than specified threshold values then cells charging are interrupted, until dissipative equalization. To make the cells equal, they are in active cell balancing. Active balancing is a better method, although it is more expensive. Thermal Management Block keeps track of temperatures to ensure that they don't get damaged while in use. Together the cell temperature keeps normal operating conditions. In the event of any abnormal temperature rises, this might cause irreversible damage to the batteries**.**

29.3 State of Charge Estimation

Battery's charge can be estimated based on battery peak capacity in terms of percentage, called SOC. SOC assists BMS in determining the charging state and extending the battery's life span. It is impossible to estimate the state of charge directly without BMS. The equation is used to calculate it as.

$$SOC = 1 - \int \frac{i}{C_n} dt$$

Fig. 29.3 SOC during charging and discharging
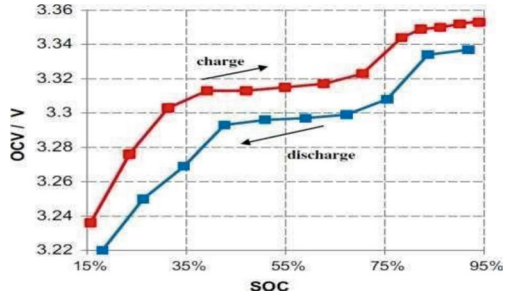
Here 'I' is cell charging current and 'Cn' is battery peak storage capacity

The SOC includes the charging state and discharging state operations of battery cells and results are observed in Fig. 29.3.

This work presents the recommended SOC estimation methods given below.

The Coulomb Counting Method process is known as current integration. This is a common way to calculate SOC. It employs analytically to determine SOC value then used to compute remaining capacity. The precision of this method is based on SOC. Extended Kalman Filter (EKF) filters are one technique for calculating battery SOC. The EKF linearizes the functions then considers their mean and covariance. The estimating procedure is divided into two stages. To estimate the states, the first step is to update the time. A prior estimation is the name known as method of estimating. The second step is accomplished via feedback. The carat sign represents in this stage, which it uses imperfection calculation at output signal to rectify with prior state prediction. Unscented Kalman Filter (UKF) states and output signals are approached based on prior using nonlinear model functions, which communicate all sigma points. The mean of various variables is estimated based on their statistical data.

29.4 BMS Design Model

The BMS design model has two major component parts viz. controller and plant. The controller is operating the BMS parameters based on conditional situations. The plant is extracting the full strength to drive the electric vehicle system.

Every BMS must have cell balancing. The battery pack's total energy is limited by the block with the lowest charge. Each block necessitates the same quantity of energy. Charge in order to provide the maximum amount of energy and how the capacity of an overall capacity is accessible after recharging the balanced battery packs. Figure 29.4 Depicts a circuit that demonstrates passive balancing. Passive balancing methods function in the lowest cells that have the same charge. There are two ways to apply passive balancing methods. A fixed shunt resistor is used in the first technique and a controllable shunt resistor is used in the second method. Despite

Fig. 29.4 Passive balancing
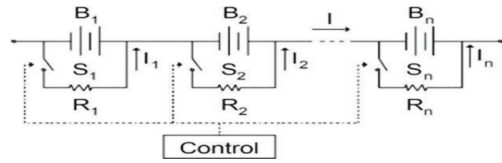
the fact that it does not appear to be a beautiful way if that may be expanded and keep the battery temperature pack under control.

29.5 Simulation Process

Cell balance is described in this section. For this technique, a battery pack with six lithium cells is considered and depicts a typical lithium battery arrangement. They are derived using a cell design and appropriate sensors. A single battery module is shown in this model as well as six single cells formed as a series combination to design a single battery module. The purpose of these modules is to improve the system's efficiency and to simulate different setups. Each cell's temperature, voltage and current are recorded.

This aids in the accurate calculation of SOC and decreases the likelihood of junk value based on coulomb counting method and unscented Kalman filter. The battery's hub currents and temperature parameters are used for input values of the coulomb counting method, whereas cell voltages are used for inputs of Kalman filters, which forecast SOC values. Figure 29.5 depicts the CC arrangement, compute SOC and equation expression.

Figure 29.6 depicts the two variants of the Kalman filter and they are fed cell voltages. The unit cell model is used in the non-linear Kalman filter variants, therefore anticipating the internal cell states. In order to estimate SOC, the steps are the same in both methods viz. state update and measurement update based on all inputs in both approaches. Similarly, the state transition and measurement function evaluate the block result using the prior SOC, temperature and current. The BMS input port receives all of the outputs.
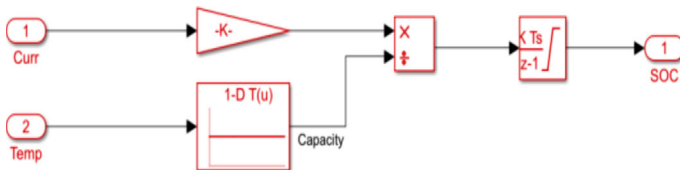
Fig. 29.5 Coulomb counting method
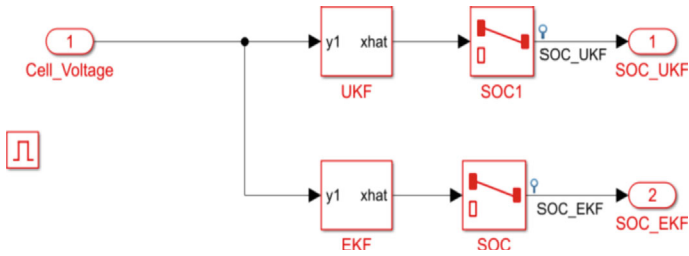
Fig. 29.6 Kalman filter model

29.6 Results and Analysis

The parameters are simulated when the BMS design model is organized. The various blocks are estimated and simulated with necessary tool data. The parameters results are observed and analyzed with tabular form comparisons.

Cell Voltage

Figure 29.7 shows cell voltages fluctuates due to battery current flowing in and out. The graph is described in detail in Table 29.1. The battery pack's cell voltage rapidly reduced from 3.8 to 3.48 V throughout the discharging phase. The voltage of the cells is standby at 3.6 V then BMS is incorporated for charging the cells voltage towards rises.
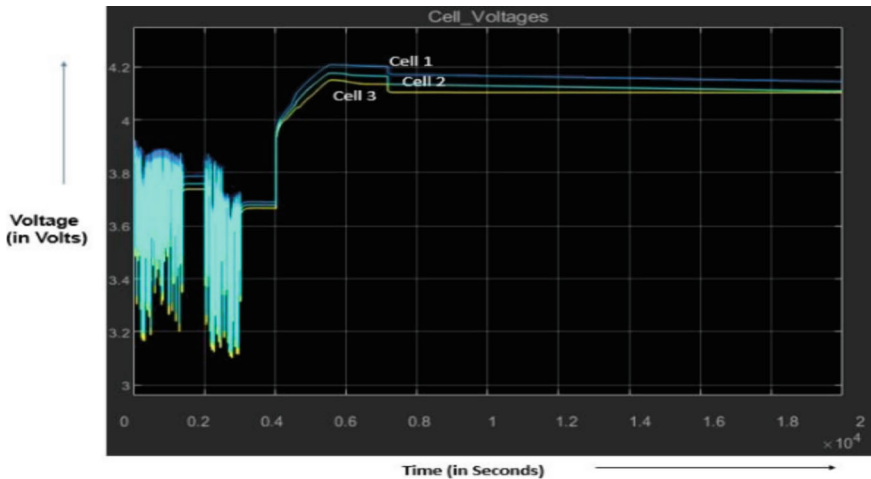
Fig. 29.7 Cells voltage

Table 29.1 Cells voltage upon BMS state

S. No	States	Obtain voltages (V)
1	Starting point of discharging	3.80
2	Last point of discharging	3.48
3	Under standalone	3.60
4	Starting point of charge	4.021
5	Last point of charge	4.171
6	Maximum voltage	4.208

Cell Temperature

The temperature variations are observed while battery charge and discharge conditions. Figure 29.8 depicts the battery temperatures as a function of BMS statuses. Generally, the inside part cells are insulating to accommodate thermal conditions. In this movement, the first cell gets much more temperature than the sixth cell. So, there is a large temperature difference between the edge cells and active thermal management is preferred. The graph is described in terms of data in Table 29.2.
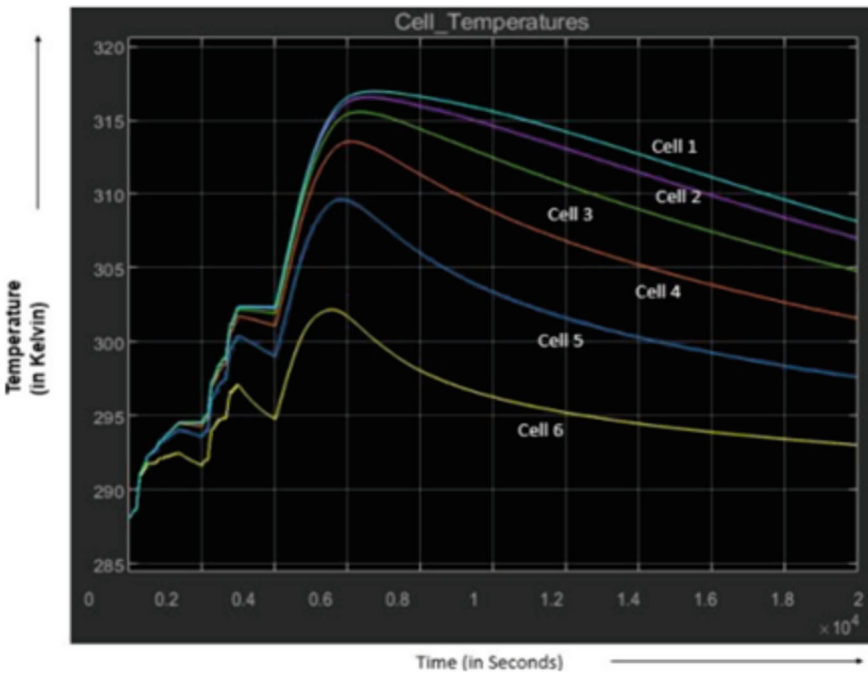
Fig. 29.8 Cellular temperature variations

Table 29.2 Cell 1 temperature variations upon BMS state

S. No	States	Obtain temperatures (K)
1	Starting point of discharging	288.1
2	Last point of discharging	297.1
3	Under standalone	295.7
4	Starting point of charge	297.3
5	Last point of charge	297.6
6	Maximum temperature	302.8

Pack Current

Cell current variations are observed at various BMS modes in Fig. 29.9. A peak resistance of battery cells is used to compute the current limit. During discharging mode, the current swings at the starting stage due to a variety of variables. The graph goes to zero when it is in standby mode. Later charging mode, the cell's voltage hits its maximum value then the current climbs and reaches its peak value. The current is reduced by limiting circuit and voltage spikes are removed. Table 29.3 describes the graph in detail.

State of Charge (SOC) and Cell Balancing

Battery cell SOC is maintained 80% by using various methods at the time of discharging. In conclusion, EKF is more precise and accurate in estimating the SOC. Cells balancing operations are a critical aspect of a battery management system. The graph is described in complete analytical data in Table 29.4. By rebounding from the initial error, EKF outperformed the UKF.
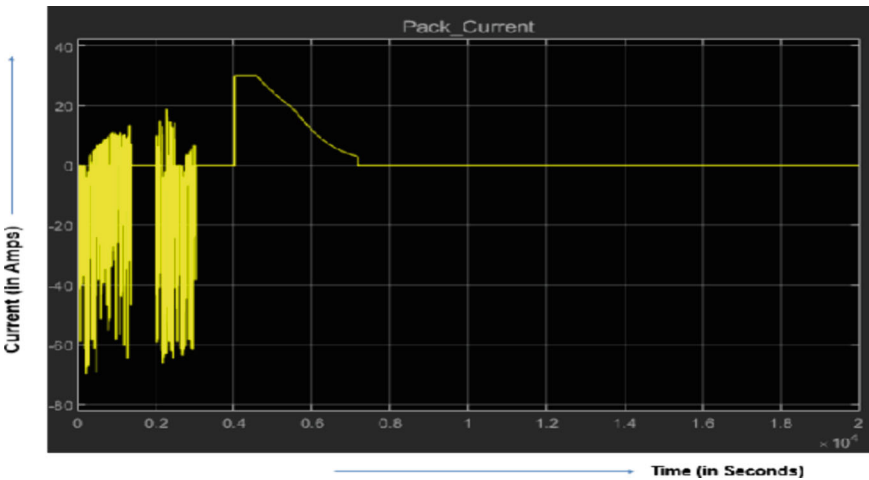
Fig. 29.9 Battery pack current

Table 29.3 Cells current upon BMS state

S. No	States	Obtain currents (A)
1	Starting point of discharging	0
2	Last point of discharging	−19.16
3	Under standalone	0
4	Starting point of charge	30
5	Middle point of charge	10.28
6	Last point of charge	0

Table 29.4 Using BMS states to compare SOC estimation methods

S. No	States	Coulomb counting's (%)	EKF's (%)	UKF's (%)
1	Starting point of discharging	80.00	80.00	80.00
2	Last point of discharging	52.96	49.31	48.10
3	Under standalone	52.45	48.70	43.60
4	Starting point of charge	60.30	56.50	55.50
5	Last point of charge	100.00	98.65	98.46

29.7 Conclusions

MATLAB software is used to estimate and observe the BMS parameters and the Simulink simulated output graphs are used to compare in three algorithms. Initially, the SOC value is set at 75% and estimated to recover the SOC potential up to 80%. While the under process of SOC is computed with Coulomb counting, EKF and UKF methods then the results are attained the maximum values. These are almost cent percent operations viz. Coulomb counting is 100% SOC, EKF is 98.74% SOC and UKF is 98.46% SOC. The Coulomb Counting, Extended Kalman Filter and Unscented Kalman Filter are estimated accuracy from error elimination process and also recovery time is very fast with minimum SOC life value at 48.7%.

Battery management system model has been established in EV and controlling critical factors. BMS model has been prudently computed based on battery life span, consistency and security. The battery management system is investigating the appropriate and optimized power to electric vehicles. Hence, the BMS has taken working actions to control greenhouse gas emissions.

References

1. Li, S., Zhao, P., Gu, C., Huo, D., Li, J., Cheng, S.: Linearizing battery degradation for health-aware vehicle energy management. IEEE Trans. Power Syst. **38**(5), 4890–4899 (2023)
2. Singh, S., More, V., Batheri, R.: Driving electric vehicles into the future with battery management systems. IEEE Eng. Manag. Rev. **50**(3), 157–161. 1 third quarter (2022)

3. Li, S., Zhao, P., Gu, C., Li, J., Huo, D., Cheng, S.: Aging mitigation for battery energy storage system in electric vehicles. IEEE Trans. Smart Grid **14**(3), 2152–2163 (2023)

4. Jafari, S., Byun, Y.-C.: Prediction of the battery state using the digital twin framework based on the battery management system. IEEE Access **10**, 124685–124696 (2022)

5. Phan, D., Bab-Hadiashar, A., Fayyazi, M., Hoseinnezhad, R., Jazar, R.N., Khayyam, H.: Interval type 2 fuzzy logic control for energy management of hybrid electric autonomous vehicles. IEEE Trans. Intell. Veh. **6**(2), 210–220 (2021)

6. Lakshminarayana, G., Sambasiva Rao, G., Divakar, R.V.L., Malik, N.H., Alsaif, F., Alsulamy, S., Ustun, T.S.: Fuzzy-based fifteen-level VSC for STATCOM operations with single DC-Link voltage. Sustain. **15**(7), 6188 (2023)

7. Mars, N., Krouz, F., Louar, F., Sbita, L.: Comparison study of different dynamic battery models. In: International Conference on Green Energy Conversion Systems (GECS), pp. 1−6. Hammamet (2017)

8. Zhou, B., Burl, J.B., Rezaei, A.: Equivalent consumption minimization strategy with consideration of battery aging for parallel hybrid electric vehicles. IEEE Access **8**, 204770–204781 (2020)

9. Hamednia, A., Murgovski, N., Fredriksson, J., Forsman, J., Pourabdollah, M., Larsson, V.: Optimal thermal management, charging, and eco-driving of battery electric vehicles. IEEE Trans. Veh. Technol. **72**(6), 7265–7278 (2023)

10. Gao, W., Li, X., Ma, M., Fu, Y., Jiang, J., Mi, C.: Case study of an electric vehicle battery thermal runaway and online internal short-circuit detection. IEEE Trans. Power Electron. **36**(3), 2452–2455 (2021)

11. Suhail, M., Akhtar, I., Kirmani, S., Jameel, M.: Development of progressive fuzzy logic and ANFIS control for energy management of plug-in hybrid electric vehicle. IEEE Access **9**, 62219–62231 (2021)

12. Zhu, F., Liu, G., Tao, C., Wang, K., Jiang, K.: Battery management system for li-ion battery. J. Eng. **2017**(13), 1437–1440 (2017)

13. Narayana Gadupudi, L., Rao, G.S., Devarapalli, R., García Márquez, F.P.: Seven level voltage source converters based static synchronous compensator with a constant DC-Link voltage. Appl. Sci. **11**(16), 7330 (2021)

14. Narayana Gadupudi, L., Sambasiva Rao, G.: Recent advances of STATCOM in power transmission lines-a review. Turk. J. Comput. Math. Educ. **12**(3), 4621–4626. e-ISSN 1309–4653 (2021)

15. Viswanatha Rao, J., Lakshminarayana, G., Obulesu, D.: Reduction of torque ripples in PMSM using fuzzy controller based driving converter. Turk. J. Comput. Math. Educ. **12**(3), 4652–4658. e-ISSN 1309–4653 (2021)

16. Kaushik, P.S.K., Lakshminarayana, G.: Voltage enhancement using solid state devices. Int. J. Eng. Res. ISSN: 2319–6890. **5**(3), 197–202 (2016)

17. Narayana Gadupudi, L., Sambasiva Rao, G.: 9-Level VSC based STATCOM for reactive power management and voltage stability improvement. J. Green Eng. (JGE). **10**(11), 10275–10288. ISSN: 1904–4720, 2245–4586 (online) (2020)

18. Narayana Gadupudi, L., Sambasiva Rao, G.: 7-level transformers integrated voltage source converter based STATCOM for voltage profile enhancement. J. Solid State Technol. **63**(5), 3134–3141. ISSN: 0038–111X (2020)

19. Lakshminarayana, G., Ravi kumar, B.: Improvement of voltage profile in a radial distribution system with variable speed wind turbines. Int. J. Eng. Res. Appl. **2**(3), 880−884. ISSN: 2248–9622 (2012)

20. Viswanatha Rao, J., Lakshminarayana, G.: Soil health monitoring system using IOT. In: (ICET) International Conference on Innovations in Clean Energy Technology, p. 29. Bhopal, India (2020)

21. Viswanatha Rao, J., Lakshminarayana, G.: Estimation and moderation of harmonics in distribution systems. In: Reddy, A., Marla, D., Simic, M., Favorskaya, M., Satapathy, S. (eds.) Intelligent Manufacturing and Energy Sustainability. Smart Innovation, Systems and Technologies, vol. 169. Springer, Singapore

22. Miao, Z., Xu, L., Disfani, V.R., Fan, L.: An SOC-based battery management system for microgrids. IEEE Trans. Smart Grid. **5**(2), 966–973 (2014)
23. Cheng, K.W.E., Divakar, B.P., Wu, H., Ding, K., Ho, H.F.: Battery-management system (BMS) and SOC development for electrical vehicles. IEEE Trans. Veh. Technol. **60**(1), 76–88 (2011)
24. Gan, N., Sun, Z., Zhang, Z., Xu, S., Liu, P., Qin, Z.: Data-driven fault diagnosis of lithium-ion battery over discharge in electric vehicles. IEEE Trans. Power Electron. **37**(4), 4575–4588 (2022)
25. Anselma, P.G., et al.: Economic payback time of battery pack replacement for hybrid and plug-in hybrid electric vehicles. IEEE Trans. Transp. Electrification **9**(1), 1021–1033 (2023)

Chapter 30
Medicinal Leaves Classification Using Hyperspectral Imaging

**Anuj V. Pophali, Arun G. Bhangdiya, Varad B. Bhagalkar,
Prathamesh P. Sharma, and Swati Hira**

Abstract India is notable for its traditional medicinal leaves. These leaves are not just utilised in home cures which involve use of leaves in regular household ingredients, they are utilised as an emergency treatment for a few common ailments like cough, fever, cold, etc. in Ayurveda. These medicinal leaves hold a superior healing power which varies significantly according to their maturity. The practices that were taken to detect the maturity level of leaves using chemical analysis were much more expensive and tedious. To develop a system to classify the medicinal leaves with their maturity, we have used hyperspectral imaging. We have used three medicinal leaf species which are Tulsi, Neem, and Pudina. For capturing hyperspectral images, we have used a specialised hyperspectral camera. These images give more information about leaves. After collecting the required information for each leaf, some ML classifier models like Support Vector Machine (SVM), Random Forest Classifier, and Logistic Regression were implemented with accuracy scores obtained was 99.76%, 66.91%, and 87.77% respectively. Each species has a subclass of Dry and Fresh which will be verified and tested properly with multiple runs along with its maturity level.

Keywords Hyperspectral imaging · Yolov5 · Support vector machine · Random forest · Logistic regression

30.1 Introduction

Leaves identification can be a tedious process and their scientific names are difficult to remember, which is why it is important to have a system that automates the task of identifying the leaves. The number of plant species on earth runs into tens of thousands of species because of which the classification becomes difficult as the number of classes is too many. We decided to focus purely on a subset of leaves

A. V. Pophali (✉) · A. G. Bhangdiya · V. B. Bhagalkar · P. P. Sharma · S. Hira
Shri Ramdeobaba College of Engineering and Management, Nagpur, India
e-mail: pophaliav@rknec.edu

that are useful to people and hence decide to use the set of medicinal leaves as our dataset.

We aim to use the hyperspectral camera to obtain the hyperspectral images of the leaves for the training and testing datasets, hyperspectral images are used because they can capture electromagnetic wavelengths that are outside the visible region of the electromagnetic waves, a simple RGB image has only 3 channels to capture details but a hyperspectral image has 164 channels which is helpful in extracting a greater number of features (Patil et al. [1]) from the given image, which results in higher accuracy when classifying the images.

The motivation behind this research is to explore hyperspectral imaging because hyperspectral imaging is a promising technology that will be used for various tasks in the future as it captures electromagnetic wavelengths outside the visible region which results in higher accuracy since the amount of data captured in a single image is larger than in a normal RGB image. As the cost of using hyperspectral images goes down its use will increase.

In addition to this, it is also important to note that leaf identification is important when finding/identifying medicinal herbs for various medicinal purposes, and in this case, high accuracy is very important. The system can also be used by people trying to find new species of trees and plants as they can match the leaf that they found to see if it already is a known tree/plant leaf.

The aim of this project is to develop a system that would identify the leaves and classify the medicinal leaves in their respective categories.

The data acquisition flow graph shows in the Fig. 30.1 process of gathering and pre-processing (Gopal et al. [2]) of data. At first multiple leaves of a specific kind are placed on a white surface and are first captured by the hyperspectral camera PIKA NIR-320 with the help of Spectronon software and the image is saved as the images of all the leaves are collected, these leaves are then given as input to a python program which creates an Excel sheet for the spectral data where a row represents the data of an individual leaf the Excel sheet is then labelled accordingly.

30.2 Literature Survey

Petersson et al. [3] introduced the link between Hyperspectral Imaging (HSI) and Deep Learning. The experiment employed many well-established ML technologies, many of which were applied for classification. It concentrated on common algorithms such as SVM (Steinwart et al. [4]) and logistic algorithms.
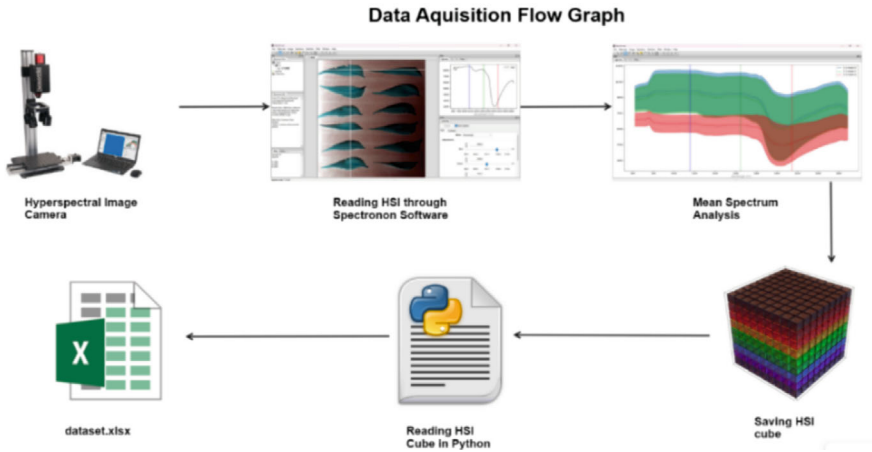
Fig. 30.1 Data acquisition flow graph

Putri et al. [5] conducted a study about using CNN to identify medicinal leaves. The CNN approach was effectively used in this study to extract characteristics from medicinal plant leaves and classify them into 9 medicinal plant leaf classes giving a resulting value between the training and testing data.

LightWeight Convolution Neural Network was utilised in this article by K-Liu et al. [6]. A lightweight convolutional neural network (CNN) is a CNN architecture that is meant to be tiny and computationally efficient while still performing well on a given job. They compared the performance of 3-band RGB photographs and Hyperspectral images to see whether there were any variations in characteristics that are relevant in the green, red, and near-infrared bands. The study also proved the Lightweight CNN model's great flexibility in identifying leaf characteristics of plant images while using a smaller number of training data than appropriate CNN models.

Jahagirdar et al. [7] conducted a study that was primarily concerned with the identification of medicinal plants using Convolutional Neural Networks (CNN)-based technology. It mostly concentrated on techniques like SVM, KNN, and others. They employed 12 different leaf classes (species): Arjun, Bael, Betel, Curry, Guava, Hibiscus, Jamun, Jatropha, Karanj, Lemon, Mango, Mint, Neem, Pomegranate, Saptaparni, and Tulsi. The system recognized the leaves and improved the accuracy of leaf categorization.

The publication by Mukherjee et al. [8] describes research that used computer vision to identify leaves as well as their maturity level. They considered three species, Neem, Tulsi, and Kalmegh, and used CNN to improve their accuracy. They used Binary Particle Swarm Optimisation (BPSO), often known as the CNN hyperparameter. This is a popular iterative search-based optimization approach utilised in a variety of domains, including CNN parameter tweaking. As a result, it was decided to investigate different medicinal leaves and how they contribute to identification and categorization.

The researchers, Naeem et al. [9] focused on Multispectral Imaging, which is comparable to hyperspectral imaging but differs somewhat in terms of spectral bands and resolution. They tested five classifiers on plant leaf species and discovered that the perceptron classifier with multiple layers had the best accuracy of 99% when compared to the other classifiers. Tulsi, Peppermint, Bael, Lemon Balm, Catnip, and Stevia were among the therapeutic plant species cited, and each received a higher categorization result.

Manoharan [10] conducted a study that was about Classification using two-stage authentication. This two-stage authentication technique improved the identification rate necessary for herbal leaf detection. The key disadvantage is that the width and other characteristics do not meet optimal precision. As a result, the aforementioned technique was mostly employed to eliminate incomplete issues in order to increase the identification rate for herbal leaves. It increased the detection rate and reduced classification mistakes. As a result, this work improved accuracy and other performance metrics.

Janani and Gopal [11] offer a technique for identifying medicinal plant leaves based on visual characteristics and Artificial Neural Networks. They correctly identified five varieties of medicinal leaves based on form, colour, and texture attributes with a 94.4% accuracy.

The majority of the articles we studied collected datasets using RGB spectral wavelengths, however, we employed Hyperspectral Images and associated spectral data to accomplish classification. All of the following articles drew conclusions from the accuracy of several classifiers on different species.

30.3 Methodology

30.3.1 Preprocess

YOLOv5 (Walchen et al. [12]; Kaur and Kaur [13]; Asker et al. [14]) is a well-known object detection method that can be trained on a variety of unique datasets. Some preparation steps are required to prepare your dataset for training. Typical steps include:

- Gathering images of the things to be detected.
- Label each image with the items' bounding boxes.

- Making a dataset.yaml file containing the specifics of your dataset.
- After you've finished these procedures, you can begin training your models.

30.3.2 Support Vector Machine (SVM)

An SVM (Sabu et al. [15]; Kleinbaum and David [16]) is a supervised machine learning algorithm that can perform classification and regression tasks. It operates by determining the optimal hyperplane between two or more classes of data points. A hyperplane (Arun et al. [17]) is a line or plane in high-dimensional space that separates data points into two or more groups.

SVMs are commonly employed for classification jobs where the data can be separated linearly. This indicates that a straight line or plane can separate the data points.

An SVM formulation can be phrased as follows:

$$-\frac{1}{2} * ||w||^2 + C * sum \left(\left(0, 1 - vi * \left(w' * ui + b \right) \right) \right) \qquad (30.1)$$

where:

- w is the weight vector
- b is the bias term
- C is a hyperparameter that tries to achieve balance between maximising the margin and minimising the number of misclassified points.
- vi is the label of data point ui.

30.3.3 Random Forest Classifier

Random forest (Cheng et al. [18]; Genuer et al. [19]) is a learning approach which helps improve machine learning results by combining several models for classification, regression, and other problems, it is trained by creating multiple decision trees (Mehl et al. [20]). The result of the random forest classifier is the class chosen by a majority voting by the decision trees.

Its key advantages include robustness against overfitting, thanks to the combination of trees and random feature subsets. Random Forest is commonly used for data analysis, pattern recognition, and predictive modelling.

$$v = argmax(average(f(u)) \, for \, d \, in \, 1 \, to \, D) \qquad (30.2)$$

where:

- v is the predicted class label for a given input u

- f(u) is the result of a random forest, which is the average of the result of voting of D decision trees
- D is the number of decision trees in the random forest.

30.3.4 Logistic Regression

Logistic regression is a simple and interpretable statistical model used for binary classification tasks. It estimates the probability of an event occurring by applying a linear combination of features to a logistic function. The model parameters are estimated using maximum likelihood estimation, and predictions are made based on a predefined threshold. Logistic regression is widely used in fields like medicine and finance, where binary classification is common, and it offers simplicity, interpretability, and insights into feature importance.

P(V = 1) is predicted using a logistic regression model as a function of U. It is one of the most fundamental machine learning approaches and may be applied to a number of classification-related problems, including spam identification, diabetes prediction, cancer diagnosis, text editing, etc.

Binary or binomial logistic regression, in which the objective or dependent variable can only have one of two possible values 0 or 1, is the most fundamental type of logistic regression. We may mimic a relationship between a number of predictor variables and a binary or binomial target variable using this technique. The linear function is primarily used in logistic regression as an input to another function, such g, in the relation shown below:

$$h_\theta(u) = g\left(\theta^T u\right) \, where \, 0 \le h_\theta \le 1 \tag{30.3}$$

In this case, g is the logistic or sigmoid function, which may be written as follows:

$$g(v) = \frac{1}{1 + e^{-v}} \, where \, v = \theta^T u \tag{30.4}$$

30.4 Algorithm

ALGORITHM:

1.	*Input: Hyperspectral images (say) I1, I2, I3, I4, ..., In*
2.	*Variables: i, x, w, z, s, t, u*
3.	*for i in I1......In do*
4.	$w \leftarrow yolo(i)$ *(Apply yolo [leaf detection] on each image i)*
5.	$x \leftarrow x \, U \, w$ *(Create a set of all images of leaves)*
6.	*end*
7.	$z \leftarrow writeDataToExcel(x)$ *(Collecting all the information in excel with labels)*
8.	$s \leftarrow SupportVectorMachine(z)$ *(Apply SVM on dataset to classify leaves)*
9.	$t \leftarrow RandomForestClassifier(z)$ *(Apply RF on dataset to classify leaves)*
10.	$u \leftarrow LogisticRegression(z)$ *(Apply LR on dataset to classify leaves)*
11.	*Output: Comparing accuracy {s, t, u} along with confusion matrix.*

The above algorithm describes the process of the project right from the collection of hyperspectral images mentioned in the algorithm with *I1, I2, I3, I4, ..., In as input*. The variables mentioned are *for* each image, after that *w* is the array of all leaf images obtained after applying yolo on a single image *i* and *x* is the set of all images.

Then we have collected all the spectral information in an excel sheet *z* along with their labels (class name) for each image. After applying the ML classifier models giving accuracy for each of them is *s, t, u* respectively. Output is the set of the accuracy along with the confusion matrix obtained from each of them.

30.5 Experimentation and Result

Above Table 30.1 consists of names of all the classes formed in the first column and their respective sample (.bil) image in the second column.

Above Table 30.2 shows the accuracy score obtained (in percent) against the respective classifier model used for classification.

Table 30.1 Classes in the dataset

Classes	Dry neem	Fresh neem	Dry tulsi	Fresh tulsi	Dry pudina	Very dry pudina	Fresh pudina
Image							

Table 30.2 Accuracy from different classification models

	Accuracy score (in %)
SVM (support vector machine)	99.76
Random forest classifier	66.91
Logistic regression	87.77

Above Figs. 30.2, 30.3 and 30.4 shows the confusion matrix obtained from the respective classification model, the diagonal entries in these matrices are the number and percent of leaves correctly predicted and the non-diagonal entries show the number and percent of leaves incorrectly predicted.

Fig. 30.2 Confusion matrix of SVM
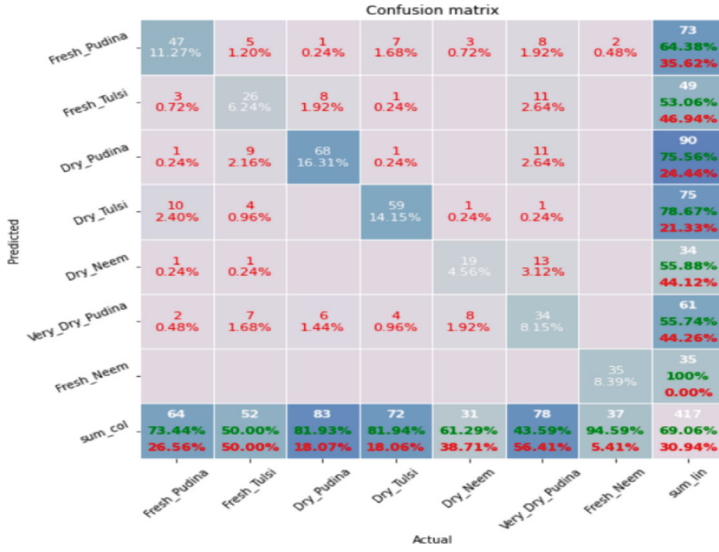
Fig. 30.3 Confusion matrix of RF

Fig. 30.4 Confusion matrix of LR

30.6 Conclusion

We have implemented three ML classifier models to determine the best algorithm for hyperspectral image classification like SVM (Support Vector Machine), Random Forest, and Logistic Regression. Comparing an accuracy obtained from each of the models, we have inferred that SVM provides better accuracy for classification. The classification model that is developed here can be used by people associated with Ayurveda to discover new sources of medicinal leaves. It can also be used to reduce the probability of misidentification and hence reduce the unintended side effects of a treatment due to misidentification of the type of plant used. The model can be further expanded by adding images of different types of leaves to the dataset to classify more leaves. In addition to this, the number of images of the classes currently present in the dataset set can be increased.

References

1. Patil, C., Sharma, A., Shimpi, M., Patil, A., Rajeswari, K.: Identification of leaf for extraction of medicinal values using machine learning. IJSRD- Int. J. Sci. Res. & Dev. **7**(8) (2019)
2. Gopal, A., Prudhveeswar Reddy, S., Gayatri, V.: Classification of selected medicinal plants leaf using image processing. In: 2012 International Conference on Machine Vision and Image Processing, MVIP 2012, pp. 5–8 (2012)
3. Petersson, H., Gustafsson, D., Bergstr¨om, D.: Hyperspectral Image Analysis using Deep Learning-a Review 978–1–4673–8910–5/16/$31.00 ©2016 IEEE (2016)
4. Steinwart, I., Chrismann, A.: Support vector machine. Information Science and Statistics, pp. 287–330. Springer (2008)
5. Putri, Y.A., Djamal, E.C., Ilyas, R.: Identification of Medicinal Plant Leaves Using Convolutional Neural Network (2020). https://doi.org/10.1088/1742-6596/1845/1/012025
6. KLiu, K.H., Yang M.H., Huang S.T., Lin, C.: Plant species classification based on hyperspectral imaging via a lightweight convolutional neural network model. Front. Plant Sci. **13**, 855660 (2022). https://doi.org/10.3389/fpls.2022.855660
7. Jahagirdar, B., Munot, D., Belhekar, N.: Identification of Indian medicinal leaves using convolutional neural networks. Int. Res. J. Eng. Technol. (IRJET) **08**, 1777–1779 (2021)
8. Mukherjee, G., Tudu, B., Chatterjee, A.: A convolutional neural network-driven computer vision system toward identification of species and maturity stage of medicinal leaves: case studies with Neem, Tulsi and Kalmegh leaves. https://doi.org/10.1007/s00500-021-06139-9
9. Naeem, S., Ali, A., Chesneau, C., Tahir, M.H., Jamal, F., Sherwani, R.A.K., Ul Hassan, M.: The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach (2021). https://doi.org/10.3390/agronomy11020263
10. Samuel Manoharan, J.: Flawless detection of herbal plant leaf by machine learning classifier through two stage authentication. Proced. J. Artif. Intell. Capsul. Netw. (2021). **3**(2) (2010)
11. Janani, R., Gopal, A.: Identification of selected medicinal plant leaves using image features and ANN. In: 2013 International Conference on Advanced Electronic Systems (ICAES), pp. 238–242 (2013). https://doi.org/10.1109/ICAES.2013.6659400
12. Wäldchen, J., Mäder, P.: Plant species identification using computer vision techniques: a systematic literature review. Arch. Comput. Methods. Eng. **25**, 507–543 (2018). https://doi.org/10.1007/s11831-016-9206-z
13. Kaur, S., Kaur, P.: Plant species identification based on plant leaf using computer vision and machine learning techniques. J. Multimed. Inf. Syst. **5**, 49–60 (2019). https://doi.org/10.33851/JMIS.2019.5.2.49

14. Asker, M., El-gengaihi, S.E., Hassan, E.M., et al.: Phytochemical constituents and antibacterial activity of Citrus lemon leaves. Bull. Natl. Res. Cent. **44**, 194 (2020)
15. Sabu, A., Sreekumar, K., Nair, R.R.: Recognition of ayurvedic medicinal plants from leaves: a computer vision approach. Fourth Int. Conf. Image Inf. Process. (ICIIP) **2017**, 1–5 (2017). https://doi.org/10.1109/ICIIP.2017.8313782
16. Kleinbaum, D.G.: Assessing discriminatory performance of a binary logistic model: ROC curves. Logist. Regres.: Self-Learn. Text. 345–387 (2010)
17. Arun, C.H., Sam Emmanuel, W.R., Christopher Duraira, D.: Texture feature extraction for identification of medicinal plants and comparison of different classifiers. Int. J. Comput. Appl. **62**(12), 0975−8887 (2013)
18. Cheng, X., Chen, Y.R., Tao, Y., Wang, C.Y., Kim, M.S., Lefcourt, A.M.: A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. Trans. ASAE **47**(4), 1313–1320 (2004)
19. Genuer, R., Poggi, J.-M., Tuleau-Malot, C.: Variable selection using random forests. Pattern Recogn. Lett. **31**(14), 2225–2236 (2010)
20. Mehl, P.M., Chen, Y.R., Kim, M.S., Chan, D.E.: Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. J. Food Eng. **61**, 67–81 (2004)

Chapter 31
Churn Modeling Using Artificial Neural Network

Nikita Chopde, Amol Deshpande, and Dayanand Ambawade

Abstract In recent years, banks have observed increased customer departures for various reasons. Churn modeling is essential to formulating effective strategies for customer retention. This study aimed to anticipate whether a customer is likely to leave the bank in the foreseeable future. Churn prediction involves identifying clients likely to discontinue service or terminate their subscription. This prediction holds significant importance for many firms, as acquiring new customers often incurs higher costs than retaining existing ones. The dataset utilized includes diverse parameters such as credit score, age, gender, estimated salary, etc. In addressing this issue, neural networks prove advantageous as they can adeptly learn and model nonlinear and intricate interactions. This capability is vital because many real-life relationships between inputs and outputs exhibit non-linear and complex patterns. Consequently, an Artificial Neural Network with two hidden layers is employed within the Python environment. The study delves into the correlation between various parameters, conducting thorough exploratory data analysis to discern the key factors influencing a customer's decision to discontinue their association with the bank.

Keywords Neural networks · Artificial intelligence · Hidden layers · Data · Churn modeling · Machine learning · Prediction

31.1 Introduction

Customer churn is where some customers exit, no longer being customers of the institution or business. Churners have always been a major problem for every business that offers services. The banks have been seeing unusual churn rates over the past few years. The banks have been seeing customers leave at a very high rate, and they wanted to know what the problem was and how they could assess and address it. B2B companies can experience sustained growth by implementing an accurate and

N. Chopde (✉) · A. Deshpande · D. Ambawade
Bhartiya Vidya Bhavan's Sardar Patel Institute of Technology, Mumbai 400058, India
e-mail: nikita.chopde@spit.ac.in

dependable churn prediction strategy. The primary objective of this study is to develop a job demographic segmentation model that informs the bank about customers at the greatest risk of leaving. This model adds value to any customer-centric organization. [1].

The banks noticed the problem of unsatisfied customers gradually leaving the bank and hence took a sample of 10,000 random customers from three European countries: France, Spain, and Germany. The sample included the Customer ID, surname, Credit score, Geography, Gender, Tenure in the bank, balance, etc. The dataset also included the number of products the customers owned, like whether they have a savings account, a credit card, a loan, etc. A boolean flag was given to two questions: whether the customer had a credit card and whether he/she was an active member of the bank or not. The activeness of a member was measured by whether or not the customer had logged into their online banking in the past month or whether they had made some transactions in the past month. The estimated salary of the customers was also appended at the end of the dataset. The final column consisted of a boolean expression of the result, that is, if the customer is predicted to leave or not.

The strategy used is Deep Learning, which has the advantage of discovering its features and learning several levels of representation on its own. Artificial neural networks (ANNs) are computer networks with biological inspiration. The component of a computing system called an ANN is made to mimic how the human brain evaluates and processes information. It serves as the cornerstone of artificial intelligence (AI) and resolves issues that would be impossible or challenging by human or statistical standards. A subset of deep learning is ANN. Deep learning is one of the most powerful branches of Machine Learning.

31.1.1 Literature Review

Agrawal et al. [2] investigated churn modeling and utilized a multilayer neural network to construct a non-linear classification model. The churn prediction model considers customer, support, use, and contextual features.

The research work carried out by Deng et al. [3] effectively uses ensemble learning algorithms to get the customer churn using four metrics on three classifiers. The paper's conclusion covers various case studies, each accompanied by its respective workflows. Yahaya et al. [3] suggest that a hybrid genetic algorithm and K means filter and artificial neural network can be used where they were able to get 85% accuracy. Performance evaluation is also done towards the end of the paper. The work presented by Jagadeesan [4] uses three different classifiers to get the most accurate classifier among those. It came out that the Random Forest Classifier is the most accurate, but when compared to ANN, the latter is more accurate. Precision, recall, and F1 values are also calculated. By contrasting various methods, the study by Sadiq et al. [5] aims to anticipate the best accuracy rate; the results are then shown on a website. The website was designed to dynamically receive user parameters to predict whether a customer would churn. Shaaban et al. [6] developed a simple model based on Data

Mining approach to aid a Customer Relationship Management (CRM) department in monitoring clients and preventing churn. In the study by Tran et al. [7], a data set of 5000 cases with 23 attributes was employed to train and test the model. The outcomes from employing three different methods—Decision Trees (DT), Support Vector Machines (SVM), and Neural Networks (NN) for classification, along with standard K Means procedures for clustering—suggest that the SVM technique yields the most favorable results based on the provided dataset. The research carried out by Mandava et al. [8] suggests that in comparison to conventional classification methods, selecting the most significant feature and classifying the dataset into distinct groups to apply several models to each cluster improves prediction. In Kaur and Kaur [9] study, preprocessing is conducted to address data imbalance and missing values. Subsequently, various machine learning algorithms such as Logistic Regression (LR), Decision Trees (DT), k-Nearest Neighbors (KNN), and Random Forests (RF) are employed for baseline feature selection and model development. Additionally, less precise ensemble methods like averaging and max voting are applied to enhance the model's performance.

31.2 Implementation

The implementation is divided into broadly five parts. Before starting the implementation, we imported some relevant libraries: NumPy, Pandas, Seaborn, and TensorFlow. NumPy, a Python library, is employed for array manipulation and functions for matrices, Fourier transforms, and linear algebra. Pandas, a high-level data manipulation tool built on NumPy, introduces the DataFrame, a key structure for storing and manipulating tabular data with rows of observations and columns of variables. Python's TensorFlow module enables rapid numerical computation and serves as a foundational library for developing Deep Learning models, either directly or through wrapper libraries aligned with TensorFlow that streamline the process.

First, we process the data, which makes the raw data machine-readable. Next, the Artificial Neural Network (ANN) model will be constructed using the training data and additional training using the test data. The final phase involves prediction and evaluation, where we forecast customers likely to leave the bank and assess the contributing factors. In our implementation, we want to predict a binary variable that can take the value one if a particular person exited from the bank and zero if the person is still the customer of the bank.

31.2.1 Data Preprocessing

The first step of the implementation is Data Preprocessing. Relevant data needs to be extracted from the dataset and polished for optimum usability. The output depends

on the input we take. Hence, Data Preprocessing is the most important step in the process, as we take raw data and make it useful.

We imported our dataset from 'Churn_Modelling.csv' and studied the dataset and its features using certain functions. While importing the dataset, we did not include irrelevant columns and rows. This is because the columns - RowNumber, Customer ID, and Surname do not affect the prediction in any manner. Also, the result row is the outcome; hence, it is omitted from the selection. As a result, we considered all the columns from CreditScore to Estimated Salary. The computer doesn't understand string values, so we need to make each string input binary. For that, Label Encoding was performed for the 'gender' column. For this, 'LabelEncoder' was imported from the scikit-learn Python library. Label Encoding involves transforming labels into numeric format, making it machine-readable in binary form. This simplification of data for computers is achieved through the process of Label Encoding.

Our dataset's 'Geography' column had three values for the countries, so Label Encoding was not a viable choice for preprocessing. To solve this problem, we used one-hot encoding, a method used for the same purpose. After performing one-hot encoding, the computer randomly assigned a three-digit code to each of the values.

The subsequent stage involves dividing the dataset into Training and Test sets. The Training set serves as a subset to train the model, while the Test set functions as a subset to evaluate the model. The Training set must be larger than the Test set. In our implementation, the Training and Test sets are divided in an 80% to 20% ratio, respectively.

The last stage in data preprocessing involves Feature Scaling. Feature scaling is a technique utilized to standardize the range of independent variables or features in the data. In the context of deep learning, feature scaling is deemed a necessary step. It is fundamental to use Feature Scaling whenever we make Artificial Neural Networks. We applied feature scaling to all the features, regardless of the value they contained.

31.2.2 Building ANN

- An Artificial Neural Network is a series of layers, commencing with the input layer and sequentially progressing through fully connected layers until reaching the output layer.
- We use the Sequential class in the models module, which we get from Tensor-Flow.
- Once the ANN is initialized through the Keras module, we use the dense class to make two hidden neural network layers.
- An output layer was also added to complete the network.
- Specific activation functions, including Relu and, were applied. Relu function was utilized for the input layers, while Sigmoid function was implemented for the output layer.
- Fig. 31.1 below represents the basic building of an Artificial Neural network and an ANN's hypothetical structure.
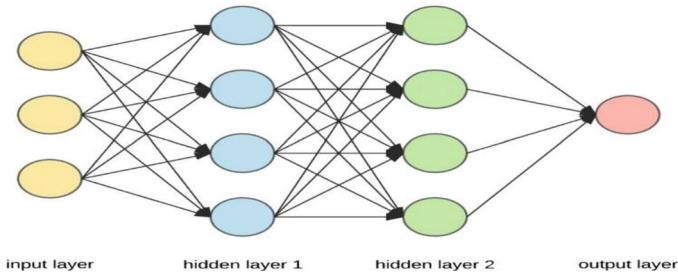
Fig. 31.1 Basic building of ANN

31.2.3   Training the ANN

We trained the ANN we made in the previous step in this step. But just like any coding assignment, we need to compile the ANN first.

Training an Artificial Neural Network (ANN) involves an iterative procedure in which individual training data examples are sequentially introduced. The weights' values are adjusted after each presentation. Once all examples have been processed through the network, one training epoch is completed, and the process may commence again. Throughout this phase, the ANN acquires the ability to predict the accurate output based on the provided input examples. Ideally, it also learns to extrapolate this prediction to new data, demonstrating its capability to generalize beyond the training set.

The 'fit' function was used to train the ANN on a training set. And to improve the accuracy of time certain amounts of epochs were applied. An epoch represents the count of complete passes that a machine learning algorithm has made through the entire training dataset.

31.2.4   Making the Prediction and Evaluating the Model

For predicting the values, we started with predicting for a single case. Taking out the data of a random row, we attempted to predict the churning of that customer. For this, the transform method was used inside the prediction method, and the data (the preprocessed data) was inserted in the arguments.

We developed the confusion matrix and accuracy using the scikit-learn python library for the test set. Accuracy of 87% was achieved. Also, statistical parameters like mean, variance and standard deviation were calculated. We compared various features of the dataset and evaluated the model based on the observations.

```
print(ann.predict(sc.transform([[1, 0, 0, 600, 1, 40, 3, 60000, 2, 1, 1, 50000]]))) > 0.5)

[[False]]
```

Fig. 31.2 Single observation prediction

31.3 Results

For predicting the values using ANN, we started with predicting for a single case. From Fig. 31.2, it is observed that by taking out the data of a random row, we attempted to predict the churning of that customer. We tried a prediction for a 40-year-old male customer from France who had a credit score of 600, 3 years of tenure, a balance of $60,000 who had 2 numbers of products and had a credit card, and was also an active member. His estimated salary was $50,000. So, we predicted if this customer would leave the bank or not.

31.4 Exploratory Data Analysis

31.4.1 Correlation Matrix

As depicted in Fig. 31.3, a correlation matrix is a table displaying correlation coefficients between variables, used for data summary, input in advanced analyses, and as a diagnostic tool.

From the above correlation matrix, it can be inferred that Age, from all the other parameters, is significantly the only factor with a high correlation with the expected parameter. This means that as customers age, the expected churn rate also increases.

31.4.2 Tenure and Exited

According to Fig. 31.4 factor plot, it can be said that in the mid-tenure period there are less exits. The drop-off rate is high in the initial years.

31.4.3 Gender and Exited

Gender correlation shown in Fig. 31.5, depicts that females tended to leave more than males. But overall, this parameter has very less correlation with churn rate.
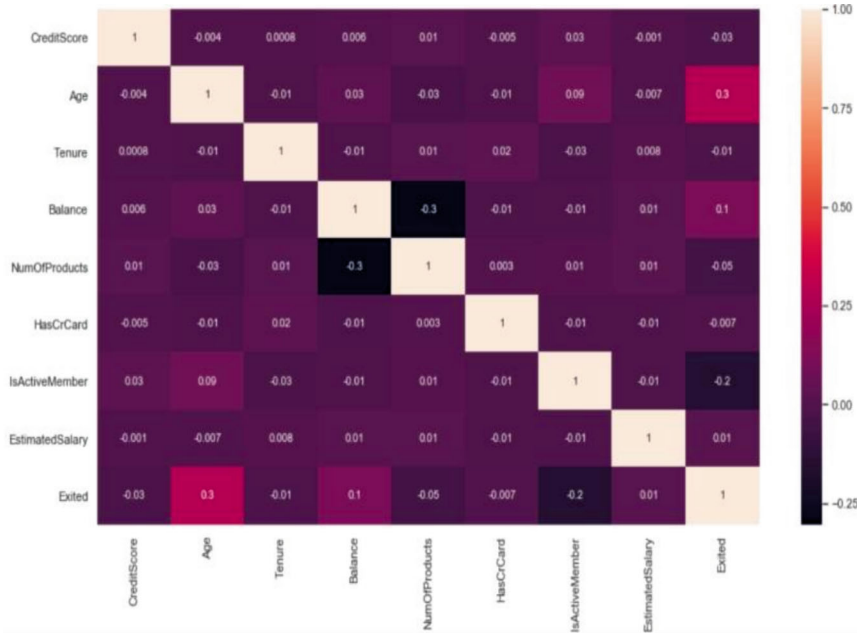
Fig. 31.3 Correlation matrix

Fig. 31.4 Factor plot for
tenure versus exited
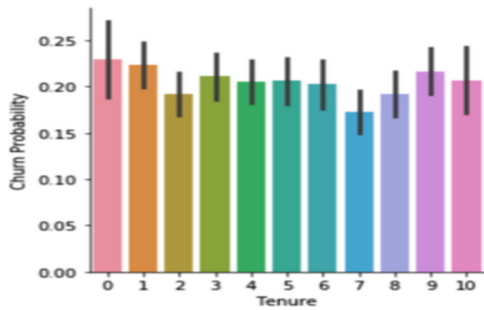
31.4.4 Age and Exited

It can be inferred from Fig. 31.6, that elderly people tended to leave more than the younger ones which can be attributed to the risk-taking abilities of the young.

Fig. 31.5 Factor plot for
gender versus exited
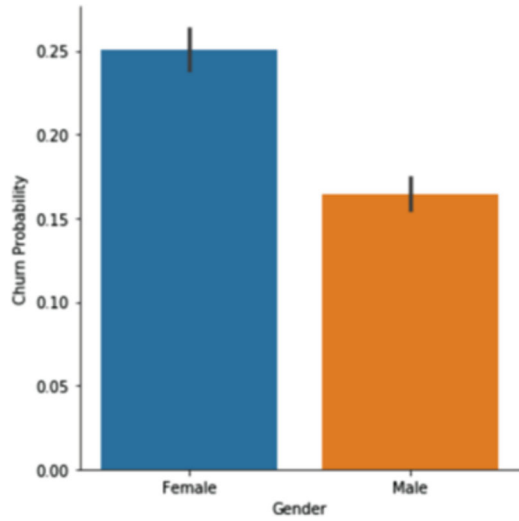
Fig. 31.6 Box plot for age
versus exited
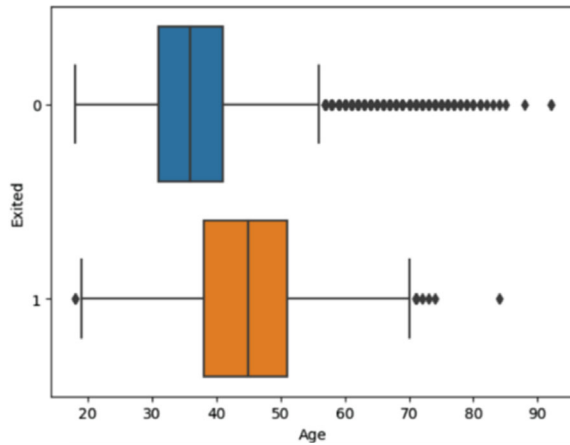
31.4.5   *Estimated Salary and Exited*

For getting the correlation between the Estimated Salary and Exited parameters, a
box plot is plotted as shown in Fig. 31.7. It is observed from the box plot that there
seems to be slightly higher salaries for the Excited customers.

Fig. 31.7   Box plot for estimated salary versus exited

31.5   Conclusion

In this research paper, we conducted churn analysis for a bank encompassing approximately 10,000 customers. The utilization of the Artificial Neural Network for this analysis was motivated by its numerous advantages, particularly its capacity to comprehend and model nonlinear and intricate relationships. Churn modeling proves advantageous for the bank, allowing managers to identify potential churners and formulate strategies for customers exhibiting features like those prone to churning.

Our investigation delved into the correlation among factors influencing the decision-making process regarding customer retention. Comprehensive scrutiny of all factors revealed their impact on the churning rate. The correlations indicated that Age emerged as the most influential factor, showing a significant positive correlation; in other words, older individuals are more inclined to leave the bank.

To enhance overall accuracy, future studies could consider employing a greater number of epochs. Additionally, incorporating data from diverse banks would provide a more comprehensive understanding of churners within the banking sector.

References

1. Pradeep B, Sushmitha Vishwanath Rao, Swati M Puranik, Akshay Hegde.: Analysis of Customer Churn Prediction in the Logistic Industry using Machine Learning, International Journal of Scientific and Research Publications, Volume 7, Issue 11, November 2017, ISSN 2250- 3153Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) Conference 2016, LNCS, **9999**, pp. 1–13. Springer, Heidelberg (2016)
2. Sanket Agrawal, Aditya Das, Amit Gaikwad, Sudhir Dhage. In: Customer churn prediction modeling based on behavioral patterns analysis using deep learning. In: 2018 International

Conference on Smart Computing and Electronic Enterprise (ICSCEE), pp.1–6, 2018.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1– Publisher, Location (2010)

3. Yihui Deng, Dingzhao Li, Lvqing Yang, Jintao Tang, Jiangsheng Zhao, In: Analysis and prediction of bank user chum based on ensemble learning algorithm. In: 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA).

4. Yahaya, R., Abisoye, O. A., Bashir, S. A..: In: An enhanced bank customers churn prediction model using a hybrid genetic algorithm and KMeans filter and artificial neural net. In: Proceedings of the 2020 IEEE 2nd International Conference on Cyberspace Cyber Nigeria.

5. Dr. Jagadeesan, A.P., Ph.D, Indhuja, In: Bank customer retention prediction and customer ranking based on deep neural networks, International Journal of Scientific Development and Research (IJSDR), September 2020 IJSDR. **5**:9

6. Dr. Md Jaffar Sadiq, Devashish Jobanputra, Tadanki Gayithri Sai Kaushik, J V V Satya Vrath Rao, In: Bank customer Churn prediction using machine learning. Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol., **8**:3 pp. 334–341

7. Essam Shaaban, Yehia K. Helmy, Ayman Elsayed Khedr and Mona Mohamed Nasr, In: A Proposed Churn Prediction Model", International Journal of Engineering Research and Applications (IJERA), **2**:4, pp. 693–697, (2012)

8. Hoang Dang Tran: Ngoc Le, Van-Ho Nguyen, In: Customer Churn Prediction in the Banking Sector Using Machine Learning-Based Classification Models. Interdiscip. J. Inf. Knowl. Manag. **18**, 087–105 (2023)

9. Sreenitya Mandava, Anvita Gupta, Komanduri Srikar, In: Customer Churn Prediction in Banking Sector—A Hybrid Approach, Ijraset Journal For Research in Applied Science and Engineering Technology, **10**, (2022)

10. Ishpreet Kaur, Jasleen Kaur, In: Customer Churn Analysis and Prediction in Banking Industry using Machine Learning, 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC).

Author Index