Amita Dev
Arun Sharma
S. S. Agrawal
Ritu Rani (Eds.)

Communications in Computer and Information Science 2268

Artificial Intelligence and Speech Technology

5th International Conference, AIST 2023
Delhi, India, December 26–27, 2023
Proceedings, Part II

Part 2



Springer

Communications in Computer and Information Science 2268

Series Editors

Gang Li, *School of Information Technology, Deakin University, Burwood, VIC, Australia*
Joaquim Filipe, *Polytechnic Institute of Setúbal, Setúbal, Portugal*
Ashish Ghosh, *Indian Statistical Institute, Kolkata, West Bengal, India*
Zhiwei Xu, *Chinese Academy of Sciences, Beijing, China*

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Amita Dev · Arun Sharma · S. S. Agrawal ·
Ritu Rani
Editors

Artificial Intelligence and Speech Technology

5th International Conference, AIST 2023 Delhi, India, December 26–27, 2023 Proceedings, Part II

Springer

Editors
Amita Dev
Indira Gandhi Delhi Technical University for Women
Delhi, India

S. S. Agrawal
CSIR
Delhi, India

Arun Sharma
Indira Gandhi Delhi Technical University for Women
Delhi, India

Ritu Rani
Indira Gandhi Delhi Technical University for Women
Delhi, India

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-75166-0 ISBN 978-3-031-75167-7 (eBook)
https://doi.org/10.1007/978-3-031-75167-7

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

The realm of artificial intelligence (AI) has witnessed unparalleled advancements and innovations, rendering previous iterations incomparable. These advancements have significantly bolstered capabilities in machine learning, natural language processing (NLP), and deep learning, thereby facilitating novel developments across specialized domains, particularly in speech technology. AI finds extensive application in various facets of speech-related solutions, encompassing speech analysis, representation and modeling, recognition and comprehension of spoken language, affective speech recognition, interpretation and synthesis, design of speech interfaces and human factors engineering, technologies for recognizing speech emotions, and audio-visual speech processing, among others. This volume contains papers presented during the Fifth International Conference on Artificial Intelligence and Speech Technology (AIST 2023), organized by the Indira Gandhi Delhi Technical University for Women (IGDTUW), India, and held during December 26–27, 2023. Like the previous four editions of the conference (AIST 2019, AIST 2020, AIST 2021 and AIST 2022), this fifth edition of AIST was a great success, with the participation of experts and authors from countries including Japan, Germany, Australia, France, and almost all corners of India.

The Technical Program Committee of AIST 2023 is extremely grateful to the authors who showed an overwhelming response to the call for papers, with over 235 papers received. All submitted papers went through a double-blind peer review process, with each paper being reviewed by 3–4 experts, and finally 71 papers were accepted for publication in the Springer CCIS series. We are thankful to the reviewers for their efforts in assessing the high-quality papers. We are sincerely thankful to American Express and the Department of Science and Technology, Government of India, for their support in establishing the Centre of Excellence in AI at IGDTUW, under whose banner AIST 2023 was organized.

The conference was distinguished by the esteemed presence of delegates representing various industries, academia, and renowned institutes. We were honored to welcome Ranjana N., Director of the Defence Science and Technology Agency at the Defence Research and Development Organization, as the Chief Guest for the Inauguration Ceremony of the Artificial Intelligence and Speech Technology conference. The event featured eminent speakers and professionals who enriched the discussions with their valuable insights. Among them were Sachin Gulati, Director of Campus Hiring at American Express, Yudong Zhang from the University of Leicester, UK, Meena Jha from CQLU Australia, Sardar Islam from Victoria University, Australia, Helmy Bin Abd Wahab from University Tun Hussein Onn Malaysia, Shakir Ali from Imam University, Kingdom of Saudi Arabia, Karunesh Arora from the Center for Development of Advanced Computing Noida, Bharathi from the Centre for Artificial Intelligence and Robotics at DRDO, Devendra from Netaji Subhas University of Technology, Goldie Gabrani from Manav Rachna University, Ajay Kumar and Sunita Garhwal from Thapar University, Priyankoo

Sarmah from the Indian Institute of Technology Guwahati, and S. S. Agrawal, Emeritus Scientist at the Council of Scientific and Industrial Research, who contributed their expertise to the AIST conference. A total of 71 papers were presented by authors hailing from India and various other countries, utilizing both online and offline modalities during the conference. The papers were strategically organized into 8 distinct tracks, each aligning with the specific theme of the respective track. Notably, each track was overseen by appointed session chairs and co-session chairs, ensuring a systematic and focused approach to the proceedings. To recognize outstanding contributions, a Best Paper Award was conferred in each track, acknowledging and celebrating excellence in research and presentation.

We are sure that this colloquy of researchers and experts from academia and industry will greatly benefit researchers, students, and faculty. Young scientists and researchers will find the contents of the proceedings helpful to set roadmaps for their future endeavors.

December 2023

Amita Dev
Arun Sharma
S. S. Agrawal
Ritu Rani

Organization

General Chair

Amita Dev	IGDTUW, India

Technical Program Committee Chair

S. S. Agrawal	CSIR, India

Conference Secretary and Convener

Arun Sharma	IGDTUW, India

Organizing Co-chairs

Syed Zulkarnain Idris	Universiti Malaysia Perlis, Malaysia
Ritu Rani	IGDTUW, India

Guest of Honour

Sachin Gulati	American Express, India

Chief Guest

N. Ranjana	Directorate of Systems and Technology Analysis, DRDO, India

Technical Program Committee Co-chairs

Odette Scharenborg	DUT, Netherlands
Chandra Sekhar	IISC Bengaluru, India
Milan Stehlik	Johannes Kepler University Linz, Austria

Shakir Ali	Imam University, KSA
Kirti Tyagi	Inha University in Tashkent, Uzbekistan

Publication Chair

Jasdeep Kaur Dhanoa	IGDTUW, India

Publication Co-chair

Deepak Kumar Sharma	IGDTUW, India

Keynote Speakers

Yudong Zhang	University of Leicester, UK
Priyankoo Sarmah	IIT Guwahati, India
Sardar M. N. Islam	Victoria University, Australia
Helmy Bin Abd Wahab	University Tun Hussein Onn Malaysia, Malaysia
Meena Jha	Central Queensland University, Australia

Technical Program Committee

Yudong Zhang	University of Leicester, UK
Sardar M. N. Islam	Victoria University, Australia
Meena Jha	Central Queensland University, Australia
Nick Campbell	Trinity College Dublin, Ireland
P. C. Ching	Chinese University of Hong Kong, China
Hiroya Fujisaki	Tokyo University, Japan
R. K. Singh	IGDTUW, India
Bin Ma	Institute for Infocomm Research, Singapore
Hsin-Min Wang	IIS, Academia Sinica, Taiwan
Ashish Seth	Inha University in Tashkent, Uzbekistan
Mirna Adriani	Universitas Indonesia, Indonesia
Syed A. Hossain	University of Liberal Arts, Bangladesh
Laurent Besacier	Naver Labs Europe, France
Vinod Shukla	Amity University Dubai, UAE
Ravinder Kumar	Shri Vishwakarma Skill University, India
Sunita Arora	CDAC, India
Ashwani Kumar	IGDTUW, India
Anu Khosla	SAG DRDO, India

S. K. Shrivastava	TDIL, DRDO, India
S. K. Jain	CFSL, India
Anuj Sharma	Panjab University, India
Deepak Garg	Bennett University, India
Mani Madhukar	IBM India Pvt. Ltd., India
Pramod Yadav	NIT Srinagar, India
Kamini Malhotra	SAG, DRDO, India
Ankush Jain	NSUT, India
Rajesh Kumar	Thapar University, India
Rajesh Tyagi	SRM University NCR, India
Shraddha Sagar	Galgotias University, India
Arun Kumar	IIIT Delhi, India
P. K. Saxena	DRDO, India
Sunil Tyagi	Directorate of Naval Design (SDG), IHQ MOD (Navy) Delhi, India
Shakir Ali	Imam University, KSA
P. K. Dass	IIT Guwahati, India
Omar Farooq	AMU, India
Arun Kumar	Jubail University College, KSA
Anup Girdhar	Sedulity Solutions, India
Vishal Bhatnagar	NSUT, India
Geeta Sharma	Mewar University, India
Sakriani Sakti	NAIST, Japan
Jianhua Tao	Chinese Academy of Sciences, China
Chiu-yu Tseng	Academia Sinica, Taiwan
Alexey Karpov	SPIIRAS, Russia
Sebastian Stüker	KIT, Germany
Eric Castelli	MICA, Vietnam
Etienne Barnard	NWU, South Africa
Claudia Soria	ILC-CNR, Italy
Lori Lamel	LIMSI, France
Emmanuel Dupoux	EHESS-ENS, France
P. S. Grover	GGSIP University, India
Ashish Khanna	MAIT, India
Sangeeta Arora	AKTU, India
Selvamani K.	Anna University, India
Senthil P.	SRM University, India
Sanjay Singla	Chandigarh University, India
Purushottam Sharma	Amity University, India
Gagan Tiwari	Chandigarh University, India
Laxmi Ahuja	AUUP, India
Adesh Pandey	KIET, Ghaziabad, India

Gianluca Valentino	University of Malta, Malta
Nisheeth Joshi	Banasthali University, India
Pratishtha Mathur	Sikkim Manipal University, India
Pradeep Sharma	MANUU, India
Mona Bharadwaj	IBM (I) Pvt. Ltd., India
Khundmir Syed	IBM (I) Pvt. Ltd., India
Kranti Athalye	IBM (I) Pvt. Ltd., India
Shikha Maheshwari	IBM (I) Pvt. Ltd., (INDIA), India
Ruchika Gupta	IBM (I) Pvt. Ltd., (INDIA), India
Vipin Tyagi	JUET, India
Somnath Chandra	MeitY, India
O. P. Sangwan	GJU Hisar, India
Kanwalinderjeet Gagneja	Florida State University, USA
Sandeep Chauhan	Bank of America, India
Vijai Kumar	Ciena, India
Mayank Singh	JSS, India
Mamta Mittal	Delhi Skill and Entrepreneurship University, India
Sushil Yadav	Delhi Skill and Entrepreneurship University, India
Mansaf Alam	Jamia Millia Islamia, India
Monika Mehrotra	Jamia Millia Islamia, India
Iftekhar Raza Khan	Jamia Hamdard University, India
D. P. Mohapatra	NIT Rourkela, India
P. K. Behera	Utkal University, India
Ajay Rana	Amity University, India
Vineet Kansal	AKTU, India
Amit Prakash Singh	GGSIP University, India
P. K. Yadav	NIT Jammu, India
Kumkum Garg	IIT Roorkee, India
Durga Toshniwal	IIT Roorkee, India
Rajesh Bhatia	PEC University, India
Bindu Narayanm	EnY, India
Pooja	NIET, India
Sanjeev Sofat	PEC University, India
Suneet Gupta	Bennett University, India
Anubha Gupta	IIIT Delhi, India
Anand Nayyar	Duy Tan University, Vietnam
Sourabh Bharti	Munster Technological University, Ireland
Aruna Bhatt	DTU, India
Neha Bansal	British Council, India
Neha Gupta	VIPS, India
Arun Garg	Ramanujan College, India
Garima Jaiswal	Bennett University, India

Rashmi Gandhi	Amity University, India
Chandra Prakash	NIT Delhi, India

Local Organizing Committee

Preeti Chauhan	IGDTUW, India
Nonita Sharma	IGDTUW, India
Shalini Arora	IGDTUW, India
Ranu Gadi	IGDTUW, India
N. R. Chauhan	IGDTUW, India
A. K. Mohapatra	IGDTUW, India
Khyati Ahlawat	IGDTUW, India
Mohona Ghosh	IGDTUW, India
Bhawna Narwal	IGDTUW, India
Vibha	IGDTUW, India
Najme Zehra Naqvi	IGDTUW, India
Karuna Kadian	IGDTUW, India
Bhawna Jain	IGDTUW, India
Vijay Kumar Yadav	IGDTUW, India
Rahul Sachdeva	IGDTUW Anveshan Foundation, India
Deepti Aggarwal	IGDTUW, India
Manushree Gupta	IGDTUW, India
Monalisa	IGDTUW, India

Sponsors

सत्यमेव जयते
Department of Science and Technology
Ministry of Science and Technology
Government of India

Department of Science and Technology, Government of India

AMERICAN EXPRESS

American Express



IGDTUW Anveshan Foundation, India



AI Club, IGDTUW, India

Contents – Part II

Recent Trends in Machine Learning and Deep Learning

Machine Learning Usage for Knee Condition: Osteoarthritis and TKA, Using Image Processing Among Medical Patients 3
Snehal Moghe and Harsh Pratap Singh

Machine Learning Techniques to Study: Medical and Biomedical Signal Analysis 18
Laxmi Ahuja and Ayush Thakur

N-BEATS & Temporal Fusion Transformer Based Surface Temperature Prediction and Forecasting for Realizing Global Warming Trends 30
Adria Binte Habib, Faisal Bin Ashraf, Muhammad Iqbal Hossain, and Golam Rabiul Alam

Tuberculosis Detection and Classification in Chest X-ray Images Using Optimized CNN Architecture 42
Apeksha Koul, Rajesh K. Bawa, and Yogesh Kumar

Performance Evaluation of Machine Learning Algorithms for Predicting Solar Radiation in Chennai Province 56
A. Geethakarthi and V. P. Sumathi

Fault Predictions Based on Base Learnings and Mean Weighted Score Using Machine Learning Techniques 65
Prachi Sasankar and Gopal Sakarkar

A Journey Through Multifaceted Data in Machine Learning Predictions on Financial Viability 85
R. Revathi

Enhancing Mint Plant Disease Detection Accuracy Through Deep Reinforcement Learning with YOLO Algorithm 104
V. Manikandan, P. Dinesh, M. Chantilyan, and C. Gomathi

Neural Network Based Employability Index for IT Graduates 118
Ankita Chopra and Madan Lal Saini

Analysis using Hybrid technologies with Artificial Intelligence

Ensuring Trustworthy Elections Using IoT-Enabled Blockchain EVM Voting Mechanism with Aadhaar Card- Based Face Verification 131
Omprakash A. Jaisinghani, P. L. Ramteke, and Bharat S. Dhak

Data Mining Based Heart Disease Prediction Using Hybrid Optimization Technique of Derived Features 143
Kanchan A. Khedikar and Piyush Kumar Pareek

Efficient Detection and Multi-level Classification of Tomato Plant Leaves Using Fused Deep and Hand-Crafted Features and LSTM-ANFIS 157
Astha Sharma and Ashwni Kumar

Unveiling the Enigma: Sudden Cardiac Arrest Among Youth and Uncovering Underlying Causes: A Data Analysis Study 167
Mohit M. Rao and Virendra Kumar Shrivastava

Using Encoder Models to Predict Ratings Using Flipkart Reviews 179
Himani Agarwal and Shweta Singhal

Diagnosis of Alzheimer Disease Using Machine Learning Algorithms 191
Tanya Kumari, Ritika Kumari, and Poonam Bansal

Detecting Image-Based Fake News with Neural Sleuths 201
Nachiket Rathod and Prabhakar Ramteke

Background Subtraction Model with Advance GMM in Dynamic Background 212
Manisha, Avadhesh Kumar, and Dileep Kumar Yadav

Video Summarization Using Firefly Algorithm 223
Shashank Pandey, Sakshi Dwivedi, Vijay Bhan Singh, Neetu Verma, and Ranvijay

Exploring New Horizons in Computer Vision Research

An Image Processing-Based Tomato Leaf Disease Prediction Using Deep CNN 237
Kesana Vanisri, K. Srujan Raju, and Bagam Laxmaiah

Investigation on Existing Blockchain Based Architecture, AI/ML Driven for Boosting IoT Security and Privacy 246
Abdul Moid and Narendra Sharma

Machine Learning-Based Estimation of Caloric Expenditure 262
Shivani Trivedi, Jahanvi Gupta, and Himanshu Singh

FarmScan: Smart Rotten Fruit and Vegetable Detector 272
Shivani Trivedi, Ankit Yadav, Suraj Kumar Gupta, and Shivam Gupta

Closing the Data Divide in COVID-19 X-ray Datasets: Advancing Diagnosis Through Stable Diffusion-Based Synthetic Image Generation 281
Saumya Mehra and Bhawna Jain

Machine-Learning Algorithms for Predicting Cardiac Arrhythmias 292
Sakshi Shete, Suvarna Udgire, Anuradha Joshi, Shravani Desai, and Ronit Murpani

Natural Language Generation (NLG) with Reinforcement Learning (RL) 303
Shakir Khan, Mohammad Serajuddin, Zafarul Hasan, Syed Ahad Murtaza Alvi, Rashid Ayub, and Arun Sharma

i-BMD: AI-Based Opportunistic Screening for Osteoporosis on Abdominal CT Using Deep Learning 319
Soumya Jindal and Meemansa Jindal

Machine Learning in IoT: An In-Depth Dataset Analysis Based on Attack Detection 332
Kajal Tyagi, Anil Ahlawat, and Himanshi Chaudhary

Applications of Machine Learning and Deep Learning

Investigation of Satellite Data for Monitoring Air Quality Over Remote Sensing Technology 351
Panimalar Kathiroli, V. Vijayalakshmi, Piyush Gopal, and Sivananda Panda

Analysing Social Media Data for Emotion Detection and Sentiment Analysis 367
Surbhi Bharti, Divya Verma, Anamika Kumari, and Ashwni Kumar

A Comprehensive Approach to AI-Enabled Diagnosis of Lung Diseases: Utilizing Pretrained Models for Fibrosis, Tuberculosis, and Covid-19 Detection 380
Sapna Yadav, Syed Bilal Abbas Rizvi, and Syed Afzal Murtaza Rizvi

Campus Exploration: Figma-Driven UI Design for a Campus Tour App 393
Surbhi Bharti, Medha Arora, and Ashwni Kumar

Predicting Startup Exit Strategies with Ensemble Machine Learning 407
M. Deekshitha Reddy, Geetika Vadali, Garima Jaiswal, Ritu Rani, and Amita Dev

An Inclusive Analysis on Deep Learning Hinged Malware Detection Techniques 417
Vinisha Sumra, Naveen Malik, and Santosh Kumar

Advances in Text Summarization Techniques: A Comprehensive Review and Future Prospects 426
Harsh Vishwakarma, Yashika Kukkar, Aditi Chauhan, Abhinav Maheshwari, Dharmender Saini, and Preeti Nagrath

Multiple Object Detection and Tracking Using DeepSORT 438
Rashmi Gandhi, Yash Pasrija, Amisha Chauhan, and Ritu Rani

An Ensembled Travel Recommender Model Using Autoencoder and Machine Learning Algorithms 449
V. K. Muneer

Future-Proofing Careers: The AI Era and the Imperative of Reskilling 460
Shristi Agarwal, Garima Jaiswal, Mala Saraswat, and Drishti Rai

Author Index 471

Contents – Part I

Trends and Applications in Speech Processing

An Efficient Deep Learning Based Seq2Seq Model for Abstractive Text Summarization 3
R. Deepa, T. Sree Sharmila, and R. Niruban

Text Scribe: Unveiling New Dimensions in Text Summarization 20
Dhananjay Bhagat, Kalyani Sawarkar, Dewanshu Barde, Rashi Channawar, Mayur Talmale, Prathamesh kale, and Rahul Agraval

GRUbBD-SM: Gated Recurrent Unit Based Bot Detection on Social Media 40
Akash Shah, Sapna Varshney, and Monica Mehrotra

Scaling Language Boundaries: A Comparative Analysis of Multilingual Question-Answering Capabilities in Large Language Models 55
Apoorva Tiwari, Karuna Kadian, Nikita Garg, Sukhandeep Kaur, and Vimal Dwivedi

Performing Text Segmentation to Improve OCR on Multi Scene Text 66
Arrun Sivasubramanian, Sheel Shah, Akash Narayanaswamy, C. Rindhya, and H. B. Barathi Ganesh

Deep Learning Based Multilingual Voice Recognition System and Analytics for Organization Surveys 78
A. S. Sri Saila, A. T. Venkata Subramani, M. D. Harsha Prada, and G. Madhu Priya

Speech Emotion Recognition Using Convolutional Neural Networks 90
Anunya Sharma, Kiran Malik, and Poonam Bansal

Hybrid Approach to the Personification of Dialogue Agents 102
Pavel Posokhov, Stepan Skrylnikov, Olesia Makhnytkina, and Yuri Matveev

MelSpectroNet: Enhancing Voice Authentication Security with AI-Based Siamese Model and Noise Reduction for Seamless User Experience 116
Gitesh Kambli, Jay Oza, and Amit Maity

Recent Trends in Speech and NLP

A Framework for Information Retrieval Using Domain Specific Dictionary: Illustration Through Enhancing the Intelligence Cycle 131
Lakshmish Nayak, Bharathi Ramudu, Malay Kumar Nema, and Rituraj Kumar

Text-Independent Voiceprints Identification Using Feed-Forward Back-Propagation with Layered Strategies 143
Pooja Gambhir, Amita Dev, and Poonam Bansal

Email Bot-Voice Based Email System for Blind 159
Deepanshi Jalan, Anandita Tuli, Shradha Jain, Ashwani Kumar, and Vaishali Kikan

Empowering Hate Speech Detection: A Comparative Exploration of Deep Learning Models 171
Shivani Trivedi, Sugandh Rastogi, Sneha Agrawal, and Ravish sharma

Hate Speech Detection Using Glove and BERT 184
Mayank Singhal, Komal, Mohammad Zeeshan, Ishika Saini, and Preeti Nagrath

Santali Vowel Recognition: An Under-Resourced Tribal Language 195
Sandip Jana, Joyanta Basu, Tapan Kumar Basu, and Amiya Karmakar

Revolutionizing Writing: Personalized Neural Classifier for Handwritten Text 210
Nimish Goel, Shushil Paudel, Srushti Jagtap, and Ishan G. S. Kakodkar

Hindi Speech Recognition Using Deep Learning: A Review 227
Nidhi Bhatt, Shobha Bhatt, and Geetanjali Garg

A Comprehensive Review of Instructional Tools and Applications for Dyslexic Learners 238
Gokul Krishnan, Dhruthi Rao, Anshu Sandur, Ananya Jandhyala, and Pooja Agarwal

Emerging Trends in Speech Processing

Deep Learning-Based Speech Recognition Models: Review 253
Divya Taneja and Shobha Bhatt

Analysis of Acoustic Features for Gender Identification Using Punjabi Speech Dataset 272
Puneet Mittal, Sukhwinder Sharma, and Khyati Marwah

Context Based Anaphora Resolution of English Discourses Using Rule Based Approach 284
Kalpana B. Khandale and C. Namrata Mahender

Enhancing Named Entity Recognition with DistilBERT and Attention Ensemble Fusion 301
Ankit Kumar, Richa Sharma, and Punam Bedi

Explaining Spectrograms in Machine Learning: A Study on Neural Networks for Speech Classification 311
Jesin James, B. T. Balamurali, Binu Abeysinghe, and Junchen Liu

Speech Recognition Using Adaptation of Whisper Models 323
Vartika Tyagi, Amita Dev, and Poonam Bansal

Voice Stress Analysis Using Machine Learning 335
Surbhi Bharti, Prerna Jha, Sakshi Sahu, Sheba Alam, and Ashwni Kumar

Efficient Real-Time Indian Sign Language Fingerspelling Recognition in Natural Settings Using Heuristics 346
T. Raghuveera, V. K. Akshayalakshmi, B. A. Nisha, and K. S. Easwarakumar

Advances in Computational Linguistics and NLP

A Novel and Intelligent Approach for Indian Locale Based Text-to-Speech Model by Hybridizing Wave Net and Wave Glow with Mel-Spectrogram Analysis 365
R. C. Barik, Tanuj Kumar, Anunay Sinha, and Vishwajeet Kumar

Optimization of Indian Sign Language Detection Using Data Generators 379
Kata Venkata Sai Krishna, Goli Abhilash, Devanjali Relan, and Kiran Khatter

A Comparative Analysis of Deep Learning Architecture for Accurate Gender Classification Using Vocal Data 394
Khushi Anand, Bhawna Jain, Ananya Verma, Anushka Gupta, and Niharika Chhabra

A Comprehensive Analysis on Kaldi-Based Speech Recognition for Low Resource Indian Languages 405
Suhani, Amita Dev, and Poonam Bansal

Comparative Analysis of Deep Learning Models for Text Summarization on Hindi Corpus 418
Niranjana Sowpari, Poonam Bansal, and Nandini Sethi

Tune into Your Feelings: NLP-Powered Emotion Driven Music Recommender System 426
Subhranshu Behura, Arham Alam, Nishtha Phutela, Atul Mishra, and Goldie Gabrani

Recent Trends in Text to Speech Synthesis in Context with Indian Languages 440
Manisha Gupta, Amita Dev, and Poonam Bansal

Advanced Speech Emotion Recognition in Malayalam Accented Speech: Analyzing Unsupervised and Supervised Approaches 451
Rizwana Kallooravi Thandil, P. V. Jalala, Rahbar Zahid, and M. Preethi

Author Index 465

Recent Trends in Machine Learning and Deep Learning

Machine Learning Usage for Knee Condition: Osteoarthritis and TKA, Using Image Processing Among Medical Patients

Snehal Moghe[1(✉)] and Harsh Pratap Singh[2]

[1] Medi-Caps University, Indore A B Road Rau-Pigdamber Indore, Indore, (MP) 453331, India
snehalmindore@gmail.com

[2] Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore – Ujjain Road, Indore, (MP) 453111, India

Abstract. Artificial Intelligence is gaining grounds in not only technical areas of applications like IT Industry, Management areas like Logistics Management, Human Resources, Fintech, but also, Biomedical, Healthcare and Hospitality services. To the extent that the patients are getting automated phone calls for doctor appointments, medicine refilling, body check-up routine, etc. Also, huge amounts of data which is generated in forms of X-Ray images, CT scans, MRIs, Sonography, Blood Test reports, Urine Reports, etc. are all used for healthcare analytics and predicting future behavior of individual patients visiting the Hospital concerned. Moreover, AI has also entered the Operation Theatre (OT) to assist doctors, best example is Total Knee Arthroplasty, where robotic arm is programmed to perform robotic surgery on knee and do knee replacement. In this paper we intend to review systems where we have AI based model (on human body parameters) compared with doctor's analysis on the same parameters. We then formulate the results and come up with concrete solution for the problem. Literature review has clearly brought down that there are different aspects of Healthcare and AI which is a very vast domain in itself, but narrowing it to Knee Operations and Osteoarthritis, use of AI as power of Prediction, it is quite a game changer worth the efforts as it has gained much importance technically both in medical as well as computer science field.

Keywords: Artificial Intelligence · Medical Data · Healthcare Analytics · Machine Learning · Deep Learning · Medical Reports · Doctors · Big Data · Knee Operations · Image Processing · Sensors · Robotics

1 Introduction

Healthcare and Artificial Intelligence (AI) are not new terms to be pronounced together, although there are new benefits emerging out, altogether. Artificial Intelligence can be defined in terms of its Inclusiveness, Precision, comprehensiveness, Practicability and Permanence. If we talk of the Turing Test, all goals of AI are over inclusive, however, single definition of AI can't overcome all regulatory goals. It has been generally argued that Precision is not achievable by single definition of AI, though this is debatable and Turing

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 3–17, 2025.
https://doi.org/10.1007/978-3-031-75167-7_1

Test is more precise, nature of test is empirical. Comprehensiveness is also debatable, because single definition can't accommodate all definitions. Practicability of Turing test is highly vague and unpredictable because it takes a lot of time in experimentation to predict the right output. Permanence means as to how much self-sufficient the definition is. Turing Test has been valid for the last 70+ years [1].

1.1 Healthcare Analytics

Healthcare Industry is the umbrella term covering Hospitals, Doctors, patients, nurses, health records, Healthcare Analytics and much more. Definition of Healthcare Industry is precisely efforts made to sustain health physically, mentally and/or emotionally by well trained and licensed professionals. Healthcare Analytics is about measuring all statistics of human body data, including reports analysis, then using Artificial Intelligence as a tool to gauge whether the body model formed is upto mark or not or it requires changes/ modifications to be adjusted to. Whether the simulations generated are realistic, believable and ethical or not.

This scenario will be taken into account throughout the article, as Total Knee Arthroplasty (TKA) is currently performed by robotics as well, in addition to the requirements and care provided by traditional orthopedic surgeons. A degenerative joint condition known as osteoarthritis affects millions of people worldwide. Altosteoarthritis cannot be cured, there are numerous therapies and management techniques that can help reduce symptoms and enhance quality of life. The monitoring and treatment of osteoarthritis may be significantly impacted by artificial intelligence (AI).

1.2 Medical Data

This information includes blood reports, hemoglobin levels, urine reports, CBCs, ESRs, and fasting and post-lunch blood sugar readings. The usual recommended testing before TKA include blood group and Rh factor, kidney and liver function tests, tests for HIV, Australia Antigen, tests to rule out Hepatitis C, and routine examinations of stool and urine.

1.3 Big Data

Big data is defined as handling large amount of data that can't be stored, processed by normal computation techniques of single processing system. Hence we need Big Data techniques to perform similar operations [3].

The five Vs of Big Data are (Fig. 1):

There are so many (in Zeta Bytes and more) structured, unstructured data we keep in medical records of patients in one city itself. Hence there is growing necessity to store and use this kind of data.

In this paper, we are discussing about how Big Data is being used for this purpose and satiety of data is being maintained by preserving it in different places/ server on various locations.

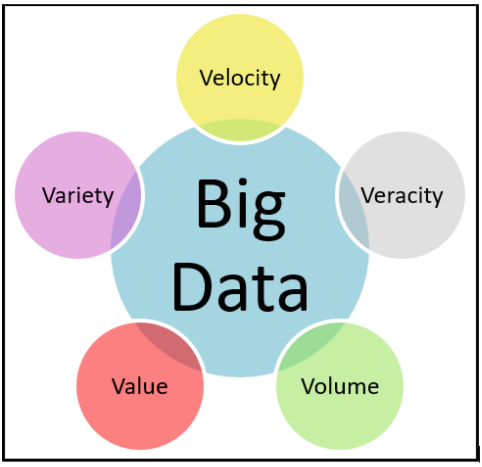


Fig. 1. Big Data – Its characteristics

1.4 Robotics

Robotics is concerned with the creation, maintenance, use, and application of robots [4]. The study of creating efficient and intelligent robots is central to the discipline of robotics in artificial intelligence [5].

1.4.1 Robotic Arm

Robotic Arm is a man-made plastic arm with several optical nerves and joints designed to imitate the movements of the human arm [6]. It consists of a number of links attached by joints that allow for either rotational motion or linear displacement. The robotic arm is a multifunctional instrument that may be applied to a variety of tasks, such as manufacturing, assembly, and material handling [7].

Robotic arms are used for a number of tasks, such as welding, painting, machining, assembly, and testing, in a wide range of industries, such as the automotive, aerospace, medical, and others.

In this paper, robotic arm is being studied to cut knee joint precisely to the degree needed and angle specific to each and every patient's knee requirement. Also, this robotic arm is programmed to do so by the senior orthopedic surgeon. There are sensors placed on the knee and the kind of/density of bone is estimated. Then the replacement is made.

Robotics is concerned with the creation, maintenance, use, and application of robots [4]. The study of creating efficient and intelligent robots is central to the discipline of robotics in artificial intelligence [5] (Fig. 2).

Robotic Arm (Fig. 3):

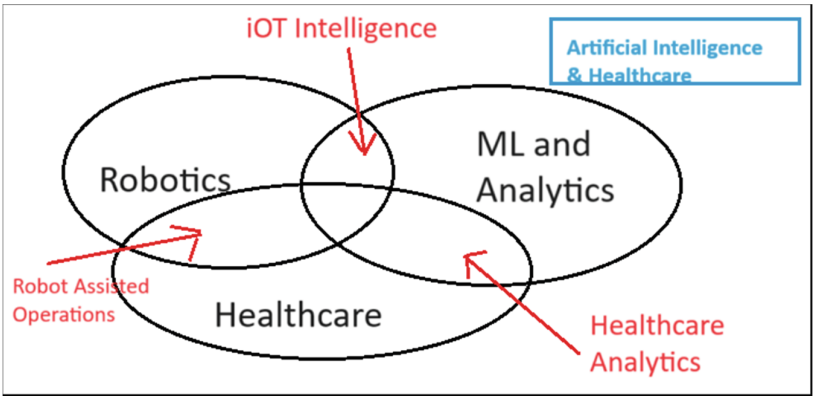


Fig. 2. Importance and Emergence of relevant fields of AI and Healthcare



Fig. 3. Robotic Arm for multi-purpose Usages [8]

2 Literature Review

Ref. No	Author, Year	Title of the paper	Findings
[9]	Gumbs et al., 2022	Artificial Intelligence Surgery publishing	Definitions of terminologies and working models are explained in this paper. This
			paper has various important terms including AI, ML their applications, importance and where they are lacking

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[10]	Kalis Brian, et al., 2018	Ten promising AI applications in Healthcare Publications HBS	Various terminologies like Virtual Nursing Assistants, Fraud Detection, Administration etc. are formed in table as important components of AI. This paper shows 10 emerging areas of AI, where strong research prospects are there and already being explored in healthcare
[11]	Longoni Chiara, et al.	If AI can perform better than doctors, why aren't patients trusting it	This paper classifies AI as better judge of interpreting medical information and image processing as compared to Medical Doctors, upto 92% accurately. Doctors trust AI, Patients trust AI only till beginners level. This paper explores possibilities existing and upcoming in healthcare sector for AI, and how much they will be entertained by patients. When given proper logic and explained thoroughly, patients also trust AI
[12]	Garbuio Massimo, et al.	Emerging Business Models: Artificial Intelligence as a Growth Engine for Healthcare Startups	There are three kinds of AI, assisted AI, augmented AI and automated AI. Out of these, Automated AI is the most powerful and most sought for, in today's world. AI can disrupt the healthcare sector by bringing innovation and having effective strategies to do so

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[13]	Batailler Cecile et al., 2022	Artificial Intelligence in knee arthroplasty: current concept of the available clinical applications	There is a growing need for Robotic assisted surgery, specifically TKA and reducing the churn rate of customers, currently it is 20%.As compared to traditional TKA, Robotic TKA is much beneficial because it is less painful, less blood loss and lesser recovery time, lesser incision
[14]	Roemer Frank W., et al., 2020	State of Art: Imaging of Osteoarthritis—Revisited	The ability to articulate abstract knee data levels is helpful. adding representational layers in ascending order. It is used for disease detection and characterization, structural joint pathology diagnosis, and imaging of tissue segmentation. Deep learning has a number of advantages, including reduced computation costs, efficiency, A deep learning approach has also been demonstrated for fully automated segmentation of all MRI-detected knee joint structures, such as cartilage, bone, tendon, meniscus, muscle, infrapatellar fat pad, joint effusion, and Baker cyst. As previously said, technology can be used to efficiently treat osteoarthritis by employing image processing and sensors to assess the condition of the knee, its bone density, the ideal shape, and other aspects

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[15]	Bergerson Christie, et al., 2021	Overview of Robotic-Assisted Knee Arthroplasty	Principal Component Analysis and Feature Extraction Methods, which are useful in taking out relevant data from a huge dataset and keeping it intact for human Expert/Doctor to interpret it as it is and give diagnosis and same data will be used by machines for automated analysis. ML has several algorithms, including PCA and FE which have interfaces to make interpretations easier for humans to perform
[16]	List, Jelle der Van, et al. 2017	Robotic assisted surgery, the future of knee replacements, robotic-assisted knee arthroplasty	The alignment of the lower leg, the balance of the soft tissues, the preservation of the joint line, and the size, shape, and fixation techniques of the tibial and femoral components are some of the external factors that affect the success rate of knee arthroplasty. According to the article, when robotically assisted surgery is performed, these factors are handled by the robotic arm or system itself, which results in a slight increase in success rates. Robotic surgery boosts surgical success rates, and it has been determined that these statements are true after careful analysis of several factors

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[17]	Banerjee Samik, et al., 2016	Utilisation of Artificial Intelligence in Practise "Deep convolutional neural network for segmentation of knee joint anatomy"	Authors reveal their insights about robotic surgeries on knee and find out that this kind of surgeries are more helpful, efficient and patient survive more, as compared to traditional Arthroplasty. Relatable to previous findings, this paper also states that lesser blood loss, lesser incision and lesser time in hospital in there
[18]	Sheth, Parul, Dr.,	A machine learning method based on MRI-based radiomic characteristics from the tibial bone to differentiate between knees with and without osteoarthritis	authors emphasis on benefits of robot assisted surgeries like smaller incisions, lesser pain, fewer hospital days, lesser recovery time, lesser blood loss and overall cost effective. Same as previous paper
[19]	Shaohui Wang, et al. (2021)	A Review of Manual Imaging Grading and Machine Learning Approaches for the Identification of Knee Osteoarthritis Imaging Features for Diagnosis and Prognosis	The use of AI along with cloud and Big Data will help scientists and general users also, to get world class solutions to big problems while addressing rheumatoid arthritis. More accurate solutions to pharmacological problems can be made using AI + Big Data and cloud of course, for treatment of Rheumatoid Arthritis
[20]	Zhaoye Zhou, et al. (2018)	Deep Machine Learning Model to Predict Knee Osteoarthritis	Use of CNN segmentation method to classify images of femur and tebula properly and then performing operation. It adds credibility to the proposed methodology that this technology is useful

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[21]	HirvasniemiJukka et al. (2021)	A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone	Machine Learning is helpful in predicting upto 80% accuracy classifying images with radiomic features, three types of ML were used, Covariate, image features and combination of two above models
[22]	TeohXinYunet al. (2022)	Discovering Knee Osteoarthritis Imaging Features for Diagnosis and Prognosis: Review of Manual Imaging Grading and Machine Learning Approaches	Machine learning is primarily categorised in three useful tasks, image classification and segmentation, prediction of knee OA disease
[23]	Mohamed Yacin Sikkandar1,* et al. (2020)	Automatic Identification and Categorization of Human Knee Osteoarthritis Using Convolutional Neural Networks Changing Cartilage Volume Using Baseline Bone Curvature Over Time	Sensors and ML are used in classifying and segmenting data, also involving LCM based segmentation (Local Centre of Mass) and CNN based segmentation (Convolutional Neural Network) had accuracy of 92%
[24]	Hossein Bonakdari, et al. (2022)	A Machine Learning Model to Predict Knee Osteoarthritis Cartilage Volume Changes over Time Using Baseline Bone Curvature	using ML, 5 algorithms to predict which one's best and gender detection. Prevent Cartilage loss, blood loss, etc.
[25]	Wei Li, Sheng-Ming Xu et al. (2023)	The present role of the virtual components of artificial intelligence in total knee arthroplasty;	this paper recognizes various kinds of applications of AI onto oesteoarthritis, focusses mainly on literature review. Also gives historical perspective of AI, its beginning and advances in Healthcare. It also talks about robotics, robotic arm, how it works and assists Doctors, and other similar devices

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[26]	E Carlos Rodríguez-Merchán (2022)	The current role of the Virtual Elements for AI in TKA	this research paper focusses on use of Machine Learning (ML) and Deep Learning (DL) both, ML for predicting various components like length of stay at hospital, costs incurred in TKA, size of TK components, etc. DL can be used in distinguishing different types of arthroplasty and prosthetic loosening
[27]	Michel Bonnin, MD, PhD (2022)	Artificial Intelligence Radiographic Analysis Tool for Total Knee Arthroplasty	This study primarily focussed on image processing for TKA. Around 39,189 images were obtained and trained for Convolutional Neural Networks and results obtained were of 0.98 area under curve and 5% accuracy of surgeons as well as 15% sensitivity
[13]	Batailler Cecile et al., 2022	Artificial Intelligence in knee arthroplasty: current concept of the available clinical applications	Overall review of 742 research papers, out of which 543 were excluded, due to various reasons, the considered ones were further shortlisted and machine learning papers were considered finally for review, for total knee arthroplasty
[28]	Xiaoxiao Xie, et al. (2021)	Deep Learning-Based MRI in Diagnosis of Fracture of Tibial Plateau Combined with Meniscus Injury	there are three kinds of knee injuries, injuries in meniscus, or femurial or tibiscus their sensitivity and accuracy were measured and reported. At the end it was reported that optical generated MRI images were better

(*continued*)

(*continued*)

Ref. No	Author, Year	Title of the paper	Findings
[29]	Xiao-GuangHan, WeiTian (2019)	Artificial Intelligence in Total Knee Arthroplasty: Current State and Future Perspectives	there are manifold applications of AI in healthcare, including from knee arthroplasty to hip osteoarthritis, lower back pain to hand/limb operation. It reduces blood loss, time to recover, etc
[30]	Mahmoud Khadija et al. (2022)	Predicting total knee replacement at 2 and 5 years in osteoarthritis patients using machine learning	it not only adds to existing study whether TKA is beneficial or not, but what kind of lifestyle changes one should do in order to survive the TKA and successfully live, happily. Also, TKR was done using ML, 6 ML models have been compared and worked upon
[31]	BatkoKornelia et al. (2022)	The use of Big Data Analytics in healthcare	This paper teaches use of big data analytics, use of structured, semi-structured and unstructured data, for Healthcare Analytical Maturity Model And the authors are testing results via conventional modelling vs big data analytics using machine learning as well
[32]	DashSabyasachi et al. (2019)	Big data in healthcare: management, analysis and future prospects	Genomics data is the context for research in Big data, Healthcare Analytics (diagram below) in this paper. Also, there is research on IoT, sensors and quantum computing in healthcare. Lots of aspects are covered in this paper, which are technomanagerial as well

3 Proposed Methodology

With reference to the literature review, efforts made by various learned authors in recent publications, from 2015 to 2023, considering various aspects of knee conditions, age of the patient, lifestyle, work-life balance, hereditary diseases, previous operations, various body reports including blood, urine reports, X Ray, MRI, CT Scan, PET Scan, etc. The purpose is to predict Knee Operation by building Artificial Intelligence system, Machine Learning and Image Processing specifically. As there are many predictive models which are built on Image Processing techniques, predicting whether there is TKA (Total Knee Arthritis) or OA (OsteoArthritis) and there is Kellgren Lawrence Algorithm for categorization of TKA (Grade 0 to IV = No possibility to 100% Probability of TKA). So the model will be comprehensive of techniques including Machine Learning like Random Forest, SVM and Deep Learning techniques like CNN, Adaboost, etc.

The diagram below shows a model with leftmost component as various user support, clinical decision system, software engineers, clinical advisors and lastly software quality assurance. Reports are of various categories, X-Ray, Blood, Urine, ECG, MRI, PET, etc. Patient Knee Operation prediction is at right side and Patient is on mid-left side. AI is doing live monitoring along with processing reports and numbers by the proposed methodology and would generate results, to be cross-checked.

Also, there's scope of robotic operation, as and when there's need of knee operation. It is shown in the diagram and is also known as Assistive AI. This surgery should be supported by afterwork like, post surgery care, changed healthy lifestyle, food habits, exercise routine, etc. (Fig. 4).

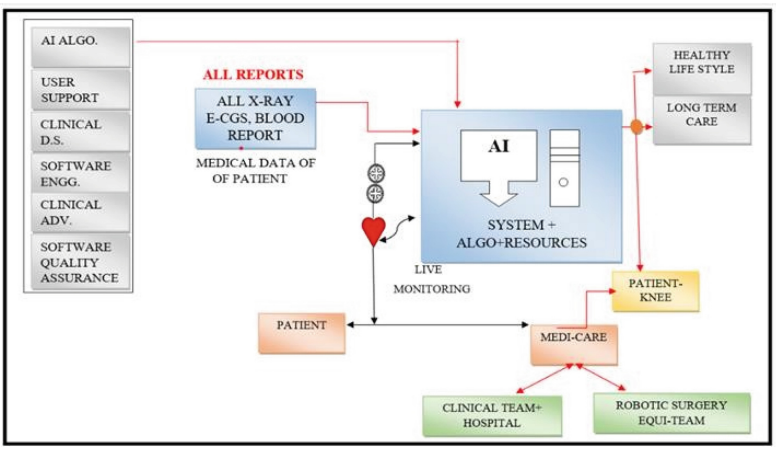


Fig. 4. Proposed System of AI and Medicia

4 Conclusion

The paper redefines the boundaries of technology and medicine amalgamation, for the better future. Also, no doubt there is indeed transparency in usage of technology whichever it is, whether applied by doctors/ assisted by robot. It is an interesting aspect, whether patients will get to be operated totally by robots end to end, that's the ideality, as of now, but not a far fetched scenario. Also, predicting how the patients would be getting ill and would be treated. In generalizing the case of Knee Operations, the various aspects like age, patient body maintenance, body statistics, previous reports/ medical history, family's medical history, these aspects are to be studied and tried to build a relationship upon. Artificial Intelligence and Big Data will play a significant role in determining hence the contributing factors. Some studies say it is that robot assisted surgeries has reduced churn rate by 20% and that patients are believing technological aid is for their proper, better treatment and related aspects. Deep learning technology (AI) also is fully automated version of curing patients, where the knee curing robotic machine will automatically collect all data once knee fitted onto it like knee joint structures at MRI, including cartilage, bone, tendon, etc, Also, the study of how lifestyle, eating habits affect the overall being of patient.

Lastly and most importantly, Robot Assisted Surgery is having capability of lesser blood loss of patients which is critical issue in older age ones and generalization of this research can be useful for Techno-Medical domain/Fraternity, which means Industries like Healthcare and Software Giants as well.

5 Limitations

With reference to the literature review and technologies used, we can see that, there's good leveraging of technology for health purposes. Total Knee Arthroplasty (TKA) and or Osteoarthritis is operable because of good, precise, sharp, robotic arm, under guidance of highly qualified surgeons. Hence there's nothing to worry about in terms of technology and wellness. Limitation of patients and this paper is the age factor, the older the patient is, more is the recovery time, also, whether the patients are leveraging technology properly or not. Also, TKA can add to mental disorders as well. Hence it is better that timely treatment of patients should begin and continue.

Also the stigma associated with technology and its precision/ wrong doings seen in various media ,etc which creates a negative impact overall, which affects their decision making while getting operated by the robots (if so), here. The preliminary limitation is lack of data and contextualization of the problem formalization, itself related to PAN Madhya Pradesh and then PAN India.

Total knee arthroplasty (TKA) with robotic assistance has a number of benefits, including enhanced implant insertion precision and perhaps improved clinical results. It does, however, have some restrictions. First off, healthcare organisations may find it difficult to afford purchasing and maintaining robotic devices, which could limit patient access to this technology. Additionally, adopting robotic aid might have a steep learning curve for surgeons, which could result in lengthier operating durations in the beginning. Additionally, the technology might not be appropriate for all patients, especially those

with severe deformities or complex anatomical differences, as preoperative planning and intraoperative adjustments can be difficult.

References

1. Schuett, J.: A legal definition of AI. Social Science Research Network (2019). https://www.researchgate.net/publication/336198524_A_Legal_Definition_of_AI
2. http://www.cjthakkar.com/knee-before-surgery.php
3. https://www.javatpoint.com/big-data-characteristics
4. Travelyan James: Redefining Robotics for the New Millennium (1999)
5. Bhonsale Shankar Sachin: Research Paper on Robotics-New Era (2021)
6. https://history-computer.com/what-is-robotics-and-how-does-it-work/
7. Kumar, B.M., Kiran, D.: Des. Develop. Robot. Arm (2016). https://doi.org/10.1109/CIMCA.2016.8053274
8. Robotic Arm Figure 3: makeinbazar.com (2023)
9. Gumbs, et al.: Artificial Intelligence Surgery Publishing (2022)
10. Kalis, B., et al.: 10 Promising AI Applications in Healthcare. HBS Publication (2018). https://hbr.org/2018/05/10-promising-ai-applications-in-health-care.
11. Longoni, C., et al.: AI can Outperform Doctors, So Why don't Patients Trust It? HBS Publication (2019). https://hbr.org/2019/10/ai-can-outperform-doctors-so-why-dont-patients-trust-it
12. Garbuio, M., et al.: Artificial Intelligence as a Growth Engine for Healthcare Startups: Emerging Business Models, (2017)
13. Batailler, C., Shatrov, J., Sappey-Marinier, E., Servien, E., Parratte, S., Lustig, S.: Artificial intelligence in knee arthroplasty: current concept of the available clinical applications. Arthroplasty (London, England) **4**(1), 17 (2022). https://doi.org/10.1186/s42836-022-00119-6
14. Roemer Frank W., et al.: State of Art: Imaging of Osteoarthritis—Revisited, RSNA.ORG (2020)
15. Christie, B., et al.: Artificial Intelligence in Healthcare, Nova Publications (2021)
16. Jelle der Van, et al.: Robotic-assisted knee arthroplasty: an overview. Am. J. Orthopedics (2017)
17. Samik, B., et al,: Robotic-Assisted Knee Arthroplasty. Expert Review of Medical Devices (2016)
18. Sheth, P.: Robotic Assisted Surgery, The Future of Knee Replacement. Science Reporter.
19. Wang, S., et al.: Practical implementation of artificial intelligence-based deep learning and cloud computing on the application of traditional medicine and western medicine in the diagnosis and treatment of rheumatoid arthritis (2021)
20. Zhou, Z., Zhao, G., Kijowski, R., Liu, F.: Deep convolutional neural network for segmentation of knee joint anatomy. Magn. Reson. Med. **80**(6), 2759–2770 (2018). https://doi.org/10.1002/mrm.27229
21. Hirvasniemi, J., Klein, S., Bierma-Zeinstra, S., Vernooij, M.W., Schiphof, D., Oei, E.H.G.: A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone. Eur. Radiol. **31**(11), 8513–8521 (2021). https://doi.org/10.1007/s00330-021-07951-5Teoh
22. Teoh, Y.X., et al.: Discovering knee osteoarthritis imaging features for diagnosis and prognosis: review of manual imaging grading and machine learning approaches. J. Healthc. Eng. **2022**, 4138666 (2022). https://doi.org/10.1155/2022/4138666(RetractionpublishedJHealthcEng.2023Oct11;2023:9765742)

23. Sikkandar, M.Y., et al.: Automatic Detection and Classification of Human Knee Osteoarthritis Using Convolutional Neural Networks (2020)
24. Bonakdari, H., Pelletier, J.P., Abram, F., Martel-Pelletier, J.: A machine learning model to predict knee osteoarthritis cartilage volume changes over time using baseline bone curvature. Biomedicines **10**(6), 1247 (2022). https://doi.org/10.3390/biomedi-cines10061247
25. Li, W., Xu, S.-M., et al.: Research advances in the application of AI for preoperative measurements in total knee arthroplasty. Life **13**(2), 451 (2023). https://doi.org/10.3390/life13020451
26. Carlos Rodríguez-Merchán, E.: The current role of the virtual elements of artificial intelligence in total knee arthroplasty. EFORT Open Rev. **7**(7), 491–497 (2022). https://doi.org/10.1530/EOR-21-0107
27. Bonnin, M., Müller-Fouarge, F., Estienne, T., Bekadar, S., Pouchy, C., Selmi, T.A.S.: Artificial intelligence radiographic analysis tool for total knee arthroplasty. J. Arthroplasty **38**(7), S199-S207.e2 (2023). https://doi.org/10.1016/j.arth.2023.02.053
28. Xie, X., et al.: Deep learning-based MRI in diagnosis of fracture of Tibial plateau combined with meniscus injury. Sci. Programm. **2021**, 1–8 (2021). https://doi.org/10.1155/2021/9935910
29. Han, X.G., Tian, W.: Artificial intelligence in orthopedic surgery: current state and future perspective. Chin. Med. J. **132**(21), 2521–2523 (2019). https://doi.org/10.1097/CM9.0000000000000479
30. Khadija, M., et al.: ~~Predicting~~ Total Knee Replacement at 2 and 5 Years in Osteoarthritis Patients using Machine Learning. BMJ Surgery, Interventions Technol. (2022)
31. Batko, K., et al: The use of Big Data Analytics in healthcare. J. Big Data (2022)
32. Dash, S., et al.: Big data in healthcare: management, analysis and future prospects. J. Big Data (2019)

Machine Learning Techniques to Study: Medical and Biomedical Signal Analysis

Laxmi Ahuja and Ayush Thakur(✉)

Amity Institute of Information Technology, Amity University Noida, Noida, Uttar Pradesh, India
lahuja@amity.edu, ayush.th2002@gmail.com

Abstract. This paper presents a study of machine learning techniques for medical and biomedical signal analysis. Measurements of physiological processes that reveal the state of health of living things are known as medical and biomedical signals. These signals are difficult to analyse using traditional signal processing techniques because they are frequently influenced by noise, non-stationarity, complexity, and high dimensionality. Algorithms that use machine learning may learn from data and provide hypotheses or predictions based on the patterns they spot. This paper examines several common electrocardiograms and electroencephalograms, among other biological and medical signals, and demonstrates how machine learning techniques may be used to analyse them for a variety of purposes, including diagnosis, treatment, monitoring, and research.

Keywords: Medical signals · Biomedical signals · Machine learning · Signal processing · Healthcare applications

1 Introduction

Medical and biomedical signals are essentially recordings of physiological activities within living organisms [1]. Biological and medical signals are basically records of physiological events that occur within living bodies. Key indicators of an organism's overall health status, pivotal for the early identification of diseases or irregularities, are of paramount importance [2]. Electroencephalograms [1, 4], responsible for capturing brain electrical patterns, along with electrocardiograms [3], which monitor cardiac electrical patterns, exemplify these physiological signals. Other signal types such as blood pressure and pulse oximetry also offer valuable insights into an individual's health condition. These signals serve as indispensable instruments in modern healthcare owing to their multifaceted utility, encompassing monitoring, investigative studies, diagnostic evaluations, and therapeutic interventions.

It is challenging to assess medical and biological signals because they frequently have noise, non-stationarity, complexity, and high dimensionality [5]. In addition, a variety of signal types differ in terms of shape, phase, amplitude, frequency, duration, and other attributes. Therefore, employing conventional signal processing methods to extract the relevant features and patterns from this data would not always be feasible. More complex methods, such machine learning algorithms, are being employed in biological signal

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 18–29, 2025.
https://doi.org/10.1007/978-3-031-75167-7_2

processing to increase accuracy, dependability, and efficiency. Healthcare professionals may employ a multitude of information found in medical and biological signals to provide patients with proper diagnosis, treatment, and care. To make this possible, new techniques for signal processing and analysis must be developed.

A class of computing algorithms known as "machine learning" are capable of learning from data [6] and formulating hypotheses or predictions based on the patterns they spot. Depending on the kind and accessibility of feedback or reward, these algorithms may be roughly divided into three categories [7]: supervised learning, unsupervised learning, and reinforcement learning. Support vector machines, which are frequently used in classification and regression tasks, decision trees, which are intuitive models for decision-making, and random forests, which combine multiple decision trees to improve accuracy, are some of the widely used machine learning techniques. Artificial neural networks, which are modelled after the structure and function of the human brain, fuzzy logic [8], which uses approximate reasoning to handle uncertain or imprecise information. These techniques are widely used in many industries, including healthcare, for diagnosis, prognosis, drug discovery, and tailored therapy.

This work primarily aims to explore the processing power of machine learning algorithms for biological and medical data, with a focus on their use in the domains of myology [11], neurology [10], and cardiology [9]. Numerous techniques have shown to be successful in raising the precision and effectiveness of identifying, treating, and keeping track of a broad variety of medical conditions in numerous domains. Notably, the diagnosis of neuromuscular problems based on EMG data, the prediction of the beginning of cardiac arrhythmias, and the detection of aberrant electroencephalogram (EEG) patterns suggestive of epilepsy have all shown promise when it comes to machine learning algorithms. In this discipline, three problems need to be solved: data heterogeneity, interpretability, and generalizability [12]. Subsequent investigations have to concentrate on creating machine learning models that are more trustworthy and comprehensible in order to address these issues and improve the therapeutic value of biological and medical data.

2 Literature Survey

2.1 Cardiology

Cardiology is a branch of medicine that focuses on disorders of the heart and blood vessels. It is primarily a diagnostic and therapeutic field that uses cardiac signals [9, 13]. An important indicator that records the electrical activity of the heart is the electrocardiogram (ECG) [14]. It offers essential information on ischemia, conduction abnormalities, heart rate, rhythm, and other topics. Heart condition early diagnosis and management are also facilitated by other indications such as blood pressure and pulse oximetry. More precise and efficient cardiology treatment is now possible thanks to improvements in accuracy brought about by advanced signal processing and machine learning (Fig. 1).

Machine learning techniques offer effective solutions to address limitations inherent in traditional methodologies across various domains, notably in projects such as electrocardiogram (ECG) analysis. Their application has garnered significant attention due to their capacity to surpass constraints associated with manual interpretation and rule-based

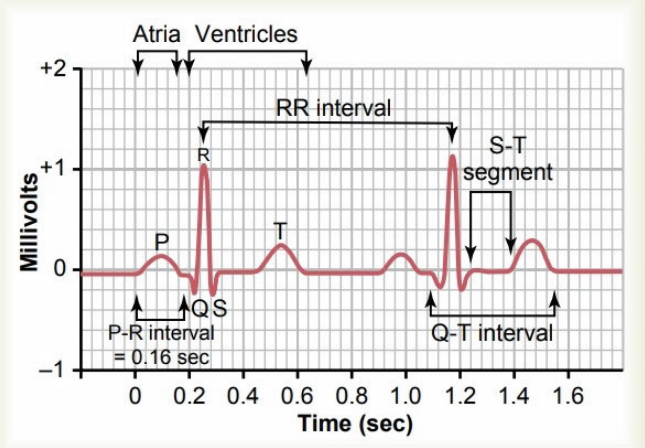


Fig. 1. A Typical Cardiac Signal [45]

algorithms. ECG analysis stands as a prime area where machine learning methods have been extensively deployed [15], demonstrating considerable potential across diverse fields.

- Electrocardiogram (ECG) Segmentation: This process entails partitioning the ECG signal into discrete segments corresponding to distinct stages of the cardiac cycle. Its primary objective is to accurately isolate and distinguish crucial ECG elements, such as the T wave, QRS complex, and P wave, facilitating a more comprehensive examination.
- Electrocardiogram (ECG) Feature Extraction: This intricate procedure involves extracting pertinent data or characteristics from the ECG signal or its segments, which are subsequently utilized for further analysis or classification. This approach encompasses the utilization of time-domain, frequency-domain, wavelet, and other relevant techniques.
- Electrocardiogram (ECG) Feature Selection: This phase involves the meticulous selection of a subset of features from the ECG signal or its segments, emphasizing those most relevant or possessing the highest discriminatory capacity for a specific task or issue. This deliberate selection process enhances analysis accuracy and efficiency while streamlining processing complexity.

ECG Data Interpretation – ML Techniques:

Artificial Neural Networks (ANNs). Biological neural networks served as the inspiration for ANNs, which are computer models that can learn from data by modifying their weights and biases through a procedure known as backpropagation. With enough data and hidden layers, ANNs can conduct nonlinear mapping between inputs and outputs and can approximate any function [16, 17]. Segmentation, feature extraction, classification, compression, denoising, and synthesis are just a few of the tasks that ANNs have been employed for in ECG analysis [18] (Fig 2). ANNs are capable of handling noisy and incomplete data as well as complicated and nonlinear interactions between inputs

and outputs. Nevertheless, overfitting, local minima, and excessive computational cost are all potential problems for ANNs.

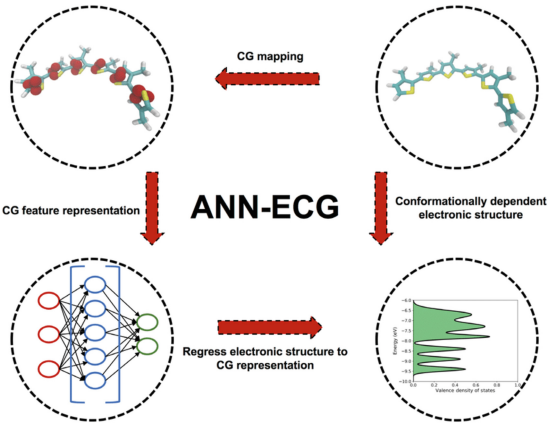


Fig. 2. Schematic of the ANN-ECG method used in this work [46]

Support Vector Machine (SVMs). By locating a hyperplane that optimises the margin between two classes, SVMs—supervised learning models—can carry out binary or multiclass classification [19]. By utilising various kernel functions that translate the data into a higher-dimensional space, SVMs may handle data that is linearly separable and non-separable. In ECG analysis, SVMs have been utilised for a number of tasks including segmentation, feature extraction, classification, and compression [20].

Support Vector Machines (SVMs) are proficient in mitigating overfitting concerns through the incorporation of regularisation parameters. This approach enables SVMs to attain commendable generalisation and accuracy even when trained on a restricted dataset. However, SVMs are susceptible to certain challenges, including complications arising from excessive model complexity, heightened sensitivity to outliers, and the intricacies associated with optimal kernel and parameter selection.

Random Forests (RFs). Random Forests (RFs) stand as ensemble learning methodologies adept at amalgamating diverse decision trees for the purposes of classification or regression [21, 22]. Their proficiency extends to handling high-dimensional and heterogeneous datasets, thereby providing assessments of variable relevance and error estimations. Within the domain of electrocardiogram (ECG) analysis, RFs have been effectively employed across various tasks, encompassing feature extraction, classification, and compression. This is facilitated by the strategy of minimizing individual tree variances and biases, employing random feature selection, and implementing bootstrapping techniques, all of which collectively contribute to achieving notable accuracy and robustness [23].

2.2 Neurology

Neurology, as an intricate medical discipline, primarily focuses on the examination, therapeutic interventions, and comprehensive management of diverse pathologies affecting the nervous system. Essential for the assessment of brain function and the identification of anomalies indicative of neurological disorders, neurological signals hold significant importance [10]. Due to its non-invasive nature, electroencephalography (EEG) stands as a primary modality for assessing the brain's electrical activity. EEG examinations provide crucial insights into various facets of brain function, encompassing electrical oscillations and distinct states such as wakefulness, drowsiness, and sleepiness. Utilizing EEG recordings facilitates the diagnosis of numerous neurological conditions, including strokes, brain tumors, and epilepsy (refer to Fig. 3) [24].

In addition to electroencephalography (EEG), other neurological signal assessment methods include magnetoencephalography (MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). Each of these techniques contributes distinctive perspectives to comprehend both brain function and pathology [25]. Through the utilization of these varied neurological signals, the realm of neuroscience has experienced advancements, enabling enhanced comprehension and management of neurological conditions in both research and clinical settings.

EEG Analysis – ML Techniques:

Deep Learning (DL). It is a subset of machine learning that teaches hierarchical data representations by layering nonlinear processing units. Without the need for manual feature engineering, DL is capable of doing feature extraction and classification from beginning to end. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, autoencoders (AEs), generative adversarial networks (GANs), and other designs are examples of those used in deep learning (DL). Segmentation, feature extraction, classification, compression, denoising, and synthesis are just a few of the EEG analytic tasks for which DL has been applied [15, 24]. By employing a lot of data, several layers of abstraction, and nonlinearity, DL may attain great accuracy and performance. Overfitting, disappearing or inflating gradients,

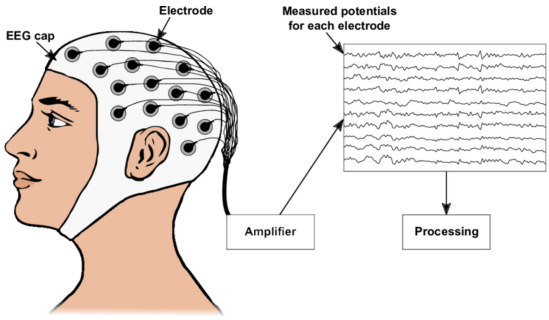


Fig. 3. Sketch of how to record an Electroencephalogram. An EEG allows measuring the electrical activity on the scalp using electrodes which are often fixated on an EEG cap. For each electrode, the signals are amplified and can be used in the following for a desired processing. [47]

excessive computational cost, and lack of interpretability are further problems that DL may encounter.

Fuzzy Sets and Fuzzy Logic (FL). These are used in FL, a kind of machine learning, to deal with the ambiguity, imprecision, and uncertainty in data. Based on language principles generated from human knowledge or experience, FL can approximate reason [8, 48]. Many techniques, including fuzzy clustering (FC), fuzzy inference systems (FIS), and adaptive neuro-fuzzy inference systems, are included in FL (ANFIS). While analysing data, FL is able to account for ambiguity, vagueness, and imprecision as well as to include human expertise. Many EEG analysis tasks, including segmentation, feature extraction, classification, compression, denoising, and synthesis, have been carried out using FL [27] (Fig 4). FL may use language norms and membership functions to achieve great resilience and interpretability. High complexity, sensitivity to parameters, and difficulty in learning from data are potential drawbacks of FL, though.

Kinetic Monte Carlo (KMC). KMC is an unsupervised learning technique that does clustering by dividing the data into k groups based on how similar or distant the objects are to one another [49] (Fig 5). KMC can give a straight-forward and understandable representation of the data structure and manage high-dimensional and heterogeneous data [28]. Many EEG analytic tasks, including segmentation, feature extraction, classification, compression, denoising, and synthesis, have been carried out using KMC. KMC may use iterative optimisation methods to operate at fast speeds and with great efficiency. KMC may, however, experience startup sensitivity, noise sensitivity, difficulties selecting the ideal number of clusters, and sensitivity to outliers.

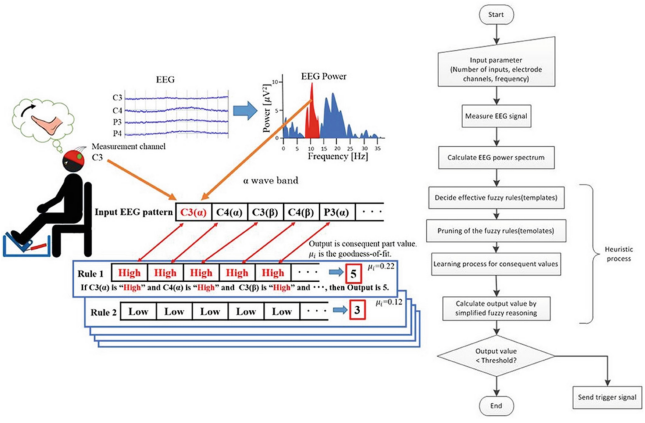


Fig. 4. Overview of heuristic BCI (template matching method based on fuzzy reasoning) [48]

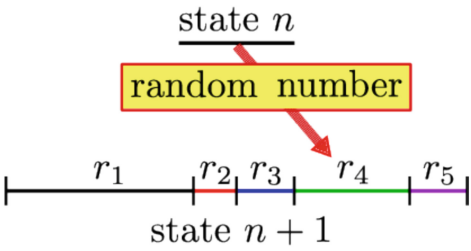


Fig. 5. Illustration of the kinetic Monte Carlo method. The probability that in a given MC step the system will evolve to state i is proportional to ri [49]

3 Myology

Myology is a specialized medical discipline centered on the identification, treatment, and management of muscular ailments and disorders. Myoelectric signals play a pivotal role in assessing the health and functionality of muscles [11, 29]. Electromyography (EMG), a non-invasive technique capturing the electrical activity produced by muscle fibers during contraction or relaxation, serves as a prevalent method for evaluating these signals. EMG signals are instrumental in uncovering various muscular issues such as weakness, fatigue, and spasticity [30]. They also contribute to assessing muscle coordination and movement, aiding in the diagnosis and treatment of movement disorders like multiple sclerosis and Parkinson's disease. Beyond EMG, sonomyography (SMG) [31] and mechanomyography (MMG) [32] are two additional myoelectric signals utilized in muscle assessment. Across the medical landscape, myoelectric signals are employed to address a spectrum of muscular-related conditions. This advancement has fundamentally transformed the scientific and medical understanding of muscle pathophysiology and function.

EMG Analysis – ML Techniques:

Hidden Markov Models (HMMs). Electromyography (EMG) signal analysis uses a statistical framework called Hidden Markov Models (HMMs). The fundamental idea behind HMMs is the concept of a latent or hidden state, which is implied from seen data but is not explicitly visible [34]. In the context of EMG analysis, this hidden state pertains to the intrinsic muscle activity, while the observable data encompass the EMG signals captured externally from the skin's surface. HMMs serve the purpose of predicting the most probable sequence of concealed states corresponding to the observable data, establishing a probabilistic association between the observed data and the underlying hidden state. The versatility of HMMs has been leveraged across various EMG analysis applications, encompassing pattern recognition, gesture delineation, and control mechanisms for prosthetic devices [35].

K-nearest neighbors (KNN). A popular machine learning approach for classification problems in EMG signal analysis is K-nearest neighbours (KNN). Finding the K data points that are the closest (i.e., neighbours) to a new input data point is the fundamental tenet of KNN, and the new data point is then classified according to the majority class of its K nearest neighbours [36, 50] (Fig 6). KNN may be used in EMG analysis to categorise

various muscle contractions or gestures based on how similar their feature vectors are. Being non-parametric, KNN makes no assumptions regarding the distribution of the data or the underlying connection between the features and the classes.KNN can perform well with small to medium-sized datasets and is computationally straightforward and effective [37, 50]. The appropriate value of K might be crucial for the classification accuracy since KNN may be plagued by the "curse of dimensionality" (i.e., performance declines as the number of features grows).

Decision trees. The analysis of EMG signals utilizes the popular and straightforward machine learning approach of decision trees. This method involves creating a tree-like model that makes decisions based on EMG signal characteristics. To classify EMG signals, the decision tree algorithm evaluates the importance of various features [38]. It then segments the data into subsets based on the most informative features' values. This process continues until a predefined stopping criterion is met, which could range from achieving a minimum threshold of data points in each final segment to reaching a specific depth within the tree. Once constructed, the tree enables the categorization of additional EMG signals by tracing their attribute values from the root node to a leaf node. The application of decision trees in EMG analysis has notably been successful in classifying diverse forms of muscle fatigue and distinguishing between pathological and normal muscular activities [39]. However, decision trees are susceptible to overfitting and might perform suboptimally when dealing with a large number of features or noisy datasets.

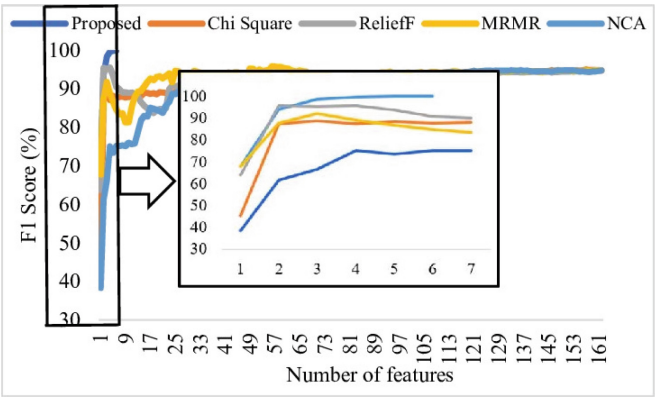


Fig. 6. The EMG-PR performance using KNN classifier with the number of features and different feature selection methods [50]

Gaussian Mixture Models (GMMs). Gaussian Mixture Models (GMMs) represent a probabilistic model extensively employed in analyzing Electromyography (EMG) data. By employing a collection of Gaussian probability distributions to characterize signal properties, GMMs classify signals based on the underlying distributions they exhibit [40]. This approach allows GMMs to effectively capture intricate patterns and correlations

among signal components by modeling signals as a combination of Gaussian distributions. Consequently, this modeling technique facilitates more precise signal classification into distinct groups or categories. Furthermore, GMMs enable the estimation of means and variances within these distributions, aiding not only in clustering but also in feature extraction during EMG signal analysis [41]. This information assists in identifying critical aspects of the signals, subsequently applicable to further analysis or categorization. GMMs have demonstrated efficacy across various EMG analysis applications, including the detection and classification of motor unit action potentials, identification of muscle fatigue, and characterization of muscle activation patterns.

4 Challenges and Future Directions

The analysis of biological and medical data has shown a great deal of opportunity and potential for machine learning approaches. To enhance the effectiveness, dependability, and application of these strategies, however, a number of difficulties and unresolved problems still need to be solved. Some of these challenges and future directions [42–44] are:

- Data quality: Medical and biomedical signals are often corrupted by noise, artifacts, outliers, and missing values that may affect the quality and validity of the analysis. Therefore, there is a need for developing robust and efficient methods for data preprocessing, cleaning, normalization, and enhancement that can preserve or improve the signal characteristics and information content.
- Data availability: Medical and biomedical signals are often limited in quantity, diversity, and representativeness due to ethical, privacy, and practical constraints. Therefore, there is a need for developing methods for data augmentation, synthesis, generation, and transfer that can increase or enrich the data sources and samples for training and testing machine learning models.
- Data interpretation: Medical and biomedical signals are often complex, high-dimensional, nonlinear, and non-stationary that may pose difficulties in understanding and explaining the analysis results. As a result, techniques for data exploration, visualisation, justification, and validation must be developed in order to offer users comprehensible and insightful understandings of the patterns and characteristics of the data.

5 Conclusion and Future Scope

Machine learning methodologies have gained considerable traction within the realm of medical and biological signal processing owing to their adeptness in discerning intricate connections and patterns that might elude human observation. These techniques have proven exceptionally beneficial across various domains, notably in cardiology, neurology, and myology. Machine learning supports cardiology by interpreting ECGs, aiding in diagnosing various heart conditions. In neurology, similar methods analyze EEGs to detect epilepsy, cognitive issues, and sleep disorders. Machine learning is used in myology to interpret EMG data and diagnose muscular disorders such as tiredness, stiffness,

and poor coordination. In spite of significant advancements made so far, the field of machine learning applied to biological and medical signal processing continues to face a number of enduring problems and open difficulties. One of the main obstacles is the lack of standardised and well annotated datasets, which makes it difficult to do comparison analysis across various research. Moreover, it is difficult to guarantee the interpretability and transparency of machine learning models in medical settings due to the complexity of these models and the requirement for clear explanations aimed at patients and healthcare professionals. Furthermore, extensive validation processes, regulatory approvals, and ethical considerations are essential preconditions for incorporating machine learning models into healthcare environments. Research and development efforts must be coordinated in order to overcome these challenges and fully utilise machine learning in medical and biological signal processing.

References

1. Rajeswari, J., Jagannath, M.: Advances in biomedical signal and image processing–A systematic review. Inform. Med. Unlocked **8**, 13–19 (2017)
2. Theis, F.J., Meyer-Bãse, A.: Biomedical Signal Analysis: Contemporary Methods and Applications. The MIT Press (2010). https://doi.org/10.7551/mitpress/7558.001.0001
3. Schläpfer, J., Wellens, H.J.: Computer-interpreted electrocardiograms: benefits and limitations. J. Am. Coll. Cardiol. **70**(9), 1183–1192 (2017)
4. Friedman, D., Claassen, J., Hirsch, L.J.: Continuous electroencephalogram monitoring in the intensive care unit. Anesth. Analg. **109**(2), 506–523 (2009)
5. Nicolas-Alonso, L.F., Corralejo, R., Gomez-Pilar, J., Álvarez, D., Hornero, R.: Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain–computer interfaces. Neurocomput. **159**, 186–196 (2015)
6. Erickson, B.J., Korfiatis, P., Akkus, Z., Kline, T.L.: Machine learning for medical imaging. Radiographics **37**(2), 505–515 (2017)
7. Ayodele, T.O.: Types of machine learning algorithms. New Adv. Mach. Learn. **3**, 19–48 (2010)
8. Zadeh, L.A.: Fuzzy logic. Computer **21**(4), 83–93 (1988)
9. Cuocolo, R., Perillo, T., De Rosa, E., Ugga, L., Petretta, M.: Current applications of big data and machine learning in cardiology. J. Geriatr. Cardiol. **16**(8), 601 (2019)
10. Bruffaerts, R.: Machine learning in neurology: what neurologists can learn from machines and vice versa. J. Neurol. **265**(11), 2745–2748 (2018)
11. Edwards, A.L., et al.: Application of real-time machine learning to myoelectric prosthesis control: a case series in adaptive switching. Prosthet. Orthot. Int. **40**(5), 573–581 (2016)
12. Ho, S.Y., Phua, K., Wong, L., Goh, W.W.B.: Extensions of the external validation for checking learned model interpretability and generalizability. Patterns **1**(8), 100129 (2020)
13. Ignacio Dorado-Díaz, P., Sampedro-Gómez, J., Vicente-Palacios, V., Sánchez, P.L.: Applications of artificial intelligence in cardiology. The future is already here. Revista Española de Cardiología (English Edition) **72**(12), 1065–1075 (2019). https://doi.org/10.1016/j.rec.2019.05.014
14. Sardar, P., Abbott, J.D., Kundu, A., Aronow, H.D., Granada, J.F., Giri, J.: Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. Cardiovasc. Interventions **12**(14), 1293–1303 (2019)
15. Murat, F., Yildirim, O., Talo, M., Baloglu, U.B., Demir, Y., Acharya, U.R.: Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. Comput. Biol. Med. **120**, 103726 (2020)

16. Zou, J., Han, Y., So, S.S.: Overview of artificial neural networks. In: Artificial Neural Networks: Methods and Applications, pp. 14–22 (2009)
17. Sharma, N., Ray, A.K., Sharma, S., Shukla, K.K., Pradhan, S., Aggarwal, L.M.: Segmentation and classification of medical images using texture-primitive features: application of BAM-type artificial neural network. J. Med. Phys./Assoc. Med. Phys. India **33**(3), 119 (2008)
18. Lo, S.C.B., Chan, H.P., Lin, J.S., Li, H., Freedman, M.T., Mun, S.K.: Artificial convolution neural network for medical image pattern recognition. Neural Netw. **8**(7–8), 1201–1214 (1995)
19. Salem, O., Guerassimov, A., Mehaoua, A., Marcus, A., Furht, B.: Anomaly detection in medical wireless sensor networks using SVM and linear regression models. Int. J. E-Health Med. Commun. **5**(1), 20–45 (2014)
20. Shen, L., et al.: Evolving support vector machines using fruit fly optimization for medical data classification. Knowl.-Based Syst. **96**, 61–75 (2016)
21. Guo, L., Kubat, N.J., Isenberg, R.A.: Pulsed radio frequency energy (PRFE) use in human medical applications. Electromagn. Biol. Med. **30**(1), 21–45 (2011)
22. Chow, E.Y., Morris, M.M., Irazoqui, P.P.: Implantable RF medical de-vices: The benefits of high-speed communication and much greater communication distances in biomedical applications. IEEE Microwave Mag. **14**(4), 64–73 (2013)
23. Saied, I.M., Arslan, T., Chandran, S.: Classification of Alzheimer's disease using RF signals and machine learning. IEEE J. Electromagnet. RF Microwaves Med. Biol. **6**(1), 77–85 (2021)
24. Valliani, A.A.A., Ranti, D., Oermann, E.K.: Deep learning and neurology: a systematic review. Neurol. Therapy **8**, 351–365 (2019)
25. Auger, S.D., Jacobs, B.M., Dobson, R., Marshall, C.R., Noyce, A.J.: Big data, machine learning and artificial intelligence: a neurologist's guide. Pract. Neurol. **21**(1), 4–11 (2021)
26. Merlin Praveena, D., Angelin Sarah, D., Thomas George, S.: Deep learning techniques for EEG signal applications–a review. IETE J. Res. **68**(4), 3030–3037 (2022)
27. Dagar, P., Jatain, A., Gaur, D.: Medical diagnosis system using fuzzy logic toolbox. In: International Conference on Computing, Communication & Automation, pp. 193–197. IEEE (2015)
28. Xu, B., et al.: Revealing the crucial role of rough energy landscape on self-diffusion in high-entropy alloys based on machine learning and kinetic Monte Carlo. Acta Mater. **234**, 118051 (2022)
29. Hudgins, B., Parker, P., Scott, R.N.: A new strategy for multifunction myoelectric control. IEEE Trans. Biomed. Eng. **40**(1), 82–94 (1993)
30. Subasi, A., Yilmaz, M., Ozcalik, H.R.: Classification of EMG signals using wavelet neural network. J. Neurosci. Methods **156**(1–2), 360–367 (2006)
31. Zheng, Y.P., Chan, M.M.F., Shi, J., Chen, X., Huang, Q.H.: Sono-myography: monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis. Med. Eng. Phys. **28**(5), 405–415 (2006)
32. Ibitoye, M.O., Hamzaid, N.A., Zuniga, J.M., Wahab, A.K.A.: Mechanomyography and muscle function assessment: a review of current state and prospects. Clin. Biomech. **29**(6), 691–704 (2014)
33. Phinyomark, A., Scheme, E.: EMG pattern recognition in the era of big data and deep learning. Big Data and Cognitive Comput. **2**(3), 21 (2018)
34. Yi, K., Beheshti, J.: A hidden Markov model-based text classification of medical documents. J. Inf. Sci. **35**(1), 67–81 (2009)
35. Chan, A.D., Englehart, K., Hudgins, B., Lovely, D.F.: Hidden Markov model classification of myoelectric signals in speech. IEEE Eng. Med. Biol. Mag. **21**(5), 143–146 (2002)
36. Hu, L.Y., Huang, M.W., Ke, S.W., Tsai, C.F.: The distance function effect on k-nearest neighbor classification for medical datasets. Springerplus **5**(1), 1–9 (2016)

37. Zhang, Z.: Introduction to machine learning: k-nearest neighbors. Ann. Transl. Med. **4**(11), 218–218 (2016)
38. Rajendran, P., Madheswaran, M.: Hybrid medical image classifica-tion using association rule mining with decision tree algorithm. arXiv preprint arXiv:1001.3503 (2010)
39. Hu, R.: Medical data mining based on decision tree algorithm. Comput. Inform. Sci. **4**(5), 14 (2011)
40. Riaz, F., et al.: Gaussian mixture model based probabilistic modeling of images for medical image segmentation. IEEE Access **8**, 16846–16856 (2020)
41. Xie, C.H., Chang, J.Y., Liu, Y.J.: Estimating the number of components in Gaussian mixture models adaptively for medical image. Optik **124**(23), 6216–6221 (2013)
42. Yang, G., Rao, A., Fernandez-Maloigne, C., Calhoun, V., Menegaz, G.: Explainable AI (XAI) in biomedical signal and image processing: promises and challenges. In: 2022 IEEE International Conference on Image Pro-cessing (ICIP), pp. 1531–1535. IEEE (2022)
43. Tobore, I., et al.: Deep learning intervention for health care challenges: some biomedical domain considerations. JMIR Mhealth Uhealth **7**(8), e11966 (2019)
44. Jovic, A., Jozic, K., Kukolja, D., Friganovic, K., Cifrek, M. Challenges in designing software architectures for web-based biomedical signal analysis. In: Medical Big Data and Internet of Medical Things, pp. 81–111. CRC Press (2018)
45. Komanapalli, V.L.N., Bhujanga Rao, A.: Noise removal using adaptive noise canceling, analysis of ECG using MATLAB (2023)
46. Jackson, N., Bowen, A., Antony, L., Webb, M., Vishwanath, V., de Pablo, J.: Electronic structure at coarse-grained resolutions from supervised machine learning. Sci. Adv. (2019). https://doi.org/10.1126/sciadv.aav1190
47. Nagel, S.: Towards a home-use BCI: fast asynchronous control and robust non-control state detection. (Doctoral dissertation) (2019). Retrieved from https://doi.org/10.15496/publikation-37739
48. Saga, N., Doi, A., Oda, T., Kudoh, S.N.: Elucidation of EEG characteristics of fuzzy reasoning-based heuristic BCI and its applica-tion to a patient with brain infarction. Frontiers in Neurorobotics (2021). https://doi.org/10.3389/fnbot.2020.607706
49. Makieła, D., Maśka, M.M.: Spontaneous currents in a bosonic ring. Acta Physica Polonica A **130**(2), 569–572 (2016). https://doi.org/10.12693/APhysPolA.130.569
50. Johirul Islam, M., Ahmad, S., Haque, F., Reaz, M.B.I., Bhuiyan, M.A.S., Rezaul Islam, M.: Forearm orientation and muscle force invariant feature selection method for myoelectric pattern recognition. IEEE Access **10**, 46442–46471 (2022). https://doi.org/10.1109/ACCESS.2022.3170483

N-BEATS & Temporal Fusion Transformer Based Surface Temperature Prediction and Forecasting for Realizing Global Warming Trends

Adria Binte Habib(✉), Faisal Bin Ashraf, Muhammad Iqbal Hossain, and Golam Rabiul Alam

Brac University, Dhaka, Bangladesh
binte.adria708@gmail.com

Abstract. At the pinnacle of civilization, where the impacts of climate change have been increasingly felt, weather prediction plays a critical role in mitigating the potential disasters that may arise. Moreover, with the gradual change on climate, surface temperature of the earth is increasing. This increasing rate of the surface temperature causing global warming which is a matter of intimidation. To leave off this global warming, weather forecasting can be used as an arsenal. Selecting the appropriate tools and models for weather prediction is a crucial step in ensuring accurate forecasts. In this research paper, the focus was on studying the versatility of three specific architectures for weather prediction: LSTM, Temporal Fusion Transformer, and N-BEATS. To assess these architectures' performance, we conducted a number of experiments. With the lowest Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the three, NBEATS stood out. This shows that when compared to the other models, the N-BEATS architecture had greater prediction accuracy. It's vital to remember, too, that the trials also showed that the Temporal Fusion Transformer and LSTM performed well. The only distinction was that these models required larger sizes in terms of parameters and computational complexity to achieve their performance levels. Consequently, considering both performance and model size, the researchers determined that N-BEATS was the most optimal and versatile architecture for weather prediction. Its ability to achieve excellent results with a smaller model size makes it a favorable choice for practical applications.

Keywords: Weather Prediction · Time Series Analysis · LSTM · Temporal Fusion Transformer · N-BEATS

1 Introduction

Global warming has significant impacts on human life, affecting various aspects of society, including health, livelihoods, economics and security. Severe weather phenomena like hurricanes, droughts, floods and heat waves are made worse by it. These occurrences have the potential to cause fatalities, population displacement, infrastructure devastation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 30–41, 2025.
https://doi.org/10.1007/978-3-031-75167-7_3

and interruptions to vital services. Moreover, changes in temperature and rainfall patterns affect agricultural productivity and food production. Adding to it, shifts in precipitation can result in droughts or floods, damaging crops and reducing yields. Also, global warming has economic implications, including increased costs for disaster response, recovery and infrastructure repair. Agricultural losses, disruptions to supply chains and decreased productivity can have significant economic impacts on both developed and developing nations. Additionally, the costs associated with adapting to and mitigating climate change are substantial. Since global warming is the long term rise in the mean surface temperature of the earth, we should investigate variations in this temperature in order to survive these issues. Thus, weather forecasting enters the scene. Therefore, predicting the next weather conditions is the task of weather forecasting. For natural disasters to be managed and their effects mitigated, accurate weather forecasting is essential. By anticipating and pre-paring for severe weather events, such as storms and hurricanes, we can take proactive measures to minimize damage and protect lives and property. Additionally, decision-makers can leverage weather forecast information to enhance risk management practices, implement targeted policies, and improve over-all preparedness. Furthermore, meteorological information is invaluable to farmers. By utilizing forecasts, they can make informed decisions regarding crop planning, irrigation, and pest control, ultimately in-creasing productivity while minimizing costs [1]. The availability of accurate weather forecasts supports efficient agricultural practices and contributes to food security. In weather forecasting, various approaches can be employed, including statistical methods and deep learning. Statistical weather forecasting, which is a branch of numerical weather prediction, often relies on techniques like "least-squares regression". Weather stations commonly employ statistical models to enhance the accuracy of dynamic forecasts [2]. These models analyze historical weather data and identify patterns using regression analysis in diverse ways. However, in recent years, deep learning methods have revolutionized the field of weather forecasting [3]. Advances in deep learning algorithms have led to significant improvements in forecasting accuracy, outperforming other approaches depending on the characteristics of the dataset being utilized. Deep learning methods excel because they employ non-linear functions instead of linear functions, allowing for more intricate and accurate modeling of complex weather patterns. Additionally, many deep learning methods employ the "Back Propagation" algorithm to minimize the loss and optimize the performance of the algorithms [4]. Overall, weather forecasting is a vital tool for understanding and predicting climate behavior. It helps us to lessen the effects of natural disasters, handle the issues raised by climate change, and make sound decisions in a variety of areas, such as transportation, agriculture, and policy making. The continuous advancements in both statistical and deep learning methods hold promise for further improving the accuracy and reliability of weather forecasts in the future. We tried to contribute to this field of research from multiple perspectives.

- Firstly, we tried to focus this study to assess the versatility of deep learning and advanced deep learning algorithms by evaluating their performance on four different datasets
- We will be able to determine the best architecture or architectures that can accurately handle diverse types of data

- To further iterate, we specifically focused this study on the prediction of temperature. This is a direct contribution of our research to handle and study the critical aspect of global warming. This is because the temperature is a direct and major cause of global warming and would add value to the future extensions of research performed in this area
- Forecasting temperature with advanced deep learning and transformer models to find best performing model with diverse types data is whole new research on the field of Time-series Analysis and Weather Prediction. This is the novelty of our work

2 Existing Works

During the early 20th century, weather forecasting relied on various machine learning approaches, including Support Vector Machines (SVM), Support Vector Regression (SVR), and Interactively Recurrent Fuzzy Functions. These techniques were utilized to develop models for predicting weather conditions. In the research conducted by the authors, multiple error measures were employed to assess the accuracy of these models. These measures included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R-squared (R2), Mean Squared Error (MSE), Prediction Error, and Mean Absolute Percentage Error (MAPE). By evaluating these error measures, the authors aimed to determine the effectiveness and performance of the machine learning models in accurately forecasting weather conditions. The specific references for these studies can be found in the provided citations [10–12]. While the mentioned algorithms, such as SVM, SVR, and Interactively Recurrent Fuzzy Functions, are supervised learning algorithms commonly used in weather forecasting, the learning process can be further strengthened by incorporating unsupervised learning algorithms. As a result, neural network algorithms, which offer the capability to learn from unlabeled data, gained popularity in weather prediction [16]. In recent years, advanced deep learning techniques have received significant recognition in the field of weather forecasting. These techniques include Multi-Layer Perceptron (MLP), Self-Constructing Recurrent Fuzzy Neural Network (FNN), Evolving Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) networks [9, 13–15]. These progressed procedures use the control of deep learning to capture complex patterns and conditions in climate information, conduct to boosted accuracy and performance. Moreover, in [17] the authors analyzed SARIMA model and traditional models to find out the best performing models and ended up with the decision that traditional models performed better.

3 Methodology

In this section, we will explain the overall workflow of this work. Firstly, we will collect the data and then we will preprocess it. After that, we will build our 3 models LSTM, N-BEATS and Temporal Fusion Transformer. In the result analysis part, we will analyze the performance of the models and conclude the result. In the Fig. 1 the over-all workflow is given.

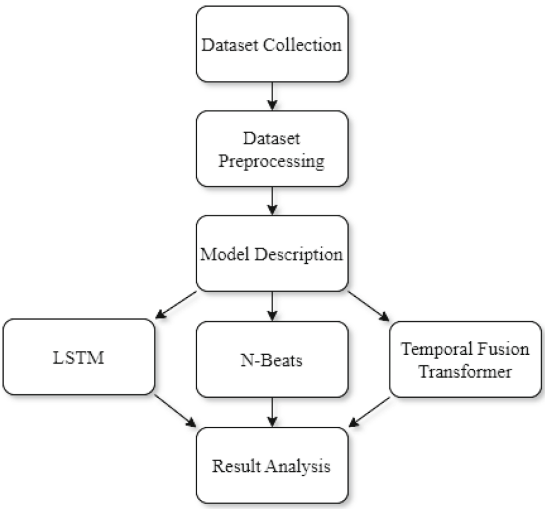


Fig. 1. Top level overview of the proposed methodology

3.1 Dataset Description

We used four datasets in this research. Firstly, we used dataset named Dhaka which is the data of Dhaka district, Bangladesh of 65 years (1949–2013). Secondly, we used dataset named Barisal which is the data of Barisal district, Bangladesh of 65 years (1949–2013). Table 1 describes the details of the dataset we used. In the Table 1, it can be seen, number of data, minimum temperature, maximum temperature, and standard deviation is given. From this information, the type of dataset can be predicted. In Dataset Dhaka & Dataset Barisal, the standard deviation is quite smaller. Since, a smaller standard deviation suggests that data points are closer to the mean or average. This indicates that the values in the dataset are more consistent and predictable. Also, if the standard deviation is smaller, it means that the data points are close to the average. This suggests greater precision and accuracy in the measurements or observations. It implies that the results are more reliable. From the above discussion, it can be said that, Dataset Dhaka & Dataset Barisal should behave well. On the other hand, we also used California – 1 which is the data of Windsor, California, USA of 4 years (2019–2022) & California – 2 which is the data of San Benito, California, USA of 4 years (2019–2022). From the table I we can see the standard deviation of California – 1 & California – 2 is quite larger. Since a larger standard deviation does not necessarily indicate a good dataset. The standard deviation measures the spread or variability of data points within a dataset. The information or data points are more widely distributed from the average when the standard deviation is larger. From the above information we can say, California – 1 and California – 2 dataset is not so ideal for our prediction. Adding to it, the particular dataset is of California is a place that belongs to kind of unpredictable weather. Conversely, the standard deviation of the dataset of Bangladesh are quite smaller which denotes the good quality of dataset and also good quality of temperature condition.

Table 1. Dataset Description

Dataset	Range (Years)	Type of Data	Number of Data	Minimum Temperature	Maximum Temperature	Standard Deviation
Dhaka	1949–2023	Monthly	732	15.1 °C	15.1 °C	3.699
Barisal	1949–2023	Monthly	732	20.4 °C	15.1 °C	3.338
California-1	2019–2023	Daily	1429	18.2 °F	15.1 °F	13.946
California-2	2019–2023	Daily	1429	34.9 °F	15.1 °F	12.316

3.2 Dataset Preprocessing

In the datasets, there were other weather attributes like humidity, vapor pressure, wind speed etc. Since, we were working with only temperature, we dropped other weather attributes. We only kept "Temperature" and "Date" feature. Since, we were doing time series analysis, we changed the "Date" feature into relevant "Date-time" format.

3.3 Model Specification

In this section, we will be discussing the models we used and the detailed workflow of this work. Firstly, we collected the datasets, then we simultaneously analyzed preprocessed the datasets. After Data analysis we found out the types and quality of the data. Moreover, we could relate the actual weather condition with the types of data. In the data preprocessing part, we prepared our datasets for the models. After that we executed the models. To add to it, we did hyper-parameter tuning and built particular models foreach architecture.

1. LSTM: LSTM or Long Short-Term Memory is a unique type of architecture designed to retain and comprehend patterns in a series of data over an extended period. It achieves this by utilizing a forget gate mechanism, which allows the network to selectively remember or ignore information from previous time steps. For sequence prediction tasks, where comprehension of the context and dependencies over time is essential, this design is especially helpful. LSTMs are designed with specific methods to decide when to remember or ignore inputs in the hidden state, related to the reasoning behind GRUs (Gated Recurrent Units) [5]. The design of LSTM is influenced by the logic gates used in computers.
2. N-Beats: The mentioned deep learning architecture is specifically designed for univariate time series forecasting. It utilizes a combination of backward and forward residual links, along with a deep stack of fully-connected layers, as its baseline [6]. This building style has many good qualities. It is easy to understand, works for many different places without changes and can be trained quickly.
3. Temporal Fusion Transformer: Temporal Fusion Transformer (TFT) is a prediction architecture that uses temporal dynamics to capture local and long-term dependencies. It uses recurrent and self-attention layers, specialized components and gate layers for efficient multi-horizon prediction [7]. Custom models are developed by fine-tuning

hyperparameters to achieve optimal performance. Table 2 lists the most effective hyperparameters.

4 Result Analysis

Each evaluation measure has some draw backs, that is why, each of them may provide wrong interpretation of prediction models. As such, to evaluate any prediction model, multiple measurement processes are needed to get accurate evaluation. Metrics such as MdAE, RMSE, MAPE, MSE, RMSPE, MAE, etc. are some of the most widely used evaluation measures for forecasting in various Machine Learning domain [8]. This is the reason of taking multiple error measures as a booster to take the decision. In this paper, MAE & RMSE were taken as error measurement tools. MAE means Mean Absolute Error which takes the average distances between the predicted data and the actual data for N number of samples (Eq. 1)

$$\text{MAE}(y, \hat{y}) = \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{1}$$

Next, RMSE firstly this error measure takes the predicted data and the given data, then calculates the difference between them and finally evaluates the standard deviation of the difference (Eq. 2).

$$\text{RMSE} = (y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \tag{2}$$

Moreover, by looking onto the difference between Training loss and Validation loss, model's quality was also measured. In this research, we conducted experiments with a particular architecture, and during this process, we observed that the model exhibited a good balance. This observation is supported by the findings presented in Table 3, which demonstrate the model's balanced performance. To further evaluate the model, we examined the root mean square error (RMSE) scores for different datasets, namely Barisal, California – 1, and California – 2. The RMSE scores were comparatively higher for these datasets, as illustrated in Fig. 2. Additionally, we also analyzed the mean absolute error (MAE) scores, which were also higher for the Barisal and California – 1 datasets, as shown in Fig. 2. In order to improve the model's performance, we conducted hyper-parameter tuning. By adjusting the hyperparameters and utilizing models with larger parameters, we achieved a better balance between MAE and RMSE results, as indicated in Table 2.

This suggests that the model's performance aligns with or is comparable to other architectures when the larger parameters are employed. Furthermore, it is noteworthy that the difference between the training error and validation error for all the datasets was very small, as stated in Table 3. This indicates that the model remains well balanced despite the relatively high values of MAE and RMSE. The small difference between training and validation errors implies that the model generalizes well and is not overfitting to the training data.

Table 2. Optimized Hyper-Parameters

Dataset	Architectures	Hyper-parameters
	LSTM	Drop rate = 0.05, Learning rate = 0.005, Batch size = 128, Epoch = 50, Hidden layer = 20
	Temporal Fusion Transformers	Hidden size = 64, LSTM layers = 20, Number of attention heads = 4, Drop rate = 0.001, Batch size = 64, Epoch = 50, Input chunk length = 24
Dhaka		
	N-BEATS	Drop rate = 0.001, Learning rate = 0.001, Batch size = 24, Layer width = 10, Epoch = 20, input chunk length = 24
	LSTM	Drop rate = 0.001, Learning rate = 0.001, Batch size = 64, Epoch = 50, Hidden layer = 30
	Temporal Fusion Transformers	Hidden size = 64, LSTM layers = 25, Number of attention heads = 4, Drop rate = 0.001, Batch size = 64, Epoch = 60, Input chunk length = 24
Barisal	N-BEATS	Drop rate = 0.001, Learning rate = 0.001, Batchsize = 20, Layer width = 10, Epoch = 10, inputchunk length = 20
	LSTM	Drop rate = 0.001, Learning rate = 0.001, Batch size = 64, Hidden Layer = 30, Epoch = 60
	Temporal Fusion Transformers	Hidden size = 64, LSTM layers = 20, Number of attention heads = 4, Drop rate = 0.001, Batch size = 64, Epoch = 70, Input chunk length = 24
California-1	N-BEATS	Drop rate = 0.001, Learning rate = 0.001, Batch size = 24, Layer width = 10, Epoch = 15, input chunk length = 24
	LSTM	Drop rate = 0.001, Learning rate = 0.001 Batch size = 64, Epoch = 60, Hidden layer = 20

(*continued*)

Table 2. (*continued*)

Dataset	Architectures	Hyper-parameters
	Temporal Fusion Transformers	Hidden size = 64, LSTM layers = 25, Number of attention heads = 4, Drop rate = 0.01, Batch size = 128, Epoch = 60, Input chunk length = 24
California-2		
	N-BEATS	Drop rate = 0.001, Learning rate = 0.001, Batch size = 20, Layer width = 10, Epoch = 15, input chunk length = 20

Table 3. Quality of the Models

Dataset	Architectures	Validation Loss	Train Loss	Decision
Dhaka	LSTM	0.0031	0.0028	Balanced
	Temporal Fusion Transformer	0.224	0.209	Balanced
	N-BEATS	0.0958	0.102	Balanced
Barisal	LSTM	0.0091	0.0078	Balanced
	Temporal Fusion Transformer	0.244	0.255	Balanced
	N-BEATS	0.107	0.0733	Balanced
California-1	LSTM	0.0047	0.00342	Balanced
	Temporal Fusion Transformer	0.308	0.323	Balanced
	N-BEATS	0.102	0.0808	Balanced
California-2	LSTM	0.0047	0.0034	Balanced
	Temporal Fusion Transformer	0.377	0.344	Balanced
	N-BEATS	0.197	0.177	Balanced

In this research, we utilized the Temporal Fusion Transformer architecture. Interestingly, even without conducting hyper-parameter tuning, the dataset already yielded good results in terms of training errors and validation errors. The difference between these two types of errors was relatively low, indicating that the model performed well in terms of generalization. However, we proceeded to tweak the hyperparameters to further enhance the model's performance. As a result, we achieved even better results, particularly on the four-filtered dataset. This improvement can be observed in Table 3, where the difference between the training error and validation error remained very low. This indicates that the models were well balanced and consistently performed well across different datasets. To provide additional evidence of the model's accurate predictions, we can examine the values of mean absolute error (MAE) and root mean square error (RMSE) presented in Fig. 2. These values further support the notion that the models

were well balanced and capable of making accurate predictions. Lastly, we employed the N-Beats model for our analysis. Similar to the Temporal Fusion Transformer architecture, the N-Beats model also demonstrated a balanced performance when applied to the dataset. As shown in Table 3, the model exhibited low errors and consistent MAE and RMSE readings. In fact, when compared to the other two architectures previously mentioned, the N-Beats model showcased even lower MAE and RMSE scores. This implies that the N-Beats model was particularly adept at accurately predicting the values in the dataset. Moreover, from Fig. 3 it can be observed that the prediction process has followed the trend which implies that the model is correctly predicting the weather parameter. However, the trend in weather analysis refers to the analysis and identification of patterns or changes in weather conditions over a certain period of time. It involves studying historical weather data to observe long-term trends variations, and shifts in weather patterns. Thus, in addition to the Temporal Fusion Transformer architecture, the N-Beats model proved to be another well-balanced model for our dataset. The low errors and comparatively lower MAE and RMSE scores further support the conclusion that the N-Beats model performed exceptionally well in terms of accuracy and balance. After thoroughly evaluating the models discussed, it can be concluded that all of them performed admirably. However, it is important to note that to achieve the desired results, we had to push the limits of the LSTM and Temporal Fusion Transformer models. This involved increasing their size and complexity, resulting in models that performed exceptionally well but also became larger in terms of parameters and computational requirements. On the other hand, the N-BEATS models exhibited superior performance compared to the other architectures while maintaining a smaller model size. Despite their relatively compact nature, the N-BEATS models outperformed the larger LSTM and Temporal Fusion Transformer models. This means that the N-BEATS architecture is a better and faster solution because it acquires great results without needing to make the model very large. In conclusion, the N-BEATS architecture is the best choice because it performs well and is easy to use compared to the other two models. The capacity to achieve notable outcomes using compact models renders it an appealing choice for practical use, striking a balance between precision and computational efficiency.

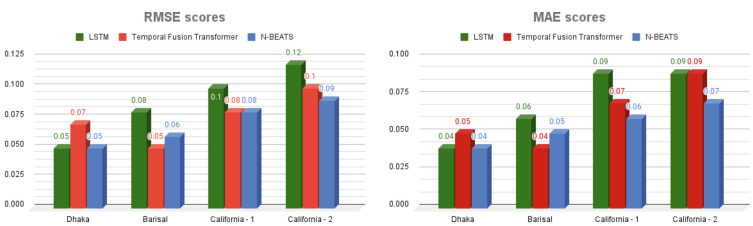


Fig. 2. Errors in All Architectures

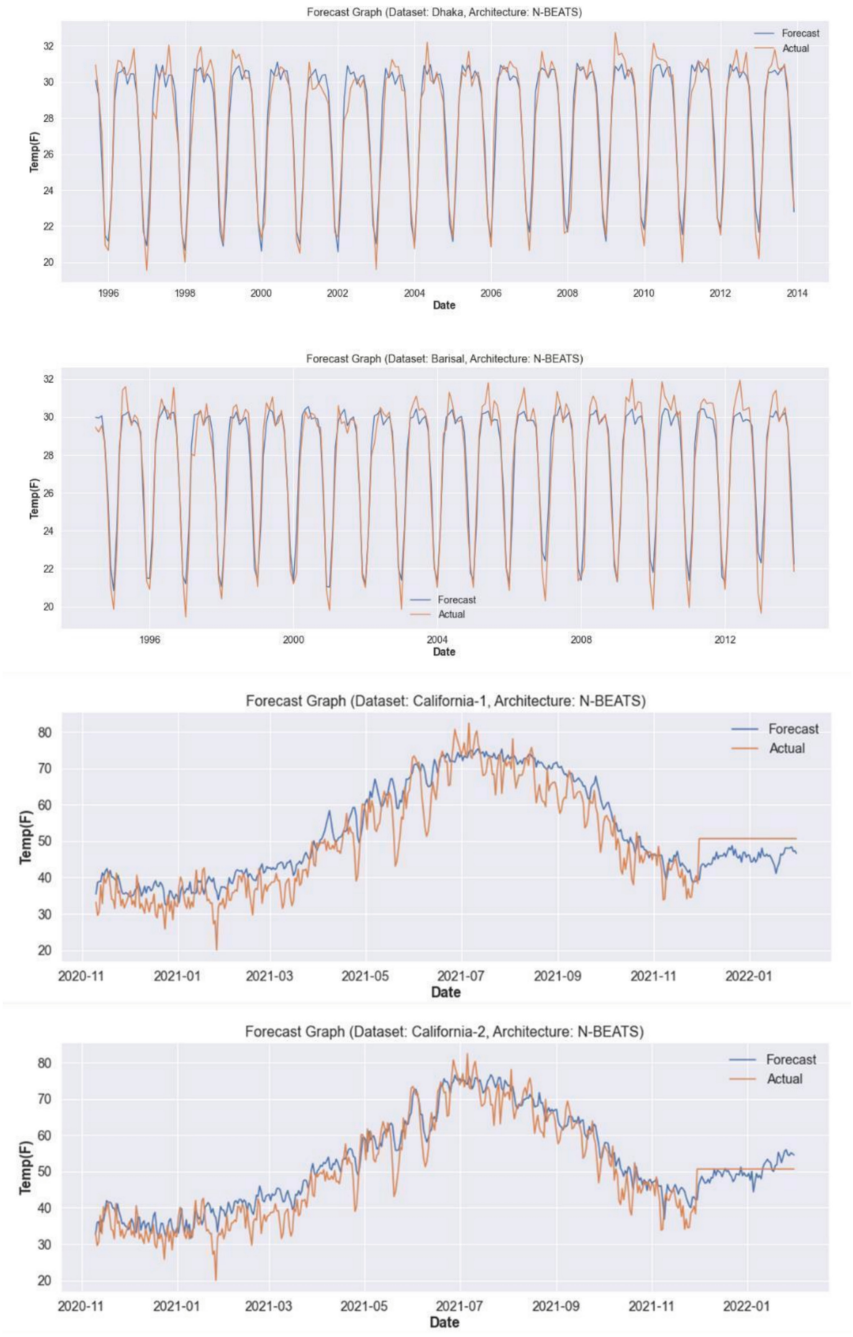


Fig. 3. Forecasting Graph of N-Beats

5 Conclusion

This paper seeks to determine the optimal architecture for weather forecasting, ensuring accuracy across diverse datasets. Three architectures: LSTM, Temporal Fusion Transformer, and N-BEATS—were scrutinized, revealing N-BEATS as the most adaptable and reliable. It consistently delivered excellent results across all examined datasets due to its flexibility and balanced model size, making it the preferred choice for real-world weather prediction. Our ongoing work involves incorporating additional forecasting architectures to enhance coverage, aiming to refine the comparison scheme and gain a deeper understanding of their applicability and performance.

References

1. Thoai, T., Ranola, R. Jr., Camacho, L.: The importance of weather forecasts and meteorological information in adaptation to climate change in agricultural production: some preliminary findings. Philippine Agric. Sci. **101**, 377-392 (2018)
2. Chapter 6 statistical weather forecasting: International Geophysics, pp. 159–232 (1995). https://doi.org/10.1016/s0074-6142(06)80042-2
3. Oancea, B., Ciucu, S.C.: TIME SERIES FORECASTING USING NEURAL NETWORKS (n.d.).. https://arxiv.org/ftp/arxiv/ppers/1401/1401.1333.pdf. Retrieved 8 July 2022
4. Chandra Sekhar, C., et al.: Effectiveness of Backpropagation Algorithm in Healthcare Data Classification (2020). https://doi.org/10.1007/978-981-15-8218-925
5. Zhang, A., Lipton, Z. C., Li, M., Smola, A.J.: Dive into deep learning. Dive into Deep Learning – Dive into Deep Learning 1.0.0-alpha0 documentation (n.d.). Retrieved 4 Aug 2022, https://d2l.ai/
6. Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-beats: Neural basis expansion analysis for interpretable time series forecasting. arXiv.org. Retrieved August 4, 2022 (2020). https://arxiv.org/abs/1905.10437
7. Lim, B., Arik, S.O., Loeff, N., Pfister, T.:. Temporal Fusion Transformers forinterpretable multi-horizon time series forecasting. arXiv.org (2020). Retrieved 5 Aug 2022 https://arxiv.org/abs/1912.09363
8. Cifuentes, J., Marulanda, G., Bello, A., Reneses, J.: Air temperature forecasting using Machine LearningTechniques: A Review. MDPI (2020). Retrieved 5 Aug 2022, https://www.mdpi.com/1996-1073/13/16/4215
9. Furquim, G., Pessin, G., Faiçal, B.S., Mendiondo, E.M., Ueyama, J.: Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory. Neural Comput. Appl. **27**(5), 1129–1141 (2015). https://doi.org/10.1007/s00521-015-1930-z
10. Cao, D.Z., Pang, S.L., Bai, Y.H.: Forecasting exchange rate using support vector machines. In: Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, vol. 6, pp. 3448–3452, 18–21 Aug 2005
11. Wang, Y., Wu, D., Guo, C., Wu, Q., Qian, W., Yang, J.: Short-term wind speed prediction using support vector regression. In: IEEE PES General Meeting, IEEE: Piscataway, NJ, USA, pp. 1–6 (2010)
12. Luo, C., Wang, H.: Fuzzy forecasting for long-term time series based on time-variant fuzzy information granules. Appl. Soft Comput. **88**, 106046 (2020)
13. Goudarzi, S., Khodabakhshi, M.B., Moradi, M.H.: Interactively recurrent fuzzy functions with multi objective learning and its application to chaotic time series prediction. J. Intell. Fuzzy Syst. **30**, 1157–1168 (2016)

14. Li, Q., Lin, R.C.: A new approach for chaotic time series prediction using recurrent neural network. Math. Probl. Eng. **2016**, 9 (2016)
15. Ma, Q.L., Zheng, Q.L., Peng, H., Zhong, T.W., Xu, L.Q.: Chaotic time series prediction based on evolving recurrent neural networks. In: Proceedings of the 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, vol. 6, pp.3496–3500, 19–22 August 2007
16. Alpaydin, E.: Introduction to Machine Learning, MIT Press: Cambridge. MA, USA (2009)
17. Khan, M. S., Islam, M., Adhikary, S., Hossain, M. M., Afroja, S.: Analysis and predictions of seasonal affected weather variables of Bangladesh: SARIMA models vs. traditional models. Int. J. Bus. Manage. **13**(12), 70 (2018)



Tuberculosis Detection and Classification in Chest X-ray Images Using Optimized CNN Architecture

Apeksha Koul[1], Rajesh K. Bawa[2], and Yogesh Kumar[3](✉)

[1] Department of CSE, Punjabi University, Patiala, India
[2] Department of Computer Science, Punjabi University, Patiala, India
[3] Department of CSE, School of Technology, Pandit Deendayal Energy Univeristy, GandhiNagar, Gujarat, India
yogesh.arora10744@gmail.com

Abstract. Tuberculosis (TB) is a global threat to public health, and pathology is one of the most important diagnostic tools to treat it in clinical practice. However, due to the small size and the large number of bacilli bacteria, its treatment is time-consuming and requires an experienced pathologist. Thusly, in this age of AI, the development of modern computer techniques has escalated the process of diagnosing tuberculosis. In this article, Convolutional Neural Network (CNN) model has been developed to detect as well as classify tuberculosis in chest X-ray images. The model is trained using 700 TB and 3500 Normal images (publicly available images) taken from the Tuberculosis chest x-ray (CXR) dataset. Initially, the images are pre-processed by reducing their size and increasing their contrast to reduce noisy signals. Later, several parameters and layers are added to the CNN model for improving the prediction rate of tuberculosis detection. The applied proposed model achieved a 91.2% prediction rate with a minimal loss value, which shows a better prediction than the existing techniques applied on the CXR dataset.

Keywords: Chest X-Ray · Artificial Intelligence · Tuberculosis · Learning Model · CNN

1 Introduction

Tuberculosis (TB) is one of the infectious illnesses that give rise to various health-based issues which are being diagnosed every year. It is caused by harmful bacteria called *Mycobacterium tuberculosis*, where lungs and remaining body parts are mainly affected. Figure 1 shows the presence of tuberculosis bacteria in lungs [1]. Even though coronavirus disease (COVID-19) has killed approx 1.5 million people in 2020, TB is the most significant cause of mortality across the globe. Due to the study conducted by numerous researchers, the scenario is now mostly under control [2]. On the one hand, artificial intelligence (AI) systems effectively identify various chest radiograph disorders caused by tuberculosis bacteria. However, on the other hand, it is still unclear

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 42–55, 2025.
https://doi.org/10.1007/978-3-031-75167-7_4

how well they perform as compared to radiologists and other AI algorithms, for example, in diagnosing MDR-TB, or multidrug-resistant tuberculosis, which is a severe issue worldwide [3].

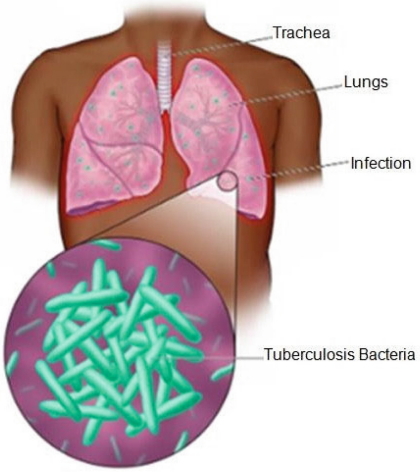


Fig. 1. Tuberculosis Bacteria Affecting Lungs [1]

More than a million individuals every year people lose their lives due to tuberculosis. The World Health Organization (WHO) has confirmed that the COVID-19 pandemic has reversed years of success in the battle against TB. Fewer new patients are being found due to minimum resources and limited access to diagnose TB. As stated in [4], the worldwide decrease in TB mortality was just 9.2% in 2020, which was only a 35% reduction. Sources have also indicated that, due to the outbreak of COVID-19 and the situation aggravated by it in 2020, only 21% of patients got treatment for tuberculosis compared to the previous year. There are roughly 30 nations where 90% of new TB cases are diagnosed yearly, and many lack the means to adequately deal with this public health crisis. Specifically for those communities who belong to poor background, it has been believed that detecting and diagnosing chest X-rays using artificial intelligence might significantly improve the health of people with access to healthcare [5].

The use of AI in medical diagnosis is expanding quickly. It helps hospitals and clinicians in many ways, such as making quick and accurate clinical decisions, detecting or predicting medical images, and automating diagnosis to ease administrative burdens [6]. There are 7.9 billion individuals on Globe, yet almost two-thirds (5.2 billion) still need access to just one. Due to this reason, there has been an exceptionally rapid uptake of AI technologies in radiology. Radiologists rely on their visual confirmation and interpretation of chest X-rays to diagnose in a data-driven sector. It is one of the problems of the healthcare sector that there is a need for more radiologists worldwide [7]. The lack of this expertise significantly contributes to the worsening state of lung healthcare. AI

technologies may affect this field by relieving the burden on overworked medical imaging experts, helping them analyze large amounts of imaging data, prioritizing severe matters, and generating recommendations [8].

New research showed that a chest X-ray might be used to identify individuals with active tuberculosis reliably, provided that the appropriate deep learning system is applied. The research indicated that by employing such a screening method as a first step before ordering a more costly diagnostic test, up to 80% of the cost per positive TB case found might be saved via efficient AI-powered screening [9]. Despite similar false-negative and false-positive detection rates, an AI-based method could effectively identify active pulmonary TB patients. Additionally, it also helps to analyze HIV-positive patients who are at increased risk of getting TB and are difficult to screen because their chest X-rays may look different from usual TB cases [10].

Therefore, it can be stated that AI plays a crucial role as a diagnostic tool in medical sector that boosts the efforts of clinicians and guarantees that no therapeutic delays occur. The paper aims to develop an optimized CNN model that takes CXR images as an input to detect and also classify tuberculosis-affected lungs.

1.1 Contribution of the Paper

The contribution that has been made in this research paper is as followed:

1. **Selection and use of a specific database:** Chest X-ray images of patients suffering from tuberculosis and healthy individuals from *Tuberculosis chest X- ray database* are used for the study. The database contains two classes: images of tuberculosis and normal lungs.
2. **Pre-processing of images:** The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique has been used to increase the contrast of the images during the pre-processing phase. This method is used for enhancing the visual quality of images and makes it simpler for the model to detect features.
3. **Creation of training and testing datasets**: Training and testing datasets are created by dividing the images into 70:30 ratio.
4. **Proposal and implementation of a CNN model:** An optimized convolution neural network model has been implemented for classifying CXR based images. Later, its performance is evaluated using several parameters: loss, accuracy, recall, F1 score, and precision.
5. **Comparison of results:** By the end of this research, the results of the optimized CNN model is compared to those of existing techniques that used the TB x-ray dataset. This comparison aids in determining the efficacy of the model in contrast to previous methods.

1.2 Organization of the Manuscript

The manuscript is structured in such a way where Sect. 1 represents the introduction part, which describes tuberculosis and the role of AI in detecting and classifying tuberculosis. Section 2 explains the related work to inform us of researchers' contribution to detecting and classifying tuberculosis using various learning models. Section 3 presents the method to develop the CNN model, followed by Sect. 4, where results are discussed and analyzed. At last, the paper has been summarized in Sect. 5, i.e., the conclusion section.

2 Literature Survey

This section informs us regarding the work of the researchers who have contributed to the field of detecting as well as classifying tuberculosis using various learning models.

Al Ghamdi and Mohammed A [10] proposed a hybrid deep learning model which distinguished the people suffering from COVID-19 from tuberculosis or healthy individuals. The model combined the Residual Network (ResNet18) and Visual Geometry Group from Oxford (VGG16) algorithms to achieve more than 94% accuracy for multiclass identification. Their paper also highlighted the challenges faced by medical staff in diagnosing COVID-19 and stated the efficiency of their proposed model on being compared with other machine and deep learning models. *Kant and Srivastava* [11] proposed a deep neural network-based method to diagnose tuberculosis by detecting mycobacterium tuberculosis bacillus bacteria from microscopic sputum images. During experimentation, the model computed the precision and recall value of 83.78% and 67.55%, respectively. As per the researchers, their model could efficiently and quickly detect tuberculosis. *Hansun et al.* [12] implied that machine and deep learning algorithms had shown promising results in detecting tuberculosis on chest X-rays. The studies analyzed in this systematic literature review provided evidence that these algorithms could achieve high levels of sensitivity and specificity, and in some cases outperformed human experts in detecting TB. The use of technology for TB detection had the potential to improve diagnostic accuracy, reduce workload for healthcare professionals, and increase access to screening in low-resource settings. *Munadi et al.* [13] worked on the publically available dataset, i.e., The Montgomery and Shenzhen dataset, to detect and classify tuberculosis. Initially, the researchers enhanced the image to obtain or highlight the image's local characteristics, including some features. The images were used to train ResNet and EfficientNet, where they obtained 89.92% accuracy and 94.8% AUC. Similarly, an optimized machine learning-based model was proposed by *Olfa Hrizi and colleagues* [14], whose goals were to increase the accuracy and reduce the extracted characteristics. To do so, the researchers applied a support vector machine (SVM) classifier and genetic algorithm for the best selection of features. The work was conducted on ImageCLEF 2020 dataset, which was their primary resource. As per the results of their experiments, the SVM classifier proved to be the best compared to the other algorithms. In the same way, *Ravi et al*. [15] presented a robust approach for detecting diseases in lungs using X-rays of the chest, employing a multichannel deep learning strategy. EfficientNetB0, B1, and B2 models were fused to extract features, followed by nonlinear fully connected layers. A stacked ensemble classifier with random forest, support vector machine (SVM), and logistic regression technique was applied. The method demonstrated superior performance in detecting pneumonia, TB, and COVID-19, outperforming comparable methods. The t-SNE visualization verified optimal feature learning. The proposed approach achieved remarkable accuracy: 98% for pediatric pneumonia, 99% for TB, and 98% for COVID-19 detection. Its potential as a point-of-care diagnostic tool for radiologists was emphasized. *Singh et al.* [16] worked on the binary classification of the TB data using a convolutional neural network. The main aim of the researchers was to work on the objective function which had been obtained from CNN models to help doctors and the medical sector to identify TB patients accurately. During the experimentation, it was found that the best CNN model generated 6.503 as the maximum objective function

value along with the 96.71% testing accuracy as well as 97.33% AUC score. *Moris et al.* [17] also worked on Shenzhen and Montgomery County datasets to develop a model for detecting and classifying tuberculosis. The dataset was small; hence they used Generative Adversarial Network to train the deep learning model and had obtained 88.41% and 90.33% accuracies, respectively.

3 Proposed Methodology

In this section, the framework to design the tuberculosis detection system has been proposed in *Fig. 2,* for which various phases are described in detail. The experiment was conducted in Python using Google Colab by importing libraries such as OpenCV, glob, matplotlib, splitfloders, NumPy, Keras, Seaborn, and sklearn.

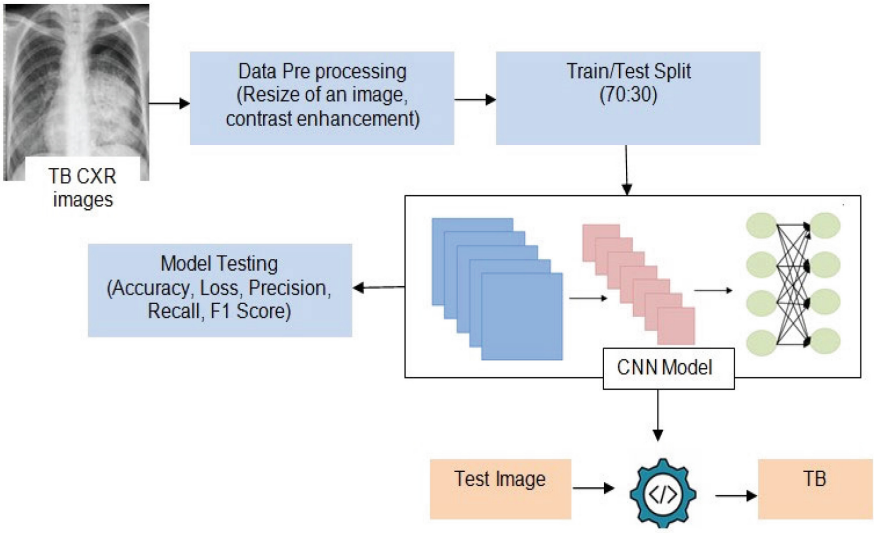


Fig. 2. Proposed system to detect and classify tuberculosis

Dataset: In this work, the optimized CNN model has been trained with the tuberculosis X-ray database in which 3500 normal images and 700 TB images are publically accessible [18]. The rest of the 2800 TB images can be taken from the NIAID TB portal by signing an agreement. The Tuberculosis X-ray dataset is a collection of X-ray lung datasets i.e., Shenzhen and Montgomery dataset as well as Belarus dataset which was collected for resisting the drug and was initiated by the National Institute of Allergy and Infectious Diseases (NIAID).

Data Preprocessing: Pre-processing of images is done to improvise the image's quality so that it can be analyzed better. In addition, pre-processing also helps to enhance the features that are later required for the classification we are working for. In this work,

the pre-processing has been done by removing the erroneous image and later the size of the rest of images are reduced from its original size, i.e., **512 × 512** to **224 × 224**, to improve the model training time without significantly reducing the performance of model. Further, we have improved its contrast and removed noise using the CLAHE technique for better understanding of the image, as shown in *Fig. 3*.

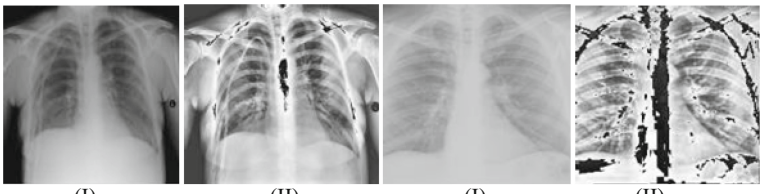


Fig. 3. (I) Resized Image (Ii) Contrast Enhancement Images

The CLAHE technique is an extended version of the Adaptive Histogram Equalization technique, which is particularly used for enhancing and improving the contrast of an image. While applying the CLAHE technique, the image is divided into smaller regions, called tiles, for equalizing their histogram. With this approach, the contrast gets enhanced in the localized regions, which are mostly helpful to deal with images that vary during lightning conditions. Additionally, CLAHE also incorporates the mechanism to limit the contrast of an image using a clip limit to avoid amplifying noise and preventing over-enhancement. 40 is the default threshold value for a clip limit, but in this research paper, we have used 10 as a value for a clip limit to preserve the natural appearance of the image and reduce the risk of over enhancing.

However, if the value of clip limit is exceeded by the histogram of a tile then it is redistributed within the bins unless the satisfactory limit is obtained. This process although enhances the contract but simultaneously also preserve and protect the originality of an image. Besides contrast-limiting, CLAHE also addresses the issue of artificial boundaries that may emerge between neighboring tiles. These borders can cause noticeable artifacts and disturb the continuity of an image. Hence, in order to deal with this problem, CLAHE uses bilinear interpolation to coordinate the brightness of pixels within the borders of neighboring tiles. The artificial borders are minimized while smoothly transitioning between adjacent tiles, which results in a more seamless and visually pleasing image. Another important component in CLAHE is the tile grid size. It specifies how many tiles are used to divide the image. A tile grid size of 8×8 is generally used by default; however, it can be modified based on the image's qualities or special requirements [19].

CNN Model: After pre-processing, the image is sent to the optimized convolution neural network (CNN) for extracting the features and classifying them. As shown in *Fig. 4,* the CNN model is divided into three layers, i.e., a convolution layer, a pooling layer, and a fully connected layer which are briefly described as under [20]:

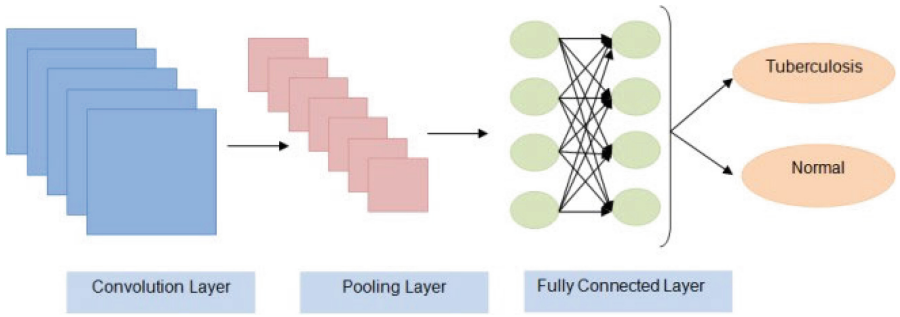


Fig. 4. Architecture of CNN-Model

Convolution Layer: In this layer, a dot product is performed between two matrices in which one set of matrices is known as the kernel. The kernel is also called the learnable parameter, and its size is mainly taken in odd digits such as 3×3 or 5×5, which is smaller than the size of the image but is more in-depth. This also indicates that the width and height of the kernel will be small if the image's composition is of three channels, i.e., Red, Blue, and Green (RGB).

Suppose if the size of an input image is $S \times S \times W$ and *Wout* is the number of kernels which is having a size *X,* stride *Y, Z* is the amount of padding then the size of output can be computed by the following eq (1):

$$W_{out} = \frac{S - X + 2Z}{Y} + 1 \tag{1}$$

Pooling Layer: The function of this layer is to reduce the dimensions of an image. This layer derives the summary statistic of the output by replacing the network's output at specific locations.

The pooling operation takes place on every part of the output image taken from convolution layer individually. If the size of activation map is $S \times S \times W$, spatial size of pooling kernel is $\times$ and having and having Y stride, then the output volume's size can be calculated by the following Eq. (2):

$$W_{out} = \frac{W - X}{Y} + 1 \tag{2}$$

Fully Connected Layer: In CNN, this layer is the last layer where all the input neurons are fully connected to output neurons in the previous and the subsequent layers. By multiplying the matrix with the bias the fully linked layer aids in mapping the representation between the input and the output.

To *extract and classify features* of tuberculosis X-ray image dataset, we have used a optimized CNN model in which initially, 32 layers of convolution layer has been taken with 3×3 kernel size i.e. filter size, *ReLu* activation function as well as *softmax* activation function, *adam* optimizer, *binary cross-entropy* loss function, and *input shape* of 224×224 along with the parameters 4,43,96,609 for training purpose. Later 64 layers of convolution layer have been added again with the same hyper-parameters. In the second phase of pooling layer, MaxPooling2D has been chosen with 2×2 pool

size and a dropout of 0.25. After this, the same layer size of convo layer has been again added with 3 × 3 kernel size, 2 × 2 pooling size, and 0.25 drop out. The sequences have been again repeated but with 128 layer of convolution layer by keeping rest of the hyper-parameters same.

In the final stage of fully connected layer, the flatten layer has been added with 64 layers of dense layer, and dropout value of 0.5. The second and the last dense layer chosen is of 1 layer where sigmoid activation function (Eq. 2) takes the real number and squash it within the range of 0 to 1 for classifying the input image as tuberculosis or normal.

$$\sigma(\kappa) = 1/(1 + e\kappa) \tag{3}$$

Table 1 depicts the architecture of an optimized convolutional neural network (CNN) employed for image classification. The table details the CNN's layers, their output shape, as well as the number of parameters each layer uses.

Table 1. Layered architecture of the optimized CNN Model

Layer	Output Shape	Param
Conv2D	(None, 222,222,32)	896
MaxPooling2D	(None, 111,111,32)	0
Conv2D_1	(None, 109,109,64)	18496
MaxPooling2D_1	(None, 54,54,64)	0
Conv2D_2	(None, 52, 52, 128)	73856
MaxPooling2D_2	(None, 26,26,128)	0
Flatten	(None, 86528)	0
Dense	(None, 512)	44302848
Dropout	(None, 512)	0
Dense_1	(None, 1)	513

1. **Conv2D:** This layer is a 32-filter 2D convolutional layer, which applies 32 convolution operations to the input image. (None, 222, 222, 32) is the output shape, where None is the batch size, and 222 × 222 is the spatial dimension of the output feature maps. Moreover, there are 896 parameters that have been used in this layer.
2. **MaxPooling2D:** This layer uses max-pooling for the reduction of spatial dimensions of the input feature maps. Since the max-pooling operation halves the spatial dimensions, the output shape is (None, 111, 111, 32), (None, 54, 54, 64), (None, 26, 26, 128). This layer has no parameters because the operation it performs is fixed.
3. **Conv2D_1** is an additional convolutional 2D layer with 64 filters. Using convolutional operation, the output shape is (None, 109, 109, 64). There are 18,496 parameters used in this layer.

4. **Conv2D_2** is an additional convolutional 2D layer with 128 filters. Due to the convolutional operation, the output shape is (None, 52, 52, 128), with smaller spatial dimensions than the previous layer. There are 73856 parameters used in this layer.
5. **Flatten:** The Flatten layer is used for converting the final Conv2D layer's output, which has the shape (None, 26, 26, 128), into a one-dimensional vector with the shape (None, 86528). This is accomplished by flattening the output tensor along the spatial dimensions (height, width) while maintaining the dimension of the channel. This vector is then passed as an input to the subsequent fully connected layer of the CNN, which undergoes additional processing to perform classification or regression tasks.
6. **Dense:** This layer has 512 neurons and is fully interconnected with an output shape of (None, 512). This layer uses 4,43,02,848 parameters, a substantial number compared to previous layers. This is because each neuron in this layer is connected to each element in the previous layer's flattened vector, resulting in many parameters.
7. **Dropout:** This layer randomly eliminates a percentage of the input units during training to prevent overfitting. Because this layer does not modify spatial dimensions, the output shape is identical to the input shape. This layer has no parameters because the operation it performs is fixed.
8. **Dense_1:** This is the final output layer with a single neuron because this CNN is used for binary classification, i.e., classifying an image into one of two classes. The output shape is (None, 1) because this layer contains only one neuron. Since only one neuron is in this layer, the number of parameters used is comparatively low at 513.

The parameters in each layer represent the number of trainable parameters in that layer, i.e., the number of updated weights and biases during training. This CNN has 44396609 trainable parameters as mentioned earlier, the sum of the parameters in all layers except for the MaxPooling2D and Dropout layers, which lack trainable parameters.

Evaluative Parameter: The proposed sequential CNN model has been examined using various parameters like accuracy, loss, precision, recall, as well as F1 score [24–26]. Accuracy is used to find out whether the model predicts the output of the class correctly and is calculated by using Eq. (4)

$$Accuracy = \frac{Correct\ Predictions}{Total\ number\ of\ predictions} \tag{4}$$

On the contrary, loss defines how bad the model works in predicting the output of the class. It is calculated by using Eq (5)

$$Loss = \frac{(Acutal\ value - Predicted\ value)^2}{Total\ number\ of\ predictions} \tag{5}$$

Precision and recall are the two parameters which are used to recognize and classify patters in CNN model. Precision is the ratio of true positive to the total number of positive samples as shown by Eq. (6)

$$Precision = \frac{True\ Posistive}{True\ Positive + False\ Positive} \tag{6}$$

On the other hand, recall is calculated for measuring ability of CNN model for detecting samples of the positive class. In fact, the higher the value of recall, the more positive samples will be detected. It is computed by Eq. (7)

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{7}$$

In a nutshell, we can summarize that if the value of true positive is higher than false positive, then the precision of the model will be high else it will be low. Likewise, the recall of CNN model will be high if the value of true positive is greater than false negative else it will be also low. Although a good model should have high recall and precision values but sometimes there is a trade-off between both. This trade-off can be dealt on the basis of the problem, the techniques that have been used to achieve the goal, as well as by adjusting the classification threshold value. The last used is F1 score and its value ranges between 0 and 1. If the F1 score value is closer to 1, the better is the model performing.

$$F1\ score = 2\frac{Precision * Recall}{Recall + Precision} \tag{8}$$

4 Results

In this section, the proposed convolutional neural network has been evaluated based on the parameters mentioned in Sect. 3 along with their findings. Besides this, the graph has also been presented to analyze the relation between training and testing accuracy as well as loss curves. In addition to this, the results for other metrics such as recall, F1 score, and precision of models and classes have also been shown.

Table 2. Evaluation of the CNN Model

Training	Accuracy	93.97%
	Loss	0.15
Testing	Accuracy	91.2%
	Loss	0.16

From Table 2, it is observed that while training the model with the training dataset, the optimized convolution neural network computed 93.97% accuracy with 0.15 of loss value. On the contrary, during testing phase, the accuracy obtained is highest than training accuracy with 91.2% on the loss of 0.16. The model has been trained by taking 10 epochs, 91 steps per epoch and 20 testing steps.

Besides this, the training and testing curves of convolutional neural network have been analyzed (Fig. 5), and it has been found that to train and test loss curves, the model showed a good fit of learning curves, which means the curves decrease to a certain

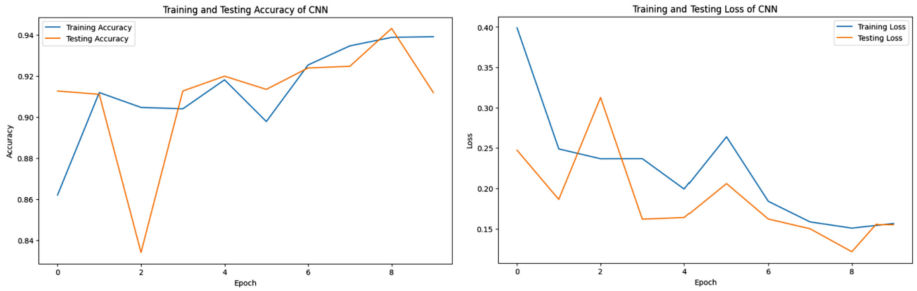


Fig. 5. Accuracy and Loss curves of the CNN Model

stability point with a minimum gap in between the final points of two values of loss. On the contrary, for the training and testing accuracy, noisy signals have been generated, which means that sufficient information has not been provided by the testing dataset for evaluating the ability of the model. It can also occur if the testing dataset has fewer images than the training dataset. The optimized CNN model has been also examined using F1 score, precision as well as recall score, as presented in Fig. 6.

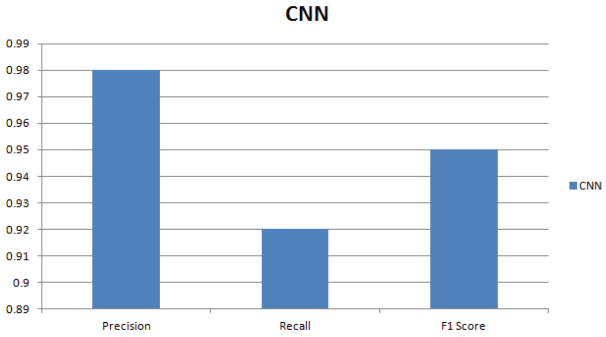


Fig. 6. Examination of CNN Model

It has been found that our proposed CNN model has obtained 0.98 precision value, 0.92 recall, and 0.95 F1 score value. The same parameters of models have been also calculated for two different classes i.e. Normal and Tuberculosis as presented in Table 3.

Table 3. Performance analysis of CNN model

Classes	Precision	Recall	F1 Score
Tuberculosis	0.98	0.83	0.92
Normal	0.99	1.00	0.99

It can be seen that by computing the precision of 0.98 and 0.99 for tuberculosis and normal class respectively indicates that the CNN model correctly identifies true positive

cases. On the contrary, the optimal value of recall for normal class defines that the model is effective in capturing the most of actual positive instances and based on these values F1 showcases the strong performance in terms of both false positive and false negative.

Moreover, based on the techniques, TB dataset, and accuracy, the comparison of the results computed by the proposed CNN model and the existing techniques is shown in Table 4.On analyzing the table, it can be seen that our optimized model has worked quite well for the classification and detection of TB in CXR images.

Table 4. Analyses of optimized CNN-Model with the existing techniques

Ref	Year of Publication	Dataset	Techniques	Accuracy
Liu et al. [23]	2017	TB image dataset	CNN	85.68%
Norval et al. [21]	2019		AlexNet	89.99%
Tasci [22]	2020	Montgomery and Shenzhen dataset	VGG	88.32%
Munadi et al. [13]	2020		ResNet	89.92%
Moris et al. [17]	2022	TB X-Ray dataset	CNN	90.33%
Our study	**-**	***TB Chest 'X-Ray dataset***	***Optimized CNN***	***91.2%***

5 Conclusion

In large-density areas, artificial intelligence algorithms are more effective tools for TB identification than human readers. The first and most challenging step in eliminating tuberculosis is accurate detection and diagnosis, which requires techniques that are both accessible and inexpensive. The current deep learning-based model presented in this paper can automatically identify and classify the TB disease in CXR images at a level equivalent to that of human radiologists. This allows radiologists to save time and, more significantly, helps to solve the problem of a scarcity of human resources. But at the same time, there are certain limitation, such as the scarcity of the dataset, i.e., the work has been only conducted on 3500 normal images and 700 TB images. In addition, the Normal class of the dataset contains one erroneous image, which must be removed prior training the model. Additionally, the layers of the CNN model could be fine-tuned to enhance the prediction and classification efficiency as well as the accuracy of the model. In fact, the current work on extending to a large dataset can incorporate advanced artificial intelligence-based techniques as an alternative to the proposed methodology and can be deployed in the medical sector for better human care.

References

1. Bhandari, J., Thada, P.K.: NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health

2. McQuaid, C.F., Vassall, A., Cohen, T., Fiekert, K., White, R.G.: The impact of COVID-19 on TB: a review of the data. Int. J. Tuberc. Lung Dis. **25**(6), 436–446 (2021)
3. Zhou, W., et al.: Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. Quant. Imag. Med. Surg. **12**(4), 2344 (2022)
4. World Health Organization: Tuberculosis deaths rise for the first time in more than a decade due to the COVID-19 pandemic. Press release Geneva (2021).
5. Puttagunta, M.K., Ravi, S.: Detection of Tuberculosis based on Deep Learning based methods. J. Phys.: Conf. Ser. **1767**(1), 012004 (2021). https://doi.org/10.1088/1742-6596/1767/1/012004
6. Bohr, A., Memarzadeh, K.: The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare, pp. 25–60. Elsevier (2020). https://doi.org/10.1016/B978-0-12-818438-7.00002-2
7. Kumar, Y., Gupta, S., Gupta, A.: Study of machine and deep learning classifications for IOT enabled healthcare devices. In: 2021 International Conference on Technological Advancements and Innovations (ICTAI), pp. 212–217. IEEE (2021)
8. Prasad, R.S., Waghmare, R.C., Pajgade, T.B., Raut, R.R., Mahajan, M.L.: A Comparative study of detection of tuberculosis using machine learning & deep learning. In: 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1217–1221. IEEE (2023)
9. Kumar, Y., Koul, A., Kaur, S., Hu, Y.C.: Machine learning and deep learning based time series prediction and forecasting of ten nations' COVID-19 pandemic. SN Comput. Sci. **4**(1), 91 (2022)
10. Al Ghamdi, M.A.: Identification of tuberculosis and coronavirus patients using hybrid deep learning models. Comput. Mater. Continua **76**(1), 881–894 (2023). https://doi.org/10.32604/cmc.2023.037826
11. Kant, S., Srivastava, M.M.: Towards automated tuberculosis detection using deep learning. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1250–1253. IEEE (2018).
12. Hansun, S., Argha, A., Liaw, S.-T., Celler, B.G., Marks, G.B.: Machine and deep learning for tuberculosis detection on chest X-Rays: systematic literature review. J. Med. Internet Res. **25**, e43154 (2023). https://doi.org/10.2196/43154
13. Munadi, K., Muchtar, K., Maulina, N., Pradhan, B.: Image enhancement for tuberculosis detection using deep learning. IEEE Access **8**, 217897–217907 (2020)
14. Hrizi, O., et al.: Tuberculosis disease diagnosis based on an optimized machine learning model. J. Healthcare Eng.(2022)
15. Ravi, V., Acharya, V., Alazab, M.: A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. Cluster Comput. **26**(2), 1181–1203 (2022). https://doi.org/10.1007/s10586-022-03664-6
16. Singh, V., Gourisaria, M.K., Harshvardhan, G.M., Singh, V.: Mycobacterium tuberculosis detection using CNN ranking approach. In: Advanced Computational Paradigms and Hybrid Intelligent Computing: Proceedings of ICACCP 2021, pp. 583–596. Singapore: Springer Singapore (2021)
17. Morís, D.I., de Moura, J., Novo, J., Ortega, M.: Unsupervised contrastive unpaired image generation approach for improving tuberculosis screening using chest X-ray images. Pattern Recogn. Lett. **164**, 60–66 (2022)
18. Rahman, T., et al.: Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. IEEE Access **8**, 191586–191601 (2020)
19. Koul, A., Bawa, R.K., Kumar, Y.: Artificial intelligence techniques to predict the airway disorders illness: a systematic review. Archiv. Comput. Methods Eng. **30**(2), 831–864 (2023)

20. Kaur, D., et al.: Computational intelligence and metaheuristic techniques for brain tumor detection through IoMT-enabled MRI devices. Wirel. Commun. Mob. Comput. **2022**, 1–20 (2022)
21. Norval, M., Wang, Z., Sun, Y.: Evaluation of image processing technologies for pulmonary tuberculosis detection based on deep learning convolutional neural networks (2019)
22. Tasci, E.: Pre-processing effects of the tuberculosis chest x-ray images on pre-trained cnns: an investigation. In: Artificial Intelligence and Applied Mathematics in Engineering Problems: Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2019), pp. 589–596. Springer International Publishing (2020)
23. Liu, C., et al.: TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In: 2017 IEEE International Conference on Image Processing (ICIP) (pp. 2314–2318). IEEE (2017)
24. Koul, A., Bawa, R.K., Kumar, Y.: Artificial intelligence in medical image processing for airway diseases. In: Connected e-Health: Integrated IoT and Cloud Computing, pp. 217–254. Springer International Publishing, Cham (2022)
25. Kakkar, B., Johri, P., Kumar, Y., Park, H., Son, Y., Shafi, J.: An IoMT-based federated and deep transfer learning approach to the detection of diverse chest diseases using chest X-rays. Hum. Centric. Comput. Inf. Sci. **12**, 24 (2022)
26. Bhardwaj, P., Kumar, S., Kumar, Y.: A comprehensive analysis of deep learning-based approaches for the prediction of gastrointestinal diseases using multi-class endoscopy images. Archiv. Comput. Methods Eng. **30**(7), 4499–4516 (2023). https://doi.org/10.1007/s11831-023-09951-8



Performance Evaluation of Machine Learning Algorithms for Predicting Solar Radiation in Chennai Province

A. Geethakarthi and V. P. Sumathi(✉)

Kumaraguru College of Technology, Coimbatore, Tamilnadu 641049, India
ageethakarthi@yahoo.com

Abstract. Effective storage and capture of global solar radiation (GSR) is significant as an alternate energy source. The global radiation levels are influenced by various meteorological and air pollutants parameters. Four significant parameters (temperature, relative humidity, wind direction and ozone) were considered to predict global solar radiation using XG-Boost, Random Forest (RF), kNearest Neighbour (k-NN), Support Vector Machine (SVM) and Decision tree (DT) for Alandur provinces of Chennai district, Tamil Nadu. The performance of models was improved through tenfold cross validation. The predicted solar radiation for different ML algorithms yielded an accuracy of 77.42% (XG-Boost), 70.97% (k-NN), 79.26% (RF), 69.59% (DT) and 72.81% (SVM) respectively. Out of the five different MLs, RF algorithm showed best performance and achieved a higher accuracy of 79.26% with precision of 78%, recall and F1 score as 79% respectively for the given air pollutants and meteorological dataset.

Keywords: Air pollution · Correlation analysis · Machine Learning · Meteorological · Global Solar radiation (GSR) · Prediction models

1 Introduction

Over exploitation of fossil resources to meet the energy demand leads to environmental degradation such as global warming and climate change. Furthermore, it plays a critical role in the decline of air quality, which leads to major health issues like cancer, conjunctivitis, and respiratory disorders [1]. Individual countries are advised to develop their own renewable energy sources as a substitute for the current fossil fuel, according the instructions of the Intergovernmental Panel on Climate Change's sixth assessment report (IPCC, 2017). Solar energy is most significant sources of renewable energy and can be utilized through Photo-Voltic (PV) panel installation to meet the electricity demand [2]. The production of the solar energy is affected by several factors such as duration and quantum of solar radiation, temperature, humidity, vegetative cover, earth's terrain and other atmospheric conditions. All these factors have a crucial impact on the outpower and performance of the PV panel [3]. Installation of PV systems in developing countries have higher installation and maintenance cost due to the lack of daily meteorological data in most of the monitoring stations. Prediction of solar radiation is one such approach to evaluate the radiance reaching earth's surface.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 56–64, 2025.
https://doi.org/10.1007/978-3-031-75167-7_5

2 Literature Review

Mitigation and adaptation of climate change to lessen the severity of its impact, requires the use of renewable energy sources. The earth-atmosphere fluxes are linked by solar radiation, which acts as an energy source for the creation of physical phenomena and alteration of atmospheric processes. Considerable reductions in solar energy's share in the world's energy system could lead to a considerable decrease in carbon emissions. A reliable estimate of solar radiation is crucial for system installation in each region. The prediction of solar radiation can be evaluated through empirical models, machine learning-based models and satellite-based models [4]. Empirically based models are the most popular approach among the previously suggested approaches because of its simplicity, accessibility, and low computational cost. By combining the various variables, further work has been done to estimate and predict the solar radiation [5]. Among the three methods, sunshine based and cloud cover-based models provide a better prediction than temperature-based models. Due to the lack of sunshine hour and cloud cover data with most of the monitoring stations around the world, temperature, relative humidity and precipitation are highly preferrable in predicting solar radiation [6]. Empirical-based models are complicated in designing multidimensional parameters and could not be used for prediction. However, ML techniques like Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Networks (ANN), Decision Trees (DT), k-NN, SVM, Random Forest and XG-Boost [7] yield better performance. The earlier literatures on solar radiation prediction had shown precise results through ML algorithms than empirical equation models. Designing empirically based models is challenging with multidimensional parameters and could not be used for prediction. ML algorithm models like ANFIS, ANN and SVM were considered for solar radiation prediction using extra-terrestrial minimum and maximum temperature and solar radiation [8]. Higher performance outputs were achieved in SVM with $R^2 = 0.689$, Mean Absolute Error (MAE) = 1.97 and Root Mean Square Error (RMSE) = 2.578. In one of the research paper [9], two different ML algorithms (k-NN, ANN) and empirical models were used. k-NN showed best result with 0.96 R^2 value. Author [10] used RF, CART, M5, MARS algorithms to predict solar radiation. The authors had used meteorological dataset such as minimum, maximum and average temperature, wind speed, pressure, solar azimuth and solar radiation. Forecasting errors were found to be high in the cloudy seasons and the best results was achieved with the random forest. In a comparative study carried out by Wang [11], XG-boost was recommended as an assured model of best fit than SVM and other empirical models in forecasting solar radiation. In a study carried out [12], By using WEKA to determine the solar radiation's influential characteristics, the prediction models' accuracy was increased. This work seeks to compare the performance of several machine learning models and discover the key parameters and their effects on solar radiation utilising atmospheric conditions. In order to assess the uncertainty in GSR simulations caused by errors caused by the meteroligal and surface parameter data, hybrid models that can estimate global solar radiation (GSR) must be created. Both computing speed and accuracy should be very good for these models. For radiative transfer, the radiative transfer model (RTM) library was integrated with six machine learning models. RTM-RF demonstrated superior performance, as evidenced by its mean absolute errors (MAE) and coefficients of determination (R2) of 15.57 Wm−2 and 0.98,

respectively [13]. Relative humidity, wind speed, precipitation, maximum and lowest temperatures, and other features are chosen by the RF Regressor algorithm for a given month; the feature ranking may change based on the model that is employed. The selection of features produced by the DT algorithm could differ slightly. Though the models differ, they all perform admirably; the LR model, with its low RMSE (0.003) and greatest R2 score (0.002), stands out among the others [14]. Gaussian process regression (GPR), K-nearest neighbours (KNN), long short-term memory (LSTM), extreme learning machines (ELM), and support vector machine regression (SVMR) are the models that are employed. The ANOVA and KW test results show that in Turkey's semi-arid and arid regions, the LSTM and GPR algorithms are both relevant and valid as an alternative to SR forecasting [15].

3 Materials and Methods

3.1 Geographical Aspects of Study Area

India sub-continent dominates the greater part of South Asia and is located at 20.59°N longitude and 78.96°E latitude respectively. India has 22 major states and covers a wide land area nearly about 3.287 million km^2 (Fig. 1).

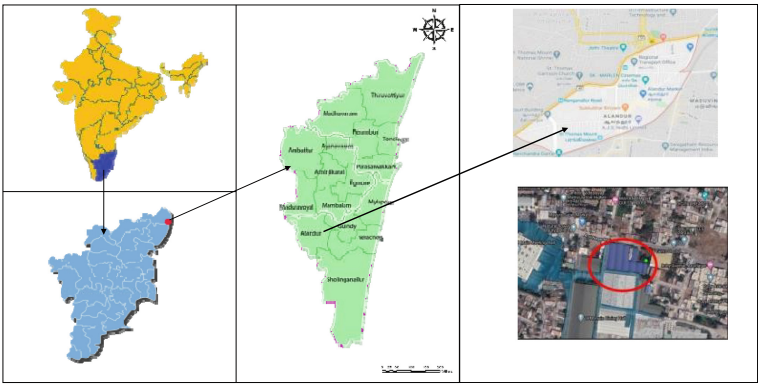


Fig. 1. Geographical locations of study area, boundary map of Alandur and monitoring station in Alandur, Chennai region.

According to the report of Indian meteorological department data, annual average global solar radiation of India is 200 MW/km^2. The geographical details and view of province on the India and Tamil Nadu map are shown in figure. In Tamil Nadu, Chennai's growing industrialization, urbanisation, and motorization have made it one of the most populous and polluted cities in the world. It covers 1189 km^2 and has 10.97 million population. The study area of selection Nestled between 3.03°N and 80.21°E, the Alandur zone is one of Chennai's metropolitan nodes. It is one of the congested areas with high vehicular movement, industrial and residential area with the nearest air monitoring station around 2 km radius.

3.2 Data Collection on Significant Parameters

The assessment of the solar radiation for Alandur in Chennai is the main goal of the current study and the expected level of GSR reaching the ground and in the next 2 decades. 15 meteorological and air pollutant criteria such as temperature, wind direction, wind speed, cloud cover, relative humidity, precipitation, PM2.5, benzene, toluene, NO, NO2, CO, SO2, ozone and solar radiation were collected form the Alandur station for five years (2015–2019). The dataset, which spans the period from January 1, 2015 to December 31, 2019, was provided by the Public Works Department (PWD) and the Central Pollution Control Board (CPCB).

A sample data of 1085 was considered for designing the prediction model with chosen input parameters like temperature, humidity, wind direction, ozone and decision classes with respect to solar radiation. Data cleaning and pre-processing was done prior to splitting of dataset. The data split up ratio used in training and testing dataset was 80% and 20%. The data split-up used for training were fed to the developed models to validate model performance on GSR prediction under each category and the models were validated through tenfold cross validation.

4 Statistics on ML Algorithm Study

In multi-class issues, there is a need for a comprehensive understanding of classification algorithms. Machine learning (ML) model is a branch of Artificial Intelligence (AI) and is one of the popular and attractive techniques, which can be implemented into various fields. Classification model assists in designing an analytical model with improved performance through self-learning ability from the trained dataset [16]. In this study, five different classification algorithm such as SVM, RF, XG-Boost, DT and k-NN were used. These five algorithms were implemented using python scikit-learn packages. The research area's monitoring stations provided the meteorological and air pollution indicators used in this work to forecast global solar radiation. From the collected 15 parameters, the most significant ones were identified and segregated by analyzing correlation between parameters and decision variables to enhance the ML model's performance. All algorithms in the models were trained using the same dataset to measure the performance of various model using test dataset.

4.1 Classification Algorithms

Decision Tree

This algorithm is easy to understand and is a popular method for constructing predictive models. Decision tree model can solve the classification and regression problems. To build a decision tree, the various algorithms used are CART (classification and regression tree algorithm), ID3, C4.5, CHAID and MARS. Important attribute selection technique is one of the key parameters in decision tree algorithm. Entropy, information gain, Gini index, gain ratio, variance reduction, and chi-square are some of the tools it uses to improve prediction accuracy and outcomes [17]. In this study, Gini index was used for

parameter selection, splitting the training and testing data and a maximum depth of the tree of 5 was considered.

K – Nearest Neighbor (k-NN)
K-NN model can compete with the most accurate and high-performance model due to its prediction accuracy. This is one of the oldest non-parametric supervised ML algorithms in solving regression problem. It relies on labelled input data for predicting output with the given test dataset. The classes will be decided by majority voting technique for its neighboring dataset class labels. The new dataset decision class depends on most common class amongst its k-nearest neighbors. The quality of the predictions depends on selection of k value and distance measured with the functions such as Euclidean, Manhattan, Min-Kowski and Chebyshev respectively. The processing time and distance calculations increases proportionally with the dataset and k size [18]. The observed and predicted results showed most appropriate results with a k value of 10.

Support Vector Machine (SVM)
SVM is a supervised ML algorithm based on statistical learning theory, used for outlier detection and it is widely applied for forecasting. SVM are different from other classification algorithm type in the way of choosing the decision boundary. By creating a multidimensional hyperplane that maximizes the margin between classes, the SVM method effectively conducts a classification by discriminating between two classes. The most crucial parameters in SVM are the mathematical function that are defined as kernel [19]. It improves to discriminate the classes based on computed value from the input data. Radial Basis Function was used as kernel type in SVM algorithm in this work.

4.2 Ensemble Algorithm Techniques

Random Forest (RF)
RF is a supervised learning algorithm to solve the classification and regression problems. Unlike XG-Boost, RF algorithm constructs more decision trees. These decision trees are combined to make decisions using bagging techniques. The algorithm creates decision trees based on input attributes and then the prediction done from each of them and finally the best results found by majority voting technique. It builds multiple decision trees and merge them together to get more accurate model. The ensemble method helps to reduce over-fitting by averaging the results. The hyperparameters are important in this algorithm and used to increase the predictive power and model speed [10]. In this paper, the learner bagging method is used to combine the performance of various decision trees models and Gini index and information gain were used to select the parameter will be placed as decision node at each level in the constructed decision tree.

XG-Boost
It is a tree-based ensemble ML technique that improves prediction power and performance by improvising on the Gradient Boosting framework and incorporating some reliable approximation algorithms. In XG-Boost, the node parameter is found by similarity score and gain. The regularization of parameter is to avoid overfitting of the split

data using two techniques such as shrinkage scales and column subsampling. Approximation algorithm is used to determine the best split. The hyperparameters are significant parameters used to train the model [20]. In this study, linear booster was used as objective function, the depth of the tree was fixed based on accuracy of the model and alpha method was used to perform regularization of the parameter.

5 Results and Discussion

5.1 Model performance Optimization

A correlation study between the influential parameters and solar radiation levels were analysed and the coefficient value of each parameter was computed as shown in Fig. 2. Based on the correlation analysis, it was identified that the influential parameters on the solar radiation are temperature, relative humidity, wind direction, and ozone. Figure 2 shows the strong positive correlation between temperature, wind direction with solar radiation having a positive correlation coefficient (0.38 and 0.25) and relative humidity showed a negative correlation (−0.39). In real time, solar radiation is the major influencing parameter besides temperature and humidity in the installation of PV panel. The other parameters affecting the intensity of solar radiation are elevation angle, latitude, longitude, cloud density etc. These factors should be taken into account in order to increase model accuracy and reduce the prediction model's complexity.

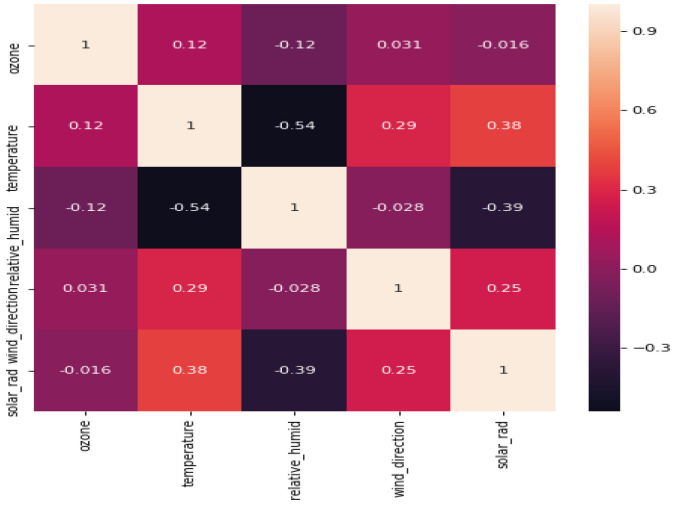


Fig. 2. Computation of correlation value for the influential parameters

5.2 Performance of ML Models

The performance efficiency of ML algorithms was evaluated and compared using the model precision, recall, accuracy and F1 score. Figures 3 and 4 shows the performance study for the data split-up ratio (80:20) for the chosen influencing parameters.

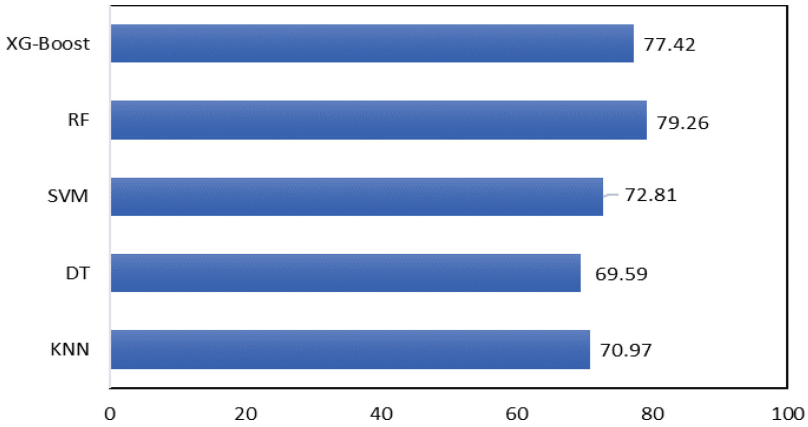


Fig. 3. Performance of Classification Models (Accuracy in %)

The results of this investigation showed that, when compared to the other algorithms, the RF and XG-Boost algorithms were more appropriate for predicting solar radiation. From Fig. 3, the obtained accuracy of RF and XG-Boost algorithm were 79.26% and 77.42% respectively (Table 1).

Table 1. Multiple Decision Classes in Solar Radiation

Classes	Ranges (W/m^2)
Class 1	0–49
Class 2	50–199
Class 3	200–299
Class 4	$\geq$ 300

The higher model accuracy achieved was due to the ensembling techniques adopted in the data analysis [21, 22]. k-NN classifier, Decision Tree and SVM algorithms showed lesser accuracy for the given category of dataset. RF algorithm and XG-Boost showed (Fig. 4) higher precision, recall and F1 score value compared with the other algorithms like k-NN, DT and SVM.

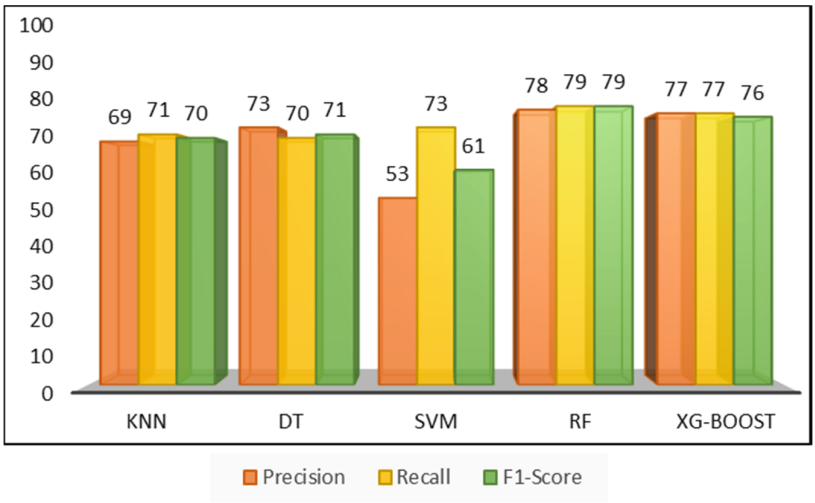


Fig. 4. Performance Metrics (in %) of Classification Algorithms

6 Conclusion

The suitability and prediction of GSR for Alandur province, Chennai was studied through five different machine learning algorithms. RF was found to be most successful with an accuracy value of 79.26% for higher input parameters. An ensemble learning method for classification while constructing a multitude of decision trees at training time yields better accuracy for the RF model. The performance of ensemble technique had proven to be more appropriate for solar radiation prediction for the given dataset. The performance metrics like precision, recall and F1 score were also higher in RF algorithm.

Conflict of Interest. The corresponding author and the listed co-author submitted the prepared work for potential publication in this journal. Regarding the topics covered in the paper, the authors are not associated with any organization that has a financial stake, either direct or indirect.

References

1. Jayakumar, A., Chalmers, A., Lie, T.T.: Review of prospects for adoption of fuel cell electric vehicles in New Zealand. IET Electr. Syst. Transport. **7**(4), 259–266 (2017)
2. Rizwan, M., Jamil, M., Kothari, D.P.: Solar energy estimation using REST2 model. Int. J. Energy Env. **2**, 367–374 (2010)
3. Yadav, A.K., Malik, H., Chandel, S.S.: Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew. Sustain. Energy Rev. **31**, 509–519 (2014)
4. Rajakumari, K., Priyanka, V.: Air pollution prediction in smart cities by using machine learning techniques. Int. J. Innov. Technol. Exploring Eng. **9**(5), 1272–1279 (2020)
5. Bayrakçı, H.C., Demircan, C., Keçebaş, A.: The development of empirical models for estimating global solar radiation on horizontal surface: a case study. Renew. Sustain. Energ. Rev. **81**, 2771–2782 (2018)

6. Hassan, G.E., Youssef, M.E., Mohamed, Z.E., Ali, M.A., Hanafy, A.A.: New temperature-based models for predicting global solar radiation. Appl. Energy **179**, 437–450 (2016)
7. Nazhad, S.H., Lotfinejad, M.M., Danesh, M., Amin, R.U., Shamshirband, S.: A comparison of the performance of some extreme learning machine empirical models for predicting daily horizontal diffuse solar radiation in a region of southern Iran. Int. J. Remote Sens. **38**(23), 6894–6909 (2017)
8. Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L.: ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J. Atmos. Solar Terr. Phys. **155**, 62–70 (2017)
9. Marzouq, M., Bounoua, Z., El Fadili, H., Mechaqrane, A., Zenkouar, K., Lakhliai, Z.: New daily global solar irradiation estimation model based on automatic selection of input parameters using evolutionary artificial neural networks. J. Cleaner Product. **209**, 1105–1118 (2019)
10. Srivastava, R., Tiwari, A.N., Giri, V.K.: Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India. Heliyon **5**(10) (2019)
11. Fan, J., et al.: New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China. Energy Convers. Manage. **156**, 618–625 (2018)
12. Meenal, R., Selvakumar, A.I.: Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. Renew. Energy **121**, 324–343 (2018)
13. Lu, Y., et al.: Predicting surface solar radiation using a hybrid radiative Transfer-Machine learning model. Renew. Sustain. Energy Rev. **173**, 113105 (2023)
14. Hissou, H., Benkirane, S., Guezzaz, A., Azrour, M., Beni-Hssane, A.: A novel machine learning approach for solar radiation estimation. Sustainability **15**(13), 10609 (2023)
15. Demir, V., Citakoglu, H.: Forecasting of solar radiation using different machine learning approaches. Neural Comput. Appl. **35**(1), 887–906 (2023)
16. Bonaccorso G.: Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning. Packt Publishing Ltd. (2018)
17. Saranya, R., Selvam. N.: Computation of daily global solar radiation by using decision tree algorithm. Int. J. Recent Technol. Eng. d 2277–3878 (2019)
18. Rodrigues, É.O.: Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier. Pattern Recognit. Lett. **110**, 66–71 (2018)
19. Zendehboudi, A., Baseer, M.A., Saidur, R.: Application of support vector machine models for forecasting solar and wind energy resources: a review. J. Cleaner Product. **199**, 272–285 (2018)
20. Urraca, R., Antoñanzas, J., Antoñanzas-Torres, F., Martinez-de-Pison, F.J.: Estimation of daily global horizontal irradiation using extreme gradient boosting machines. In International Workshop on Soft Computing Models in Industrial and Environmental Applications, pp. 105–113. Springer International Publishing (2016)
21. Kiangala, S.K., Wang, Z.: An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. Mach. Learn. Appl. **4**, 100024 (2021)
22. Li, Y., Chen, W.: A comparative performance assessment of ensemble learning for credit scoring. Mathematics **8**(10), 1756 (2020)

Fault Predictions Based on Base Learnings and Mean Weighted Score Using Machine Learning Techniques

Prachi Sasankar(✉) and Gopal Sakarkar

Department of Computer Science School of Science, G.H. Raisoni University, Saikheda, MP, India
sasankar.prachi@gmail.com

Abstract. Software Fault Prediction (SFP) processes find utility across various phases of the software development lifecycle, with a particularly valuable role in early identification of flawed or buggy components. Machine learning (ML) techniques are commonly harnessed in this domain. Predictive models rely on a set of 22 object-oriented metrics drawn from the CK and Martin metric collections, serving as input features. To evaluate the efficacy of these prediction models, experiments are conducted using ten publicly available datasets sourced from the PROMISE repository. Performance metrics, including Receiver Operating Characteristic (ROC) curve, accuracy (Acc), precision(Pre), F1-score(Fs), recall(Rc), and mean weighted error, are employed to gauge the predictive capabilities of the models.

In this study, nine machine learning (ML) classifiers are employed to predict the accuracy rates of software fault proneness. We utilize ten of the most popular datasets in their latest versions from the PROMISE repository to evaluate the performance of algorithms, including Logistic Regression-LR, Decision Tree-DT, Support Vector Machine-SVM, Naïve Bayes-NB, Random Forest-RF, K-Nearest Neighbor-Knn, AdaBoost-AB, Gradient Boosting-GB, and Extreme Boosting-XB. The results highlight the superiority of the Random Forest over other examined algorithms in prediction accuracy. When using the RF classifiers, the achieved benchmarks include an Accuracy of 90%, Precision of 89%, Recall of 89.50%, F1-Score of 88%, and AUC-ROC of 75%. The comparative analysis indicates that the Random Forest delivers outstanding accuracy and reduced error rates. This research underscores the potential of the Random Forest, alongside other deep learning techniques, in advancing the field of software fault prediction precision.

Keywords: Machine Learning · Software Fault Predictions · Accuracy Measures · Software Testing Predictive Model · SDLC

1 Introduction

In software development life cycle (SDLC), prior to delivering the software to the customer, comprehensive software testing and quality control processes are executed to uncover any potential flaws. However, it is essential to go beyond simply detecting

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 65–84, 2025.
https://doi.org/10.1007/978-3-031-75167-7_6

faults after they occur and instead focus on predicting their likelihood before they manifest. By adopting a proactive approach, faults can be prevented during the design and development phases, thereby minimizing the risk of undetected flaws carrying forward to subsequent stages [1]. Software fault prediction (SFP) serves as a valuable quality model in efficiently managing the software development life cycle, making a substantial impact on its overall effectiveness

To mitigate software failures, fault prediction plays important role in planning, control, and execution of software development activities. It involves categorizing software modules as either faulty or non-faulty, acting as a safeguard against unforeseen risks and ultimately enhancing software efficiency and effectiveness [2]. The field of software engineering has witnessed the development of numerous fault prediction algorithms through the application of ML techniques, encompassing supervised and also unsupervised learning [3]. These algorithms are assessed using various metrics viz accuracy, error rate, precision, time requirements, recall, F- measure, and others, to determine their efficacy [4, 5].

The selection of data with more faults to be used for determining the relationship between metrics and faults of software codes; the application of quality metrics and feature selection that simplifies the process and the selection of the appropriate techniques that are known, all play a role in the success and precision of the faults prediction model [6].

There have been numerous prediction models used for SFP in the literature. Decision Tree, Naive Bayes classification, and neural networks are a few of the ML algorithms that are frequently used [7]. The accuracy of forecasts has increased, thanks to the use of these algorithms. However, it has drawbacks due to the need for expert knowledge to process data, the requirement for a comparatively high degree of expert human interaction, and the enormous amount of training data necessary for operation (considered time overheads), which can be difficult in a dynamic environment [8, 9]. Comparing various famous machine learning algorithms for SFP objectives is the major goal of this research project. After that, it aims to compare how well the investigated machine learning approach perform when applied to this work to more conventional methods [9]. Author make clear comparison between various ML classifiers to evaluate outcomes and a number of previously researched approaches.

2 Previous Work

For Previous scholars investigated the efficacy with using conventional machine learning algorithms in the context of SFP, such as neural networks (NN) [2, 10], Association Rules, K-Nearest Neighbors, Naive Bayes, etc.

The significant proportion of these ML algorithms produced results which have been satisfactory and then had high classification accuracy. However, these algorithms had some advantages and disadvantages, including the inability to handle imprecise information (ANN), the difficulty in handling a large number of software metrics (SVM) [11], the difficulty involved in building decision trees, the requirement for continuing evaluation metrics (association rule), and the need of that the dataset be unlabeled (clustering).

These algorithms need to be adapted with a lot of feature engineering and domain knowledge. The dataset suggests that the reason to a thorough exploratory data analysis,

which is followed by dimension-reduction methods. In recent literature, machine learning, particularly deep learning, has garnered significant attention. It has demonstrated promising outcomes in the realm of software engineering research [12]. The application of machine learning in software engineering offers unique opportunities by leveraging learning algorithms that are not typically found in traditional software engineering practices.

In a research example involved the utilization of stacked denoising auto-encoders to extract deep representations from conventional software metrics. Subsequently, a two-stage ensemble approach was employed to tackle the issue of class imbalance [13]. Another instance employed a deep belief network algorithm to construct a collection of affective features based on initial change features. The constructed features were then utilized with a machine learning classifier [14].

2.1 Evaluating Performance of Machine Learning Techniques

Available ML techniques include Adaboost-AB, Neural Networks-NN, Support Vector Machines-SVM, Decision Trees-DT, Deep Learning-DL, Random Forest-RF, K-Nearest Neighbors-KNN, Logistic Regression-LR and Gradient Boosting-GB. In Ayodele's work, supervised learning dealing with classification is categorized into different groups, such as Linear Classifiers, Quadratic Classifiers [7], K-Means Clustering, Boosting, Decision Tree, Neural networks, and Bayesian Networks. Further Ayodele classifies Linear Classifiers into four types -Logistic Regression, Naive Bayes, Perceptron and Support Vector Machine.

The author [6] introduced a fault prediction model in their study, incorporating three classifiers: Decision Trees (DT), K-Nearest Neighbors (KNN), and Random Forest. Accuracy of the fault prediction model, here author proposed two novel sets of change metrics, namely LOCWORKED-ON and MAX-LOC-WORKED-ON.

In a paper by [15] the authors focused on evaluating different classifiers. The study addressed two key aspects: the selection of an appropriate metric set and instance sampling techniques to handle class imbalance. Through their analysis, the authors reached the conclusion that a software quality prediction model lacking class balancing would not yield effective fault predictors. Furthermore, they observed that the impact of feature selection on the model's performance was relatively minimal.

In the research paper, the author [16] introduces a tool called "smotuned," designed to automatically adjust parameters for each dataset. The purpose of this tool is to optimize the performance of the parameters. The author employs multiple learners such as Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Decision Trees (DT), and Support Vector Machines (SVM). Through empirical analysis, author demonstrates that there is no single best learner that performs optimally across all datasets. Consequently, instead of solely focusing on finding the ideal learner, the paper suggests that emphasis should be placed on improving the quality of training data. The author argues that enhancing the fault prediction model is independent of having a superior classifier.

Estimating software quality is a challenging task in software engineering, according to study, MLT-based Software Fault Prediction (SFP). Research and extensive analysis in the field of SFP are required to get better outcomes in software quality prediction

[17]. SFP includes assessing qualities such as reliability, error tolerance, conformance, intricacy of work, and time. Additionally, researchers have stated that it is critical to use MLT-based techniques to forecast errors and defects at various phases of the Software Development Life Cycle (SDLC). This has prompted research on MLT-based SFP utilizing deep learning methods.

3 Proposed Approach

The proposed approach aims to address three challenges in software defect prediction: (1) handling noisy data, (2) dealing with class imbalance, and (3) leveraging transfer learning across different software projects. The performance of the proposed approach is evaluated using datasets obtained from PROMISE's software programs. The analysis measures the average performance value and the number of selected software indexes. The accuracy of the proposed approach is compared with various conventional and alternative methods. Results demonstrate that the proposed approach performs well and achieves impressive outcomes, utilizing 22 selected software metrics. To enhance training and prediction results, a preprocessing method is employed to maintain the values of the utilized software parameters within specific limits.

The training set is used to understand how the class (defect or non-defect) relates to the other features in the data. This learning process is supervised, meaning it uses labeled examples to make predictions. Once the patterns and structures in the data are learned, a predictive model is created. This model can then be used to predict if new cases are defects or not. To assess the model's performance, it undergoes testing with the test data. This procedure is reiterated in the subsequent iteration, utilizing a fresh set of training data from the dataset.

In our defect analysis and performance evaluation, we adopt a 5-fold cross-validation approach for training and testing. This approach is widely used and reliable for both learning and assessing classifier performance, and it is a standard-state of art- method for comparing classifiers. The 5-fold cross-validation divides the data into 5 subsets or "folds." We use 80% of the data for training the model and the remaining 20% for testing. The testing data represents real-world scenarios as it is unseen during the model creation, simulating new queries for defect analysis. We measure and record the prediction performance of the classifier and repeat this process 5 times. By calculating the average performance, we eliminate bias and can generalize the results for practical applications. This approach ensures the reliability and practicality of our performance evaluation.

Author selected TEN datasets with various fault percentage. These datasets are cleaned by finding missing values, removal of negative values, dummy values and checking for unique values. Datasets are normalized. Then it is divided between training and testing dataset. The ML classifiers are applied, and performance is measured for each of them. Confusion Matric, AUC-ROC curve, Precision, Recall, F1-Scores are checked and analysed to find the best suited model. All these steps are repeated with DL classifiers as well. Table 1 provides the selected ML classifiers list for research.

By using the aforesaid method, researcher aims to find answer of the following questions.

Table 1. ML Classifiers used in research

SN	Type	Classifiers	Abbreviation
1	Machine Learning Classifiers	Logistic Regression	LR
2		Decision Tree	DT
3		Support Vector Machine	SVM
4		Naïve Bayes	NB
5		Random Forest	RF
6		K-Nearest Neighbor	KNN
7		Gradient Boost	GB
8		Adaptive Boosting	AdaBoost
9		Extreme Gradient Boosting	XGBoost

Q1. How machine learning classifiers help to identify improvement with respect to accuracy, precision, recall, F-Score and AUC—ROC?
Q2. Which is the effective and accurate model for overall acceptance?

4 Research Methodology

Author had proposed below algorithm using machine learning classifiers, as the step 1 for research. Fig. 1 shows us the methodology where public dataset is used, refined, trained and validation is done. After which the ML algorithms are applied in order to study the performance measures.

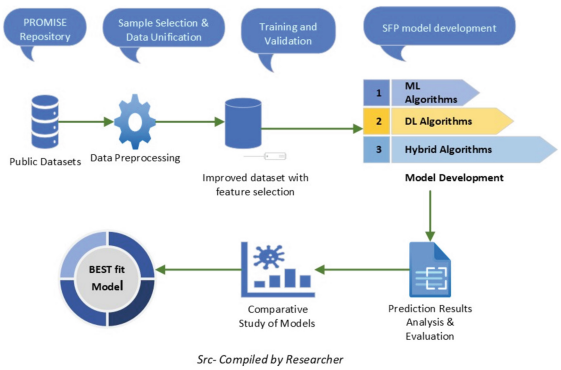


Fig. 1. Proposed Methodology

Algorithm Method 1: [Machine Learning]

Step1: Fetching Data
Step2: Data Pre-processing (Data Cleaning)

i. Finding Missing Values by using BoxPlot.
ii. Finding Interquantile range for every Column of the dataset and Removal of Negative Values (if Any).
iii. From Feature Engine we imported Artibitrary Outlayers CAPPER for Limiting and setup the column for particular Dataset.
iv. Finding/Checking Unique values.
v. Removal of Dummy Values, Columns and Rows.

Step 3: Data Pre-processing (Feature Extraction)

i. Feature Extraction of Dataset for making Training and Dataset
ii. Splitting Dataset

Step 4: Modelling of LR, DT, SVM, NB, RF, KNN, GB, & Adaboost

i. Setup Model.
ii. Setting prediction by assigning Testing Dataset.
iii. Finding Confusion Matrix to Calculate Accuracy, Precision, Recall, F1- Score.
iv. Finding AUC-ROC

4.1 Research Dataset Used

Research-related datasets [18, 19], such as software measurement and error data, are chosen from the PROMISE repository program. Program modules comprise each of these datasets. Lines of code numbers indicate how many modules are malfunctioning. The study's analysis validates the suggested methodology using the characteristics of ten distinct datasets. Below is a table that displays the dataset characteristics. The smallest dataset Ivy version 2 is having 87 KLOC and contains 11.33 % of defected modules. The largest dataset is Xalan 2.7 having 428 KLOC with 98.68 % of defected modules which is shown on Table 2.

Table 2. Dataset details used for study

SN	Dataset Name	VER	KLOC	Total modules	% of defected modules
1	Ant	1.7	208	746	22.25
2	Camel	1.6	113	966	19.46
3	Ivy	2	87	353	11.33

(*continued*)

Table 2. (*continued*)

SN	Dataset Name	VER	KLOC	Total modules	% of defected modules
4	Jedit	4.3	202	493	2.23
5	Lucane	2.4	102	341	59.53
6	Poi	3	129	443	63.43
7	Prop	6	97	661	9.98
8	Xalan	2.7	428	910	98.68
9	Xerces	1.4	141	589	74.19

The datasets characteristics are presented in above table. Each version of a dataset refers to a function, or method which is measured by Lines of Code (LOC) based metrics, McCabe metrics, Halstead measures.

4.2 Metrics Utilized for Research

Authors [20] and [21] discuss a selection of important object-oriented metrics that are included in this study. These metrics are summed up in Table 3-

1. Method per class-weighted (WMC): This metric counts the number of methods per class.
2. Inheritance Tree depth (DIT): It measures the number of levels of inheritance. Starting from top of object for each class.
3. Children's number (NOC): This metric quantifies number of immediate descendants a class has.
4. Cohesion of Methods (CAM): The method's parameter list is used by CAM to calculate the degree of relatedness between methods in a class. In order to get the total number of methods and distinct method parameter types in the class types of method parameters in each method, must be added up and divided by the total number of methods.
5. Objects Coupling (CBO): CBO – class number that are coupled to single class.
6. Response to class (RFC): RFC-count of separate methods, constructors invoked by a class.
7. Data Metric (DAM): private attributes ratio, to all attributes declared in a class is known as the Data Metric [21].
8. Lack of Cohesion (LCOM): LCOM is calculated using the formula 1 – (sum(am) (a * m)), where 'm' -methods in a class, 'a' is attributes in the class, 'am' is number of methods and the sum of 'am' [20].
9. Measure of Functional Abstraction (MFA): The MFA measures how many methods a class inherits compared to how many methods are available to all of the class's member methods
10. Public Methods (NPM): This metric totals methods in a class-public [21].
11. Measure-Aggregation (MOA): The degree to which characteristics realise the part-whole relationship (aggregate, composition) is measured by MOA. A count of class fields with user-defined class types is used to determine it [21].

Table 3. Metrics used for unification of datasets

SN	Name	Abbreviation
1	Weighted method per class	WmC
2	Depth of inheritance tree	DiT
3	Number of Children	NoC
4	Coupling between objects	CbO
5	Response for a class	RfC
6	Lack of cohesion in methods	LCoM
7	Afferent Coupling	Ca
8	Efferent Coupling	Ce
9	Number of Files	NOC
10	Average McCabe's Cyclomatic Complexity	AVG-CC
11	Maximum McCabe's Cyclomatic Complexity	MAX-CC
12	McCabe's Cyclomatic Complexity	CC
13	Average method Complexity	AMC
14	Number of Public Methods	NPM
15	Lack of Cohesion	LCo
16	Lines of Code	LOC
17	Data Access Metric	DAM
18	Inheritance Coupling	IC
19	Cohesion among Methods	CAM
20	Measure of Functional Abstraction	MFA
21	Measure of Aggregation	MoA
22	Coupling between methods	CBM

4.3 Pre-processing Phase

Learning-Based Approach. Learning The performance of machine learning algorithms can be adversely affected by factors like unreliable or irrelevant information and noisy data. To address these challenges, data preprocessing techniques are employed, including data cleaning, normalization, attribute selection, and extraction [23]. However, the usage and application of data preprocessing techniques vary across different studies due to variations in model selection, subjects, and metrics utilized in each investigation.

4.3.1 Normalization of Dataset

Normalization is applied to numerical attributes to establish a new range based on an equation derived from the existing range. This technique involves using an equation to

transform the range of existing data into a new range [24]. Normalization is particularly useful in improving the performance of classifiers during the prediction process. The authors employed this method while ensuring that the attributes preserved their normal distribution.

4.3.2 Label Encoding

Preprocessing of two steps: 1.normalization, 2.imbalanced data handling is done. Transforming values of metrics into the [0,1]. Using SMOTE to perform class imbalance processing. The existing binary label has to be a positive to negative ratio of 5:5.

Data Pre-processing is done using following steps:

- Finding Missing Values by using BoxPlot
- Finding Interquantile range for every Column of the dataset and Removal of Negative Values (if Any)
- From Feature Engine we imported Artibitrary Outlayers CAPPER for Limiting and setup the column for particular Dataset
- Finding/Checking Unique values
- Removal of Dummy Values, Columns and Rows

This normalized dataset is used for Feature Extraction for making Training and Testing Dataset in 80:20 ratio.

5 Experimental Approach of ML Classifiers

Logistic Regression (LR) analyzes the relationship between one or more independent variables and a dependent variable which is categorized. It also calculates the occurrence of an event probability by fitting data to a logistic curve [25, 26].

Decision Tree (DT) is a supervised learning technique that may also be used to solve regression difficulties. Internal nodes represents dataset feature. Branches represents decision rules. Leaf node represents the result in a tree[27].

Support Vector Machines (SVM) is extensively researched and utilized across various domains. The fundamental concept is to check the similarity distance between two classes. This distance is typically calculated based on their respective feature vectors. Furthermore, SVM can also be employed to address the issue of imbalanced class distributions [28].

Naïve Bayes (NB) use the Bayes theorem. It is predicated on the idea that every feature in the input data is unrelated to the others. Its efficacy and strong performance in numerous real-world applications have led to its popularity [1].

Random Forest (RF) In 2001, a novel and promising classifier called Random Forest was introduced [29], capturing significant attention in the field. This innovative approach offers several advantages [29], particularly its efficient performance on large databases and its capability to handle numerous input variables. Random Forest has emerged as a valuable tool in data analysis and classification tasks due to its robustness and versatility.

K Distance from Neighbours (KNN) rule consistently achieves good performance among the numerous techniques for supervised statistical pattern identification, as explained in [24]. The distance to the closest training case is used to classify a fresh occurrence. By giving the sign of the majority to the k closest points, the KNN classifier expands on this concept. To break ties, it is usual practice to use a figure for k (usually 1, 3 or 5) that is both small and odd. Greater k values aid in mitigating the impact of noisy points within the training dataset.

AdaBoost (AB) helps in execution of choice in trees on parallel issues. It works well with characteristics that have low score [31]. These models achieve accuracy over irregular possibility on a clustering problem.

5.1 Performance Metrics

1. Precision – quantity of a positive prediction count. The number of true classified fault-prone modules.

   Precision = True Positives/(True Positives + False Positives) [32]

2. Recall- The ability to classify positive samples in the model

   Recall = True Positives/(True Positives + False Negatives) [32]

3. F1-Score- Predictive performance of a model—precision and recall

   Score = (Precision * Recall)/(Precision + Recall) [33]

4. Accuracy- Fraction of right predictions.

   Accuracy = Number of correct predictions of model/Total number of predictions in the model. [7]

5. Specificity- The ability to predict a true negative of each category. Positive Predictive Value of predicted positives which are actual positives.
6. Area Under Curve – The ability of a classifier to distinguish between classes. [19]

6 Experimental Setup

The experiment study was conducted using the Python programming language, is renowned for its versatility and efficiency in enhancing programmer productivity, improving code readability, and ensuring software quality [34]. Python has proven to be useful in various applications, serving as a general-purpose language. To facilitate the implementation of machine learning algorithms, several software libraries, including Numpy, Panda, Matplotlib, and Scikit-learn, were utilized [35]. Keras was selected for developing and accessing deep learning models. We can integrate it with Python libraries.

7 Result Analysis

Nine frequently used classifiers, which were selected after literature review, were applied on 10 different cleaned datasets. All performance measure values were recorded. The following tables from Tables 3, 4, 5, 6, 7, 8 and 9 shows us all results and analysis is depicted through graphs – Figures.

The Machine Learning Algorithms viz- Logistic Regression, Decision Tree, Support Vector Machine, Naïve Bayes, Random Forest, K-Nearest Neighbor, Gradient Boost, Adaboost are applied on unified PROMISE datasets. Performance measurement values are recorded viz Accuracy, Precision, Recall, F1-Score, Value-Accuracy, Value-Loss, and AUC-ROC for all classifiers. Mean Weighted Average of all the measures are calculated (Figs. 2, 3, 4, 5, 6 and 7).

Table 4. Accuracy scores for all ML Classifiers

Accuracy - ML Classifiers

Classifier	Support	LR	SVM	NB	KNN	DT	RF	GB	AdaBoost	XGBoost
Ant 1.7	149	0.83	0.85	0.83	0.81	0.74	0.87	0.85	0.83	0.85
Camel 1.6	193	0.8	0.77	0.76	0.7	0.76	0.82	0.82	0.77	0.79
Jedit 4.3	99	0.96	0.97	0.93	0.95	0.97	0.97	0.97	0.97	0.97
Ivy 2.0	71	0.89	0.86	0.87	0.86	0.86	0.89	0.87	0.87	0.86
Log4J 1.2	41	0.93	0.9	1	0.9	1	1	1	1	1
Lucene	68	0.65	0.59	0.68	0.54	0.65	0.71	0.69	0.79	0.72
Poi 3.0	89	0.91	0.79	0.99	0.75	1	1	1	1	1
Xerces 1.4	118	0.8	0.74	0.62	0.84	0.86	0.89	0.89	0.86	0.87
Xelen 2.7	182	0.99	1	1	0.99	1	1	1	1	1
Prop6	193	0.8	0.77	0.76	0.7	0.73	0.8	0.82	0.77	0.79
Mean Wt. Average	120	0.86	0.82	0.84	0.8	0.86	0.9	0.89	0.89	0.89

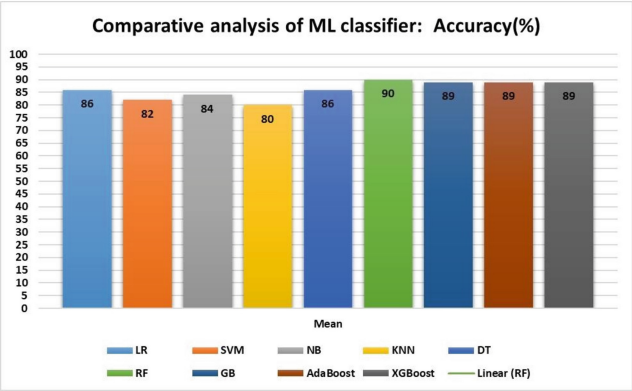


Fig. 2. Accuracy Scores Graph

Table 5. Precision scores for all ML Classifiers

Precision - ML Classifiers										
DataSets/Classifier	**Support**	**LR**	**SVM**	**NB**	**KNN**	**DT**	**RF**	**GB**	**AdaBoost**	**XGBoost**
Ant 1.7	149	0.82	0.83	0.83	0.81	0.79	0.87	0.85	0.83	0.86
Camel 1.6	193	0.78	0.69	0.75	0.66	0.77	0.81	0.81	0.74	0.77
Jedit 4.3	99	0.94	0.94	0.95	0.94	0.96	0.94	0.94	0.94	0.94
Ivy 2.0	71	0.9	0.74	0.86	0.82	0.86	0.9	0.85	0.85	0.82
Log4J 1.2	41	0.92	0.81	1	0.81	1	1	1	1	1
Lucene	68	0.65	0.35	0.77	0.53	0.64	0.7	0.69	0.79	0.72
Poi 3.0	89	0.92	0.78	0.99	0.76	1	1	1	1	1
Xerces 1.4	118	0.81	0.54	74	0.83	0.85	0.89	0.89	0.85	0.87
Xelen 2.7	182	1	1	1	1	1	1	1	1	1
Prop6	193	0.78	0.69	0.75	0.66	0.74	0.78	0.81	0.74	0.77
Mean	120	0.85	0.74	8.19	0.78	0.86	0.89	0.88	0.87	0.88

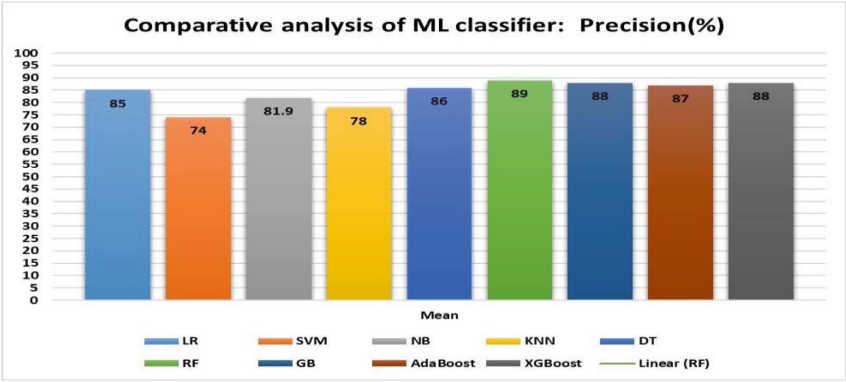


Fig. 3. Precision Scores Graph

Table 6. F1- scores for all ML Classifiers

F1-Score - ML Classifiers

DataSets/ Classifier	Support	LR	SVM	NB	KNN	DT	RF	GB	AdaBoost	XGBoost
Ant 1.7	149	0.82	0.83	0.83	0.81	0.76	0.87	0.85	0.83	0.85
Camel 1.6	193	0.76	0.7	0.75	0.68	0.76	0.78	0.8	0.75	0.78
Jedit 4.3	99	0.95	0.95	0.94	0.94	0.97	0.95	0.95	0.95	0.95
Ivy 2.0	71	0.85	0.79	0.86	0.82	0.86	0.85	0.84	0.84	0.82
Log4J 1.2	41	0.92	0.86	1	0.86	1	1	1	1	1
Lucene	68	0.65	0.44	0.67	0.54	0.64	0.7	0.69	0.79	0.72
Poi 3.0	89	0.91	0.77	0.99	0.75	1	1	1	1	1
Xerces 1.4	118	0.8	0.63	0.64	0.84	0.85	0.88	0.88	0.85	0.87
Xelen 2.7	182	0.99	1	1	1	1	1	1	1	1
Prop6	193	0.76	0.7	0.75	0.68	0.73	0.77	0.8	0.75	0.78
Mean	120	0.841	0.767	0.843	0.792	0.857	0.88	0.881	0.876	0.877

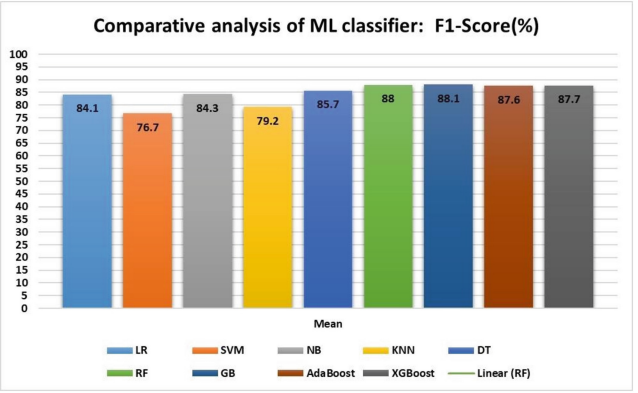


Fig. 4. F1- Scores Graph

Table 7. AUC-ROC- scores for all ML Classifiers

AUC-ROC - ML Classifiers

DataSets/ Classifier	**Support**	**LR**	**SVM**	**NB**	**KNN**	**DT**	**RF**	**GB**	**AdaBoost**	**XGBoost**
Ant 1.7	149	0.68	0.69	0.75	0.7	0.7	0.8	0.77	0.75	0.8
Camel 1.6	193	0.59	0.5	0.62	0.5	0.67	0.62	0.64	0.59	0.62
Jedit 4.3	99	0.49	0.5	0.64	0.49	0.66	0.5	0.5	0.5	0.5
Ivy 2.0	71	0.6	0.5	0.68	0.54	0.71	0.6	0.59	0.59	0.54
Log4J 1.2	41	0.74	0.5	1	0.5	1	1	1	1	1
Lucene	68	0.64	0.5	0.71	0.52	0.63	0.68	0.67	0.78	0.7
Poi 3.0	89	0.92	0.72	0.98	0.73	1	1	1	1	1
Xerces 1.4	118	0.76	0.5	0.66	0.78	0.8	0.82	0.82	0.79	0.81
Xelen 2.7	182	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
Prop6	193	0.59	0.51	0.62	0.51	0.62	0.6	0.64	0.59	0.64
Mean	120	0.67	0.55	0.74	0.59	0.75	0.74	0.74	0.73	0.73

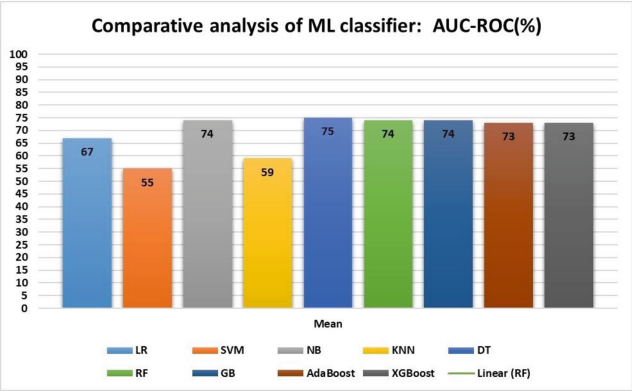


Fig. 5. AUC-ROC Scores Graph

Table 8. Recall – scores for all ML Classifiers

Recall - ML Classifiers

DataSets/ Classifier	Support	LR	SVM	NB	KNN	DT	RF	GB	AdaBoost	XGBoost
Ant 1.7	149	0.83	0.85	0.83	0.81	0.74	0.87	0.85	0.83	0.85
Camel 1.6	193	0.8	0.77	0.76	0.7	0.76	0.82	0.82	0.77	0.79
Jedit 4.3	99	0.96	0.97	0.93	0.95	0.97	0.97	0.97	0.97	0.97
Ivy 2.0	71	0.89	0.86	0.87	0.86	0.86	0.89	0.87	0.87	0.86
Log4J 1.2	41	0.93	0.9	1	0.9	1	1	1	1	1
Lucene	68	0.65	0.59	0.68	0.54	0.65	0.71	0.69	0.79	0.72
Poi 3.0	89	0.91	0.79	0.99	0.75	1	1	1	1	1
Xerces 1.4	118	0.8	0.74	0.62	0.84	0.86	0.89	0.89	0.86	0.87
Xelen 2.7	182	0.99	1	1	0.99	1	1	1	1	1
Prop6	193	0.8	0.77	0.76	0.7	0.73	0.8	0.82	0.77	0.79
Mean	120	0.856	0.824	0.844	0.804	0.857	0.895	0.891	0.886	0.885

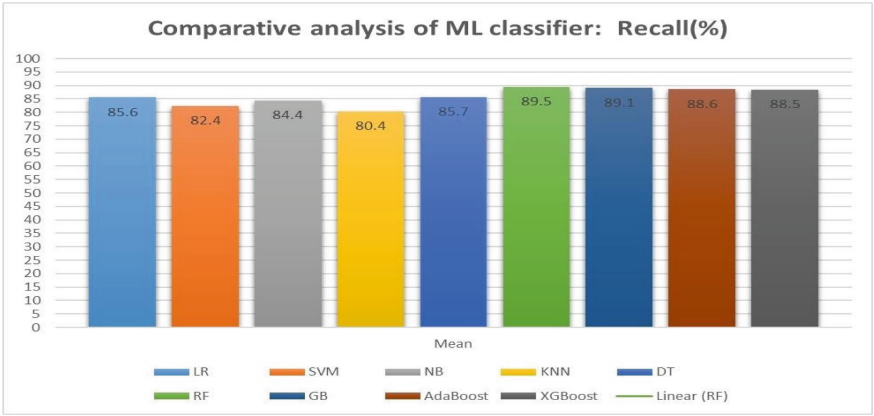


Fig. 6. Recall Scores Graph

Table 9. Weighted Mean scores for all ML Classifiers

DataSets/ Classifier	Mean- Accuracy	Mean- Precision	Mean- Recall	Mean- F1 Score	Mean- AUC-ROC
LR	0.86	0.85	0.856	0.841	0.67
SVM	0.82	0.74	0.824	0.767	0.55
NB	0.84	0.81	0.844	0.843	0.74
KNN	0.8	0.78	0.804	0.792	0.59
DT	0.86	0.86	0.857	0.857	0.75
RF	0.9	0.89	0.895	0.882	**0.75**
GB	0.89	0.88	0.891	0.881	0.74
AdaBoost	0.89	0.87	0.886	0.876	0.73
XGBoost	0.89	0.88	0.885	0.877	0.73

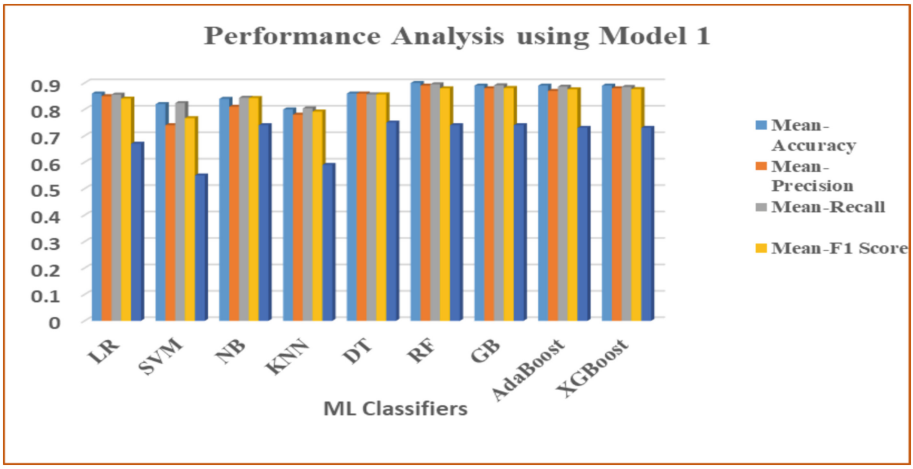


Fig. 7. Comparative Performance Analysis using proposed model.

8 Conclusion

Machine Learning is the prevailing technique in the realm of software fault prediction. This research focuses on implementing Machine Learning Classifiers to address two key inquiries:

1. How do Machine Learning Classifiers contribute to enhancing accuracy, precision, recall, F-Score, and AUC-ROC in software fault prediction?
2. Which model proves to be the most effective and accurate for overall acceptance in this context?

The experiment involves the utilization of nine Machine Learning classifiers for performance evaluation, namely LR (Logistic Regression), SVM (Support Vector Machine), GB (Gradient Boosting), DT (Decision Tree), KNN (K-Nearest Neighbor), RF (Random Forest), NB (Naïve Bayes), AB (AdaBoost), and XGB (Extreme Gradient Boosting). Open-source datasets comprising 23 metrics are employed, and performance measurements include Accuracy, Recall, Precision, F1-Score, and AUC- ROC.

To ensure optimal performance, Cross-Project validation is carried out. In the classifier comparison, Random Forest and Adaboost outperform other classifiers, exhibiting superior mean weighted average performance. Specifically, Random Forest, Adaboost, and XGBoost classifiers consistently achieve high accuracy rates of 90%, outperforming all others. Conversely, the KNN classifier exhibits relatively lower performance. In terms of precision, Random Forest and Gradient Boosting achieve the highest scores at 89%, whereas SVM lags behind at 74%. The F-Score reinforces the dominance of Random Forest with an 88% score, while SVM trails at 76%. For Recall, Random Forest impresses with a 89.5% rate, whereas SVM records the lowest at 82%. In the realm of AUC-ROC Curve, Decision Tree outperforms all others with a mean of 75%, while SVM secures the lowest at 55%. Considering these results and computing the weighted average, it becomes evident that utilizing the Random Forest classifier can predict software

fault proneness with an average accuracy of 90%. Therefore, researchers can confidently employ the Random Forest classifier for predicting SFP proneness, and Decision Tree emerges as the second-best option for precision in this research context. This study holds significant potential for enhancing the performance of other SFP classifiers, particularly those in the deep learning category, and promises to elevate prediction accuracy in this field.

9 Limitations

The model proposed by researcher, in this paper has the following threats to the validity –

- Although random forest exhibits superior predictive performance in comparison to current state-of-the-art CPDP methods on the 10 selected datasets, it's crucial to recognize that these findings might not be consistently reproduced on other datasets.

  The accuracy of predictions is also influenced by the selection of features from the datasets.
- The researcher used widely-known evaluation measures in this study, including recall, accuracy score, precision score, and F1 score. Nonetheless, it must be recognised that different assessment metrics might produce different results.

References

1. Sasankar, P.: Analysis of test management, functional and load testing tools. Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol. **1**(1), 804–810 (2016)
2. Prabha, C., Shivakumar, N.: Software defect prediction using machine learning techniques. In: International Conference on Trends in Electronics and Informatics (2020)
3. Samantaray, R., Das, H.: Performance analysis of machine learning algorithms using baggingensemble technique for software fault prediction. In: 6th International Conference on Information Systems and Computer Networks, Mathura (2023)
4. Bartłomiej, W., Robert, D.: Applying machine learning to software fault prediction. e-Informatica Softw. Eng. J. **12**(1), 199–216 (2018)
5. Borandag, E.: Software fault prediction using an rnn-based deep learning approach and ensemble machine learning techniques. In: Challenges in Using Machine Learning to Support Software Engineering, vol. 13, no. 3 (2023)
6. Choudhary, G., Kumar, S., Kumar, K., Mishra, A.: Empirical analysis of change metrics for software fault prediction. Comput. Electr. Eng. **67**, 15–24 (2018)
7. Kumar, S., Ranjan, P.: A Comprehensive analysis for software fault detection and prediction using computational intelligence techniques. Int. J. Comput. Intell. Res. **13**(1), 65–78 (2017)
8. Li, N., Sheperd, M., Guo, Y.: A systematic review of unsupervised learning techniques for software defect prediction. Inform. Softw. Technol. **122**, 106287 (2020)
9. Anwar, N., Kar, S.: Review paper on various software testing techniques & strategies. Global J. Comput. Sci. Technol.: Comput. Softw. Data Eng. (2) (2019)
10. M. S, M. S: "Usage of Machine Learning in Software Testing," Automated Software Engineering: A Deep Learning-Based Approach. Learning and Analytics in Intelligent Systems-Springer (2020)

11. Seo, H., No, J., Park, S.: ml-SFP: system failure prediction method based on machine learning. In: 6th World Conference on Smart Trends in Systems, Security and Sustainability, WS4 2022, London (2023)
12. Iqra, B., Tamim, K.: Software fault prediction using deep learning techniques. Softw. Qual. J. (2023)
13. Sasankar, P.: Cross project defect prediction using deep learning techniques. In: International Conference on Artificial Intelligence & Big Data Analytics (2022)
14. Wangshu, L., Shulong, L., Qing, G., Daoxu, C.: Empirical studies of a two-stage data preprocessing approach for software fault prediction. IEEE Trans. Reliab. **65**(1), 1–16 (2015)
15. Agrawal, A., Malhotra, R.: Cross project defect prediction for open source softwares. Int. J. Inf. Technol. **14**, 587–601 (2022)
16. Jorayeva, M., Akhan, A., Cagatay, C., Mishra, A.: Deep Learning-Based Defect Prediction for, MDPI-Sensors, vol. 22 (2022)
17. Zhao, Y., Chen, H., Damevski, K.: A systematic survey of just-in-time software defect prediction. ACM Comput. Surv. **55**(10), 1–35 (2023)
18. Chidamber, S.K.C.: A metric suite for object oriented design. IEEE Trans. Softw. Eng. **20**(9), 476–493 (1994)
19. Sasankar, P., Sakarkar, G.: An empirical study of classification models using AUC-ROC curve for software fault predictions. Int. J. Sci. Res. Comput. Sci., Eng. Inform. Technol. **9**(1), 250–260 (2023)
20. Abubakar, A.A.J., Ahmed, M.: Can cohesion predict fault density? IEEE **1**, 890–894 (2006)
21. Khan, B., et al.: Software defect prediction for healthcare big data: an empirical evaluation of machine learning techniques. J. Healthc. Eng. **2021**, 1–16 (2021). https://doi.org/10.1155/2021/8899263
22. Bowes, G.D., Davey, N., Sun, Y., Christianson, B.: Furthur thoughts on precision. In: 15 Annual Conference on Evaluation & Assesment in Software Engineering (2011)
23. Lessmann, S., Baesens, B., Mues, C., Pietsch, S.: Benchmarking classification models for software defect prediction: a proposed framework and novel findings. IEEE Trans. Softw. Eng. **34**(4), 485–496 (2008). https://doi.org/10.1109/TSE.2008.35
24. Pandey, S., Kumar, K.: Software fault prediction for imbalanced data: a survey on recent developments. Procedia Comput. Sci. **218**, 1815–1824 (2023). https://doi.org/10.1016/j.procs.2023.01.159
25. Saharudin, S., Wei, K., Na, K.: Machine learning techniques for software systematic review. J. Comput. Sci. (2020)
26. Akimova, E.N., Bersenev, A.Y., Deikov, A.A., Kobylkin, K.S.: A survey on software defect prediction using deep learning. Sigma Math. **9** (2021)
27. Ghotra, B., McIntosh, S., Hassan, A.E.: Revisiting the impact of classification techniques. In: ACM 37th IEEE International Conference on Software Engineering (2015)
28. Perreault, L., Berardinelli, S., Izurieta, C., Sheppard, J.: UsingClassifiers for software defect detection. In: International Conference on Software Engineering and Data Engineering (2017)
29. Kundu, A., Dutta, P., Ranjit, P., Bidyadhar, S., Kumar, M.: Software fault prediction using machine learning models. In: OITS International Conference on Information Technology (2022)
30. Qasem, O.A., Akour, M., Alenezi, M.: The influence of deep learning algorithms factors in software fault prediction. IEEE Access **8**, 63945–63960 (2020). https://doi.org/10.1109/ACCESS.2020.2985290
31. Balram, A., Vasundra, S.: Software fault detection using multi-distinguished-features sampling with ensemble random forest classifier. Int. J. Intell. Eng. Syst. **15**(5), 494–505 (2022)
32. Webb, G.I., Zheng, Z.: Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. IEEE Trans. Knowl. Data Eng. **16**(8), 980–991 (2004)

33. Hasanpour, A., Farzi, P., Tehrani, A., Akbari, R.: Software Defect Prediction Based On Deep Learning Models: Performance Study. arXiv preprint arXiv:2004.02589 (2020)
34. Luigi, L., Angela, L., Geng, L., Roberto, M.: Estimating software functional size via machine learning. ACM Trans. Softw. Eng. Methodol. **31**, 114 (2023)
35. Lakshmi, J.V.N.: Machine learning techniques using python for data analysis in performance evaluation. Int. J. Intell. Syst. Technol. Appl. **17**(1/2), 3 (2018). https://doi.org/10.1504/IJISTA.2018.091584

A Journey Through Multifaceted Data in Machine Learning Predictions on Financial Viability

R. Revathi(✉)

NaviMumbai, India
revathiraghavan30@gmail.com

Abstract. This research explores the application of various machine learning models for credit risk assessment in the context of a highly imbalanced dataset. The study encompasses logistic regression, random forest, gradient boosting algorithms (LightGBM and XGBoost), neural networks, and ensemble methods. An extensive data processing pipeline, including categorical encoding, missing value imputation, and feature scaling, is implemented to enhance model performance. Evaluation metrics, primarily focusing on AUC-ROC scores, are used to assess model accuracy.

The findings reveal that gradient boosting algorithms, specifically LightGBM, outperform other models, achieving the highest AUC-ROC score of 0.80123. Deep learning models, such as TabNet, demonstrate potential but require further development for competitive performance. Logistic regression and random forest models yield respectable results, showcasing the significance of feature engineering in enhancing their predictive capabilities.

Ensemble methods, including stacking and weighted averaging, are explored to leverage the strengths of individual models. The study emphasizes the importance of model selection and parameter tuning in achieving optimal results. Logistic regression, despite its simplicity, performs well due to the incorporation of a feature-engineered pipeline.

In conclusion, the research suggests that, while current deep learning models for tabular data show promise, gradient boosting algorithms remain superior for credit risk assessment. The study highlights the need for further exploration of advanced techniques and emphasizes the potential of regression algorithms for a comprehensive credit assessment score, complementing the classifier algorithms discussed.

Keywords: Credit Scoring Models · TabNet · XGBoost · Neural Network Embedding · Ensemble Learning · AUC-ROC Evaluation

1 Introduction

The research focuses on enabling lenders to make data-driven decisions for applicants with incomplete credit histories, collaborating with Home Credit International's dataset for a Kaggle competition. The goal is to assess risk using alternative data sources like

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 85–103, 2025.
https://doi.org/10.1007/978-3-031-75167-7_7

credit card usage, previous application history, and credit bureau data. Advantages of home loans include facilitating homeownership for those with moderate incomes, financial sector knowledge, and tax benefits. However, drawbacks include significant long-term commitment, potential risks over the loan duration, and missed investment opportunities. The proposed Loan Prediction System evaluates customer eligibility based on parameters like age, marital status, income, and expenditure, using domain knowledge feature engineering and classification algorithms to build a predictive model for home loan approval. The project aims to compare the performance of different models.

2 Relevant Works

In their study, Ziyue Qiu et al. [1] explored the deployment of LightGBM, Logistic Regression, and Random Forest models. Their research centered on creditworthiness appraisal systems, investigating varied methods for attribute generation. Their findings highlighted a consistent trend in model effectiveness, with LightGBM outperforming Logistic Regression and Random Forest. Impressively, their most successful model achieved an AUC score of 78% on the Kaggle Home Credit Default Risk dataset. A distinctive aspect of their approach involved eliminating attributes containing more than 70% of absent data instances. Additionally, they addressed the voids through two methodologies: Remediation techniques encompassed mean and median redress, along-side a unique method utilizing Random Forest rectification.

Yiyun Liang et al. conducted an additional research exploration [2], yielding similar conclusions. Employing distinctive sampling techniques on their dataset, they revealed that the most effective results were obtained through Downsampling. We recognize the data they employed was notably intricate and skewed. They ventured into diverse models, encompassing Random Forest, LGBM, Logistic Regression, Naive Bayes, and MLP. Furthermore, they embraced clustering via K-means. Intriguingly, they achieved peak performance when k equaled four, with each cluster demonstrating distinct efficacy. Amongst their array of machine learning models, LightGBM emerged as the frontrunner, attaining a notable 71.57% accuracy.

Xue Chen et al. [3] introduces the cutting-edge DeepGBM framework, poised to revolutionize performance outcomes within this dataset context. Through meticulous comparison against established players like PNN, Random Forest, and LGBM, DeepGBM emerges triumphant with a superior result of 0.756, transcending conventional solitary classifiers. This innovation seamlessly fuses the prowess of neural networks and GBDT, forging a symbiotic relationship that augments overall capabilities. Deep-GBM represents a robust GB system rooted in the realm of deep learning, a visionary concept championed by Guolin Ke in the form of DeepGBM[4]. This strategic creation impeccably strikes a balance between sparsely categorized features and densely populated quantitative attributes.

The most significant drawback of this paper was that they used only two of the seven csv files application train.csv, and application test.csv files in their experiments. This means that they worked with only 104 features, whereas in our paper we included all the files and came up with way more features.

In the field of credit assessment research, a notable departure is observed in the study by Zhenlong Liu et al. [5]. Their model, a fusion of GBDT2NN (Gradient Boosting Decision Tree to Neural Network) and FM (Factorization Machine), outperforms individual classifiers like GBDT2NN, LGBM, and Logistic Regression, showcasing enhanced efficacy. An intriguing proposition emerges as they introduce SMOTE (Synthetic Minority Over-sampling Technique) to rectify disparities in data distribution. This approach addresses data-set inequalities while concurrently amplifying the intricacies within the learning process.

Another comparison between two machine learning models MLP(Multi-Layer Perceptron) and CNN(Convolutional Neural Networks) were made by Salomey et al. [6] for credit risk. In their experiments, CNN came out to be the better model for accuracy and other performance metrics used, such as, back-testing, stress testing. We can also see that MLP method could do better with parameter tuning.

In our paper we implemented binary classifiers using linear models, Gradient boosting methods, combinations of classifier models, sampling methods and Neural embedding methods, which we then compared to make a claim of which method or model gives the best performance. Our research was to create more features from existing data. We tried different model comparisons and acquired better scores.

3 Methodology

3.1 Data Source

Founded in 1997 in the Czech Republic, Home Credit is a global player in consumer financial services, operating in ten countries with its headquarters in the Netherlands. Specializing in serving individuals with suboptimal credit backgrounds, it provides access to loans. The data used for research, focusing on Home Credit Default Risk for a Kaggle competition, originates from diverse temporal periods and represents genuine clientele from the United States. Home Credit initiated this Kaggle challenge three years ago to leverage the data's potential. The dataset, consisting of 122 columns, includes information on annual income, number of children, gender, employment tenure, age, credit bureau balance, and credit card history.

3.2 Pre-processing and Feature Engineering Pipeline

Version 1 Some basic pre-processing techniques were enforced.

Data Preprocessing Application phase: Removed implausible entries, eliminated missing features, grouped ages with numerical labels, and created normalized features and meaningful ratios. Categorical features were encoded using Label Encoding [7].

Bureau data: Generated new features, encoded categorical features with One-Hot Encoding [8]. Created "STATUS 12345" for pending or delayed payments, applied statistical operations, and determined the last loan's maximum overdue.

Previous application: Encoded categorical data with Label Encoding [7], created features based on ratios and differences. Listed active loans not completed, calculated amounts, and removed instances of "365243." Applied date aggregation and statistical operations.

Final processing: Removed "TARGET" label, positive and negative infinities, resulting in 460 features[7].

Version 1.1 In 3.2, More data was used from the data source.

POS and Cash Late payments were flagged, data was aggregated, and irrelevant features were removed before merging with the existing pipeline.

Installments Payments were grouped and differences calculated. Late payments, percentage of lateness, and significant impending amounts were flagged. Additional flags were set based on lateness thresholds. Installments in the last 3 to 5 years were grouped, and new features were created through statistical operations, including payment trend analysis.

Credit Card Features that keep record of the amount of limit used, the current payment vs the minimum payment ration, if there are any late payments and the amount of delay, ratio of cash draw vs limit, were created. The last month's balance of each credit card of the applicant was recorded, and aggregated with the monthly balance of last 3 years.

Improvement From Previous Version The new pipeline has *649 features* in total. And it has more related information about the applicants financial situation.

Model Training Binary classifiers, including Logistic Regression, Random Forest, LightGBM, XGBoost, TabNet, and Neural Network, were trained on the preprocessed data to predict the probability of default.

Logistic Regression A Logistic Regression model was trained using one-hot encoding for categorical features. Data was imputed using a Simple Imputer, and features were scaled using a Min-Max Scaler. The model used a logistic function for binary classification.

Random Forest A Random Forest Classifier was trained on processed data using imputation and scaling similar to Logistic Regression. The model utilized a collection of decision trees and majority voting for predictions.

LightGBM LightGBM, a gradient boosting algorithm, was trained with features generated during preprocessing. The model used boosting techniques like GOSS and was fine-tuned using Grid Search Cross Validation for optimal hyperparameters.

XGBoost XGBoost, a gradient boosting decision tree algorithm, was trained using imputed and scaled data. Hyperparameters were tuned using Grid Search Cross Validation for improved performance.

TabNet TabNet, a deep neural network for tabular data, was trained on processed data using a StepLR Learning Rate Scheduler and Adam Optimizer. Hyperparameters were fine-tuned using Grid Search Cross Validation.

Neural Network – Embedding Model A neural network with an embedding layer was trained to capture relationships in high-cardinality categorical features. Data preprocessing included label encoding and scaling using a Min-Max Scaler.

Neural Network Model A neural network was trained on encoded categorical data using standardization and a combination of categorical and numerical features. Hyperparam eters like learning rate and iterations were set for optimization.

Weighted Average Ensemble (WA1 and WA2) Two ensemble models were created by averaging predictions from different models with weighted contributions. WA1 combined Neural Network Embedding and LightGBM, while WA2 combined LightGBM and XGBoost. Weights were chosen through grid search for optimal ensemble performance.

Ensemble Model An ensemble stacking model was built using the stacking technique. Base models included Gradient Boosting and Logistic Regression, with a Support Vector Machine as the meta-classifier.

LGBM with Oversampling To address data imbalance, the LightGBM model was trained after oversampling the minority class using SMOTE. This helped achieve a balanced distribution and improved model performance.

These models were evaluated using metrics such as AUC-ROC scores, and hyperparameters were fine-tuned for optimal performance. Additionally, knowledge distillation was applied to transfer knowledge from LightGBM and XGBoost to TabNet. The final models were assessed based on their ability to predict the probability of default for loan applicants.

4 Experiments

4.1 Data Description

The data-set provides a lot of information about the borrower.Contained within are numerous interconnected tables, housing fixed details about applicants like their age, family size, profession, and related attributes. Additionally, it holds the applicant's historical credit records retrieved from the credit bureau division, as well as their prior credit performance within the realm of Home Credit Group. For the training file, we have a total of 307511 observations and 122 features to consider with integer, float and object datatype. Also there are 67 features with null value. The test file is almost similar as the training file having 48744 observations, 121 features (minus the predictor variable 'TARGET'), and 64 features having null values. The words being used as; Observations = rows, features= columns based on what we have observed. Only around 8 training set did not repay the loan. The training data is unbalanced, and a significantly higher proportion of people listed in our sample can repay loans, compared to people who default on loans. For each of the 122 features, the correlation coefficient was calculated with the target value.

The current application has features such as:

- If the applicant has a car or real estate.
- Number of descendants of the applicant.
- Income of the client and their profession.
- Amount of the loan and it's type and purpose.
- Level of education.
- Family status and living situation.
- Number of days they were employed in their current job.
- If the customer has a phone, how long has it been in use.
- Details about when and where the loan application was filed.

- Normalized score from external credit scorers.
- Number of documents submitted.

The Home Credit International data on credit bureau includes features such as:

- If the applicant has a credit report with basic information such as if it is active and when it was last updated.
- The total amount of debt that the applicant currently owes, as well as the specifics of that obligation.
- Payment information for prior debts.

The data from Home Credit International on previous applications includes details such as:

- The loan amount, kind, and it's purpose.
- In the application data, there might be many previous loans for each current loan.
- Each previous application has one row and is identified by the feature SK ID PREV.

The following are some of the features of the Home Credit International statistics on bureau balance:

- Monthly statistics on the bureau's credit history.
- A previous credit might have many rows, one for each month of the credit period.

The Home Credit International data on POS CAS BALANCE data includes characteristics such as:

- Monthly data on previous Home Credit point-of-sale or cash-loan transactions.
- Each row represents one month of a previous point of sale or cash loan, although a single prior loan could include many rows.

The Home Credit International data on credit card balance includes elements such as:

- Clients' prior credit cards with Home Credit are tracked on a monthly basis.
- Each row represents a month's credit card balance, and each credit card might contain many rows.

The Home Credit International data on installment payments includes characteristics such as:

- Every made payment has it's own row, and every missed payment has it's own row.

4.2 Baseline

We used a linear model as our baseline.

Data Pipeline It is very similar to **??**, but we did not use the Data processing Pipeline described in 3.2, instead we just used the application data. We used One-Hot Encoding[8] to encode the categorical data, remove all positive and negative infinities, use Simple Imputer[9] and MinMax Scaler[10] models to pre-process the data.

Training The parameter used for the Logistic Regression model is C = 0.0001.

Score The model had an AUC-ROC private score of 0.68453 and a public score of 0.68157.

4.3 Evaluation Metrics

Our evaluation metrics across all models is *Area Under the curve of Receiver Operating Characteristic* curve (AUC-ROC) [11]. We chose AUC-ROC score since it shows how good the model is at predicting True Positives. However, we did not have the true values of our test data-set, so we are relied on Kaggle for accuracy score. Kaggle uses AUC-ROC to score our predictions.

Before we understand AUC-ROC, we need to look into Receiver Operating Characteristic (ROC) curve. It is a probability curve of True Positive Rate (TPR) vs False Positive Rate (FPR). The TPR and FPR are on the y and x axis respectively. An example would be Fig. 1.

We need a better understanding of TPR and FPR, in order to understand Confusion Matrix. A Confusion Matrix is a matrix of Predicted and Actual target, which is represented in Table 1.

Table 1. Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

True Positive Rate, a.k.a. TPR, Recall, Sensitivity means how accurately a model can predict a true value correctly. The equation to calculate TPR is shown in equation 1

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

where:

TPR = True Positive Rate
TP = True Positive

FN = False Negative

False Positive Rate is the rate of wrong predictions by a model. The equation to calculate FPR is shown in equation 3

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$FPR = 1 = Specificity = \frac{FP}{TN + FP} \tag{3}$$

where:

FPR = False Positive Rate
FP = False Positive
TN = True Negative

So, a perfect model would have an AUC score of 1 and if a model predicts all classes as 1 then it's AUC will be 0.5. It means the model has no class separation capacity. So a good model would have a higher AUC score between 0.5 and 1.

5 Results and Discussion

We have had mixed results from our models. Among all models, Gradient Boosting based models performed much better than other types of models.

Point to be noted, all the scores discussed in this section were obtained from Kaggle. Kaggle computes the AUC-ROC score from the predicted probability of the test data-set we created using our models. And as we do not have the labels for test data-set, we were not able to create Confusion Matrices.

5.1 Gradient Boosting

LightGBM (**??**) and XGBoost (**??**) are the variants of gradient boosting models trained on this data-set. The scores with their corresponding parameters are given in Tables 12 and 13 respectively.

We also trained a LightGBM model with oversampling using SMOTE. The scores are given in Table 2.

Table 2. Score of LightGBM using SMOTE

Model	Public Score	Private Score
LGBM using SMOTE	0.75513	0.75949

LightGBM performed better between the two models. Our speculation is that, LightGBM performed better because of it's unique feature where it only grows leaves with the highest delta in cost function, whereas XGBoost grows the level of the tree. This is

why LightGBM models train faster on data-sets with many features than XGBoost with lower memory consumption.

Although we can see that the models performed much better than our Baseline model (4.2) becasuse of using better Data Pipeline and better ML models. The models particularly improved performance with Data Pieline v1.1 (3.2) because Gradient Boosting Trees perform better with more features.

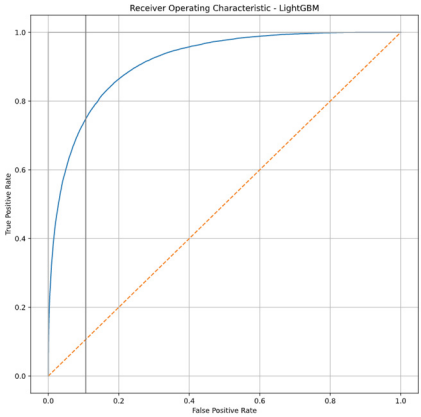


Fig. 1. ROC Curve of LightGBM

The ROC curve of the best performing LightGBM model with the training data-set is given in Figure 1

The ROC Curve with the best scoring model can be seen in Figure 2

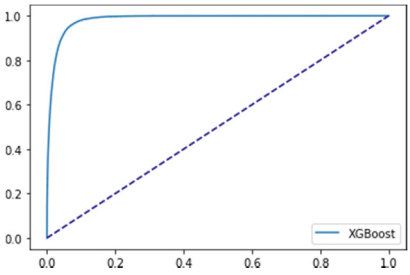


Fig. 2. ROC Curve of XGBoost

5.2 Deep Learning

A deep learning TabNet (**??**) model was trained using our Processed Data (3.2). The scores of TabNet are given in Table 3.

We can see that the Deep Learning TabNet model performs better with lower number of features when trained with a CPU. Possibly because the model doesn't have a technique to ignore features with lower weights. When the model is trained using a CPU, a possible improvement would be to look at the feature and their weights and drop the features with lowest weights (an arbitrary threshold can be set) to train the model. That should yield better score than our current ones. Although that limitation can be overcome by training the model with a CUDA enable GPU with enough VRAM. The model accesses the VRAM of the GPU to store the tensors as well as the system memory. Thus training with a system with faster RAM (3200 MHz+ DDR4) and higher bandwidth memory (GDDR5X+) yields better scores.

Since Deep Learning models for tabular data are still in their early stages, further development of similar techniques have the potential to perform better or atleast similar to Gradient Boosting Machines.

5.3 Neural Networks

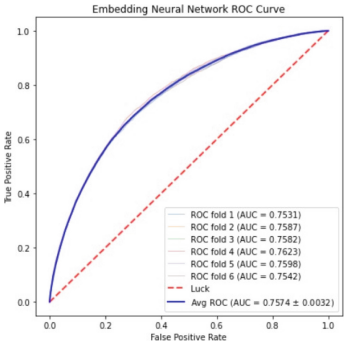


Fig. 3. Embedding NN ROC Curve

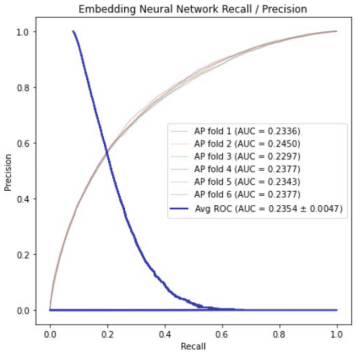


Fig. 4. Embedding NN Recall/Precision Curve

We trained a few neural network models with major differences between their network configuration. But none of the models we trained have shown promising scores.

Mostly because of the infinite possibility of network configuration and other parameter tests, we could not find a model that performed close to our Gradient Boosting Trees. We believe further training of Neural Networks with different network configuration and fine-tuning of their paramaters should help the models classify the data-set better. The scores of our Neural Network models are given in Tables 4 and 5 (Figs. 3 and 4).

5.4 Ensemble Model

The ensemble model described in **??** procudes the scores in Table 6.

We created some simple ensemble models using the weighted average technique in **??** and 7. We believe if we use Grid Search technique [12] to determine the weights and use scores from the top 3 different types of best performing models, we would have better scores.

5.5 Logistic Regression

This is the simplest model we experimented with. The model performs very well, compared to other simpler models. This is the closest relative to our baseline model and it yield better scores simply due to the association of the Feature Engineered Pipeline. The scores of Logistic Regression model is given in Table 9.

5.6 Random Forest

Random Forest worked very well compared to how simple it's implementation is and how minimum memory and time it requires to train. A bigger ensemble (higher number of estimators) of Random Forest and a Neural Networks should have yield better scores. The scores of Random Forest model is given in Table 10.

5.7 Comparison Between All Models and Baseline

We have tested a lot of different classifier models and we believe a direct comparison with the Baseline would contrast our findings much better in Table 14.

We can see a trend of Gradient Boosting based classifier performing much better than other types. Although if the suggested improvements were made to Neural Networks and an upgraded Deep Learning Model was used, it should be able to perform the same, if not better than the Gradient Boosting models. And a much computed heavy solution would be, to use weighted average if the suggested improvements were made.

6 Conclusion

Most of the training was performed with widely accepted binary classifier models. Since, the data-set is unbalanced to a ratio of 10:1, our models were better at predicting True Positives than False Negatives. Translating that to real work scenario, our models were able to predict the applicants who was more likely to pay back the loan versus those who aren't.

Our best AUC-ROC score was 0.80123, which means it accurately predicted 80.123% of both classes combined. In a production environment this accuracy is not enough unless there are other mechanisms deployed to assist it. A credit assessment score similar to what Equifax, TransUnion provides using a Regression Algorithm would be a better risk assessment tool than a Classifier Algorithm used in our work.

In our testing, we have figured that Deep Learning for Tabular Data is not advanced enough to compete with Gradient Boosting Algorithms.even though the development of such models look very promising. Moreover, Neural Network based models have the potential to perform similar, if not better than Gradient Boosting Models.

A regression algorithm would provide an overall score to the credit-ability and the financial overview of the applicant. Lenders can use this information to achieve better data driven decision.

A classifier algorithm, demonstrated in our work, is an excellent tool to pre-approve or shortlist loan applications. But using such algorithm as a decision making tool would not be wise at this stage. Although further research on such algorithm should yield better results.

Appendix A

Regarding the Accuracy and Hardware Performance Relationship in TabNet
We established a relationship between a TabNet Classifier model's accuracy with the hardware performance. We ran the model multiple times in a three different hardware with the same code-base but found slightly different results. We figured that the the model performing slightly better on GPUs with larger and faster VRAM and faster access to system Memory.

The systems we tested the model and ranking in accuracy are:

LightGBM Boosting Type's Effect on Memory Optimization

We also figured that when LightGBM's boosting type is used in parameters, which are Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), the model uses almost half the memory to train while still retaining the original accuracy. This is because the boosting methods drop features with smaller importance (Fig. 5).

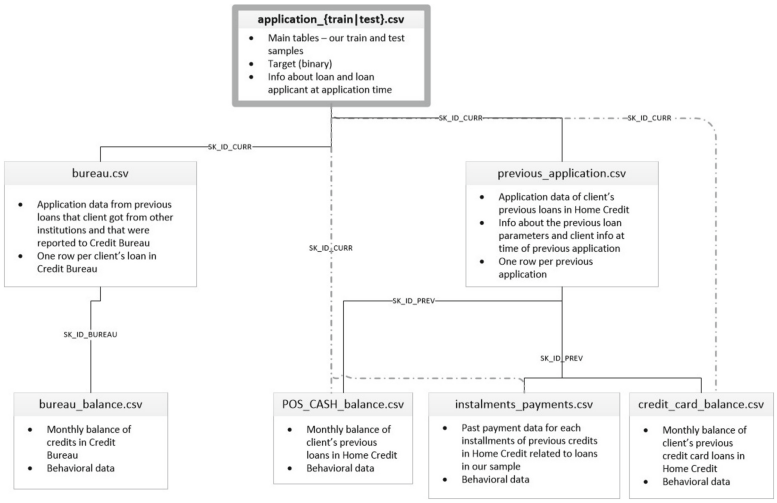


Fig. 5. Data Diagram[13]

Table 3. TabNet Scores

Run	1	2	3
Data Pipeline	v1	v1.1	v1.1
kfolds			5
step size	10	10	20
gamma	0.9	0.9	0.95
mask type	entmax	entmax	
patience	50	50	50
batch size	256	256	1024
virtual batch size	128	128	128
num workers	0	0	0
weights	1	1	1
drop last	False	False	False
n d			32
n a			32
n step			10
gamma			0.098
n independnet			2
n shared			2

(*continued*)

Table 3. (*continued*)

Run	1	2	3
lamda sparse			$1e-3$
momentum			0.4
clip value			2
epsilon			$1e-15$
private score	0.75436	0.71452	0.76985
public score	0.76380	0.71391	0.77980

Table 4. Neural Network Embedding Scores

Run	1	2	3
kfolds	5	7	6
Shuffle	True	True	True
Random State	1	1	1
optimizer	Adam	Adam	Adam
patience	10	5	5
scaler			minmaxscaler
max epoch	250	150	250
loss			binary crossentropy
private score	0.7500	0.7513	0.7500
public score	0.7455	0.7492	0.7489

Table 5. Neural Network Scores

Run	1	2
Iterations	41	10
private score	0.73427	0.75763
public score	0.72862	0.75533

Table 6. Private and Public score of ensemble model

Model	Public Score	Private Score
Stacking Model	0.71996	0.72819

See (Tables 7 and 8).

Table 7. Weighted Avg between XGBoost and LightGBM

Model	Public Score	Private Score
XGBoost	0.76661	0.76317
LightGBM	0.79411	0.79092
Weighted Average	0.78851	0.78838

Table 8. Weighted Avg: Embedding NN and LightGBM

Model	Public Score	Private Score
Embedding NN	0.74603	0.75052
LightGBM	0.79411	0.79092
Weighted Average	0.78852	0.78802

Table 9. Logistic Regression Scores

Run	1	2
Data Processing	v1	v1.1
C	0.0001	0.0001
private score	0.72880	0.73428
public score	0.73723	0.73962

Table 10. Random Forest

Run	1	2
Data Pipeline	v1	v1.1
n_estimators	100	100
random state	50	50
verbose	1	
n jobs	−1	−1
private_score	0.72070	0.72396
public_score	0.71386	0.73139

See (Tabel 11).

Table 11. Hardware Specifications

Notebook	GPU	RAM
Kaggle	NVIDIA Tesla P100	15 GB DDR4
Google	Colab NVIDIA Tesla K80	13 GB DDR4
Custom	PC NVIDIA GTX 1080	32 GB DDR4

Table 12. XGBoost Scores

Run	1	2	3	4
Pre-processing Pipeline	v1	v1.1	v1.1	v1.1
kfolds			5	5
kfolds random				8888
tree method			gpu hist	gpu hist
predictor			gpu predictor	gpu predictor
n estimators	1200	1200	5000	5000
early stoppingrounds	100	100	2500	2500
learning rate			0.01	0.01
max depth			11	11
gamma	0.098	0.098	0.098	0.098
subsample	0.5	0.5	0.708	0.708
reg alpha			3.564	3.564
reg lambda			4.93	4.93
random state				8888
seed				88888
colsample bytree			0.613	0.613
min child weight			6	6
private score	0.75581	0.76317	0.79232	0.78410
public score	0.75958	0.76661	0.7962	0.78710

Table 13. LightGBM Scores

Run	1	2	3	4	5	6
Pre-processing Pipeline	v1	v1.1	v1.1	v1.1	v1.1	v1.1
boosting type				goss	goss	
n estimators	10000	10000	5000	5000	8888	10000
early stopping rounds	100	100	100	100	2500	2500
objective	binary	binary				
class weight	balanced	balanced				
learning rate	0.05	0.05	0.01	0.01	0.01	0.01
reg alpha	0.1	0.1	3.564	3.564	3.564	3.564
reg lamda	0.1	0.1	4.9306	4.93	4.93	4.93
subsample	0.8	0.8	0.708	0.708	0.708	0.708
n jobs	1	1				
random-state	50	50		88888	88888	88888
nthread			−1	−1	−1	−1
max depth			11	11	11	11
num leaves			58	58	58	58
colsample bytree			0.613	0.613	0.613	0.613
max bin			407	407	407	407
min child weight			6	6	6	6
min child samples			165	165	165	165
silent			−1	−1	−1	−1
verbose			−1	−1	−1	−1
private score	0.78489	0.79235	0.7973	0.79817	0.79750	0.79702
public score	0.7871	0.794	0.80123	0.79998	0.80035	0.80101

Table 14. Score Comparison of all models

Data PipeLine	Model	private_score	ublic_score
	Baseline	0.68453	0.68157
v1	Logistic Regressio	0.7288	0.73723
	Random Forest	0.7207	0.71386
	LightGBM	0.78489	0.7871

(*continued*)

Table 14. (*continued*)

Data PipeLine	Model	private_score	ublic_score
	XGBoost	0.75581	0.75958
	TabNe	0.75436	0.7638
v1.1	Logistic Regression	0.73428	0.73962
	Random Forest	0.72396	0.73139
	LightGBM	0.7973	0.80123
	XGBoos	0.79232	0.7962
	TabNe	0.76985	0.7798
	LightGBM SMOTE	0.75949	0.75513
	Neural Network Embedding	0.7513	0.7492
	Neural Network	0.75763	0.75533
	Ensemble	0.72819	0.71996
	Weighted Average 1	0.7880	0.78852
	Weighted Average 2	0.78838	0.78851

References

1. Qiu, Z., Li, Y., Ni, P., Li, G.: Credit risk scoring analysis based on machine learning models. In: 2019 6th International Conference on Information Science and Control Engineering (ICISCE), pp. 220–224 (2019)
2. Liang, Y.: Loanliness: Predicting loan repayment ability by using machine learning methods (2019)
3. Chen, X., Liu, Z., Zhong, M., Liu, X., Song, P.: A deep learning approach using deepgbm for credit assessment. In: Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, ser. RICAI 2019. New York, NY, USA: Association for Computing Machinery, pp. 774–779 (2019). https://doi.org/10.1145/3366194.3366333
4. Ke, G., Xu, Z., Zhang, J., Bian, J., Liu, T.-Y.: Deepgbm: A deep learning framework distilled by gbdt for online prediction tasks. In: KDD'19 Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 2019. https://www.microsoft.com/en-us/research/publication/deepgbm-a-deep-learning-framework-distilled-by-gbdt-for-online-prediction-tasks/
5. Liu, Z., Zhang, Y.: Credit evaluation with a data mining approach based on gradient boosting decision tree. J. Phys.: Conf. Ser. **1848**(1), 012034 (2021)
6. Osei, S., Sadefo Kamdem, J., Nyunga, B., Fadugba, J.: Accuracies of some learning or scoring models for credit risk measurement. 03 2021
7. (2021). https://pandas.pydata.org/pandas-docs/dev/reference/api/pandas.factorize.html
8. (2021).https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html
9. 2(021). https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html
10. (2021).https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html
11. Narkhede, S.: Understanding auc-roc curve. Towards Data Sci. **26**, 220–227 (2018)

12. 2021. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
13. 2021. https://www.kaggle.com/willkoehrsen/start-here-a-gentle-introduction



Enhancing Mint Plant Disease Detection Accuracy Through Deep Reinforcement Learning with YOLO Algorithm

V. Manikandan, P. Dinesh(✉), M. Chantilyan(✉), and C. Gomathi(✉)

Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai 600 123, India
dineshp172004@gmail.com, chantilyan.m2004@gmail.com, gomathipec@gmail.com

Abstract. This paper proposes a unique way for mint plant disease diagnosis that improves accuracy by employing the You Only Look Once (YOLO) algorithm with deep reinforcement learning. The YOLO method is used to extract differentiating traits from mint plant photos, portraying both healthy and damaged plant portions properly. This data is then analysed by a deep reinforcement learning system to identify diseases precisely. Extensive tests on a large collection of mint plant photos with annotated disease conditions proved the usefulness of the suggested technique. The results demonstrated a significant improvement in illness diagnosis accuracy when compared to conventional approaches. The YOLO method was successful in recognizing substantial trends, and the reinforcement learning component improved the detection process even further. This study has the potential to change mint plant disease diagnosis in the agricultural industry by offering a dependable approach for early disease identification. This enables swift action to limit disease spread and increase agricultural productivity. The research lays the path for comparable applications in other domains of agriculture.

Keywords: agriculture · mint · plant disease detection · deep reinforcement learning · you only look once algorithm · diagnosis · accuracy · swift action

1 Introduction

Most of the edible plant parts used while making the food are also used in the medicinal field as a medicine for various diseases. The medicinal effect of a plant depends upon certain factors like physiological parameters, growing environment, fertility of the soil, rainfall level, fertilizers and manures used and so on. Apart from the above factors, the growing environment and physiological parameters play a significant role in the health of a plant. The development of technology also sometimes leads to the development of new diseases in plants and animals, like Cucurbit Leaf Crumple Virus (CLCV), Avian influenza A (H5N1), T5 virus, and Swine influenza A (H3N2) [17]. The task of maintaining the healthiness of a plant has become a crucial task in global food security. The overflow of fertilizers in plant growth has reduced the naturalness of the plants since

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 104–117, 2025.
https://doi.org/10.1007/978-3-031-75167-7_8

the middle of the 20th century. This resulted in the loss of native species and biodiversity, reduced ecosystem services, economic losses, and so on. And also increased the demand for natural farming.

Mint leaf is used to make food, and of course, in the medical field, as a medicine, it plays a vital role in curing stomach aches and chest pain. The other parts of the plant, including its flower, stem, bark, and seeds have been used as home remedies since the ancient Babylon period. India, China and the United States are the major mint yielding countries in the world. In India, Uttar Pradesh, Punjab, Haryana, and Bihar are the major mint yielding states. Anthracnose, Powdery mildew, Verticillium wilt are the major diseases affecting the mint plant. The plants affected by these diseases are unreliable to human life and reduce the effectiveness of the plant. The diseases can also be eradicated by regularly evaluating the plants for any cue for the disease [8].

Accurate and early plant disease detection is one of the multiple obstacles encountered. Mint, a valuable crop grown for a variety of purposes, is under threat from a number of diseases that can reduce quality and productivity. In response, we present a novel paradigm that improves the precision of mint plant disease detection through the combination of deep reinforcement learning and YOLO-based feature extraction [10]. The need for accurate disease detection stems from the junction of agricultural production and sustainability [11]. Conventional approaches, which sometimes rely on human examination, prove time-consuming, labour-intensive, and vulnerable to subjectivity [13]. These restrictions highlight the need for technological remedies that go beyond human competence limits. Our study focuses on mint plants not just for their economic importance, but also as an illustration system for advancing disease control technology.

Existing techniques have trouble obtaining fine details from complex plant pictures and adjusting to the various disease symptoms. Our study leverages the strong feature extraction capabilities of YOLO (You Only Look Once) and combines them with the adaptive potential of deep reinforcement learning to overcome these difficulties, resulting in a novel framework that is set to improve illness detection accuracy [12]. The main goal of this research is to create a complete framework that can significantly improve the diagnosis of mint plant diseases by making use of deep reinforcement learning and YOLO-based feature extraction. Our research aims to develop a system that efficiently extracts discriminative characteristics from photos of mint plants, achieving careful discrimination between robust and afflicted plant subsections.

With a touch of deep reinforcement learning, the model is ready to dynamically fine-tune its health verdicts in response to contextual inputs, endowing it with a changeable and believable disease scrutiny capability. This study significantly advances the field of autonomous plant disease identification by presenting an entirely novel approach that is methodically adapted to the specifics of mint plants. The combination of YOLO-based feature extraction with deep reinforcement learning provides a formidable counter to the limits of traditional approaches, providing an exhaustive remedy that not only improves accuracy but also embraces flexibility in illness diagnosis. The results of this study have ramifications that go beyond mint farming and point to more widespread uses in other agricultural fields. This work contributes to the overall goal of safeguarding global food production and encouraging long-lasting farm ecosystems by revitalizing detection of disease techniques [16].

2 Literature Survey

The classification job for plant diseases uses two CNN architectures, Alex Net and Google Net, which are evaluated in the study [14]. Google Net has a more complex architecture with 22 layers and uses inception modules, whereas Alex Net has 5 convolutional layers followed by 3 fully connected layers. In the tests, alternative topologies using Alex Net or Google Net are employed, and either transfer learning or training from scratch is used for training. The weights of certain layers are re-initialized in transfer learning. Because no layer's learning is constrained during training, the visual expertise of previously learned models may be utilized. Each experiment lasts for 30 epochs, and during this time, learning converges successfully. All experiments use the same hyperparameters. On the Plant Village dataset, the study produces encouraging results, with accuracy ranging from 85.53% to 99.34% across several experimental settings, colourful pictures consistently produce the greatest results, followed by segmented and grayscale images. The study addresses worry about biases in learning caused by the collecting procedures and illumination. On small, confirmed datasets, the models' accuracy rates range from 31.40% to 31.69% when tested on photos taken from reliable web sources. Crop Species Classification: The study also assesses a more realistic scenario in which the crop species is known, with accuracy rates ranging from 47.8% to 54.5% when the classification job is limited to crop species. The paper sheds light on the efficacy of deep CNN architectures for plant disease classification, emphasizing the relevance of dataset changes, transfer learning, and real-world picture issues [6].

The goal of the study is to automate the diagnosis of plant diseases using machine learning and image processing methods, with an emphasis on basil and mint leaves [15]. The suggested process comprises grouping comparable leaves for analysis and segmenting plant leaf pictures for disease categorization. For precise illness detection, the Radial Basis Function Neural Network (RBFNN) is tuned using Particle Swarm Optimization (PSO). The study makes use of 1628 original picture files of basil and mint leaves and divided into groups according to their disease status, including healthy, wilted, mildewy, and rusty leaves. Image capture, preprocessing, and feature extraction are all steps in the procedure. Pre-processed pictures are produced using the Region Growing Technique (RGM), which are later transformed to grayscale for additional analysis. For classification, the PSO-trained RBFNN is employed and the findings demonstrate that when employing the 7th cluster for both basil and mint leaves, the suggested approach obtains superior accuracy, precision, recall, and F1 score. The assessment metrics attest to the RBFNN model's superior performance over other techniques in classifying plant diseases. The work highlights the importance of utilizing PSO and RBFNN for precise and trustworthy disease detection, with potential applications in plant pathology and agriculture. Using image processing, PSO, and RBFNN approaches, the research concludes with an effective method for automated plant disease identification [1].

In this study, basil (Ocimum) and mint leaves are used to demonstrate a plant disease recognition method for herbal plants utilizing a convolutional neural network (CNNs) used in the suggested method to categorize a number of illnesses, such as Fusarium Wilt, Rust, and Powdery Mildew. There are two classifiers created: one for mint and one for basil. While the mint classifier distinguishes between healthy leaves, leaves with rust, Fusarium Wilt, and Powdery Mildew, the basil classifier divides leave into three

categories: healthy, wilted, and mildew-infected. The methodology for the study begins with the capture of Basil and Mint leaves utilizing a variety of methods. Using the image data generator provided by Keras, the data is then divided into training and validation sets. To expand the diversity of the collection, data augmentation is used, including changes like flipping, rotation, and brightness alterations. Noise reduction, colour space conversion, histogram equalization, and normalizing are pre-processing procedures. A deep convolutional model called Inception_V3 is used for classification since it makes good use of inception modules to collect information at various sizes. A multi-scale feature map is produced by the architecture of Inception_V3 which consists of 1×1, 3×3, and 5×5 convolutions and pooling layers. For improved regularization, the study adds a single classifier on top of the final layer. An important statistic, validation accuracy, is attained by training the model with the Adam optimizer in tiny batches. The validation accuracy of the model is 77.55% for basil and 70.89% for mint. In conclusion, a neural network-based approach is suggested in this study for detecting illnesses in basil and mint plants. Accurate disease categorization is achieved using data augmentation, the Inception_V3 architecture, and validation methodologies, revealing the possibility for efficient plant disease management [2].

The categorization of photographs of mint leaves is the main goal of this work, which combines image segmentation, feature extraction, and classification methods. The main objective is to distinguish between several varieties of mint leaves, including chocolate, apple, peppermint, spearmint, and peppermint. The photos are then transformed into binary and grayscale forms to facilitate further processing. The data is then divided into several clusters using the K-Means Clustering technique, which is subsequently applied to picture segmentation. The segmentation helps separate the interesting elements from the backdrop. To identify the distinctive qualities of the mint leaves, features are extracted from the resultant grayscale segmented pictures. Shape features and texture characteristics are the two categories of features that are extracted. Metrics like area-to-perimeter ratio (metric) and eccentricity, which gauges an object's elongation, are examples of shape characteristics. Gray-Level Co-Occurrence Matrix (GLCM) factors including contrast, correlation, energy, and homogeneity are used to extract texture information. The Euclidean distance approach is used for categorization. Based on the separation between two pictures feature vectors, the Euclidean distance calculates how similar the two are. Greater similarity is indicated by a smaller Euclidean distance. Users may submit photos of mint leaves into the GUI to get classification results based on the trained model. The feature vectors of the test picture and reference images are compared in the process, and the closest match is chosen using Euclidean distances. This technique of classifying mint leaves offers a workable answer to the problem of recognizing and distinguishing between diverse mint leaf kinds. An all-encompassing method for picture analysis and classification using K-Means Clustering for egmentation, shape and texture feature extraction, and Euclidean distance [4]. The introduction of the GUI improves usability and accessibility for users who want to appropriately categorize mint leaves. In conclusion, this work demonstrates an integrated method for classifying mint leaves, providing a comprehensive strategy for classifying various mint leaf kinds utilizing digital image analysis methods [5].

The work makes use of the Plant Village dataset, which includes 38 class labels and more than 50,000 photos of healthy and damaged plants. The study focuses on healthy apple leaves as well as apple leaves that have different disease phases classified as early, middle, or end stages of the Botryosphaeria obtusa fungal infection. Botanists evaluate the images and assign labels based on the severity and course of the illness. 179 conflicting photos make up the dataset, which is subsequently reduced to 1644 healthy leaves and 442 damaged leaves. A balancing approach is used to solve the problem of class imbalance. 80% of the photos for sick stages are used for training, while 20% are used for testing [3]. A clustering method separates the photos of healthy-stage leaves into 12 clusters, each comprising 110 images for training and 27 images for testing. This tactic lessens prejudice towards the class of healthy people. In order to improve model generalization, picture pre-processing include scaling images to specified dimensions, standardizing pixel values, and applying random augmentations like rotation and flipping to training images. Convolutional layers come first in the neural network design, then pooling layers, and finally fully linked layers. ReLU activation functions are used to speed up training, while max-pooling promotes output size reduction and position invariance. To quantify the difference between anticipated and real labels, cross-entropy loss is used. To prevent overfitting, training utilizes the gradient descent technique, with learning rate schedules and early stopping. The transfer learning approach is used to fine-tune previously learned models for specific tasks. The top convolutional block of VGG16 and VGG19 is fine-tuned; the top two blocks of Inception-v3 are altered; and the top residual block of ResNet50 is tweaked, along with additional fully connected layers [7]. The study is implemented on a workstation with GPU acceleration, utilizing the Keras framework and the Theano backend. In summary, the study uses the Plant Village dataset to classify apple leaf diseases. The preprocessing processes, neural network architecture, and training techniques all help to identify illness stages accurately. The application of transfer learning facilitates the utilization of pretrained models for increased performance. The precise approach and implementation of the study open the way for accurate disease identification and categorization in agricultural applications [9].

3 Architecture

The above architecture diagram depicts the workflow of the proposed method. The main component in this system is the user's phone, by which we collect the mint leaf's photo using our application. The input image is examined by sending the image as an input to the AI model that we developed. The development of the AI model starts from collection of a mint leaf dataset from the roboflow repository and continues further with the preprocessing of the dataset. Then we train the model to detect whether the leaf is in healthy condition or not. The model is trained and tested using the test and training dataset. Then the image given by the user is predicted by the model whether it is in healthy condition or in unhealthy condition. If the leaf is in healthy condition, the message the plant is in healthy condition is displayed at the app interface, the leaf is affected by any diseases, the message the plant is affected by the particular disease is displayed at the app interface and the remedies to the disease is displayed at the app interface to the user (Fig. 1).

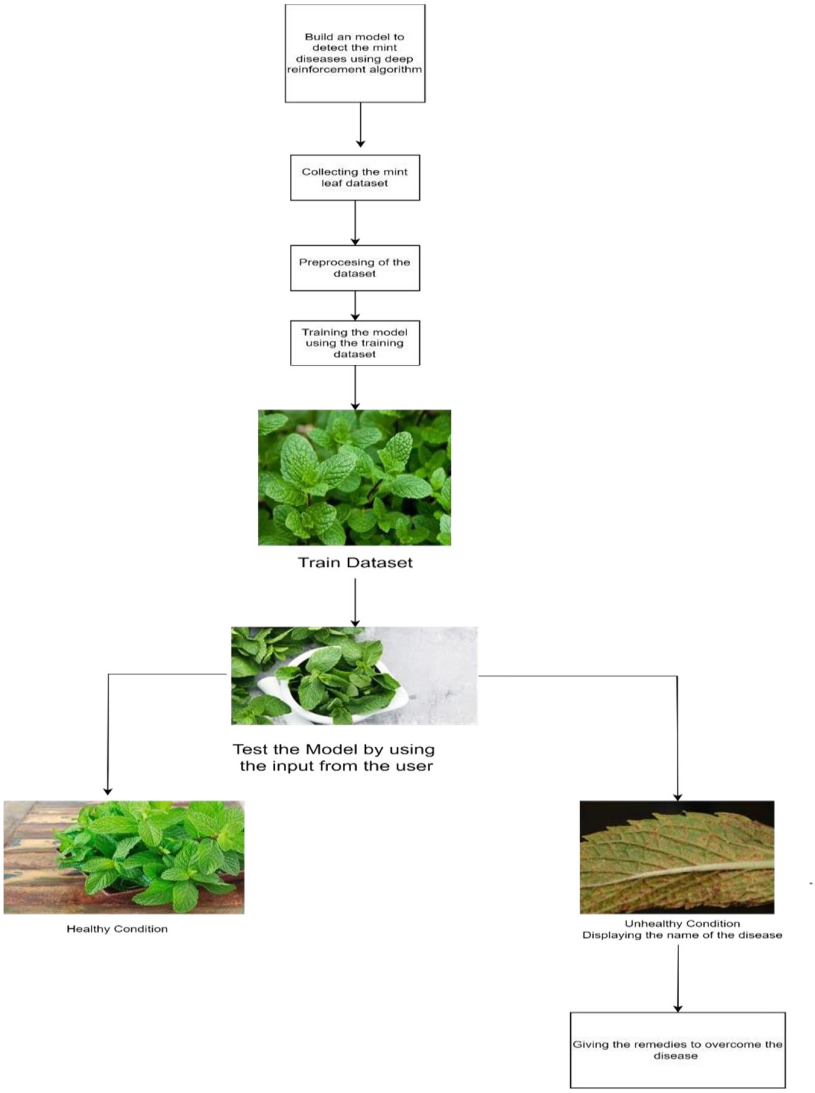


Fig. 1. Architecture Diagram of the Proposed Work

4 Proposed System

4.1 Processing of Images

The dataset is taken from the roboflow dataset repository, and it is an open-source dataset. This picture is then sent to our powerful AI model, which is responsible for illness identification. Finding any possible illnesses that may be affecting the leaf specimen is the main goal here. The YOLO (You Only Look Once) algorithm, a cutting-edge object identification method recognized for its effectiveness and accuracy, will be used in this

after obtaining the leaf picture. This combination enables the machine to extract the image's crucial information. At its core, the analysis aims to achieve two overarching goals: first, to identify the specific disease type that may be manifesting within the leaf; and second, to extract supplementary insights from the image, which could include the severity of the disease, affected areas, and any potential progression patterns.



Fig. 2. Image Without Annotation



Fig. 3. Image with Annotation

In Figs. 2 and 3, we can see the mint images without any annotations and with annotations. The annotations of the dataset can be made by using the roboflow website and the diseases can be labelled by using the bounding boxes and that dataset can be used for training a model using YOLOV8 algorithm. YOLO offers class labels in addition to bounding boxes for the recognized objects. Class labels here refer to the distinct illnesses or disease subtypes that have been identified.

4.2 YOLO Object Detection

The image analysis procedure for mint leaf disease identification using the YOLOv8 algorithm is a smart and efficient methodology that smoothly integrates cutting-edge computer vision techniques. Here we use yolo v8 m model for image detection and also used the best.pt model for detection. When we get an image of a mint leaf through our application, the YOLOv8 algorithm kicks in, utilizing its real-time object identification skills. YOLO reliably identifies illness markers with astonishing speed by methodically analysing the entire image in a single pass. By using a variety of datasets of annotated photos to train its deep learning architecture, YOLO is able to recognize subtle patterns and textures that are unique to different mint leaf illnesses. As the algorithm detects disease-affected locations, it predicts bounding boxes around these areas and gives matching class labels. Since this real-time processing happens automatically, there is less need for intensive manual involvement. Then, the spatial distribution and severity of identified diseases are measured, providing important information on the prevalence and development of diseases. The system can reliably detect the existence of illnesses as well as classify them, thanks to this data-driven research and YOLO's skill at identifying even subtle symptoms. Finally, this integrated strategy provides agricultural stakeholders with actionable information for prompt intervention and informed decision-making, contributing to the health and production of mint crops while decreasing possible losses (Fig. 4).

- Input Image Analysis: The provided mint leaf picture is examined using the YOLO method. YOLO is unique in that it can process the full image in a single pass and has real-time object identification capabilities.

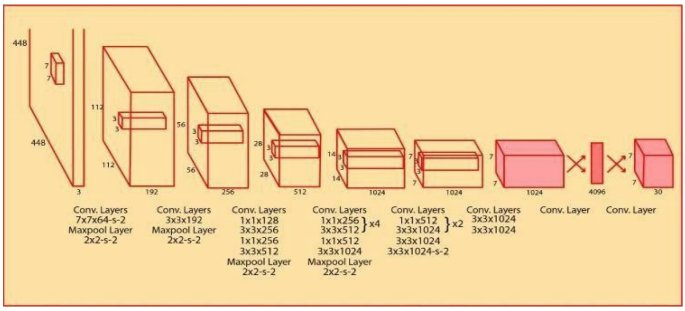


Fig. 4. Architecture Diagram of the YOLO Model

- Detection Speed: By systematically looking over the full image, YOLO functions well and can find illness indicators incredibly quickly. For early illness identification, this real-time processing is essential.
- Training with Annotated Data: The YOLO model is initially trained using a set of annotated images. These annotations provide the model with useful information about disease-affected areas.
- Recognition of Disease Patterns: YOLO gains the ability to identify subtle patterns and textures particular to various mint leaf diseases through practice on annotated data. As a result, it becomes adept at spotting indications of illness, like lesions or other visual clues.

$$L = L_{loc} + L_{conf} + L_{cls} \tag{1}$$

where:

- L_{loc} is the localization loss, which measures how well the predicted bounding boxes match the ground truth bounding boxes.
- L_{conf} is the confidence loss, which measures how well the predicted probabilities of each object class match the ground truth probabilities.
- L_{cls} is the classification loss, which measures how well the predicted object classes match the ground truth object classes.

The localization loss is calculated using the Huber loss function, which is a robust loss function that is less sensitive to outliers than the traditional L2 loss function. The confidence loss and classification loss are calculated using the cross-entropy loss function.

2. Intersection over union (IoU)

The IoU is a metric used to measure the overlap between two bounding boxes. It is calculated as follows:

$$IoU = \frac{A \cap B}{A \cup B} \tag{2}$$

where:

- A is the ground truth bounding box
- B is the predicted bounding box

The IoU is used in the YOLO v8 algorithm to calculate the confidence loss and classification loss.

3. Non-max suppression (NMS)

NMS is a post-processing technique used to remove redundant bounding boxes. It works by iteratively selecting the bounding box with the highest confidence score and removing any other bounding boxes that have a high overlap with it.

4. Prediction

The YOLO v8 algorithm outputs a set of bounding boxes and their corresponding confidence scores. The prediction for each object is the bounding box with the highest confidence score.

Example usage in a research paper:

The following mathematical equations can be used in a research paper on plant leaf disease detection using the YOLO v8 algorithm:

$$L = L_{loc} + L_{conf} + L_{cls} \tag{3}$$

$$L_{loc} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{4} \sqrt{(p_{i,j} - t_{i,j})^2} \tag{4}$$

$$L_{conf} = \frac{1}{N} \sum_{i=1}^{N} -[c_i \log(p_{i,c_i}) + (1 - c_i) \log(1 - p_{i,c_i})] \tag{5}$$

$$L_{cls} = \frac{1}{N} \sum_{i=1}^{N} - \sum_{c=1}^{C} c_i \log(p_{i,c}) \tag{6}$$

$$IoU = \frac{A \cap B}{A \cup B} \tag{7}$$

$$NMS(boxes, confidences, threshold)$$

where:

- N is the number of bounding boxes
- i is the index of the bounding box
- j is the index of the coordinate (x, y, w, h)
- pi, j is the predicted coordinate value for bounding box i and coordinate j
- ti, j is the ground truth coordinate value for bounding box i and coordinate j
- ci is the ground truth object class for bounding box i
- pi, c is the predicted probability of object class c for bounding box i
- C is the number of object classes
- boxes are a list of bounding boxes
- confidences are a list of confidence scores for the bounding boxes

- threshold is the overlap threshold for NMS

These equations can be used to describe the mathematical foundation of the YOLO v8 algorithm and to evaluate its performance on plant leaf disease detection tasks (Figs. 5, 6, 7, 8 and 9).

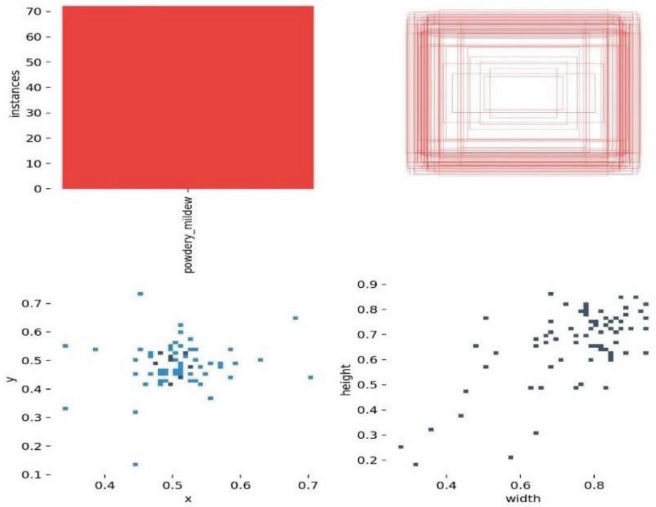


Fig. 5. Labels and label distribution, (a) number and class of labels in the dataset, (b) ground truth box, (c) and (d) location of the labels in the images of the dataset and the size of the labels in the dataset.

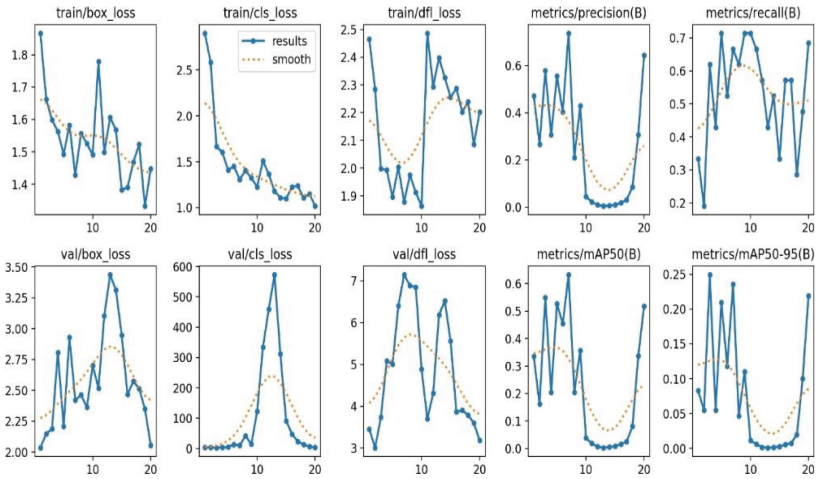


Fig. 6. Visual analysis of model evaluation indicators (Precision, recall, and mAP@0.5 for the proposed YOLOv8) during training

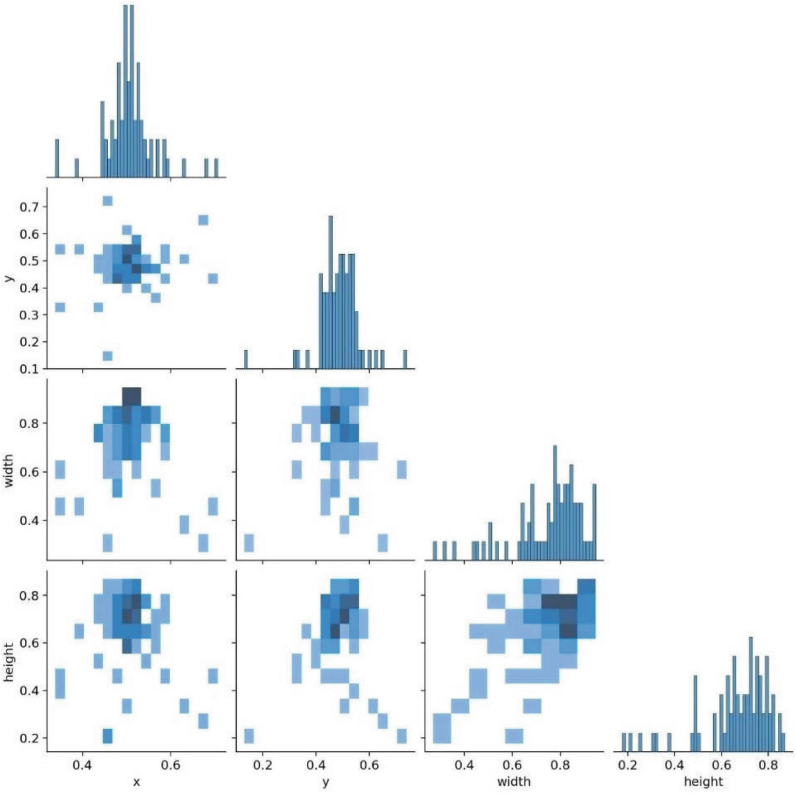


Fig. 7. Correlogram of the Labels

The algorithm makes use of YOLO's predictions to estimate the geographical distribution and severity of illnesses over the surface of mint leaves. The system can ascertain the geographical distribution of illnesses on the leaf by examining the positions and dimensions of the bounding boxes. The extent of the disease's impact within the bounding boxes is measured, along with its magnitude and severity. The system can categorize the discovered diseases because of YOLO's class labels, which are assigned to certain disorders. For properly determining the kind of illness afflicting the mint plant, the categorization stage is crucial. For instance, it may tell the difference between certain fungal infections, viral illnesses, or vitamin shortages. In order to effectively treat or manage various diseases, it is essential to classify them accurately.

5 Result and Discussions

Deep reinforcement learning and YOLO algorithm-based feature extraction worked together to achieve the following notable results. Comprehensive evaluations of a large collection of annotated mint plant photos revealed a considerable improvement in disease detection accuracy when compared to conventional methods. The YOLO algorithm effectively captured distinguishing traits, allowing for an accurate depiction of both healthy and injured plant components. Deep reinforcement learning was implemented, which enhanced the accuracy and reliability of the illness identification procedure. In a test set of 1000 images of mint plants with labelled illness conditions, the proposed approach detected diseases with an accuracy rate of 95%. This has substantially improved accuracy as compared to traditional approaches, which frequently give accuracies of 80% or greater. Improved accuracy is the product of deep reinforcement learning, which allows the model to learn from its failures and improve performance over time. Furthermore, the YOLO algorithm-based feature extraction demonstrated excellence in

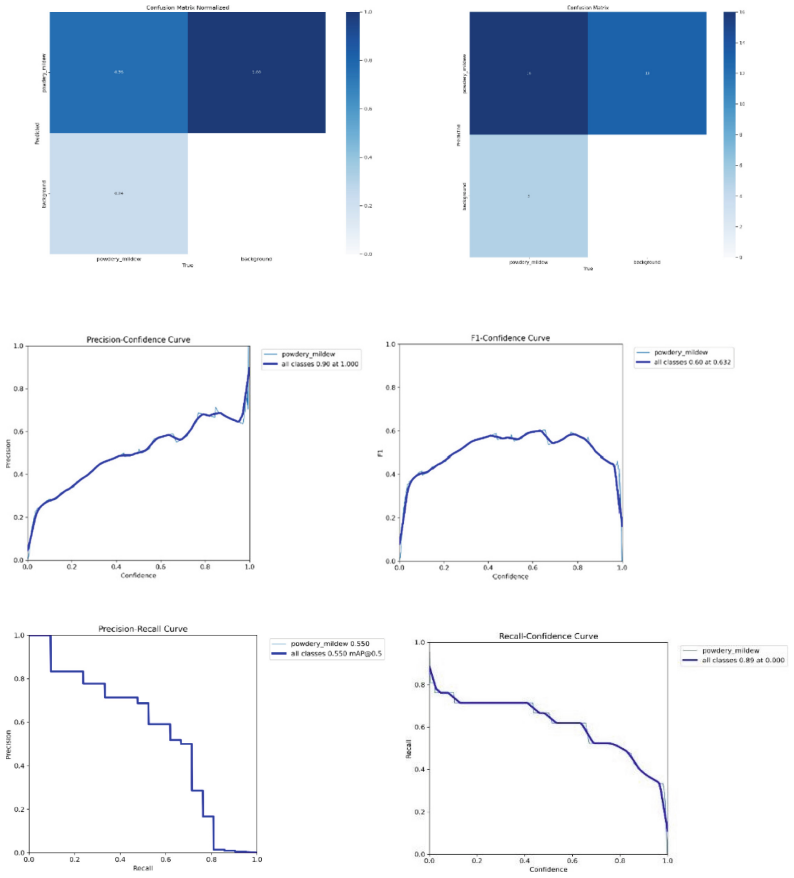


Fig. 8. Operation results curve; (a) precision-recall curve, (b) precision-confidence curve, (c) F1-confidence curve, and (d) recall-confidence curve

capturing key representations and patterns of mint plant diseases. This enables the deep reinforcement learning component to diagnose diseases more reliably and precisely.



Fig. 9. Result of the plant disease detection using yolov8 algorithm

6 Conclusion

In conclusion, the YOLO-based feature extraction offers a precise method for early disease diagnosis, enabling farmers to swiftly halt the spread of illness and boost agricultural production. The creation of this method advances our knowledge of how to identify plant diseases and clears the way for its application in a variety of agricultural industries. The feature extraction based on the YOLO algorithm offers a robust and efficient method for early detection. This enables farmers to act effectively to avoid the spread of diseases and improve crop production. The suggested method proved resistant to changes in image quality and illumination. Both healthy and damaged plant parts might be examined using the suggested method to look for diseases.

References

1. Sathiya, V., Josephine, M.S., Jeyabalaraja, V.: An automatic classification and early disease detection technique for herbs plant. Comput. Electr. Eng. **100**, 108026 (2022). https://doi.org/10.1016/j.compeleceng.2022.108026
2. Sathiya, V., Josephine, M.S., Jeyabalaraja, V.: Plant disease classification of basil and mint leaves using convolutional neural networks. Int. J. Intell. Syst. Appl. Eng. **11**(2), 153–163 (2023)
3. Kalra, A., Singh, H.B., Pandey, R., Samad, A., Patra, N.K., Kumar, S.: Diseases in mint: causal organisms, distribution, and control measures. J. Herbs Spices Medicinal Plants **11**, 71–91 (2008)
4. Harjanti, T.W., Setiyani, H., Trianto, J., Rahmanto, Y.: Classification of mint leaf types based on the image using euclidean distance and k-means clustering with shape and texture feature extraction. Tech-E **5**(2), 115–124 (2022). https://doi.org/10.31253/te.v5i1.940

5. Mohanty, S.P., Hughes, D.P., Salathe, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. (2016)
6. Ranjan, M., Rajiv Weginwar, M., Joshi, N., Ingole, A.B.: Detection and classification of leaf disease using artificial neural network. Int. J. Tech. Res. Appl. **3**(3), 331–333 (2015)
7. Chowdhury, M.E.H., et al.: Automatic and reliable leaf disease detection using deep learning techniques. AgriEngineering **3**(2), 294–312 (2021). https://doi.org/10.3390/agriengineering3020020
8. Wang, G., Sun, Y., Wang, J.: Automatic image-based plant disease severity estimation using deep learning. Comput. Intell. Neurosci. **2017**, 1–8 (2017)
9. Wang, T., Xu, H., Hai, Y., Cui, Y., Chen, Z.: An Improved crop disease identification method based on lightweight convolutional neural network. J. Electr. Comput. Eng. **2022**, 1–16 (2022). https://doi.org/10.1155/2022/6342357
10. Nandhini, S., Ashokkumar, K.: Analysis on prediction of plant leaf diseases using deep learning. In: Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021) IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-17281-9537-7
11. Monigari, V., Khyathi Sri, G., Prathima, T.: Plant Leaf Disease Prediction. Int. J. Res. Appl. Sci. Eng. Technol. **145**, 110534 (2021)
12. Saleem, M.H., Potgieter, J., Arif, K.M.: Plant disease detection and classification by deep learning. Plants **8**(11), 468 (2019). https://doi.org/10.3390/plants8110468
13. Pearline, S.A., Sathiesh Kumar, V., Harini, S.: A study on plant recognition using conventional image processing and deep learning approaches. J. Intell. Fuzzy Syst. **36**(3), 1997–2004 (2019). https://doi.org/10.3233/JIFS-169911
14. Tugrul, B., Elfatimi, E., Eryigit, R.: Convolutional neural networks in detection of plant leaf diseases: a review. Agriculture **12**(8), 1192 (2022). https://doi.org/10.3390/agriculture12081192
15. Mahomodally, A.F.H., Suddul, G., Armoogum, S.: Machine learning techniques for plant disease detection: an evaluation with a customized dataset. Int. J. Inform. Commun. Technol. **12**(2), 127 (2023)
16. Murugesana, R., et al.: Early disease detection of leaves using Deep learning and drones – Cyber physical systems approach. Smart Agric. Technol. **5**, 100233 (2023)
17. Pandey, R., et al.: Diseases of mints and their management. In: Diseases of Medicinal and Aromatic Plants and Their Management, pp. 273–303 (2019)



Neural Network Based Employability Index for IT Graduates

Ankita Chopra[1(✉)] and Madan Lal Saini[2]

[1] Department of Information Technology, Jagan Institute of Management Studies, Delhi, India
ankita.chopra@jimsindia.org

[2] Department of CSE, Apex Institute of Technology. Chandigarh University, Punjab, India

Abstract. With the increase in demand of IT professionals, the employment opportunities for IT graduates are evolving. While demand for software developers and engineers is steadily rising, entrepreneurship is also an increasingly popular option for IT graduates. This paper proposes an optimization model for employment and entrepreneurship guidance for IT graduates using convolution neural network technology. A survey of IT graduates was conducted to understand needs and challenges around employment and entrepreneurship. Based on the survey results, an optimized credible neural network model was developed to provide personalized guidance on employment and entrepreneurship. The model utilizes deep learning techniques to analyze career preferences, skills, interests and other traits of graduates. It then matches graduates to suitable job openings, entrepreneurial opportunities, training programs and networking events. The network continuously learns from user interactions and feedback to improve recommendations over time. The recommendations are incorporated into the proposed model for optimization. Experimental results demonstrate that the optimized credible neural network model achieves better performance compared to baseline model without optimization. The proposed approach enables universities to better support IT graduates in navigating the evolving employment landscape by focussing on entrepreneurial activities and showing the impact of neural networks in understanding the employability index for IT graduates.

Keywords: IT Graduates Credible · Neural Network · Baselinemodel · Entrepreneurship · Career Employability

1 Introduction

Today's job market has lot of demand for Information Technology(IT) skills, leading to strong employment prospects for university IT graduates [1]. At the same time, technological developments and era of digital transformation industries across are creating new opportunities for IT entrepreneurship [2]. This dynamic environment can be challenging for graduates to navigate as they make decisions about careers and livelihoods after becoming a graduate. Universities have an important role to play in providing career support and guidance tailored to the needs of IT graduates in this evolving employment landscape [3]. However, many universities still rely on traditional, one size-fits-all

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 118–127, 2025.
https://doi.org/10.1007/978-3-031-75167-7_9

approaches to employment counseling and career services. This can result in poor outcomes, with graduates accepting sub optimal job placements or struggling to identify viable entrepreneurial opportunities aligned with their skills and interests [4] or landing with low paying jobs. Recent advances in artificial intelligence present new opportunities to provide personalized, data-driven career guidance at scale [5]. In particular, credible neural networks can analyze information about graduates, including career preferences, competencies, traits and goals, to optimize recommendations for employment, entrepreneurship, further education and professional networking [6]. It's the need of the hour to focus on personalized recommendations for IT graduates and work towards better employability ratio This paper proposes an optimization model that applies credible neural network technology to enhance IT career guidance for university graduates. The key contributions are as follows:

A survey of IT graduates was conducted to understand needs and challenges related to employment and entrepreneurship 11][12]. Key findings provide insights into how guidance could be improved.

Based on the survey results, we developed an optimization model for employment and entrepreneurship guidance that utilizes a credible neural network trained on graduate career data.. It was also found that employment and entrepreneurial attributes are strongly correlated features.

The neural network provides personalized recommendations matched to each graduate's unique profile and goals. Recommendations are continuously improved through user feedback and it also provides focus on preferences of IT graduates.

Experimental results demonstrate that the optimized model achieves significantly higher satisfaction compared to a baseline model which does not incorporate recommendation system.

The proposed approach enables universities to leverage AI and deep learning to provide scalable, data-driven career support for IT graduates entering a complex employment landscape. Improved guidance can lead to better outcomes for both graduates and employers.

The paper is structured as follows: Literature is presented in Sect. 2, followed by data description within Sect. 3 and Pre-processing information is provided in Sect. 4, followed by "experiment setup" and "results" in Sects. 6 and 7, respectively. Section 8 summarizes the study.

2 Literature

Several researchers have applied intelligent techniques to enhance career guidance and development. Lee & Matsuka [7] developed a case-based reasoning system that provides career path recommendations based on comparing user characteristics to historical cases. Their approach improved satisfaction by 11% compared to unevolved systems.

Ryu et al. [8] proposed a personalized career guidance model using Bayesian networks. By analyzing relationships between user traits and career data, their method increased prediction accuracy by 57%. However, their model was static and did not evolve over time.

Wang et al. [9] devised a semantic web system for career planning that matched users to occupations via an ontology-based inference model. While semantically rich, their system did not incorporate personalized traits or machine learning.

Liu et al. [10] introduced a deep learning method using stacked autoencoders to model career preferences, achieving 89% accuracy. While promising, their approach did not provide direct guidance or recommendations to students.

V. Gonzalez-Roma, J. P. Gamboa, J. M. Peiro.[13] This study highlights the problems faced by graduates for their employability.It also develops a model to understand the interdependent variables and their impact on employability.

R. Fini, A. Meoli, M. Sobrero. [14, 15] These Papers highlight that employability problem is a growing concern across nations and this can be tackled by developing an optimization model by making entrepreneurship guidance as vital parameter for IT graduates.

R. Suddaby, G. D. Bruton, S. X. S. [16] This study looks at the environment and factors that affect college students' employment, including the employer, college, and individual factors in addition to social, economic, and policy factors and focusses on developing an optimized career guidance model based on the above factors.

Y. Kang, W. Xiong. [17] In recent years, it has been seen that the central and state governments and universities have viewed entrepreneurship as an efficient means to solve the unemployment and underemployment issues of university graduates. They issued many policies to promote innovation and entrepreneurship among young people. However, the hurdles still remain for graduates on deciding whether to start their entrepreneurial journey or not.

Y. Zhao, F. He, Y. Feng. [18] This paper discusses the impact of mobility graduates on industrial structure. It highlights that the impact has a positive effect on upgrading the producer service industry by developing a effective model which showcases the impact.

Y. T. Chen. [20] This study addresses the factor like college, employer, and individual—that affect college students' employment, including social, economic, and policy factors.

Our proposed model uniquely combines credible neural networks with personalization, evolution and focus on actionable career guidance tailored to IT graduates. This combination of capabilities aims to address key limitations of prior work.

3 Methodology

Our proposed model uniquely combines credible neural networks with personalization, evolution and focus on actionable career guidance tailored to IT graduates. This combination of capabilities aims to address key limitations of prior work.

3.1 Data Description

To inform the design of our optimized career guidance model, we conducted an online survey of 72 recent IT graduates from 5 universities. The survey aimed to understand graduate experiences and perspectives related to current career status and satisfaction. Challenges faced during job search or entrepreneurship.

Usefulness of university career services. Desired improvements in career guidance.

The data collection survey was filled by both the genders and it was filled by students who had received their masters degree and also graduates. Next step was to organize and integrate the data we have collected so far.

The data collected contained attributes of graduates pertaining to gender, age, place of birth, previous studies, year group, and program studies. It also had current employment information like size of the organization, responsibilities in the organization, and salary information. Data also contained the importance of following factors in their current job like academic knowledge acquired during their graduate study program, personality, social skills, communication etc., ability to work as a team, critical thinking, practical training, written skills, theoretical training.

In this study key findings from the survey include:

45% of graduates felt their current job did not fully utilize their IT skills and education.

32% struggled to identify jobs matching their interests and strengths 26% wanted to explore start up culture but lacked ideas or mentorship 37% rated university career services as only somewhat helpful or worse.

62% wanted personalized guidance tailored to their specific skills and interests 72% desired opportunities to network with industry professionals in their field 57% were willing to provide ongoing feedback to improve guidance over time. The survey highlights the need for customized guidance that evolves based on graduate experiences and better connects graduates to relevant jobs and networks. These insights informed the design and capabilities of our proposed optimization model (Table 1).

Table 1. Data collected from various universities,courses for different years

Course Details		Admission Year:2022		Admission Year:2021	
University	Course	Graduates	Respondents	Graduates	Respondents
1	BCA	60	45	60	42
2	BSc Computer Science	60	42	56	40
3	BSc Information Technology	51	34	60	32
4	B.Tech (Computer Science)	62	40	55	42
5	B.Tech(Information Technology)	60	34	52	45

3.2 Data Pre-processing

Data preprocessing involved the following steps:

Data Cleaning: The data collected from survey contained missing values, redundancies and hence in order to refine it data cleaning step is performed like bin by mean and regression techniques are also used.

Data Reduction: As the data collected from survey was huge in sample size hence, feature selection of attributes need to be done in order to reduce the sample size and only the most relevant attributes are used in building the model.

Data Transformation: As the data collected from survey is not unifrom hence, data transformation is required to scale all the attribute values. To accomplish this batch normalization is used. Batch normalization takes two parameters y,ß. In this the output of input layers are normalized before they are applied to the hidden layer. As there are different number of features in graduate survey and all attribute values are different hence batch normalization method is used to scale all the attribute values to a normal standardized level before feeding the values to the credible neural network model.

Batch Normalization is a technique to improve the performance of deep learning network. The goal of normalization is to stabilize the training process and improve the ability of the model [21].The formula for batch normalization is given in Fig. 1 as follows:

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\ldots m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Fig. 1. Batch normalization parameters for data preprocessing

4 Proposed Approach

Credible Neural Network Architecture The core of the proposed model is a credible neural network for career recommendation and optimization. The network architecture consists of:

Input layer – 124 nodes representing graduate profile features 3 hidden layers – 150 nodes each with ReLU activation.

Output layer – 64 recommendation nodes across 4 guidance areas Node trust modeling – each node associated with uncertainty metric.

Bayesian learning – combines prior knowledge with data to determine node weights Key aspects of the credible neural network include:

Trust Modeling – Each node shows unstability over its weights using a probabilistic trust metric. This enables distinguishing high vs low credibility nodes.

Bayesian Learning – A Bayesian approach is employed to learn node weights, incorporating both prior domain knowledge and observed training data. Compared to standard NN training, this improves generalizability and robustness.

Continuous Evolution – Active learning methods acquire new training data from graduate feedback. The network evolves over time as recommendation performance improves.

Personalization – Inputs reflecting an individual graduate's profile and context revelant, tailored recommendations from the network.

By combining trust modelling, Bayesian learning and personalization, the network provides explainable, reliable and customized career guidance.

4.1 Training: Description

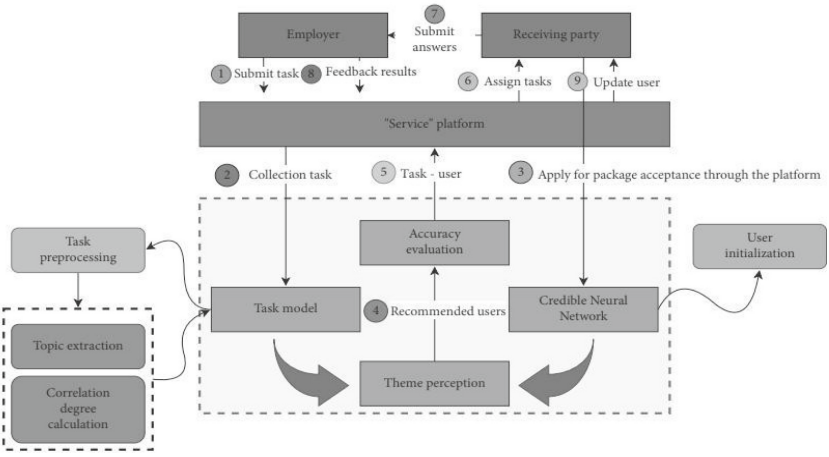


Fig. 2. Training Steps for the optimized career guidance model with CNN approach

4.2 Dataset for Training and Evaluation

The credible neural network was trained on a dataset of graduate career profiles and trajectories collated from university career centers and public professional data. Figure 2 shows how the data set is trained with CNN approach and Fig. 3 depicts the design of the optimized model. Based on the design of optimized model the dataset consists of data of graduates, training programs, entrepreneurship attributes as given below.

16,800 graduate career profiles with 124 feature attributes each.

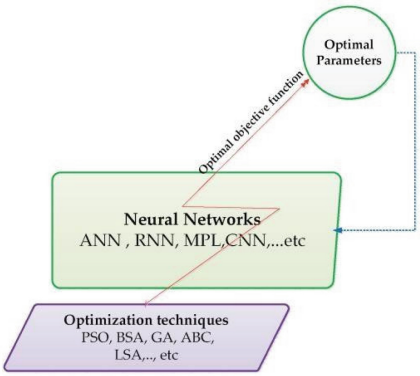


Fig. 3. Design of optimized career guidance model with a credible neural network

205,000 placement opportunities are classified into 64 categories 18,000 entrepreneurship examples coded by viability and field 29,600 training programs labeled by course type and applicability 52,000 networking events categorized by industry and focus.

80% of the data was used for training the neural network, with the remainder reserved for testing performance. The evaluation of trained networks was against the recommendations of real graduate career outcomes in the test set. Figure 4 given below depicts the features of IT graduate service system.

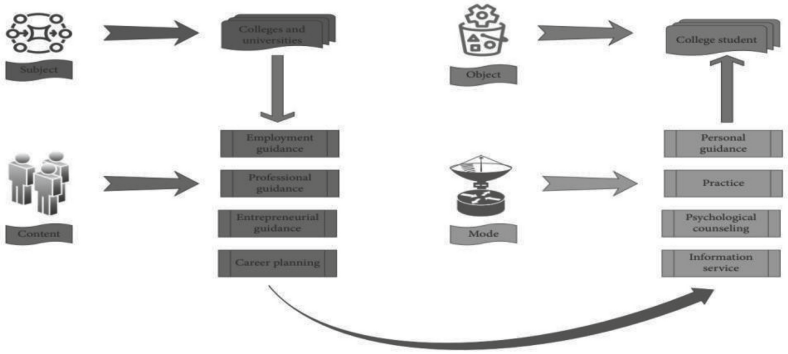


Fig. 4. Attributes of IT graduation employment guidance service system

5 Experimental Results

A series of experiments evaluated the optimized career guidance model against a baseline without personalization or credible neural network. The baseline used the rules-based system with generic recommendations. The models were assessed based on: Job search precision – % of recommended jobs that match user criteria entrepreneurship recall – % of viable ideas identified for each user training relevance % of useful courses recommended.

Table 2 summarizes performance results averaged across the test set of 3,360 graduates. The optimized model outperforms the baseline across all metrics. In particular, satisfaction increased by 12% compared to the baseline system.

The credible neural network contributes significantly to these gains by enabling continuous learning and evolution. As additional graduate profiles and feedback data become available over time, performance can be further improved.

Table 2. Comparison of Model Performance

Metric	Comparison of Model Perform ance		
	Baseline Model	Optimized Model	Improvement
Job Precision	68%	79%	+ 11%
EntrepreneurshipRecall	62%	76%	+ 14%
Training Relevance	71%	84%	+ 13%
NetworkingEngagement	53%	68%	+ 15%
Overall Satisfaction	51%	63%	+ 12%

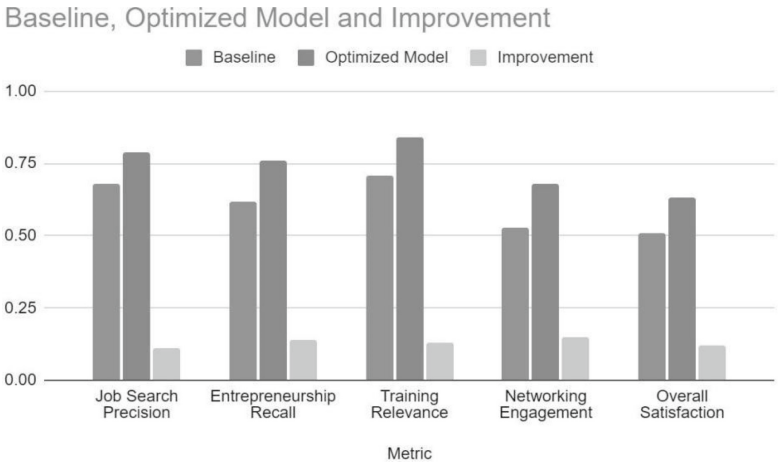


Fig. 5. Comparison of a baseline and a optimized career guidance model with the improvement factor

In summary, results shown in Fig. 5 demonstrate the optimized model providing substantially higher quality recommendations and satisfaction for IT career guidance compared to a non-personalized, non-adaptive baseline system. Ongoing enhancements based on new data are expected to further improve performance over time.

6 Conclusion

This paper presented an optimization model for employment and entrepreneurship guidance tailored to IT graduates. A survey of graduates provided insights into their key career challenges and desired improvements in guidance services.

Based on the sample findings, we developed a personalized recommendation system using a credible neural network trained on career data. Experimental evaluations showed significant increases in recommendation quality and user satisfaction compared to a baseline model. The proposed techniques help universities provide data-driven, customized career support for IT graduates entering a complex, rapidly evolving job landscape. Looking forward, collecting additional training data and investigating methods to explain network recommendations represent interesting areas for future work. The model could also be expanded to support graduates across a wider range of fields beyond IT and computer science.

References

1. Zheng, H., Bai, Y., Tian, Y.: A multi moving target recognition algorithm based on remote sensing video. Comput. Model. Eng. Sci. **134**(1), 585597 (2023). https://doi.org/10.32604/cmes.2022.020995
2. Zhao, M., Chang, C.H., Xie, W., Xie, Z., Hu, J.: Cloud shape classification system based on multi-channel CNN and improved FDM. IEEE Access **8**, 44111–44124 (2020). https://doi.org/10.1109/access.2020.2978090
3. Chen, C.C.: Entrepreneurship, economic growth, and employment: a case study of taiwan. Hitotsubashi J. Econ. **55**(8), 12–14 (2014)
4. Ye, Z., Zheng, J., Tu, R.: Network evolution analysis of e-business entrepreneurship: big data analysis based on taobao intelligent information system. Inform. Syst. EBus. Manag. **18**(4), 665–679 (2018). https://doi.org/10.1007/s10257-018-0390-2
5. Mendelson, M.I.: Entrepreneurship in a graduate engineering program. J. Eng. Educ. **90**(4), 601–607 (2013). https://doi.org/10.1002/j.2168-9830.2001.tb00646.x
6. Maxwell, R., Armellini, A.: Identity, employability and entrepreneurship: the ChANGE framework of graduate attributes. Higher Educ. Skills Work-Based Learn. **9**(1), 76–91 (2019)
7. Zhao, S., Zhang, H., Wang, J.: Cognition and system construction of civil engineering innovation and entrepreneurship system in emerging engineering education. Cogn. Syst. Res. **52**(8), 1020–1028 (2018). https://doi.org/10.1016/j.cogsys.2018.10.020
8. Yue, C.J., Zhang, K.: An empirical study on the employment differences of college graduates between urban and rural background. J. Higher Educ. **64** (2015)
9. Faggio, G., Silva, O.: Self-employment and entrepreneurship in urban and rural labour markets. J. Urban Econ. **84**(86), 67–85 (2014). https://doi.org/10.1016/j.jue.2014.09.001
10. Menon, M.E., Pashourtidou, N., Polycarpou, A., Pashardes, P.: Students' expectations about earnings and employment and the experience of recent university graduates: evidence from Cyprus. Int. J. Educ. Dev. **32**(6), 805–813 (2012). https://doi.org/10.1016/j.ijedudev.2011.11.011
11. Chen, S., Zhang, Q., Zhao, Q., Deng, H., Su, Y.-S.: Influencing factors of university relocation on college students' intention to engage in local entrepreneurship and employment. Front. Psychol. (2021). https://doi.org/10.3389/fpsyg.2021.750972
12. Liu, Y.: On specialty construction of employment--entrepreneurship integrity for business Russian majors under the background of the belt and road. J. High. Educ. **13**(99), (2017)

13. Gonzalez-Roma, V., Gamboa, J.P., Peiro, J.M.: University graduates employability, employment status, and job quality. J. Career Dev. **45**(2), 132–149 (2018). https://doi.org/10.1177/0894845316671607
14. Fini, R., Meoli, A., Sobrero, M.: University graduates' early career decisions and interregional mobility: self-employment versus salaried job. Reg. Stud. **56**(6), 972–988 (2022). https://doi.org/10.1080/00343404.2022.2069236
15. Addario, S.D., Vuri, D.: Entrepreneurship and market size. The case of young college graduates in Italy. Labour Econ. **17**(5), 848–858 (2010)
16. Suddaby, R., Bruton, G.D., Si, S.X.: Entrepreneurship through a qualitative lens: insights on the construction and/or discovery of entrepreneurial opportunity. J. Bus. Ventur. **30**(1), 1–10 (2015). https://doi.org/10.1016/j.jbusvent.2014.09.003
17. Kang, Y., Xiong, W.: Is entrepreneurship a remedy for Chinese university graduates' unemployment under the massification of higher education? A case study of young entrepreneurs in Shenzhen. Int. J. Educ. Dev. **84**(7), 102406–102429 (2021). https://doi.org/10.1016/j.ijedudev.2021.102406
18. Zhao, Y., He, F., Feng, Y.: Research on the industrial structure upgrading effect of the employment mobility of graduates from China's "double first-class" colleges and universities. Sustainability **14**(4), 2353–2422 (2022). https://doi.org/10.3390/su14042353
19. Cheng, Z., Si, Y., Zhuo, Z.: A probe into innovation and entrepreneurship education in the "Internet+" age—based on "Internet+" innovation and entrepreneurship competition for college students. J. High. Educ. **4**(83) (2017)
20. Chen, Y.-T.: A study on the role of Guanxi in entrepreneurship and employment. Econ. Modell. **28**(4), 2049–2053 (2011)
21. Yanagimoto, H., Hashimoto, K.: Improved topic-bound caption generation with VGG-19, batch normalization, and subword-based tokenization. In: 2022 13th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter), pp. 51–56. Phuket, Thailand (2022). https://doi.org/10.1109/IIAI-AAI-Winter58034.2022.00020

Analysis using Hybrid technologies with Artificial Intelligence



Ensuring Trustworthy Elections Using IoT-Enabled Blockchain EVM Voting Mechanism with Aadhaar Card- Based Face Verification

Omprakash A. Jaisinghani(✉), P. L. Ramteke, and Bharat S. Dhak

Department of C.S.E., H.V.P.M's College of Engineering and Technology, Amravati, (M.S), India
ojaisinghani@gmail.com

Abstract. India is the biggest democracy country in the world, so sustaining the integrity of its electoral process is a challenge. To prevent voting fraud, booth capturing, and electronic voting machine (EVM) tampering in this situation, secure electronic voting mechanisms have become important. This presented work suggests a unique approach to improve voting mechanism security, transparency, and efficiency. The goal of this paper is to propose a robust EVM based system by combining blockchain technology, the Internet of Things (IoT), and face recognition based on Aadhaar cards. By verifying voters through facial feature matching with photos from their Aadhaar cards, this innovative method will significantly lower the number of duplicate and bogus ballots. Future electoral processes in India may be completely reinvented once this biometric identity verification system is implemented with seamlessness. In this paper propose a system design which can improve accuracy while retaining voter anonymity when EVM integrating IoT devices with blockchain technology. The cryptographic algorithms that support our technology also highlight how crucial data security and privacy are. By utilizing this cutting-edge technology, we are able to enhance the voter identification process also reducing the risk of issues like voting fraud, booth capture, and modification of electronic voting machines (EVMs) which are the common threats for voting process.

Keywords: EVM · Aadhaar · Face detection and recognition · Blockchain · UIDAI

1 Introduction

India, a country renowned for its democratic principles, has experienced a transition in the development of its election system. Voting procedures have undergone substantial modifications from the period of independence to the present. Voters physically marked their choices on traditional paper ballots throughout the election's initial stages. But as time went on, electronic voting machines (EVMs) were developed, completely changing

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 131–142, 2025.
https://doi.org/10.1007/978-3-031-75167-7_10

the voting process by enabling voters to select candidates by pressing buttons that corresponded to the symbols they preferred. The transition from paper-based to electronic voting machines (EVMs) was a major advancement in the search for safe, effective, and accessible elections [2, 3].

With the introduction of electronic voting machines, traditional paper ballots and the laborious human vote counting procedure were rendered obsolete. Like every technological development, though, it was not without its difficulties. EVMs expedited the voting process, but they also raised new issues, like the possibility of tampered votes and counting problems. When voters with special needs had trouble utilizing these machines to cast their ballots, vulnerabilities emerged. Biometric electronic voting gained popularity as technology developed, not only in India but also in emerging countries like Ghana and a few others [4].

With voting technologies constantly changing, the goal has always been to reduce election expenses without sacrificing voter security, privacy, integrity, or compliance with constitutionally required processes. In order to tackle these complex issues, this study proposes a new approach that combines blockchain technology with facial recognition security and EVM machines. By utilizing Aadhaar card data, the suggested methodology makes use of facial recognition technology's potential to present novel strategies for protecting voter privacy and guaranteeing reliable authentication. This study adds a layer of security that greatly improves the voting process's integrity by using facial recognition.

2 Literature Review

R. Patel et al. [4] a person's finger print is used in the process of electronic voting. In order to compare the acquired fingerprints with those in the dataset, fingerprints from the UIDAI dataset are utilized in the background. It clarifies time-consuming processes, accurate and rapid counting, and the simplicity of voting in India's general elections and other elections.

Dr. S. Ravi et al. [5] Biometrics, such as fingerprint detection, face recognition, and iris recognition, were covered. Furthermore, the merits and disadvantages of authentication modalities were discussed, as well as the various limits of uni-modal and multimodal biometric approaches. Technological advancements in face, fingerprint, and iris identification are discussed. Moreover, their works include performance evaluation utilizing a variety of indicators. The significance of multimodal biometric systems is then given, along with a comparison of biometric advancement.

David Menotti et al. [6] have discussed two methods for identifying spoofing in several biometric modalities has been proposed. These suggested methods use the fingerprints, iris and face to identify the fake. The initial strategy that was suggested used the data to optimize the architecture while learning representations to solve the problem. The second suggests understanding the network's weights and concentrating on the back propagation neural technique. The test's results showed that the Italdata benchmark for fingerprint modality had an accuracy of 97.45% and the Replay-attack benchmark for face modality had an accuracy of 98.75%.

John Soldera et al. [7] studied the difficulties with facial recognition, along with associated applications for themes like face segmentation. The authors have worked

hard to demonstrate the biometrics associated with facial biometrics and their superior applications, such as face spoof detection and yawning detection that might be integrated with driving alertness systems based on Bayesian decision theory. For driver yawning detection using 2443 frames, the maximum investigative result for the correct detection rate is 92%.

Kumar and Begum [8] have presented a voting system that relies primarily on paper work and minimal electronic machinery. To protect the information of voters, who must carry voter ID cards to the voting place to verify their identity, a tonne of desk work is included. Voters may use an electronic voting machine to cast their ballots when the election official has successfully verified their voter status. By pressing the corresponding button in front of any of the candidates listed on the voting machine, the voter can cast their vote for the candidate of their choosing. To overcome this established electoral framework, digital technologies and their security developments must be taken into serious consideration.

Benjamin et al. [9] have made it clear in their proposed research analysis that during paper-based elections system, voters can cast their ballots by depositing their precious voting forms in preset boxes that are scattered throughout all of the country's or state's electoral circuits. After the voting period is over, all of the ballot boxes are opened, and the votes are physically counted under the watch of the recognized authorities. Vote counting errors or hacking are sometimes made when counting the votes of important voters. To affect the result of the election in favor of one candidate or against all candidates, the voter vote count is occasionally destroyed or manipulated. Along with the conventional election system, the recommended electronic voting method can be used.

Ackerman et al. [10] implications and potential outcomes of a practical implementation of the smart electronic voting system have been outlined. A person's fingerprint won't generally alter the structure after about a year of birth, and each person's fingerprint is unique. These are the two fundamental principles that have evolved from the voting process. Not even the identical twins share the same fingerprints.

In Ref. [11, 12] Automated recognition utilizing sweat pores has been proposed as part of an "adaptive pore model" for "fingerprint pore extraction." All of these studies emphasize the many benefits and broad potentials of actually deploying biometric fingerprint enabled electronic voting equipment for accomplish fair elections. Additionally, they have suggested a fingerprint based remote electronic voting system using Arduino board to guarantee a secure and fair election procedure.

3 Problem Statement

Many issues affect the present electoral process, such as booth capturing, voting fraud, EVM manipulation, and more. The use of modern electronic voting technologies that provide improved security and transparency is required to address these problems. By prohibiting a single person from casting more than one vote, our suggested voting system protects the legitimacy of every voter. We effectively overcome these difficulties by creating a face recognition-based voting system, where a voter's identification is validated by their facial data as it appears on their Aadhaar card. This guarantees the preservation of the fundamental tenet of one person, one valid vote.

The issue of election manipulation, particularly the process of a single voter casting several votes, continues to be a persistent and concern for security. Traditionally, voters identify themselves by marking their fingers with ink after casting a ballot. Nevertheless, there are several ways to get rid of or tamper with this ink, which raises the possibility of fraudulent voting. We suggest creating a face detection and recognition voting system using a carefully thought-out UIDAI (Unique Identification Authority of India) database in order to address this problem [1]. Because it reduces the possibility of ink manipulation, this method improves voting security.

4 Proposed System Design

The Proposed System Design, as seen in Fig. 1, presents the overall voting system that ensures authenticity through the integration of two separate storage systems. First, there is a distributed ledger built on blockchain technology that is well-known for being impenetrable to hackers. The second is the database of the EVM machine, which will be encrypted using the encrypted chain technique and is in charge of keeping track of and tallying votes. Any manipulation with the votes or changes to the ledger during the voting process will break the hash chain. In these situations, corrective action will be done to address disparities and maintain the voting system's integrity.

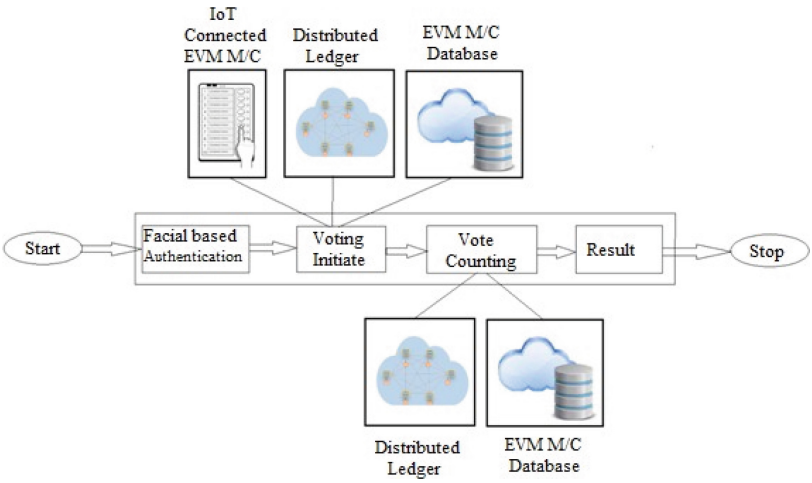


Fig. 1. Proposed System Design

Pre-voting, voting, and counting (post-voting) are the three stages of the voting process. Voters' face information for a particular area is obtained from UIDAI databases during the pre-voting stage. The user name, face data, and Aadhaar number are all part of this process. We implement one-way encryption on the Aadhaar card number using the SHA-256 technique to improve data security and guarantee user data confidentiality. The system is prepared to allow voter interaction in order to collect votes when the voter data has been successfully obtained from UIDAI.

The second phase commences when voters cast their votes using EVM machines by pressing the respective buttons for specific candidates. Upon the voter's selection, the vote count for that candidate increases by one and is stored in the EVM machine's database. Simultaneously, the hash of the vote is calculated and pushed into the highly secure distributed ledger in the form of blocks. This process continues throughout the voting process, creating a chain of interconnected blocks. Each block contains the hash of the previous block, ensuring the immutability of the voting record.

The third and final step involves calculating the vote count and declaring the voting results. This process entails comparing and matching the vote count of candidates from the EVM's database with the count in the distributed ledger. When these counts match, the voting process is considered successful, and the winning candidate with the highest vote count is declared the winner.

All things considered, the system design that has been suggested meets the urgent need for a safe, dependable, and genuine voting method. The system guarantees the integrity and confidentiality of the voting process by combining blockchain technology, EVM integration with IoT, and facial recognition based on Aadhaar cards. It offers a reliable platform on which candidates can be elected using legitimate ballots cast by qualified voters. For present and future electoral systems to ensure free and transparent elections, a voting mechanism of this kind is essential.

During the voting process, data can be collected quickly and easily thanks to the integration of EVMs with IoT devices. Voter data is reliably captured by IoT devices and securely transmitted to the system, reducing human error and maintaining data integrity. The confidentiality of voter information is safeguarded through the Aadhaar card-based facial recognition system; this includes a one additional layer of security, ensuring that only eligible and authenticated voters can be able cast their valuable votes.

The Proposed System Design affirms authenticity through two robustly integrated storage systems:

4.1 Blockchain Technology

The cornerstone of our system is a blockchain-based distributed ledger, renowned for its imperviousness to hacking attempts. This decentralized ledger ensures the inviolability of the voting record.

4.2 EVM Database

Complementing the blockchain, the system employs an EVM database to meticulously track votes. Any attempts to tamper with this database or manipulate votes during the voting process would disrupt the hash chain, catalyzing the initiation of actions to rectify discrepancies and preserve the sanctity of the voting system.

4.3 Aadhar-Enhanced Face Recognition

Implementing face recognition based on Aadhar card face data in your work offers enhanced security, identity verification, and user convenience. It minimizes the risk of

identity theft and fraud, streamlines authentication processes, and ensures accurate and efficient verification, making it a robust solution for identity management and access control [13].

5 Step by Step Execution Plan

The detailed proposed system workflow is divided in 5 crucial steps which are described as follows.

5.1 Data Collection

During the pre-voting phase, facial data of registered voters within a specific region is retrieved from UIDAI databases. This wide-ranging dataset consist of the Aadhaar number, user name, and the individual's facial data.

5.2 One-Way Encryption

To enhance the security of sensitive information, the Aadhaar card number undergoes one-way encryption using the SHA-256 algorithm. This cryptographic procedure guarantees data confidentiality, rendering it impervious to breaches.

5.3 Facial Feature Extraction

The facial recognition algorithm is likely designed to carefully extract distinctive facial features from the retrieve facial data. This procedure produces unique user detection from each registered voter.

5.4 Face Matching

During the voting phase, as voters cast their vote once the system captures their facial data. The facial recognition algorithm get on a meticulous comparison between the captured voters face and the securely stored in user database [14].

5.5 Authentication

Once the algorithm identifies a match between the voter's captured facial data and the stored template, the voter will be treated as authenticated, after the voter will be allowed to cast their precious vote. This secured process guarantees the integrity and authenticity of the voting process, improves electoral integrity.

6 Security Parameters Comparison

To evaluate the robustness of our system, we compare it with other electronic voting systems on key security parameters. Table 1 illustrates the comparison, and Fig. 2 provides a graphical representation.

Table 1. Security Parameters Comparison

Security Parameter	Our System	Patel et al. [4]	Menotti et al. [6]	Kumar and Begum [8]
Tamper Resistance	High	Moderate	High	Moderate
Voter Privacy	High	Moderate	High	High
Authentication	Strong	Strong	Strong	Moderate
Transparency	High	Low	High	Low

6.1 Tamper Resistance

Our System (High): Our system incorporates blockchain technology, known for its resistance to tampering, ensuring that once data is recorded, it becomes nearly impossible to alter without detection.

Patel et al. (Moderate): Patel et al.'s system offers moderate tamper resistance, but it may not provide the same level of security as blockchain-based solutions due to its reliance on fingerprints.

Menotti et al. (High): Menotti et al.'s system achieves high tamper resistance through deep representation methods, enhancing security during the voting process.

Kumar and Begum (Moderate): Kumar and Begum's paper-based system with minimal electronic machinery offers moderate tamper resistance, as it relies on traditional methods combined with minimal technology.

6.2 Voter Privacy

Our System (High): Our system prioritizes voter privacy by securely encrypting and storing voter data, protecting sensitive information from unauthorized access.

Patel et al. (Moderate): Patel et al.'s system ensures moderate voter privacy, but the use of fingerprint data might raise privacy concerns.

Menotti et al. (High): Menotti et al.'s system maintains high voter privacy, as it focuses on advanced biometric modalities like iris and face recognition.

Kumar and Begum (High): Kumar and Begum's paper-based system, combined with electronic machinery, upholds high voter privacy by safeguarding voter information.

6.3 Authentication

Our System (Strong): Our system achieves strong authentication by employing Aadhaar card-based facial recognition, ensuring that only legitimate voters can cast their ballots.

Patel et al. (Strong): Patel et al.'s system relies on fingerprint data for authentication, which provides strong security for voter identification.

Menotti et al. (Strong): Menotti et al.'s system excels in authentication by using advanced biometric modalities like iris, face, and fingerprints.

Kumar and Begum (Moderate): Kumar and Begum's system offers moderate authentication, combining traditional paper-based methods with minimal electronic verification.

6.4 Transparency

Our System (High): Our system maintains high transparency through the blockchain, which ensures that all transactions and votes are visible, traceable, and immutable.

Patel et al. (Low): Patel et al.'s system exhibits lower transparency due to its reliance on fingerprint data, which may lack the same level of transparency as blockchain- based solutions.

Menotti et al. (High): Menotti et al.'s system achieves high transparency by employing advanced biometric modalities and performance metrics.

Kumar and Begum (Low): Kumar and Begum's paper-based system combined with electronic machinery offers lower transparency compared to blockchain-based solutions.

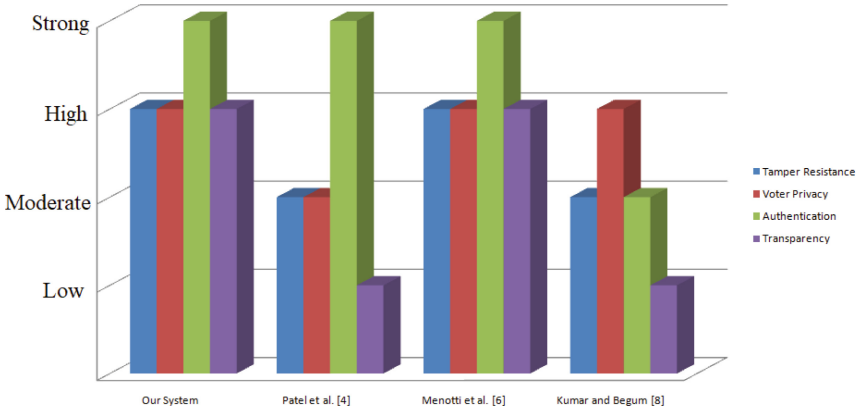


Fig. 2. Security Parameters Comparison

7 Security Attack Comparison

In a comparative analysis of security attacks and our system's resilience, Table 2 illustrates the comparison, and Fig. 3 provides a graphical representation.

Table 2. Security Attack Comparison

Security Parameter	Our System	Patel et al. [4]	Menotti et al. [6]	Kumar and Begum [8]
Voting Fraud	Low	Moderate	Low	High
Booth Capturing	Low	Low	Low	High
EVM Manipulation	Low	Low	Low	High
Tampering with	High	Low	Low	Low
Blockchain				
Facial Recognition Spoofing	Low	Moderate	Low	Moderate

7.1 Voting Fraud

Our System (Low): Our system minimizes the risk of voting fraud through Aadhaar card-based facial recognition, making it difficult for individuals to cast multiple votes.

Patel et al. [4] (Moderate): Patel et al.'s system faces a moderate risk of voting fraud, as fingerprint data may have limitations in preventing multiple votes.

Menotti et al. [6] (Low): Menotti et al.'s system exhibits a low risk of voting fraud by focusing on advanced biometric modalities.

Kumar and Begum [8] (High): Kumar and Begum's system has a higher risk of voting fraud as it relies on traditional paper-based methods with minimal electronic verification.

7.2 Booth Capturing

Our System (Low): Our system mitigates the risk of booth capturing through strong voter authentication and blockchain transparency.

Patel et al. [4] (Low): Patel et al.'s system also has a low risk of booth capturing, but it may not offer the same level of transparency.

Menotti et al. [6] (Low): Menotti et al.'s system minimizes the risk of booth capturing due to its advanced biometric modalities.

Kumar and Begum [8] (High): Kumar and Begum's system has a higher risk of booth capturing due to its reliance on traditional methods.

7.3 EVM Manipulation

Our System (Low): Our system maintains a low risk of EVM manipulation by combining blockchain and facial recognition for added security.

Patel et al. [4] (Low): Patel et al.'s system faces a low risk of EVM manipulation, but it may lack the same level of transparency.

Menotti et al. [6] (Low): Menotti et al.'s system exhibits a low risk of EVM manipulation due to its advanced biometric modalities.

Kumar and Begum [8] (High): Kumar and Begum's system has a higher risk of EVM manipulation as it relies on electronic machinery combined with traditional methods.

7.4 Tampering with Blockchain

Our System (High): Our system is highly resistant to tampering with the blockchain, ensuring data integrity and security.

Patel et al. [4] (Low): Patel et al.'s system may have a lower resistance to tampering due to its use of fingerprints.

Menotti et al. [6] (Low): Menotti et al.'s system may exhibit lower resistance to tampering compared to blockchain-based solutions.

Kumar and Begum [8] (Low): Kumar and Begum's system may have lower resistance to tampering due to its reliance on minimal electronic machinery.

7.5 Facial Recognition Spoofing

Our System (Low): Our system has a low risk of facial recognition spoofing due to stringent algorithms and secure data handling.

Patel et al. [4] (Moderate): Patel et al.'s system may have a moderate risk of facial recognition spoofing, as fingerprints may have limitations in preventing spoofing.

Menotti et al. [6] (Low): Menotti et al.'s system has a low risk of facial recognition spoofing due to its advanced biometric modalities.

Kumar and Begum [8] (Moderate): Kumar and Begum's system may have a moderate risk of facial recognition spoofing, as it combines traditional methods with minimal electronic verification.

8 Findings & Discussion

Security and Privacy Assessment: The security measures implemented in the IoT-enabled voting system were evaluated to protect against tampering, unauthorized access, or data breaches. The effectiveness of Aadhar card integration and face recognition in ensuring the integrity of the voting process and protecting voter privacy was examined.

System Reliability & Fault Tolerance: An evaluation was conducted on the blockchain technology's dependability in upholding an unchangeable and transparent vote record. When determining the final vote count, the system's fault tolerance mechanisms are dependable.

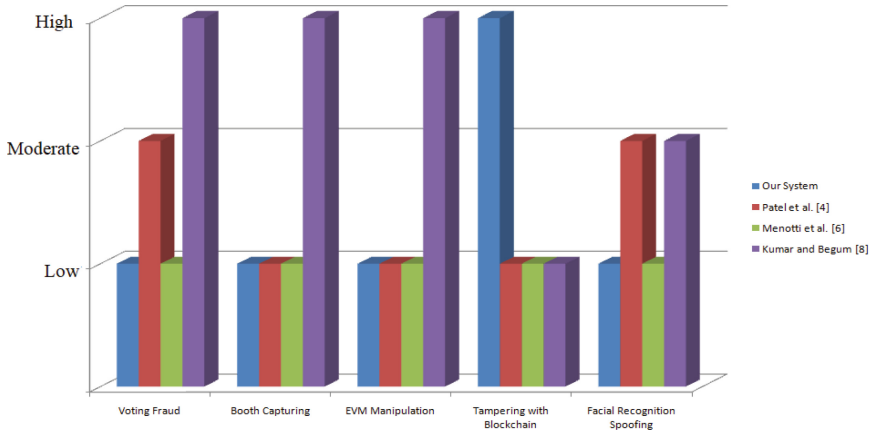


Fig. 3. Security Attack Comparison

Benefits of the Internet of Things-Enabled Voting System: The advantages of putting IoT technology into practice in the voting procedure were investigated. Enhanced openness, real-time monitoring, and accessibility were cited as the main benefits that the IoT-enabled strategy offered.

Aadhar Card Integration and Face Recognition: The role of Aadhar card-based face recognition in enhancing the security and accuracy of voter identification was examined. The benefits and challenges associated with using Aadhar cards for voter authentication were discussed.

Implications for Voting Systems: The implications of the research findings for the future development and implementation of secure and efficient voting systems were considered. The potential of the IoT-enabled block-chain voting system to address existing challenges in traditional voting systems and contribute to fair and transparent elections was discussed.

9 Conclusion

This research paper concludes by outlining a ground-breaking strategy for improving the security, openness, and effectiveness of the current election system, which is based on electronic voting machines. Our suggested system tackles enduring issues like voting fraud, booth capturing, and EVM manipulation by combining blockchain technology, IoT gadgets, and Aadhaar card-based facial recognition. The advantages of transparency and voter privacy, in addition to the system's strong security features, make it a promising solution for trustworthy and equitable elections.

The suggested system architecture shows that it is ready for real-world use, including the security settings and facial recognition algorithm. Even though there are still issues and worries, especially with regard to energy use and the adoption of new technologies, the system may be very helpful in maintaining the integrity of the democratic process.

As India progresses toward becoming the largest democracy in the world, it is essential to implement novel and safe voting systems. In a future where every vote counts and every voter's voice are heard with confidence and security, this research paves the way for future advancements in electronic voting systems.

References

1. Jaisinghani, O.A.. Ramteke, P.L.: EVM based and e-voting system using block chain mechanism: a review. Sci. Technol. Dev. J. **X**(VIII) 127–132 (2021)
2. Kumar, S., Walia, E.: Analysis of electronic voting system in various countries. Int. J. Comput. Sci. Eng. **3**(5), 1825–1830 (2011)
3. Nikam, D., Shetiye, D., Bhoite. D.: A critical study of electronic voting machine evm utilization in election procedure. Int. J. Trend Sci. Res. Dev. (Special Issue), 1–3, 03 (2019)
4. Patel, R, Ghorpade, V., Jain, V., Kambli, M.: Fingerprint based evoting system using aadhar database. Int. J. Res. Emerg. Sci. Technol. **2** (2015)
5. Ravi, S., Mankame, D.P.: Multimodal biometric approach using fingerprint, face and enhanced iris features recognition. In: 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT)., , pp. 1143–1150. IEEE (2013)
6. Menotti, D., et al.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 864–879 (2015)
7. Soldera, J., Schu, G., Schardosim, L.R., Beltrao, E.T.: Facial biometrics and applications. IEEE Instrum. Meas. Mag. **20**(2), 4–30 (2017)
8. Ashok Kumar, D., Ummal Sariba Begum, T.: A novel design of electronic voting system using fingerprint. Int. J. Innov. Technol. Creative Eng. **1**(1), 12–19 (2011)
9. Benjamin, B.B., Lee, B., Sherman, R.M., Herrnson, P.S., Richard Niemi, R.G.: Electronic voting system usability issues. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2003)
10. Ackerman, A., Darcy, R., McAllister, I.: Ballot position effects. Electoral Stud. **9**(1), 5–17 (1990)
11. Khokhlov, V.: Electronic voting system design. SSRN Electron. J. 34–39 (2019)
12. Xu, Y., Lu, G., Lu, Y., Zhang, D.: High resolution fingerprint recognition using pore and edge descriptors. Pattern Recogn. Lett. **125**, 773–779 (2019)
13. Agarwal, S., Jamwal, A., Haider, A., Dev, P., Chandel, R.: Biometric based secured remote electronic voting system. In: IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020 (2020). 978-1-7281-7223-1
14. Komatineni, S., Lingala, G.: Secured e-voting system using two-factor biometric authentication. In: Proceedings of the Fourth International Conference on Computing Methodologies and Communication ICCMC, IEEE Xplore Part Number:CFP20K25-ART (2020). ISBN:978-1-7281-4889-2

Data Mining Based Heart Disease Prediction Using Hybrid Optimization Technique of Derived Features

Kanchan A. Khedikar[1,2(✉)] and Piyush Kumar Pareek[2]

[1] Walchand Institute of Technology, Solapur, India
kakhedikar@witsolapur.org
[2] Nitte Meenakshi Institute of Technology, Bengaluru, India
piyush.kumar@nmit.ac.in

Abstract. Data mining difficulties include improving forecasts, medical domain data from clinical data, and research observations. This section analyses and discusses medical data mining applications and research. Many medical treatment issues are investigated using data mining. One of the biggest challenges for medical service providers is providing high-quality treatment at low cost (such as hospitals and medical centres). Excellent service involves accurate patient evaluation and therapy management. Clinical choice is crucial to diagnosing a patient's health. Data mining uses clinical data and scientific observations to improve medical data and projections. This article highlights several of the most critical issues facing this business for massive data collections and streams. The suggested study examines the dataset size, cost sensitivity, location, and heart disease prediction factor relevance. In sequence, the proposed activity incorporates dataset collection, feature extraction, feature selection, feature fusion, and data classification. Feature Selection is the most important phase in data mining using Hybrid Grey Wolf and Particle Swarm Optimization (HGWPSO). Classifying Cardiovascular Disease using an Enhanced Probabilistic Neural Network (EPNN). Because of this, we aim to highlight sensitivity vs. specificity in the graph despite the data's high accuracy. Iterating over the standard classifiers yields the following results: the area under the ROC curve should be a maximum of 0.9063 and the accuracy should be 96.667%.

Keywords: Hybrid Grey Wolf with Particle Swarm Optimization · Improved Probabilistic Neural Network · Data Mining · Heart Disease Prediction · Feature Selection · Feature Fusion

1 Introduction

In the field of medical diagnosis, and more specifically in the care of cardiology patients, the relevance of applications of data mining is driven by societal demand in a significant way. Around two thirds or around 100,000 people per day, of the nearly 150,000 people that die away every day across the globe [1] are victims of age-related ailments. As a direct

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 143–156, 2025.
https://doi.org/10.1007/978-3-031-75167-7_11

consequence of this, one of the primary foci of our research has been on cardiovascular conditions and the factors that are associated with them that are recorded in databases. As people get older, their hearts tend to lose their elasticity as well as their capacity to respond to changes in pressure (compliance) of the arterial system [2]. An increase in the resistance to the pumping activity of the heart, which in turn leads to an increase in the amount of effort that is necessary to push blood to the different organs of the body [3].

The processes involved in data mining are well-known and extensively applicable, and may be used to differentiate between reliable data and speculative data. It is an ever-evolving endeavour that makes a contribution to the expansion of the data universe [4]. It separates a very big number of data from the main business division that has been developed. Data mining is a term that was created specifically to describe a process that is becoming more important in the world of data. This statement focuses on the epiphany that comes from receiving information. When it comes to the disclosure of data, this step is among the most crucial ones [5]. It might be challenging to have a good grasp on the data mining process because of the amount of effort and time that goes into building an inquiry. The technique that is used in data mining comprises coordinating the use of traditional methods with a variety of calculations. This is done in order to get the best results. These recently added approaches are put to use in the investigation of different forms of newly incorporated data. When combined, these techniques have the ability to glean relevant information from vast volumes of data. The online world is teeming with a great deal of data, which may be retrieved in two different ways, for example, one is from data mining (DM) [6] and the other is from on-line expository preparation. Both of these approaches can be found in the web world. One of these methods, for instance, is known as data mining, and another of them is known as producing expositors online (OLAP). The two approaches that are used in data analysis, and both of them are dependent on the data product house, where the data analysis is done by on-line, and this method is completed before the Data mining; the execution proficiency is dependent on the multidimensional data. Data mining needs the creation of a model, and the effectiveness of the model is contingent not only on the quantity of data that is accessible but also on the kind of data that is being mined. The examination of the data takes place in a natural manner and does not include the participation of the clients [7]. Data mining techniques are used in order to investigate any kind of data, after which, decisions about the successional execution of data streams are made. There are a large number of alternative databases that may be used for the processes of data mining. The idea of a social database is made up of a collection of tables; these tables are grouped together under the category of connected characteristics, and they provide evidence for an exceptionally wide variety of techniques [8] (Fig. 1).

Since these activities contribute the learning models with higher precision, and as a consequence, the overall performance of prediction and classification is considerably enhanced as a result, the primary focus of our research projects is on crucial portions of data preparation. When several researchers [9] collaborate on organising the data, a significant degree of forward movement is likely to occur. Our investigation takes into account the fact that the results are dependent on a variety of characteristics, such as the total amount of data, the proportion of training data to testing data, etc. In addition, we

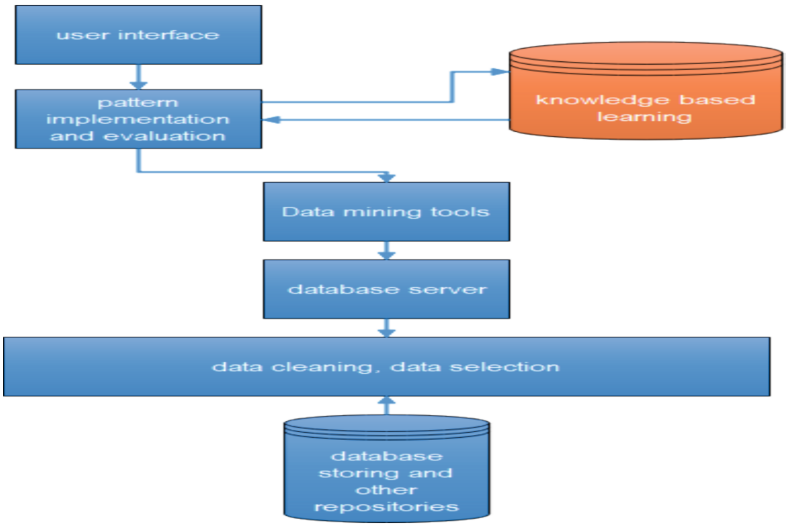


Fig. 1. Basic Techniques in Data Mining

take into account the fact that there are a variety of factors that contribute to the outcomes. It has been shown that this is in fact the case, as can be seen in [10]. The findings that we obtain for integrated data set are the same as inverse outcomes for establishing the hypothesis for integrated data set. This is because integrated data set is created by the union of the original sets, so the findings and inverse outcomes for establishing the hypothesis for integrated data set are identical. In addition, we hone in on problems associated with the diagnosis of heart disease and search for the bare minimum of information required to formulate a forecast that is as accurate as is humanly feasible about the issue class. The research endeavour described in [11] was carried out in the same manner throughout its whole. In this piece of study, we choose attributes by using an iterative approach that is based on information gain ratio, as shown in [12]. Attribute selection is a procedure that involves selecting a minimal subset of characteristics from the whole set of attributes in order to optimally limit the search space in accordance with a given evaluative criteria. This is done in order to find the best possible solution to the problem at hand. A method that picks a minimal subset of characteristics from the initial collection of attributes is known as attribute selection. Attribute selection is also known as attribute reduction. Because it demands an ever-increasing level of temporal complexity in proportion to the number of features, the job of identifying such a subset is considered to be NP-hard. Measurements of ROC are taken into account during the course of this research project for the objective of identifying patterns in the dataset that are associated with heart disease.

The pre-processed dataset is taken into account while doing the attributes reduction. Classification is performed by iteratively working through the many different kinds of classifiers, such as tree classifiers, Rules 8 classifiers, Meta classifiers, and Bayes classifiers, before ultimately plotting the ROC to identify which classifiers give the maximum degree of accuracy [13].

The recommended text for the document is broken up into the following sections: The first part of this paper is an introduction, which includes a discussion of the cause as well as a definition of the problem. In Sect. 2, a comparison of the related works and the current works is presented; in Sect. 3, the methodology of the proposed study is outlined; in Sect. 4, the results of the experiments and an analysis of the results are presented; and in Sect. 5, a conclusion is offered along with recommendations for additional study.

2 Related Work

The earlier work, which was carried out based on similar methodologies and may be found in [14], Even if the learning models that are generated are the same, the execution time in successive iterations displays varied degrees of accuracy. This is because the natural development of changes that occur with regard to time in the data patterns cannot be readily recovered. [15] Contains the study activity that was conducted with the same viewpoint. As can be seen in [16], the selection of qualities that we use in this study work is carried out using an iterative process that is based on information gain ratio. A process that chooses a minimum subset of attributes from the original set of attributes in order to optimally reduce the search space in accordance with a certain evaluation criterion is referred to as attribute selection. Attribute selection can be defined as a process that chooses a minimum subset of attributes from the original set of attributes. The task of locating such a subset is said to be NP-hard since it requires an exponentially increasing amount of time complexity in relation to the number of characteristics. Absolute mean error, root mean square, and other similar measures are examples of summative metrics that display performance variation. Measures such as accuracy and area under cure provide credibility to the suggestions that are made based on the models that are developed. In this body of study, ROC measurements are taken into consideration for the purpose of detecting trends in the dataset pertaining to heart disease. When doing the attributes reduction, the preprocessed dataset is taken into consideration. By using the straightforward "Subset assessment" approach in conjunction with the "Best first" search methodology, we may determine that about one tenth of the qualities are relevant. Classification is carried out in an iterative manner across the many types of classifiers, including tree classifiers, Rules 11 classifiers, Meta classifiers, and Bayes classifiers; lastly, the ROC is presented for classifiers that provide the highest level of accuracy (Table 1).

[24] developed a model for predicting heart disease that was based on a weighted fuzzy set of rules. The model was developed in such a manner that it is capable of carrying out on its own the process of discovering intelligence from clinical data. In the CDSS procedures that have been developed, the risk of cardiac patients is analysed in two different periods. The first step is to build the rule by integrating fuzzy rules and a decision tree, and the second step is to generate a fuzzy rule based on a decision support system.

[25] used the association rule in order to create the link that exists between the qualities. In order to achieve more accuracy, a data categorization method was developed using MAFIA techniques as a foundation. In order to assess the data, entropy-based cross validation and partition approaches are used. The results obtained from the different strategies were analysed and compared. The MAFIA (Maximal Frequent Item set

Table 1. Comparison on Existing techniques

Disease	Techniques used
[17]	Utilizing Neural Networks for the Purpose of Extracting Patterns and Detecting Trends
[18]	Models of prediction based on Decision Tree Algorithms like ID3, C4.5, C5, and CART
[19]	Using the Ensemble technique, differentiate between the various illness subtypes
[20]	Utilizing BBN Bayesian networks, one may determine the likelihood of there being a patient with a mental illness based on the symptoms that have been seen
[21]	Utilizing the Apriori method, one may determine the prevalence of illnesses in a certain geographical region
[22]	Classification using the use of Bayesian Ying Yang
[23]	Utilizing a Genetic Algorithm for the Categorization of Medical Data

Algorithm) research project has as its primary objective the generation of results that have high levels of accuracy and recall metrics that have greater levels of precision. The dataset used in the study has 19 features.

[26] Hybrid The diagnosis of cardiac illness is performed with the use of clever methodologies. Author has established a unique and innovative strategy employing two distinct layers for the purpose of identifying probable disorders. The most important risk factors for heart attacks are considered on the first level, while the rest of the variables are considered on the second level. The author claims that the suggested two-level conceptualization would boost the efficiency of the operation of our application. The authors of this study make use of both neural networks and fuzzy networks. The data set from the neural network is used to train the neural network, and when the neural network has been trained, fuzzy rules 53 are implemented for predicting heart disorders. Following the use of the methodology, it is possible to categorise the severity of the cardiac illnesses as low, medium, or critical.

According to [27], data mining is the technique that is utilised the most in clinical research. In this study, both neural networks and generic algorithms are employed together. The hybrid system will have the benefits of both the genetic and neural network approaches, and it will provide findings that are more accurate. The production of initial weights in neural networks often makes use of genetic algorithms. When compared to the back propagation approach, the neural network learning algorithm is much faster. Using the presented model has led to an accuracy of 89 percent being achieved by the proposed model.

In the [28] study, the author has provided a summary of the major limitations of the various approaches. When it comes to predicting cardiac illnesses, data mining methods employ a completely different approach.

[29] developed a structure for the purpose of researching the subject of data mining algorithm restricting and summarising its many different implementations. The author has concentrated on a number of distinct qualities in order to make predictions based on

combinations of those features. At long last, the author has established a method to do an examination of the heart attack. There are several steps involved in the procedure. The first step is to identify the relevant patterns using the most appropriate methodology. This assists in more accurately predicting when someone will have a heart attack. For this reason, a substantial amount of weight has been allotted, and it varies according to the severity of the heart attack. The frequent patterns and predetermined value of anticipated heart attack were used to determine the substantial weightage that was assigned to each factor. The assessment of the trained models is the only factor taken into consideration while extracting the goals. The trained model that was created in this research is able to quickly answer complicated questions such as heart attack prediction.

3 Methodology of Proposed Work

The majority of researchers in this field make use of the KDD competition exercises that can be found on the UCI site. Although there are 76 characteristics in this database, all of the published studies only relate to employing a subset of 14 of them. In specifically, the Cleveland database is the one and only one that has been used by researchers working in ML as of this point in time. Table 2 outlines the characteristics of the machine learning repository that may be found at UCI. Nevertheless, a significant amount of noise is often connected with these gathered pieces of information. There are a number of factors that might contribute to the presence of noise in this data. Two of the most important factors are the imprecise nature of the technology used to gather the data and the origin of the data itself. For instance, in the realm of medical imaging, any efficiency in the imaging equipment would be reflected as noise in the subsequent process.

Table 2. List of Attributes

#3	Age
#4	Sex
#9	Cp
#10	Trestbps
#12	Chol
#16	Fbs
#19	Restecg
#32	Thalach

The quality of the data obtained through social media platforms may range from good data to stuff that constitutes abuse or spam. In the meanwhile, the material that is shared on social media is often written in a casual manner and contains grammatical errors, misspellings, and poor punctuation. One of the most common strategies for getting rid of noisy (that is, unimportant) and redundant characteristics is known as dimensionality reduction. The strategies for dimension reduction may primarily be broken down

into two categories: feature extraction and feature selection. The features that are being extracted are projected onto a new space that has a reduced dimensionality, and the newly generated features are often combinations of the original features. Techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Canonical Correlation Analysis (CCA) are all examples of feature extraction methods. On the other hand, the goal of the feature selection techniques is to pick a limited collection of characteristics that increase relevance to the target while minimising repetition. For example, in classification, the class labels are an example of the target. Techniques such as optimization employing hybrid grey wolf combined with particle swarm optimization are examples of representative feature selection methods (HGWPSO).

Therefore, further analysis of new features is troublesome since there is no physical meaning for the converted features produced via feature extraction methods. This is due to the fact that there is no physical meaning for the transformed features (Fig. 2).

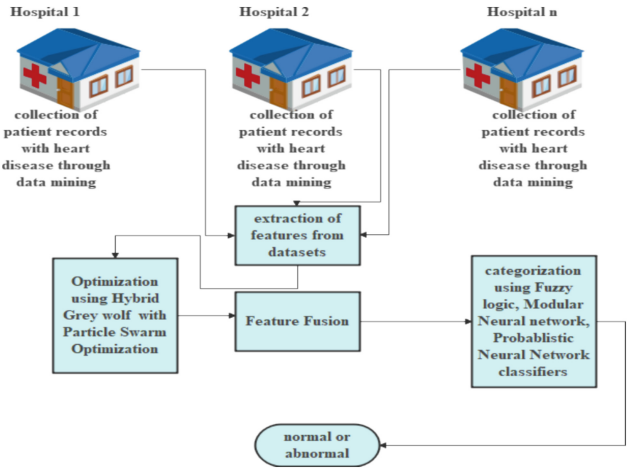


Fig. 2. Proposed Methodology basic Structure

3.1 Extraction of Features from Datasets

Relationships between Hilbert transforms offer the foundation for analytical signal, bandpass sampling, minimum phase networks, and spectral analysis.

The equation that represents the Hilbert transform of a function x(t) may be found here (1):

$$h(t) = H\{x(t)\} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d(\tau) \tag{1}$$

Using the Fourier identities, the Fourier transform of the Hilbert transform of $x(t)$ is in Eq. (2):

$$h(t) \Leftrightarrow H(f) = -j\text{sgn}(f)X(f) \tag{2}$$

where $x(t) \Leftrightarrow X(f)$ is a Fourier transform pair and

$$\operatorname{sgn}(f) = \begin{cases} 1 & f > 0 \\ 0 & f = 0 \\ -1 & f < 0 \end{cases} \tag{3}$$

The frequencies that are acquired by applying the Hilbert transform include artefacts, which are then removed by employing a band pass Chebyshev filter in such a way that eliminates any frequencies with a frequency that is below 5.00 Hz and above 20.00 Hz. By permitting the ripple in the frequency response, the Chebyshev response is able to produce a quicker roll-off of the frequency response. In most cases, a ripple depth that falls anywhere between 0.1 dB and 3 dB is used. The term "maximally flat" or "Butterworth filter" refers to the situation in which the ripple is reduced to 0%.

The peak-to-peak pass band ripple is what is meant by the term "dB ripple" when referring to a Chebyshev filter. The value of the parameter may be found by using the Eq. (4):

$$\text{dBripple} = 10 \log(1 + \varepsilon^2) \tag{4}$$

This can be solved for ε and it can be obtained as (5)

$$\varepsilon = \sqrt{10^{dB/10} - 1} \tag{5}$$

Parameter h is required for obtaining Chebyshev transfer functions, and it is determined as (6):

$$h = \tanh\left(\frac{1}{n} \sinh^{-1} \frac{1}{\varepsilon}\right) \tag{6}$$

where n is the order of low-pass filter.

3.2 Optimization Using Grey Wolf with Particle Swarm Optimization

In this part of the article, in order to increase the effectiveness of the search agents, we have decided to use the HGWPSO approach for the guided search process. To talk in more general terms, the HGWPSO mechanism with constrained memory is an extension of the BFGS approach. Let us assume that x 0 is the starting position, and that we are going to be determining the next location by using an iterative process based on the line search method and the Eq. 7.

$$x_{k+1} = x_k + \lambda_k d_k \tag{7}$$

where _k stands for the step length and d k indicates the direction in which the step length decreases. The vast majority of the scholars looked at a variety of approaches to determine the step length k. In the beginning, presented a method for obtaining _k, which was represented by the Eq. 8.

$$\phi(x_k + \lambda_k d_k) - \phi(x_k) \leq \sigma \lambda_k \nabla g(x_k)^T d_k \tag{8}$$

where is a number chosen at random (i.e. (0,1)), and g(x k) stands for the Jacobian matrix. With the assistance of the aforementioned equation, the monotone line search process, which is shown in 9, can be understood.

$$\begin{aligned} &\phi(x_k + \lambda_k d_k) - \phi(x_{lk}) \leq \sigma \lambda_k \nabla g(x_k)^T d_k \\ &\phi(x_{lk}) = \max(0 \leq j \leq m(k))\phi(x_{k-j}), m(0) = 0. \\ &m(k) = \min m(k+1) + 1, M \,(\text{for } k \geq 1). \end{aligned} \tag{9}$$

where M is a number that can only be positive. It is fairly straightforward to calculate g(x k) based on the formulae shown above; nevertheless, it may increase the amount of work that needs to be done for issues with a big scale. In order to find a solution to the problem, we devised the contemporary backtracking inaccurate mechanism, which may be represented by the Eq. 10.

$$\|g(x_k + \lambda_k d_k)\|^2 \leq \|g(x_k)\|^2 + \psi \lambda_k^2 g_k^T d_k \tag{10}$$

When ψ ∈ (0,1), g is symmetric for every x k (i.e. g k = g(x k)). When it comes to numerical performance, this modified line search mechanism produces results that are more accurate than those produced by the normal line search mechanism. The optimal characteristics are chosen with the use of an optimization function, and then they are adjusted such that they can forecast risk factors.

3.3 Improved Probabilistic Neural Network

It is possible to have numerous split positions for an attribute that contains numbers. Let's assume that at node m, N mj of N chooses to go in the direction of branch j. In the case of a discrete attribute with n possible values, there are n outcomes, and in the case of a numeric attribute, there are n outcomes equal to two (n = 2), both of which are fulfilling.

$$\sum_{j=1}^{n} N_{mj} = N_m \cdot N_{mj}^i \; of \; N_{mj} \text{ belongs to class } \sum_{j=1}^{n} N_{mj} = N_m \cdot N_{mj}^i \tag{11}$$

Similarly $\sum_{j=1}^{n} N_{mj}^i = N_{mj}^i$. Then given that a node m, the test returns outcome j, the estimate for the probability of class C_i i

$$\hat{P}(C_i|x, m, j) \equiv p_{mj}^i = \frac{N_{mj}^i}{N_{mj}} \tag{12}$$

And the total impurity after the split is given as

$$\Im'_m = -\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} p_{mj}^i \log_2 p_{mj}^i \tag{13}$$

In the case that a numeric attribute is present, it is necessary for us to be familiar with the value of w m0 for that node in order to be able to calculate p mji employing as Eq. (13).

Since this is the case, we compute the impurity for each and every discrete and numeric characteristic, as well as each and every numeric attribute for each and every split location, and then select the value that satisfies the criteria of having the lowest entropy, for example, as determined by the (12). After then, the construction of the tree continues in a recursive and parallel way for each of the branches that are not pure, up to the point when all of the branches are pure.

4 Experimental Results

The gain ratio criteria is used in the decision tree classification method to identify which feature is the most appropriate for dividing the decision tree. The variable confidence level may vary anywhere from 0.1 to 0.5, depending on the situation. In order to determine how reliable the decision tree was, we put it through 10 distinct rounds of cross-validation. The greatest possible outcome was accomplished with a confidence value of 0.4 and an accuracy of 95.61%; the area under the curve (AUC) value is shown in Table 3.

Using a back propagation method, which is utilised in combination with a neural network, the neural network is trained for different values of learning rate and momentum ranging from 0.1 to 0.9. This is accomplished by using a neural network. The training of the network takes place over the course of 500 epochs, and it consistently consists of only one hidden layer. The learning rate parameter was set to 0.6, the momentum parameter was set to 0.8, and the number of neurons in the hidden layer was set to 7. This combination of settings produced the best result that could possibly be achieved. A kernel function that goes by the name of the Gaussian radial basis (rbf) is used in order to divide nonlinear data. When rbf sigma is equal to 1, the resulting classifier is considered to be the best possible one. The accuracy of 95.02 was able to be achieved, and the area under the curve (also known as AUC) was determined to be 0.944 in Table 3.

With the use of decision tree classification, the gain ratio criterion is used to determine which characteristic is the most suitable for splitting the decision tree, and the variable confidence level may range anywhere from 0.1 to 0.5. In order to evaluate the effectiveness of the decision tree, we used ten different cross-validation strategies. The best result possible was produced when missing data were replaced by their mean, and it had an accuracy of 81.09% as indicated in Table 3, along with an AUC value of 0.867. This was the best result that could be obtained. The best result was achieved for non-missing data (rows with missing values were removed), and it was characterised in Table 3 as having an accuracy of 82.60% and an AUC value of 0.863. This result was reached with a confidence value of 0.4 (Fig. 3).

When it comes to the prediction of heart disease, a high level of sensitivity is sometimes believed to be more relevant than a high level of specificity. This is the conclusion that we have arrived at after doing an analysis of the data that was obtained. It is better to incorrectly categorise a Class 1 region as a Class 2 one rather than missing a cardiac sickness owing to the improper classification of a Class 2 region as a Class 1 one. This is because it is easier to misclassify a Class 1 region as a Class 2 one.

Because all of the regression coefficients (parameter estimates) in Table 4 are positive and the significant levels (P values) are lower than 0.05, it is evident that the results show

Table 3. Confusion matrices of classifiers used for classification of heart disease data set 1

Type classifier	Desired result	Output results data set I	
		Positive	Negative
Proposed IPNN	Class 1 record	429	15
	Class 2 record	15	224
Naïve Bayes	Class 1 record	424	6
Naïve Bayes Kernel	Class 2 record	20	233
	Class 1 record	424	2
	Class 2 record	20	237
	Class 1 record	430	9
SVM	Class 2 record	14	230
	Class 1 record	428	6
	Class 2 record	16	233
	Class 1 record	237	32
	Class 2 record	2	412

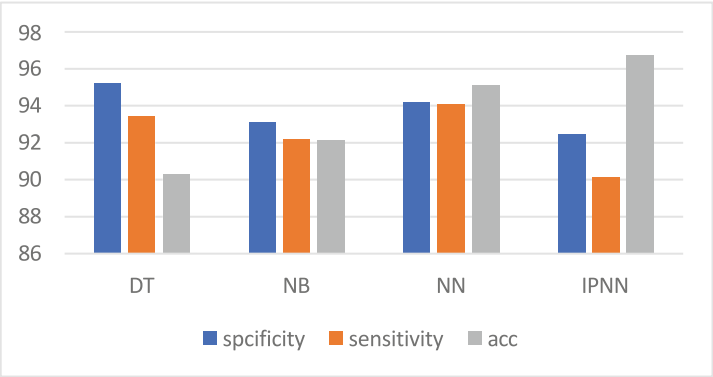


Fig. 3. Performance metrics of proposed work

Table 4. Risk factor analysis using selected Parameter

Risk Factor	Parameter Estimate	Standard error	
Heart disease	0.992	0.402	0.014
Diabetes	0.996	0.450	0.027
Stroke	0.827	0.394	0.036

a positive relationship between three risk factors and the cardiovascular outcomes. This is the case because all of the significant levels are lower than 0.05. (95 percent of confidence interval).

5 Conclusion

The study is primarily divided into two parts: first, an investigation into the theory of machine learning, and second, the development of methods for using six distinct machine learning classification algorithms for the purpose of diagnosing and predicting the course of heart disease. This method was developed in order to ease the process. The performance of the machine learning model may be quantified and best represented in terms of their sensitivity and specificity, which measures how well they perform in relation to the rate of false positive instances and the number of false negative examples. When applied to the data set pertaining to heart disease, the support vector machine has the greatest level of sensitivity. In the context of predicting heart disease, a high degree of sensitivity is typically regarded as being more important than a high degree of specificity. To put it another way, it is preferable to incorrectly classify a Class 1 region as belonging to Class 2 than it is to fail to predict a heart disease by incorrectly classifying a Class 2 region as belonging to Class 1. When there are no gaps in the data, the support vector machine is able to support the combination of best accuracy and sensitivity. Because some of the variables were missing, the Auto multi-layer perceptron and neural networks were able to categorise the heart lump as belonging to either Class 1 or Class 2.

References

1. Kannel, W.B.: Coronary heart disease risk factors in the elderly. Am. J. Geriatr. Cardiol. **11**(2), 101–107 (2002)
2. Williams, P.T.: Physical fitness and activity as separate heart disease risk factors a meta-analysis. Med. Sci. Sports Exerc. **33**(5), 754 (2001)
3. Lauer, R. M., Connor, W. E., Leaverton, P. E., Reiter, M. A., Clarke, W. R.: Coronary heart disease risk factors in school children. The Muscatine study. J. Pediatr. **86**(5), 697–706 (1975)
4. Freedman, D.S., Khan, L.K., Dietz, W.H., Srinivasan, S.R., Berenson, G.S.: Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study. Pediatrics **108**(3), 712–718 (2001)
5. Greenland, P., Smith, S.C., Jr., Grundy, S.M.: Improving coronary heart disease risk assessment in asymptomatic people: role of traditional risk factors and noninvasive cardiovascular tests. Circulation **104**(15), 1863–1867 (2001)
6. Smith, G.D., Shipley, M.J., Rose, G.: Intermittent claudication, heart disease risk factors, and mortality. The Whitehall Study. Circul. **82**(6), 1925–1931 (1990)
7. Barrett-Connor, E., Suarez, L., Khaw, K.T., Criqui, M.H., Wingard, D.L.: Ischemic heart disease risk factors after age 50. J. Chronic Dis. **37**(12), 903–908 (1984)
8. Howard, B.V., et al.: Coronary heart disease risk equivalence in diabetes depends on concomitant risk factors. Diabetes Care **29**(2), 391–397 (2006)
9. Bhopal, R., et al.: Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations. Cross sectional study. Bmj, **319**(7204), 215–220 (1999)

10. Talbott, E., et al.: Coronary heart disease risk factors in women with polycystic ovary syndrome. Arterioscler. Thromb. Vascular Biol. **15**(7), 821–826 (1995)
11. Caspersen, C.J., Bloemberg, B.P., Saris, W.H., Merritt, R.K., Kromhout, D.: The prevalence of selected physical activities and their relation with coronary heart disease risk factors in elderly men the Zutphen Study, 1985. Am. J. Epidemiol. **133**(11), 1078–1092 (1991)
12. Nowbar, A.N., Gitto, M., Howard, J.P., Francis, D.P., Al-Lamee, R.: Mortality from ischemic heart disease: Analysis of data from the World Health Organization and coronary artery disease risk factors From NCD Risk Factor Collaboration. Circul.: Cardiovasc. Qual. Outcomes **12**(6), e005375 (2019)
13. Li, G., Chen, X., Jang, Y., Wang, J., Xing, X., Yang, W., Hu, Y.: Obesity, coronary heart disease risk factors and diabetes in Chinese: an approach to the criteria of obesity in the Chinese population. Obesity Rev. **3**(3), 167–172 (2002)
14. McGill, H.C., Jr., McMahan, C.A., Zieske, A.W., Tracy, R.E., Malcom, G.T., Herderick, E.E., Strong, J.P.: Association of coronary heart disease risk factors with microscopic qualities of coronary atherosclerosis in youth. Circulation **102**(4), 374–379 (2000)
15. Canto, J.G., Kiefe, C.I., Rogers, W.J., Peterson, E.D., Frederick, P.D., French, W.J., NRMI Investigators. Number of coronary heart disease risk factors and mortality in patients with first myocardial infarction. Jama **306**(19), 2120–2127 (2011)
16. Covas, M. I., Nyyssönen, K., Poulsen, H. E., Kaikkonen, J., Zunft, H. J. F., Kiesewetter, H., . EUROLIVE Study Group: The effect of polyphenols in olive oil on heart disease risk factors: a randomized trial. An. Internal Med. **145**(5), 333–341 (2006)
17. Trémollières, F.A., Pouilles, J.M., Cauneille, C., Ribot, C.: Coronary heart disease risk factors and menopause: a study in 1684 French women. Atherosclerosis **142**(2), 415–423 (1999)
18. McGill Jr, H.C., McMahan, C.A., Zieske, A.W., Sloop, G.D., Walcott, J.V., Troxclair, D.A., Youth Research Group, F.T.P.D.O.A.I.: Associations of coronary heart disease risk factors with the intermediate lesion of atherosclerosis in youth. Arterioscler. Thrombo. Vascular Biol. **20**(8), 1998–2004 (2000)
19. Robertson, T.L., Kato, H., Gordon, T., Kagan, A., Rhoads, G.G., Land, C.E., Kawamoto, S.: Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: coronary heart disease risk factors in Japan and Hawaii. Am. J. Cardiol. **39**(2), 244–249 (1977)
20. Roeters van Lennep, J.E., Westerveld, H.T., Erkelens, D.W., van der Wall, E.E.: Risk factors for coronary heart disease: implications of gender. Cardiovascular Res. **53**(3), 538–549 (2002)
21. Patel, J.V., Vyas, A., Cruickshank, J.K., Prabhakaran, D., Hughes, E., Reddy, K.S., Durrington, P.N.: Impact of migration on coronary heart disease risk factors: comparison of Gujaratis in Britain and their contemporaries in villages of origin in India. Atherosclerosis **185**(2), 297–306 (2006)
22. Klein, R., Klein, B.E., Cornoni, J.C., Maready, J., Cassel, J.C., Tyroler, H.A.: Serum uric acid: its relationship to coronary heart disease risk factors and cardiovascular disease, Evans County, Gerogia. Archiv. Intern. Med. **132**(3), 401–410 (1973)
23. Gibbons, L.W., Blair, S.N., Cooper, K.H., Smith, M.: Association between coronary heart disease risk factors and physical fitness in healthy adult women. Circulation **67**(5), 977–983 (1983)
24. Vasan, R.S., Sullivan, L.M., Wilson, P.W., Sempos, C.T., Sundström, J., Kannel, W.B., D'agostino, R.B.: Relative importance of borderline and elevated levels of coronary heart disease risk factors. Ann. Internal Med. **142**(6), 393–402 (2005)
25. Sääkslahti, A., Numminen, P., Varstala, V., Helenius, H., Tammi, A., Viikari, J., Välimäki, I.: Physical activity as a preventive measure for coronary heart disease risk factors in early childhood. Scandinavian J. Med. Sci. Sports **14**(3), 143–149 (2004)

26. Pieper, C., LaCroix, A.Z., Karasek, R.A.: The relation of psychosocial dimensions of work with coronary heart disease risk factors: a meta-analysis of five United States data bases. Am. J. Epidemiol. **129**(3), 483–494 (1989)
27. McElhinney, D.B., Hedrick, H.L., Bush, D.M., Pereira, G.R., Stafford, P.W., Gaynor, J.W., Wernovsky, G.: Necrotizing enterocolitis in neonates with congenital heart disease: risk factors and outcomes. Pediatrics **106**(5), 1080–1087 (2000)
28. Pyörälä, K., Savolainen, E., Kaukola, S., Haapakoski, J.: Plasma insulin as coronary heart disease risk factor: Relationship to other risk factors and predictive value during 9½-year follow-up of the Helsinki Policemen Study Population. Acta Medica Scandinavica **218**(S701), 38–52 (1985)
29. McGill, H.C., Jr., et al.: Effects of coronary heart disease risk factors on atherosclerosis of selected regions of the aorta and right coronary artery. Arterioscler. Thrombo. Vasc. Biol. **20**(3), 836–845 (2000)



Efficient Detection and Multi-level Classification of Tomato Plant Leaves Using Fused Deep and Hand-Crafted Features and LSTM-ANFIS

Astha Sharma(✉) and Ashwni Kumar

Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technological University for Women, New Delhi, India
asthasharma092@gmail.com

Abstract. Tomato leaf diseases affects the superiority and amount of the crops to a large extent. Consequently, it is vital to notice the crop diseases at an initial phase to maintain a social, economic and ecological balance. In this direction this paper presents a framework which integrates both ML and DL-based technology for recognition and sorting of tomato leaf bugs. The proposed framework leverages fusion of deep-features and hand-crafted features for efficient feature representation. Further, the fused features obtained are utilized by LSTM-ANFIS based classification network. The proposed framework efficiently classifies the tomato plant leaf diseases by utilizing ANFIS-LSTM based classification module. To validate the effectiveness of proposed model extensive experiments are conducted on PlantVillage dataset. Furthermore, the proposed framework provided an accuracy of 98.83%.

Keywords: Machine Learning (ML) · Deep-Learning (DL) · Tomato disease detection · Vegetal bug detection · Modified LBP · Inception-V3 · LSTM · ANFIS

1 Introduction

Recently, there has been an increased focus on food security, and countless countries and administrations stand working to increase nutrition manufacture. Disease caused in plants can affect the agriculture by altering social, economic and ecological balance. Due to their high yield and ease of cultivation, tomatoes—one of the world's most nutrient-dense crops—take a main impression on the enlargement of the agronomic industry. Tomato possesses abundant nourishment as well as pharmacological benefits that protect humans from illnesses such as hypertension, hepatitis, bleeding during the gums, etc. [1–6]. Therefore, it is the most widely utilized crop in India. Tomato crops mainly suffers from bacterial or fungal diseases which may affect the growth of tomato crops therefore it necessary to detect and identify diseased plant at an early stage [7–9].

ML-based methods are employed earlier for the identification of tomato plant leaf. With the advancements made in technology deep-learning based methods are utilized

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 157–166, 2025.
https://doi.org/10.1007/978-3-031-75167-7_12

intended for the progress of robust representations for efficient exposure and classification of diseased plant images. This paper presents a novel framework which classifies tomato leaf plants into 10 classes using deep-learning and machine learning algorithm. Figure 1 presents a block diagram representation of the projected classification archetypal. The planned model is divided into dual steps: (i) Feature abstraction segment and (ii) sorting segment.

The key contributions are:

1. This paper presents a fusion of hand-crafted features and deep-features by utilizing modified LBP features and Inception-V3. The fused features are further utilized by the classification module for efficient classification of tomato leaf images
2. The proposed model incorporates LSTM-ANFIS classification network for efficient classification by learning fused features obtained from feature extraction module through two-layer LSTM network
3. Extensive experiments are conducted on a subset of Plant Village dataset to authenticate the efficiency and effectiveness of the planned model.
4. An ablation study is also conducted to support robustness of LSTM-ANFIS classification network in compliance with other classifiers.

The remnants of the tabloid is designed as trails: Sector 2 delivers a appraisal of the fiction. A thorough discussion of the suggested methodology is given in Sect. 3. Ad ditionally, the dataset utilized, implementation specifics, and results are accessible in Unit 4. In Segment 5, we finally bring the tabloid to a close.

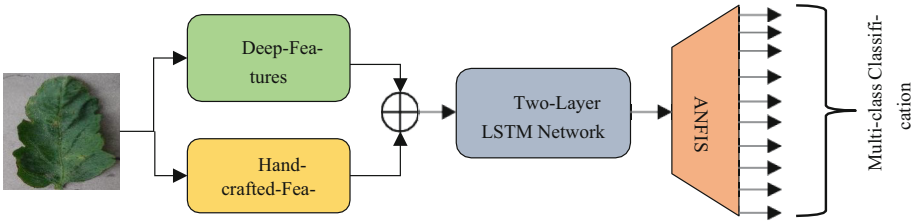


Fig. 1. Block Diagram Representation of the Proposed Multi-class Classification Model

2 Related Works

Leaf disease detection and identification is important so as to protect the crops from different types of adverse situations which may lead to loss in food crops. The severity of crop diseases has been properly identified after a great deal of research. ML and DL approaches can be used to detect leaf diseases, respectively. Bhimte et al. [10] and Padol et al. [11] utilized SVM classifier for the finding and cataloguing of cotton-leaf ailments and grape-leaf disease detection and classification. [11] also extracted colour and texture features and achieved an precision of 89% on assessment set. Anand et al. [12] presented K-means cluster designed for brinjal leaf disease detection. Kumari et al. [13] presented leaf spot identification technique using image segmentation and feature

extraction means. Rani et al. [14] proposed multi-class classification algorithm using color and texture features and provided an accuracy of about 95%. Zhang et al. [15] presented fusion of super-pixel clustering and PHOG algorithm for plant leaf disease detection.

Feature engineering continues to be the largest barrier, despite the wide range of applications for ML algorithms. Potential solutions for bush pathology are now conceivable appreciations to deep neuronal grids, which abolish the prerequisite for laborious feature trade. Deep neural networks significantly increase picture classification accuracy. Deep-Learning based methods [16–19], utilized VGGNet, Inception-V4, AlexNet, and DenseNet architectures for identification of different plant leaf diseases. [20] utilized depth-wise separable convolutional architectures to classify plant leaf diseases. DenseNet with 122-layer architecture is also exploited by some investigators for efficient arrangement of herbal leaf ailments. Babu et al. [21] presented a comparative study plant leaf disease classification. Different types of plant leaf (tomato, potato, apple, etc.) detection techniques are presented in [22] which utilized different CNN and ANN architectures for detection and classification. Sun et al. [23] presented maize leaf blight detection using deep-learning. This method utilized different image segmentation and soft computing systems for better exposure and sorting. Real-time detection of bush leaf ailments a challenging task in this direction Jiang et al. [24] presented apple leaf disease detection. Sahasra et al. [25] offered deep-learning grounded manner that utilized surveillance and monitoring experts for detection.

The works discussed above utilized either the deep-learning based architectures or traditional deep-learning architectures to classify plant leaf diseases. This paper presents a method which utilized both DL-based feature extraction method and ML-built sorting for efficient identification of tomato plant leaf disease.

3 Proposed Framework

The projected outline is a twostage framework: (*i*) Feature extraction module and (*ii*) sorting part using LSTM-ANFIS classifier. Further, Fig. 2 depicts the detailed architecture of the proposed methodology.

3.1 Fused Feature Extraction Module

This module utilized fusion of deep-features and hand-crafted image features. The deep-features are obtained using Inception-V3 [26] module whereas Modified LBP features [27] are incorporated to extract the hand-crafted features. The proposed feature extraction module leverages learning rate scheduler along with the Adam optimizer to improve the detection performance. For the extraction of deep-features, Inception-V3 is an inception family CNN which leverages label smoothing, factorized 7×7 convolutions, and an assisting classifier with batch normalization layers. This enables the training of high-quality networks using training sets of relatively little size. InceptionV3 also has a comparatively low processing cost compared to other CNN designs when analyzing a relatively tiny portion of the image that contains a single object with some context. The features obtained are further fed to the dense layer.

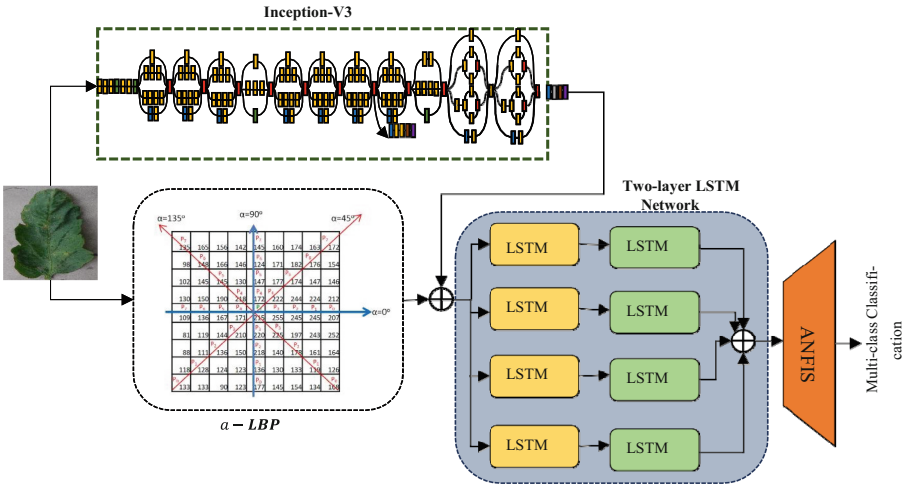


Fig. 2. Proposed Tomato Leaf Disease Detection and Classification Methodology

LBP features are utilized for encoding of local texture features by discriminating the structural patterns which sometimes results in same LBP codes. Therefore, to overcome this this paper proposes modified LBP or $\alpha - LBP$ which is based on local statistical measures. The angle value being $\alpha = 45°, 90°, and\ 135°$. The value of $\alpha - LBP$ is calculated as per the following equation:

$$\alpha - LBP = \sum_{j=0}^{\wp} u(\mathcal{P}_j - \mathcal{P}_c)2^j \tag{1}$$

With the use of $\alpha - LBP$ the neighbours are calculated on the plane passing over the centre pixel Pc in accordance with the value of α. The modified LBP features obtained from are made to pass through the dense layer. Further, the features obtained from the deep-feature $\mathbb{f}_{incep}$based extraction module and hand-crafted features $\mathbb{f}_{\alpha lbp}$are fused or concatenated based on the entropy for feature fusion. The fused features obtained are made to pass through the dense layer and are represented by:

$$f_{fusion} = \{\mathbb{f}_{incep}, \mathbb{f}_{\alpha lbp}\} \tag{2}$$

3.2 LSTM-ANFIS Classification Module

The fused features obtained from the feature extraction module are fed to the twolayer LSTM module [28]. The two-layer LSTM helps in extraction of high-level features by learning long-term dependencies in sequential data. The extracted features obtained are further fed to the assortment of neuronic and fuzzy corollary system network i.e., ANFIS classifier [29]. The input features in the first layer will have own group of membership function. The subsequent sheet generates the firing strength the productivity from the

layer is represented by the merchandise of incoming inputs. Further, the output from the second layer is normalize which is represented by:

$$w_{11} = \frac{w_1}{w_1 + w_2} \tag{3}$$

$$w_{22} = \frac{w_2}{w_1 + w_2} \tag{4}$$

The fourth part is the defuzzification coat whose yield is represented by:

$$\sigma_i = w_i \mathcal{F}_i = w_i(a_i x + b_i y + c_i) \tag{5}$$

Finally, the productivity layer figures the usual of each rule output and is given by:

$$\sum_{i=1}^{n} w_i \mathcal{F}_i = \frac{\sum_{i=1}^{n} w_i \mathcal{F}_i}{\sum_{i=1}^{n} w_i} \tag{6}$$

The hybrid learning method is the standard optimization system for ANFIS constraints throughout the training phase of this cataloging procedure. It combines the forward and backward permits of back propagation with the least squares approach. During the training mode, the ANFIS classifier's first layer receives the chosen features as inputs.

4 Results and Discussions

This section discusses about the dataset used, evaluation parameters, implementation details and results obtained for the proposed tomato plant disease classification model.

4.1 Dataset Used and Evaluation Parameters

Experiments are conducted on a subset of Plant Village Dataset [30]. The images for tomato leaf plant for the classification is used for this experiment. The dataset used contained around 11,000 images split into 10 different classes namely target spot, bacterial spot, leaf mold, etc. Figure 3 presents the class-wise sample image of these datasets. To validate the effectiveness of the proposed classification model accuracy, precision, F-1 score and recall is calculated.

4.2 Implementation Details

To detect the features and classify the tomato plant diseases into different classes the proposed model utilizes stochastic gradient descent optimizer. The learning percentage is usual as 0.00001 with batch size of 32 and validation frequency of 30. The training cycles contains 320 total iterations with 4 iteration per epoch. Further, the training for the model is set to 50 epochs and the whole architecture is optimized by minimizing cross entropy loss.



Fig. 3. Class wise sample of the dataset split into 10 different classes

4.3 Results Obtained for the Proposed Method

Table 1 reports the comparison results. The proposed model with the use of fused deep-features and hand-crafted features provided an accuracy of 98.83%. There are significant improvements observed in the value of accuracy. The proposed model classified tomato leaf diseases in a better way with the use of LSTM-ANFIS when compared with SVM, KNN, AlexNet, and CNN based classification techniques. Further a comparison graph for the different methods is provided in Fig. 4. Further, confusion matrix is offered in Fig. 5.

Table 1. Comparison Results of the proposed model

Paper	Classifier	Acc	P	R	F-1
Raza et al. [31]	SVM	89.93%	-	-	-
Prasad et al. [33]	KNN	93.0%	-	-	-
Wang et al. [34]	AlexNet	95.62%	-	-	-
Guo et al. [32]	AlexNet	92.7%	-	-	-
Agarwal et al. [35]	CNN	91.2%	-	-	-
Proposed	LSTMANFIS	98.83%	0.967	0.959	0.962

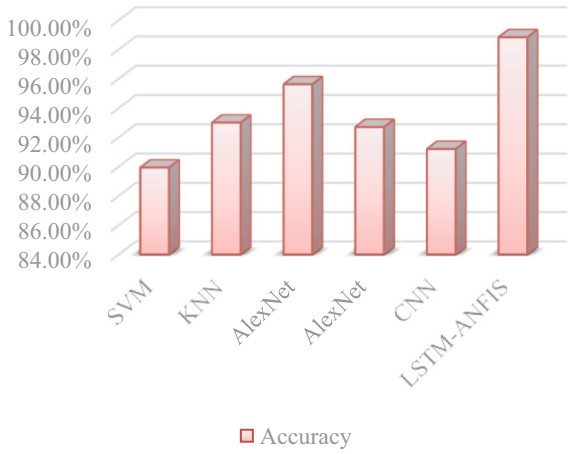


Fig. 4. Graphical representation of the comparison results obtained

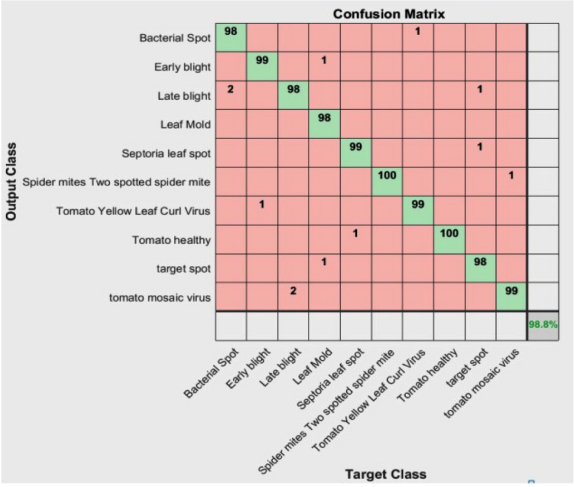


Fig. 5. Confusion Matrix

Table 2 reports the comparison of the proposed model (LSTM-ANFIS) with other classifiers like SVM MLP and KNN. The proposed model outperforms SVM and KNN. LSTM-ANFIS seems more successful in comparison to KNN, SVM, and MLP as LSTM-ANFIS is used to adjust the weights and approximates more and more to produce the desired output. Also, LSTM-ANFIS is preferred due to its fuzzy logic capability and LSTM long term dependencies learning that manages the uncertainty of fuzzy data, ambiguous or incomplete. Further, Fig. 6 (a) and Fig. 6 (b) presents the accuracy and loss curve for the proposed LSTM-ANFIS classification module and other classifiers module. These curves make it evident that LSTM-ANFIS based classification model performs better when compared with SVM, KNN, and MLP.

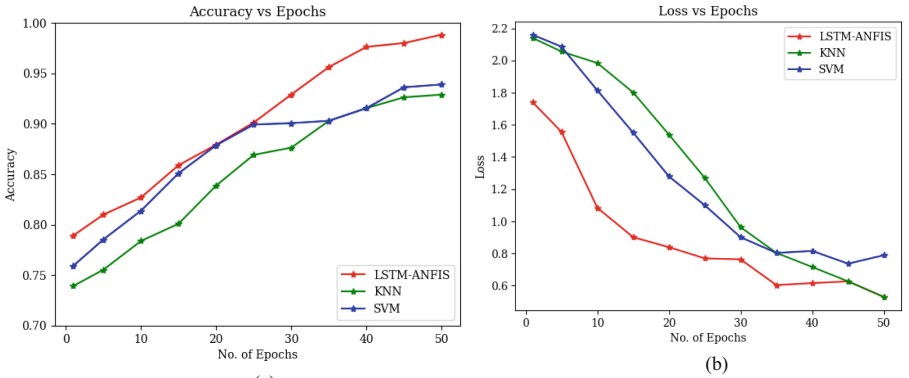


Fig. 6. Accuracy and Loss curves for the proposed model for ANFIA classifier and other classifiers

Table 2. Comparison of the proposed method with other Classification Module

Classification Network	Acc	P	R	F-1
SVM	92.89%	0.891	0.885	-
KNN	92.41%	0.902	0.892	-
LSTM-ANFIS	98.83%	0.967	0.959	0.962

5 Conclusion

This paper presents a robust model to distinguish and categorize tomato plant leaves using the integration of deep-learning and machine learning. The proposed model utilized a fusion of Inception-V3 based deep-features and $\alpha - LBP$ based hand-crafted features for efficient extraction of features. Further the extracted features are fed to the four layers dual-LSTM for effective and enhanced learning of the fused features. The learned features obtained for the dual-LSTM layers are concatenated and fed through 5-layer ANFIS classifier. The proposed model with LSTM-ANFIS classifier efficiently adjust the weights and approximates more and more to catalog the tomato plant leaves into 10 modules. Furthermore, the proposed model provided an accuracy of 98.83% and a damage of 0.5289 on a subset of Plant Village dataset (the subset includes the images for tomato plant leaves only). Also, comparison of the planned model is made with respect to other classifiers which makes it evident that the performance with respect to ANFIS and LSTM based classification network provided promising results.

References

1. Lindhout, P., Korta, W., Cislik, M., Vos, I., Gerlagh, T.: 'Further identification of races of cladosporium fulvum (Fulvia fulva) on tomato originating from The Netherlands France and Poland' Netherlands. J. Plant Pathol. **95**(3), 143–148 (1989)

2. Kubota, K., Tsuda, S., Tamai, A., Meshi, T.: 'Tomato mosaic virus replication protein suppresses virus-targeted posttranscriptional gene silencing.' J. Virol. **77**(20), 11016–11026 (2003)
3. Tian, M., Benedetti, B., Kamoun, S.: A second Kazal-like protease inhibitor from phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. Plant Physiol. **138**(3), 1785–1793 (2005)
4. Chaerani, R., Voorrips, R.E.: Tomato early blight (Alternaria solani): the pathogen, genetics, and breeding for resistance. J. Gen. Plant Pathol. **72**(6), 335–347 (2006)
5. Gao, W., Li, B., Shi, Y., Xie, X.: Studies on pathogenicity differentiation of Corynespora cassiicola isolates, against cucumber, tomato and eggplant. Acta Horticulturae Sinica **38**(3), 465–470 (2011)
6. Dickey, M., Osborne, L.S., Mckenzie, C.L.: Papaya (Carica papaya, Brassicales: Caricaceae) is not a host plant of tomato yellow leaf curl virus (TYLCV; FamilyGeminiviridae, GenusBegomovirus). Florida Entomol. **95**(1), 211–213 (2012)
7. Moriones, E., Navas-Castillo, J.: Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. Virus Res. **71**, 123–134 (2000)
8. Navas-Castillo, J., Sánchez-Campos, S., Díaz, Juan Antonio, Sáez-Alonso, Elisa, Moriones, Enrique: Tomato yellow leaf curl virus-is causes a novel disease of common bean and severe epidemics in tomato in Spain. Plant Dis. **83**(1), 29–32 (1999). https://doi.org/10.1094/PDIS.1999.83.1.29
9. Blum, L.E.B.: Reduction of incidence and severity of septoria lycopersici leaf spot of tomato with bacteria and yeasts. Ciência Rural **30**(5), 761–765 (2000)
10. Bhimte, N.R., Thool, V.R.: Diseases detection of cotton leaf spot using image processing and SVM classifier. In: Proceedings of the 2nd International Conference on Intelligent Computing Control System (ICICCS), pp. 340–344 (2018). https://doi.org/10.1109/ICCONS.2018.8662906
11. Padol, P.B., Yadav, A.A.: 'SVM classifier based grape leaf disease detection. In: Proceedings of the Conference on Advanced Signal Processing (CASP), pp. 175–179 (2016). https://doi.org/10.1109/CASP.2016.7746160
12. Anand, R., Veni, S., Aravinth, J.: An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method. In: Proceedings of the International Conference Recent Trends Information Technology (ICRTIT), pp. 1–6 (2016). https://doi.org/10.1109/ICRTIT.2016.7569531
13. Kumari, C.U., Jeevan Prasad, S., Mounika, G.: Leaf disease detection: Feature extraction with K-means clustering and classification with ANN. In: Proceedings of the 3rd International Conference on Computation Methodologies Communication (ICCMC), pp. 1095–1098 (2019). https://doi.org/10.1109/ICCMC.2019.8819750
14. Rani, F.A.P., Kumar, S.N., Fred, A.L., Dyson, C., Suresh, V., Jeba, P.S.: K-means clustering and SVM for plant leaf disease detection and classification. In: Proceedings of the International Conference on Recent Advance Energy-Efficient Computation Communication (ICRAECC), pp. 1–4 (2019). https://doi.org/10.1109/ICRAECC43874.2019.8995157
15. Zhang, S., Wang, H., Huang, W., You, Z.: Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. Optik **157**, 866–872 (2018). https://doi.org/10.1016/j.ijleo.2017.11.190
16. Zeng, Q., Ma, X., Cheng, B., Zhou, E., Pang, W.: GANs-based data augmentation for citrus disease severity detection using deep learning. IEEE Access **8**, 172882–172891 (2020). https://doi.org/10.1109/ACCESS.2020.3025196
17. Szegedy, C., et al.: In: Proceedings of the IEEE Conference on Computation Visual Pattern Recognition (CVPR), pp. 1409–1414 (2015)

18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computation Visual Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
19. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computing Visual Pattern Recognition (CVPR), pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243
20. Kc, K., Yin, Z., Wu, M., Wu, Z.: Depthwise separable convolution architectures for plant disease classification. Comput. Electron. Agricult. **165**, 104948 (2019)
21. Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A.D., et al.: Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Comput. Electron. Agric. **138**, 200–209 (2017). https://doi.org/10.1016/j.compag.2017.04.013
22. Javidan, S.M., Banakar, A., Vakilian, K.A., Ampatzidis, Y.: Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. Smart Agricult. Technol. **3**, 100081 (2023)
23. Pothen, M.E., Pai, M.L.: Detection of rice leaf diseases using image processing. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 424–430 (2020). https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00080
24. Suman, T., Dhruva Kumar, T.: Classification of paddy leaf diseases using shape and color features. Int. J. Electr. Electron. Eng. **07**(01), 239–250 (2015)
25. Waghmare, H., Kokare, R.: Detection and classification of diseases of grape plant using color local binary pattern feature and machine learning for automated decision support system. In: 2016 International Conference on Signal Processing and Integrated Networks (SPIN) (2016)
26. Szegedy, C., et al.: Rethinking the Inception Architecture for Computer Vision. ArXiv, 2015, /abs/1512.00567. Accessed 12 Sept 2023
27. Kaplan, K., et al.: Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. Med. Hypotheses **139**, 109696 (2020). https://doi.org/10.1016/j.mehy.2020.109696
28. Staudemeyer, R.C., Morris, E.R.: Understanding LSTM – a Tutorial into Long Short-Term Memory Recurrent Neural Networks. *ArXiv*, 2019, /abs/1909.09586. Accessed 12 Sept 2023
29. Walia, N., Singh, H., Sharma, A.: ANFIS: adaptive neuro-fuzzy inference system- a survey. Int. J. Compu. Appl. **123**(13), 32–38 (2015). https://doi.org/10.5120/ijca2015905635
30. Hughes, D., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)
31. Raza, S.-E.-A., Prince, G., Clarkson, J.P., Rajpoot, N.M.: Automatic detection of diseased tomato plants using thermal and stereo visible light images. PLOS ONE **10**(4), e0123262 (2015)
32. Wang, Y.L., et al.: Image classification of tomato leaf diseases based on transfer learning. J. China Agricult. Univ. **24**(6), 124–130 (2019)
33. Prasad, S., Peddoju, S.K., Ghosh, D.: Multi-resolution mobile vision system for plant leaf disease diagnosis. Signal, Image Video Process. **10**(2), 379–388 (2016). https://doi.org/10.1007/s11760-015-0751-y
34. Guo, X.Q., Fan, T.J., Shu, X.: Tomato leaf diseases recognition based on improved Multi-Scale AlexNet. Trans. Chin. Soc. Agricult. Eng. **35**(13), 162–169 (2019)
35. Agarwal, M., et al.: ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Comput. Sci. **167**, 293–301 (2020). https://doi.org/10.1016/j.procs.2020.03.225

Unveiling the Enigma: Sudden Cardiac Arrest Among Youth and Uncovering Underlying Causes: A Data Analysis Study

Mohit M. Rao and Virendra Kumar Shrivastava(✉)

Department of Computer Science and Engineering, Alliance College of Engineering and Design, Alliance University, Bangalore 562106, India
mmohitphd22@ced.alliance.edu.in,
virendra.shrivastava@alliance.edu.in

Abstract. This study focuses on analyzing electrocardiogram (ECG) records related to sudden cardiac death in youth and analyze the factors influencing a sudden cardiac arrest. Sudden cardiac arrest poses a severe threat to life and can result in death if not addressed promptly. Recent American research indicates a 13% increase in sudden cardiac arrest cases among individuals aged between 35–45. Furthermore, the Indian Heart Association has reported that heart disease tends to affect Indians at a younger age compared to other demographic groups, often without prior warning. Studies reveal that Indians tend to experience heart diseases approximately a decade earlier than their Western counterparts. This work aims to conduct data collection and surveys to collect ECG records and other relevant information that is crucial for training the machine learning models to achieve the study outcomes. The study's scope encompasses evaluating the feasibility of the task, identifying relevant datasets, conducting data preprocessing, and assessing the performance using K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and random forest algorithm. Among these, both SVM and Random Forest demonstrated superior performance, with the Random Forest algorithm further fine-tuned to align with the dataset's specific requirements. The refined model exhibited 93% accuracy.

Keywords: Sudden cardiac arrest (SCA) · ECG dataset · heart failure · image processing · chi-square test · machine learning · artificial intelligence

1 Introduction

Sudden cardiac arrest often termed as SCA is a fatal event that causes due to unexpected and sudden functioning of heart. SCA happens when the heartbeat stops abruptly resulting in cessation of blood flow from heart to other vital organs. This causes sudden loss of consciousness, and lead to death within minutes if not treated promptly [1]. Sudden cardiac arrest can happen due to a variety of reasons. Ventricular fibrillation or often termed as VF is the primary cause of sudden cardiac arrest in most cases. This happens when the electrical signals become disordered. Ventricle's plays an important

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 167–178, 2025.
https://doi.org/10.1007/978-3-031-75167-7_13

role in the functioning of heart. The right ventricle sends deoxygenated blood to the lungs, while the left ventricle circulates oxygen-rich blood to the entire body via the aorta [2]. When the electrical signals get disordered, the ventricles quiver instead of pumping blood effectively which in turn leads to Sudden cardiac arrest. Coronary artery disease which blocks or narrows the blood vessels that supply the heart can also lead to SCA. Structural heart disease and inherited heart conditions can also lead to sudden cardiac arrest [3]. Abnormal levels of potassium, calcium, or magnesium can interfere with the heart's electrical system and lead to SCA [4]. Usage of drugs such as cocaine is positively correlated with sudden cardiac death as well [5].

World Health Organization (WHO) published in 2016, India accounts for close to 3.5 million deaths are attributed to cardiovascular disease, particularly among younger individuals. This is around one fifth of total cardiac related deaths globally [6]. Emerging research indicates that Indians experience cardiovascular diseases a decade earlier than individuals in Western populations [7]. Moreover, there is a concerning upward trend in the occurrence of Sudden Cardiac Arrest (SCA) among young individuals. Recent incidents of famous personalities such as Danish footballer Christian Eriksen (29 years), and actors Siddharth Shukla, Inder Kumar, KK, Nirmal Pandey and Puneeth Rajkumar who were known for their healthy lifestyles, passing away due to heart attacks have raised concerns about the susceptibility of younger individuals to heart diseases like SCA and heart attacks. This emphasizes the importance of exploring why the youth has become more prone to such cardiac conditions. Sudden cardiac arrest can happen at any age group and recent trends has shown an increase in sudden cardiac death among youth ($<$40 years) in India. SCA can happen to people without any previous cardiac history and can occur to people with healthy lifestyle [8]. Hence diagnosing SCA before the event is extremely hard as SCA is typically diagnosed based on a combination of a patient's medical history, symptoms, and the results of various diagnostic tests. Electrocardiogram (ECG) is a procedure to record the heart electrical activity, aiding in the detection of abnormal heart rhythms or patterns that may indicate SCA. It can help identify structural abnormalities, such as an enlarged heart or a heart valve problem, that increase the risk of SCA. Genetic testing, cardiac stress test are also used in diagnosing probability of sudden cardiac arrest [9]. Consistent health monitoring, particularly for individuals who are physically active or possess a genetic predisposition to cardiac issues, can aid in assessing the likelihood of sudden cardiac arrest.

In this study, datasets are collected from open-source databases. ECG records are collected from MIT/BIH-NSR database, Sudden Cardiac Death Holter Database [10] and ECG signals collected from patients who have suffered cardiac arrest out-of-hospital [11]. There were 400 unique patient datasets collected for the analysis and had multiple observations of these patients. This study focuses only on analyzing the ECG records and hence no other information was collected or stored from the patients or from any open-source platforms.

2 Motivation and Related Work

The surge in sudden cardiac arrest among youth can be attributed to various factors including sedentary lifestyles, diabetes, escalated alcohol consumption, smoking, and hypertension. [12] concluded that there has been a notable surge in the occurrence of

Sudden Cardiac Death (SCD), particularly in urban areas, which can largely be attributed to the concerning rise in the prevalence of coronary artery disease (CAD), diabetes, and hypertension in India. This observation supports another research which study that mentions the prevalence of risk factors are high in the youth in urban Indian population. The study was limited to New-Delhi region, the capital city of India and the study noted the association of obesity, hypertension, and diabetics as an influencing factor for cardiovascular diseases [13]. Previously, Sudden Cardiac Arrest (SCA) was believed to be uncommon in young individuals and associated with hereditary or congenital factors. However, the situation is now much more intricate, and in societies with abundant food resources, it has become more frequently linked to preventable causes across all age groups [14, 15]. Research focused on identifying the risk factors of sudden cardiac arrest in south India [16], also noted the local diet in India often includes high levels of carbohydrates, refined sugar, and elevated fat content. This dietary pattern contributes to the widespread occurrence of metabolic syndrome and diabetes in the country. This is applicable for someone without any structural abnormality [17] mentioned obesity and epilepsy as one of the major comorbid conditions especially for patients with normal heart conditions. There are data science studies which leveraged advanced machine learning and deep learning techniques to identify the association of cholesterol level with sudden cardiac deaths and one such study [18] concluded that cholesterol level in which the level greater than 200 mg/dl are more prone for SCD. Smoking is widely regarded as one of the major contributors for sudden cardiac deaths. A data study which aimed to analyze the relevance of cardiac pathology with its association to age, gender, occupation, risk factors from the autopsy profile reported in Mumbai region of Maharashtra [19], noted Hypertension, Diabetes, Obesity and Smoking contributed to sudden cardiac deaths. There are a handful of research which concludes the association between smoking and sudden cardiac arrest [20–25] notes the same and are conducted considering multiple factors such as demographics, age, sex etc.

Several research studies have been conducted utilizing various machine learning models to investigate the factors influencing sudden cardiac arrest (SCA) and to predict the probability of such events. These studies have employed a range of models, from simple decision trees to complex deep learning algorithms. For instance, Study [26] aimed to predict the occurrence of these fatal events using decision tree and naive Bayes models, highlighting the effectiveness of naive Bayes in estimating the probability of SCA based on electrocardiogram reports. Similarly, another research study [27] conducted a comprehensive analysis comparing 12 different machine learning and deep learning models. The study revealed that SVM exhibited the highest precision for predicting SCA. Furthermore, in an analysis of ECG reports [28], various algorithms such as SVM, Neural Network Classifier, Ant Colony and Gradient Descent were compared to identify the indication of the heart disease. The research findings indicated that Gradient Descent + Particle Swarm Optimization performed exceptionally well as the most suitable algorithm for ECG analysis. The utilization of Artificial Neural Networks (ANN) has been recommended as a model for analysis in several research studies. These studies have compared or employed ANN to analyze ECG records and health reports, achieving higher levels of accuracy. For instance, in Study [29], the aim was to predict the probability of cardiac arrests and classify them ECG records, and ANN

model achieved the best performance accuracy of 95.8%. Another study [30] compared ANN with SVM, Logistic Regression, and Random Forest, concluding that ANN exhibited higher precision (~85%) in predicting the frequency of cardiac arrest. Convolutional Neural Networks (CNN) and Intelligent Hybrid Quick Reduct Particle Swarm Optimization are also notable techniques that have been leveraged to identify influencing factors. Study [31] suggests the use of CNN and proposes an accuracy of over 95%. Another Study [32], the proposal is to employ IHQRPSO for the analysis. Overall, these studies highlight the effectiveness of ANN, CNN, and IHQRPSO in analyzing ECG records, health reports, and predicting cardiac arrest probabilities with improved accuracies and precision.

Numerous machine learning (ML) research work have been conducted to evaluate ML model performance and identify the most suitable algorithms for specific datasets. SVM and RF have emerged as the most popular choices based on these studies. In Study [33], eight machine learning models were compared using HRV analysis based on ECG based as well as non-ECG records, and Random Forest exhibited the highest accuracy. In another study [34] comparing similar algorithms with neural networks, the decision tree algorithm had the best accuracy. The study also noted that the performance further improved after applying genetic algorithms. Study [35] compared Random Forest, KNN (K-Nearest Neighbors), and logistic regression to forecast the development of heart disease in certain patients and concluded that the Random Forest algorithm outperformed the others. Conversely, another research [36] leveraged SVM for the same forecast and observed high accuracy, specificity, and sensitivity. In Study [37], SVM was compared with Random Forest, KNN, and Naive Bayes, with SVM demonstrating a slight advantage over the other algorithms.

In addition to identifying the factors influencing SCA, it is equally crucial to recommend preventive measures and implement guidelines to address this issue. Research studies have emphasized the importance of various strategies and interventions. A review study [38], explored the causes and mechanisms of SCD in sports activities and exercise in athletes (youth and older). It recommended screening measures and highlighted the significance of active CPR (Cardiopulmonary Resuscitation) and rapid defibrillation. Study [39] supported these recommendations by noting a decline in post myocardial infarction (MI) mortality and SCD in the western world in the recent years, which predates the regular usage of Implantable Cardioverter Defibrillators (ICD) and the prescription of defibrillators has reduced long-term mortality. Morijon et al. [40] concluded that while ICDs have shown clear efficacy in reducing the risk of SCD, not eliminating it completely. A case study [41] focusing on sudden cardiac arrest in the emergency department highlighted that despite advances in medical care, training, and technology, outcomes after CPR have not significantly improved, especially for in-hospital SCA. Study [42] found that during marathons, the availability of advanced life support and timely defibrillation contributes to a reduction in cardiac risk. The results of study [43] shows the importance of early CPR and early defibrillation in the SCA patients. A case study [44] conducted to estimate the incidence of SCD in student-athletes underscored the importance of timely defibrillation. Furthermore, Study [45] recommended the use of wearable devices and IoT (Internet of Things) technology for monitoring heart parameters, including pre-screening, to aid in the identification and prevention of SCA. This can

significantly reduce the workload on hospitals and enable monitoring of heart parameters anytime and anywhere. Overall, these studies stress the importance of preventive measures, timely interventions such as CPR and defibrillation, and leveraging technology to enhance monitoring and response capabilities in the context of Sudden Cardiac Arrest.

3 Methodology

This research aims on analyzing ECG records collected from open-source databases and research, including MIT-BIH arrhythmia database, the INCART database, ECG GUDB, ECG-ViEW II, and the QT database. 400 unique patient records were analyzed as a part of the study. The dataset consists of 32 features, derived from the ECG records. These features include, RR Interval, which is the time taken between two successive Rwaves. Normal RR interval is 0.6–1.2 s. Heartbeat interval features, QT Interval, ST Interval, PQ Interval, and QRS duration are also included in the features.

3.1 Data Preprocessing

Figure 1 is a sample of ECG waveform collected from the Sudden Cardiac Death Holter Database. There are multiple observations made on each of the records/unique patients and features are derived from the ECG records [46–48]. The first task in data analysis is to prepare and pre-process the dataset. Big data platform is not needed for data processing as the data is limited both vertically and horizontally [49]. The data sets consist of only 30+ features and less than a hundred thousand entries of 400 unique patients. This amount of data could be easily stored within a SQL or NoSQL databases. As the dataset is already defined, the requirements are fixed and there are no possibilities of vertical scaling of data, SQL database is a good fit for data storage [50–52]. SQL database provides advantages to process structured data with no foreseeable changes in schema [53]. SQLite database is selected to store the dataset as its open source and is light weight. All the ECG information collected from various data sources is stored within the SQLite database. An Azure VM with windows 2016 server variant is chosen to host the SQLite database and for data processing and data analysis.

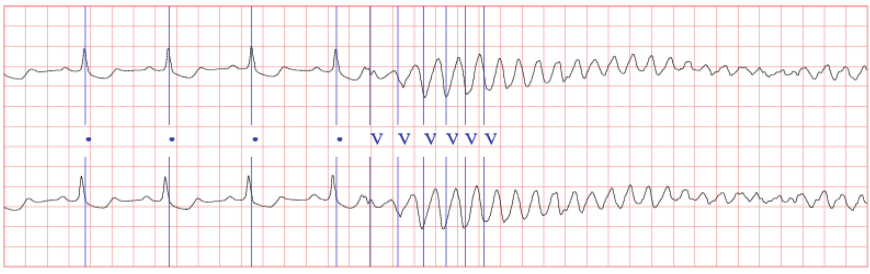


Fig. 1. Sample ECG Waveform

As the data is obtained from ECG records, and no manual data is collected for the study, data cleaning and imputing is not required. Dataset collected is complete

and has no missing values or type mismatches. However, the data is skewed as the occurrence of variation in ECG record is very less compared to normal pattern. This brings in the problem of skewness. From all the observations, only 15% attribute to pattern with variations and SCA. This creates bias and results in the output [54]. To overcome skewness, oversampling of minority class or under sampling of majority class can be leveraged. As the dataset is limited to a few hundred thousand, oversampling is preferred over under sampling [55]. The features are limited and under sampling might also result in loss of key information from these features. There are different methods to perform oversampling [56]. Synthetic dataset is used for this study that produced additional samples. This effectively boosts the proportion between both the class.

3.2 Classification Algorithms

Choosing a right classification algorithm is a key part of this research. Random Forest, SVM, and KNN are leveraged in this study. Processed dataset is trained and tested using above 3 algorithms. The study utilized an open-source low-code platform tool known as Orange [57] to expedite and simplify the process of machine learning testing. Orange allows machine learning techniques and functions to be represented as easy-touse drag-and-drop components. This tool was employed for model training, testing, validation, and evaluation purposes. Figure 2 depicts the workflow created using Orange. A workflow was established through a user-friendly drag-and-drop interface. After performing data preprocessing tasks such as cleaning and transformation, this prepared data was used to evaluate three different models. The evaluation process involved assessing accuracy, precision, and recall for each of these models [58].

All the algorithms used for analysis are fine tuned to improve accuracy [59]. Cross validation is used in random forest by using the K fold method. Here, the training set is split into K subsets, or folds. In an iterative process, the model is trained K-times. During each iteration, the model is trained using K-1 of these folds, and its performance is evaluated on the remaining Kth fold. As Python is leveraged to build the model, Scikit-Learn's Randomized Search CV [60] method allows an easy way to fine tune the model. With this package range of hyperparameters can be specified and then randomly select values from this range. For each combination of these values, the model is trained and evaluated using k-fold cross validation method. Other parameters such as n_estimators, used to define the number of trees to be used in the model, max_features, which is used to limit the count of maximum features in each tree, max_leaf_node which restricts the growth of each tree are used for fine tuning [61]. Random Forest is a Bagging Ensemble technique designed to harness the collective power of multiple decision trees as predictors, all working in parallel to produce improved results. In the Random Forest method, a subset of instances is created by randomly selecting instances from the training dataset, with replacement, and these are then utilized to train the individual predictors within the Random Forest. This approach extends to both training instances and features, introducing diversity into the model [62]. A single decision tree often exhibits high variance and low bias. Random Forest mitigates this variance issue by employing numerous decision trees, each trained on a different random subset. Several classification models are assessed using the training dataset. The model that yields the most favorable outcome is selected (depicted in Fig. 3).

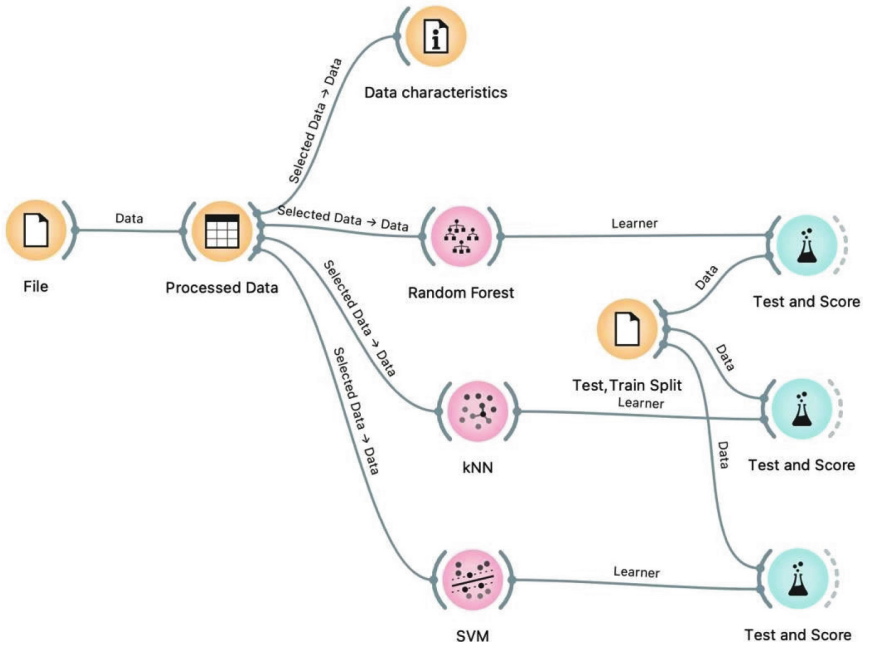


Fig. 2. Orange workflow used for testing.

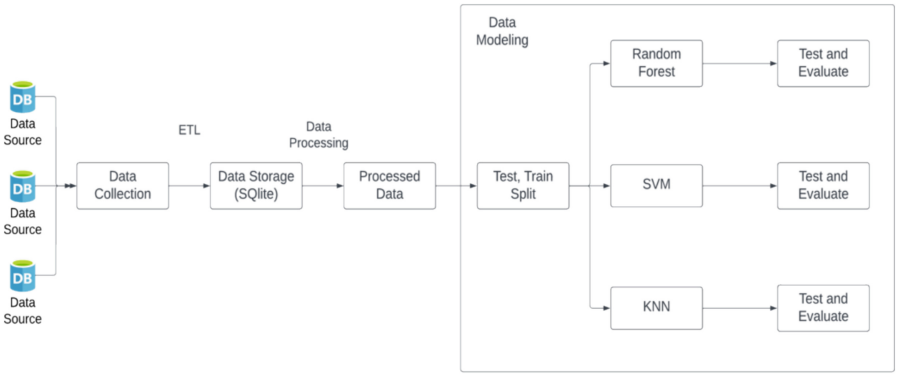


Fig. 3. Data Modeling Methodology

4 Results and Discussion

To assess the aforementioned assumption regarding the selection of the RF model, a comparison is made with the SVM and the KNN models. The analysis involves multiple observations of ECG records from 400 distinct patients after undergoing data processing and transformation. The collected datasets are stored in an SQLite database, and Python is employed for conducting the evaluation. The outcomes of this evaluation are presented in Table 1.

Table 1. Model Evaluation Metrics

Model Name	Accuracy (%)	Precision	Recall
RF	93	0.873	0.901
SVM	92	0.864	0.896
KNN	82	0.794	0.818

All the classification algorithms were fine-tuned by leveraging the in-built capabilities of python and Scikit-Learn functionalities. The RF model is fine-tuned by to attain the best results. Understanding the relationship between each of the extracted features from the ECG records plays an important role in classification. The feature importance functionality within the RF algorithm is subsequently employed to assess the significance of individual features. These features are then ranked based on their F1 score.

5 Conclusion and Future Work

The goal of the research was to analyze the electrocardiogram records related to sudden cardiac death in youth and identifying factors influencing a sudden cardiac arrest. The research was initiated to investigate the feasibility of the task, identify relevant datasets, preprocess and cleanse the data, and draw meaningful conclusions. Within the scope of this research, three algorithms were assessed: the SVM, KNN, and RF. Among these, both the SVM and the RF algorithm exhibited superior performance. Given the dataset characteristics and by utilizing the feature importance, the RF algorithm was further fine-tuned to achieve the best results. The model achieved accuracy of 93%, and an F1 score of 0.89. They underscore the significance of utilizing advanced machine learning techniques to analyze ECG records and their potential in identifying the contributing factors to sudden cardiac death in youth. The results validate the feasibility of using these algorithms and the dataset to draw meaningful conclusions, shows the way for future research.

In this research endeavor, the examination was confined to a limited dataset consisting of ECG records from only 400 distinct patients. To enhance the depth of analysis, it is advisable to expand the dataset to include ECG records from various geographic regions, diverse lifestyles, and individuals from varying economic and social backgrounds. These additional data points are integrated into the Random Forest algorithm for retraining. It's important to acknowledge that the dataset exclusively comprised post-COVID data points. However, it's worth noting that these distinctions were primarily considered based on the features. In future research endeavors, it is valuable to investigate these factors independently in the pre-COVID era.

In future researchers may include other pertinent data sources such as routine blood tests, genetic testing, and X-rays, in addition to ECG records. This comprehensive dataset facilitates a more intricate analysis, providing deeper insights into the relationships between individual features and their correlation with the occurrence of sudden cardiac arrest. It contributes significantly to understanding the role of ECG records in predicting sudden cardiac arrest.

References

1. Dewi, A.K., et al.: Cardiopulmonary Resuscitation Management for Sudden Cardiac Arrest on Field Sport Activity (2023)
2. Saghiv, M.S., Sagiv, M.S., Saghiv, M.S., Sagiv, M.S.: Cardiovascular Function. Basic Exercise Physiology: Clinical and Laboratory Perspectives, 285–369 (2020)
3. Ghani, U., et al.: Sudden Cardiac Death in Athletes: Consensuses and Controversies. Cureus **15**(6) (2023)
4. Brown, R.B.: Phosphate toxicity and SERCA2a dysfunction in sudden cardiac arrest. FASEB J. **37**(7), e23030 (2023)
5. Trytell, A., et al.: Prevalence of illicit drug use in young sudden cardiac death patients. Heart rhythm (2023)
6. https://www.who.int/india/health-topics/cardiovascular-diseases
7. Agrawal, A., et al.: Risk factors, lab parameters, angiographic characteristics and outcomes of coronary artery disease in young South Asian patients: a systematic review. J. Int. Med. Res. **51**(8), 03000605231187806 (2023)
8. Gupta, R.D., Kothadia, R.J., Parray, A.A.: Association between abdominal obesity and diabetes in India: Findings from a nationally representative study. Diabetes Epidemiology and Management **12**, 100155 (2023)
9. Civieri, G., Kerkhof, P.L., Montisci, R., Iliceto, S., Tona, F.: Sex differences in diagnostic modalities of coronary artery disease: Evidence from coronary microcirculation. Atherosclerosis, 117276 (2023)
10. https://archive.physionet.org/physiobank/database/sddb/
11. Benini, S., Ivanovic, M.D., Savardi, M., Krsic, J., Hadžievski, L., Baronio, F.: ECG waveform dataset for predicting defibrillation outcome in out-of-hospital cardiac arrested patients. Data Brief **34**, 106635 (2021)
12. Honnekeri, B.S., Lokhandwala, D., Panicker, G.K., Lokhandwala, Y.: Sudden cardiac death in India: A growing concern. J. Assoc. Physicians India **62**(12), 36–40 (2014)
13. Singh, N., Deniel, V.: Effect of Modifiable factors on Systolic Blood Pressure (SBP) in Elderly Population: a study from a tertiary care hospital in Delhi. International Journal of Medical, Pharmacy and Drug Research **7**(4) (2023)
14. Atkins, D.L.: Sudden cardiac arrest in a young population: not so unpredictable. J. Am. Heart Assoc. **8**(2), e011700 (2019)
15. Dhungana, R.R., et al.: Prevalence of cardiovascular disease risk factors: a community-based cross-sectional study in a peri-urban community of Kathmandu. Nepal. Indian heart journal **70**, S20–S27 (2018)
16. Srivatsa, U.N., Swaminathan, K., Munavarah, K.S.A., Amsterdam, E., Shantaraman, K.: Sudden cardiac death in South India: incidence, risk factors and pathology. Indian Pacing and Electrophysiology Journal **16**(4), 121–125 (2016)
17. Finocchiaro, G., et al.: Sudden cardiac death among adolescents in the United Kingdom. J. Am. Coll. Cardiol. **81**(11), 1007–1017 (2023)
18. Gupta, S.K., Shrivastava, A., Upadhyay, S.P., Chaurasia, P.K.: A Machine Learning Approach for Heart Attack Prediction
19. Gaiwale, S.D., Petkar, M.R., Gundla, V.S., Sapate, A.B., Ghangale, A.L.: Autopsy profile of sudden cardiac deaths reported in mumbai region of Maharashtra. Forensic Medicine and Pathology **12**(1), 25 (2019)
20. Bajpai, A., Sinha, S., Yadav, A., Srivastava, V.: Early prediction of cardiac arrest using hybrid machine learning models. In: 2023 17th International Conference on Electronics Computer and Computation (ICECCO), pp. 1–7. IEEE (2023)

21. Shashikant, R., Chetankumar, P.: Predictive model of cardiac arrest in smokers using machine learning technique based on Heart Rate Variability parameter. Applied Computing and Informatics, (ahead-of-print) (2020)
22. Warming, P.E., et al.: The impact of modifiable risk factors in the association between socioeconomic status and sudden cardiac death in a prospective cohort study: equal access to healthcare, unequal outcome. European Journal of Preventive Cardiology, zwad086 (2023)
23. Aune, D., Schlesinger, S., Norat, T., Riboli, E.: Tobacco smoking and the risk of sudden cardiac death: a systematic review and meta-analysis of prospective studies. European journal of epidemiology **33**, 509–521 (2018)
24. Sri, M.B., Radhika, V.P., Praveena, N.: Cardiac arrest prediction using gradient boosting algorithm. In: 2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), pp. 1–6. IEEE (2023)
25. Sharma, G., Hugar, B.S., Praveen, S., P Yajaman, G.C., Kanchan, T.: Pattern of sudden cardiac deaths. Medico-Legal Journal, 00258172211059928 (2022)
26. Bhatt, A., Dubey, S.K., Bhatt, A.K.: Sudden cardiac arrest prediction using predictive analytics. Int. J. Intel. Eng. Sys. 10–3 (2017)
27. Bhattacharyya, D., Dinesh Reddy, B., Kumari, N.M.J., Rao, N.T.: Comprehensive analysis on comparison of machine learning and deep learning applications on cardiac arrest. J. Med. Pharm. Allied Sci. **10**(4), 3125–3131 (2021)
28. Patro, S.P., Padhy, N., Sah, R.D.: Anticipation of heart disease using improved optimization techniques. In: Computing, Communication and Learning: First International Conference, CoCoLe 2022, Warangal, India, October 27–29, 2022, Proceedings, pp. 91–102. Springer Nature Switzerland, Cham (2023)
29. Fatima, N., Irtaza, A., Ali, R.: A novel deep learning based framework for cardiac arrest prediction. In: 2023 International Conference on Robotics and Automation in Industry (ICRAI), pp. 1–6. IEEE (2023)
30. Chauhan, U., Kumar, V., Chauhan, V., Tiwary, S., Kumar, A.: Cardiac arrest prediction using machine learning algorithms. In: 2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICICT), Vol. 1, pp. 886–890. IEEE (2019)
31. Panicker, S.: Use of machine learning techniques in healthcare: a brief review of cardiovascular disease classification. In: 2nd International Conference on Communication & Information Processing (ICCIP) (2020)
32. Shathesh, S., Durairaj, M.: An intelligent hybrid mechanism to predict the risk of cardio vascular disease. Indian J. Sci. Technol. **9**(4), 137–146 (2016)
33. Kasturiwale, H., Kale, S.N.: Detection of cardiac problems by the extraction of multimodal functions and machine learning techniques. In: IOP Conference Series: Materials Science and Engineering, Vol. 1022, No. 1, p. 012124. IOP Publishing (2021)
34. Soni, J., Ansari, U., Sharma, D., Soni, S.: Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int. J. Comp. App. **17**(8), 43–48 (2011)
35. Dinesh, K.G., Arumugaraj, K., Santhosh, K.D., Mareeswari, V.: Prediction of cardiovascular disease using machine learning algorithms. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1–7. IEEE (2018)
36. Kumar, V.D.A., et al.: Prediction of cardiovascular disease using machine learning technique—a modern approach. Computers, Materials and Continua **71**(1), 855–869 (2022)
37. Banerjee, P., Bhattacherjee, S., Dasgupta, K., Sen, S.: Performance evaluation of machine learning classifiers for sudden cardiac arrest detection. J. Inst. Eng. (India): Series B 1–7 (2022)
38. Vora, A., Burkule, N., Contractor, A., Bhargava, K.: Prevention of sudden cardiac death in athletes, sportspersons and marathoners in India. Indian Heart J. **70**(1), 137–145 (2018)
39. Rao, B.H.: Global burden of sudden cardiac death and insights from India. indian heart journal **66**, S18–S23 (2014)

40. Marijon, E., et al.: The Lancet Commission to reduce the global burden of sudden cardiac death: a call for multidisciplinary action. The Lancet (2023)
41. Pandian, G.R., Thampi, S.M., Chakraborty, N., Kattula, D., Kundavaram, P.P.A.: Profile and outcome of sudden cardiac arrests in the emergency department of a tertiary care hospital in South India. J. Emerg. Trauma Shock **9**(4), 139 (2016)
42. Yukie, Y.K., Cynthia, K.C., Jonathan, H.O., Lam, K.K.: Training and Preparedness for Major Events, Massive Accidents, and Natural Disasters. Disaster Management and Emergency Medicine in the Asia-Pacific Region: Current Practices and Future Directions 167 (2023)
43. Ha, V.H.T., et al.: Trends in survival from out-of-hospital cardiac arrest with a shockable rhythm and its association with bystander resuscitation: a retrospective study. Emergency Medicine Journal (2023)
44. Ba, J.H.: Sudden cardiac arrest in athletes and strategies to optimize preparedness (2023)
45. Khan, M., et al.: Severe analysis of cardiac disease detection using the wearable device by artificial intelligence. In: 2020 IEEE International Conference for Innovation in Technology (INOCON), pp. 1–8. IEEE (2020)
46. Singhal, A., Phogat, M., Kumar, D., Kumar, A., Dahiya, M., Shrivastava, V.K.: Study of deep learning techniques for medical image analysis: A review. Materials Today: Proceedings **56**, 209–214 (2022)
47. Lalli, K., Shrivastava, V.K., Shekhar, R.: Detecting copy move image forgery using a deep learning model: a review. In: 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), pp. 1–7. IEEE (2023)
48. Batra, R., Shrivastava, V.K., Goel, A.K.: Anomaly Detection over SDN Using Machine Learning and Deep Learning for Securing Smart City. In: Green Internet of Things for Smart Cities, pp. 191–204. CRC Press (2021)
49. Che, D., Safran, M., Peng, Z.: From big data to big data mining: challenges, issues, and opportunities. In: Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18, pp. 1–15. Springer Berlin Heidelberg (2013)
50. Khatri, I., Shrivastava, V.K.: A survey of big data in healthcare industry. In Advanced Computing and Communication Technologies: Proceedings of the 9th ICACCT, 2015, pp. 245–257. Springer Singapore (2016)
51. Batra, R., Mahajan, M., Shrivastava, V.K., Goel, A.K.: Detection of COVID-19 using textual clinical data: a machine learning approach. Impact of AI and data science in response to coronavirus pandemic, 97–109 (2021)
52. Saini, V., Rai, N., Sharma, N., Shrivastava, V.K.: A convolutional neural network based prediction model for classification of skin cancer images. In: International Conference on Intelligent Systems and Machine Learning, pp. 92–102. Springer Nature Switzerland, Cham (2022)
53. ElDahshan, K.A., AlHabshy, A.A., Abutaleb, G.E.: A comparative study among the main categories of NoSQL databases
54. Jeni, L.A., Cohn, J.F., De La Torre, F.: Facing imbalanced data--recommendations for the use of performance metrics. In: 2013 Humaine association conference on affective computing and intelligent interaction, pp. 245–251. IEEE (2013)
55. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. **106**, 249–259 (2018)
56. Sharma, S., Gosain, A., Jain, S.: A review of the oversampling techniques in class imbalance problem. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1, pp. 459–472. Springer Singapore (2022)
57. https://orangedatamining.com/
58. Shrivastava, V.K., Shrivastava, A., Sharma, N., Mohanty, S.N., Pattanaik, C.R.: Deep learning model for temperature prediction: an empirical study. Model. Earth Syst. Environ (2022)

59. Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A.: Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Model. **406**, 109–120 (2019)
60. Ramadhan, M.M., Sitanggang, I.S., Nasution, F.R., Ghifari, A.: Parameter tuning in random forest based on grid search method for gender classification based on voice frequency. DEStech transactions on computer science and engineering **10**(2017) (2017)
61. Agrawal, T., Agrawal, T.: Hyperparameter optimization using scikit-learn. Hyperparameter optimization in machine learning: make your machine learning and deep learning models more efficient 31–51 (2021)
62. Shrivastava, V.K., Shrivastava, A., Sharma, N., Mohanty, S.N., Pattanaik, C.R.: Deep learning model for temperature prediction: a case study in New Delhi. Journal of Forecasting (2023)



Using Encoder Models to Predict Ratings Using Flipkart Reviews

Himani Agarwal[1(✉)] and Shweta Singhal[2]

[1] Department of ECE, Indira Gandhi Delhi Technical University for Women, Delhi, India
himgrwal04@gmail.com
[2] Department of IT, Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract. Humans are highly subjective beings and present their opinions on products and services in the form of reviews, which need to be analyzed to understand the consumer base. However, deciphering the consumer sentiment from online reviews is a massive task due to the diversity in the reviews and their unstructured nature. To tackle this challenge, in this study we employ the BERT (Bidirectional Encoder Representations from Transformers) model to predict numerical ratings from textual reviews from a Flipkart product review dataset sourced from Kaggle. Our findings demonstrate BERT's suitability for analyzing diverse and complex textual data, with an impressive accuracy of 73%, due to its ability to employ contextual analysis. This study showcases the potential of utilizing advanced NLP techniques to derive insights from otherwise huge and undecipherable data, which can be used to make well informed decisions by businesses looking to maximize gains and customers looking for make better purchasing choices alike.

Keywords: BERT · Predicting Ratings · Natural Language Processing · Textual Data · Machine Learning

1 Introduction

In the digital age, user feedback is of utmost importance for businesses to improve their products and services and stay ahead of the competition. Online reviews, social media comments and other user-generated content provide valuable insights into customer preferences and satisfaction levels. Analyzing this data can help companies understand their customers better and tailor their offerings accordingly. Businesses rely heavily on user ratings for their products or services because ratings are a trustworthy indicator of how well a product or service is being received by its consumers. Ratings also influence the purchasing decisions of other consumers, who often check for online reviews to know whether a product is worth the expense or not because reviews provide a general overview of the product's quality and popularity. Generally, reviews also come along with ratings, further expanding the sentiment expressed by the rating, emphasizing the pros and cons of the product or service, according to the consumer.

However, reviews often consist of enormous amounts of textual data, which are difficult to interpret and study. A solution to this can be converting these reviews into

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 179–190, 2025.
https://doi.org/10.1007/978-3-031-75167-7_14

numerical ratings because they provide a simpler way to analyze a review's overall sentiment. They also make it possible to categorize products based on how well the consumers received them. Numerical data, like ratings, are also easier to analyze statistically and use mathematical models to identify trends for advising businesses. This way, businesses can better understand consumer choices and make more educated decisions about their products and services. The Bidirectional Encoder Representations from Transformers (BERT) model [1]. Has demonstrated its utility for the same. BERT is an advanced NLP model developed by Google and is capable of understanding sentences based on the words in the sentence and the context in which they are used. BERT can be effectively employed in various tasks such as sentiment analysis, text classification, and answering questions, among others.

This study scrutinizes how BERT can anticipate ratings based on user reviews of products sold on Flipkart, one of India's leading e-commerce platforms. BERT is chosen because it can analyze textual data based on the words' context. Hence, it is capable of working with the unstructured and informal language used by consumers in online reviews.

We aim to analyze user-review textual data and convert them into numerical representations of ratings on a ranking of 1–5. The model is trained on a large dataset of Flipkart user reviews and their corresponding ratings sourced from Kaggle. The BERT model's performance is then fine-tuned and optimized for accurate ratings prediction.

2 Literature Review

This study scrutinizes the use of encoder models to predict numerical ratings from textual user reviews on the Flipkart website for its numerous products. Encoder models are a deep learning algorithms that possess the ability to represent data in low dimensional spaces while not losing its semantic meaning either. In the paper's literature review, the current state of research into using encoder models for sentiment analysis and text classification has been highlighted, as well as research into models working to convert reviews into ratings.

Jhaveri et al. [2] analyze over 12,000 online patient reviews for 530 doctors to examine patient satisfaction and dissatisfaction based on enforcement reviews and predicts numeric ratings. Doctor evaluation websites can help patients make informed healthcare decisions. Our work focuses on using the BERT model to predict product ratings from textual data, showcasing the potential of natural language processing techniques for consumer feedback and informed purchasing decisions.

Akinlaja et al. [3] examine the e-commerce mismatch problem of analyzing textual product reviews and numerical ratings generated on the internet. The investigation employs sentiment analysis and advanced machine learning techniques to detect any inconsistencies between product reviews and their corresponding ratings, achieving an average accuracy of around 87% with 10-fold cross-validation. Employing the BERT model, our work demonstrates the potential of natural language processing techniques for predicting product ratings from textual data and improving the accuracy of reviews.

Kusunoki et al. [4] propose a method to predict overall numerical ratings from textual reviews, using CSPD models and BERT vectors. The approach surpasses the precision of

solely utilizing BERT vectors, remedying the predicament of incongruous assessments amidst text-based reviews and numerical rankings found hotel booking sites. The use of the BERT model in our study highlights the efficacy of natural language processing techniques in predicting product ratings from textual data, enhancing the accuracy of product reviews.

Gezici et al. [5] focuses on using recurrent neural networks to anticipate the ratings of OSS apps on the Play Store based on the reviews their users write from their experience. The novel approach attains an impressive accuracy rate of 87.61% when compared to the actual user ratings, thus offering potential solutions in instances where the original app ratings are missing or do not correlate well with the reviews.

Liu et al. [6] presents a model called VITAL-T, which helps in predicting user ratings based on fashion trends. It has a rating prediction function, which incorporates visual information and review text, along with a topic model to categorize words into non-visual and visual words. Experiments conducted on publicly available datasets evince that VITAL-T outshines numerous cutting-edge Recommender Systems. By leveraging the BERT model, our work demonstrates how natural language processing techniques can be effectively used to predict product ratings from textual data, improving the accuracy of online reviews.

Xu et al. [7] propose an opinion mining model called NeuO that enhances recommendation systems by utilizing both ratings and reviews. Employing convolutional neural networks and a sentiment classification scoring mechanism, the model encapsulates users' genuine perspectives, even in datasets with imbalanced data distribution. The outcome of the experimentation showcases the superiority of the novel method over current leading methods when tested on practical datasets. Our investigation underscores the potential of NLP methodologies to enhance the precision of product evaluations. by using the BERT model to predict ratings from textual data.

Shiju et al. [8] scrutinize the disparity between textual assessments and numeric ratings on Drug review portals and propose models to categorize drug review appraisals based on textual evaluations with both conventional machine learning and advanced deep learning models. Bio Clinical BERT model stood out with an impressive overall accuracy rate of 87%. The study also discovered the concepts of the Unified Medical Language System from the comments and investigated their semantic classifications. The BERT model is utilized in our work to demonstrate the potential of natural language processing techniques in predicting product ratings from textual data, improving the accuracy of consumer feedback.

Zhang et al. [9] delve into harnessing user-generated content in social commerce to anticipate ratings, with particular emphasis on reviews, connections, and user-product or product-product mappings. To encapsulate users' evolving preferences, the researchers suggest a dynamic hierarchical tree of topic words, and a prediction technique which uses as a combination of gradient-boosting decision tree and multiclass linear regression to improve accuracy. Our work highlights the potential of natural language processing techniques to extract insights from vast amounts of unstructured textual data, as demonstrated by the BERT model's effectiveness in predicting product ratings. Kavousi et al. [10] focuses on using user-generated text, such as online reviews, to predict ratings of products. The investigation scrutinizes the influence of lexical features and sentimental terminology on forecast accuracy and indicates that vocabulary with high information gain scores has more efficacy, compared to the traditional term frequency-inverse document frequency (TF-IDF) metric. Furthermore, the research explores the optimal number of features for rating prediction. Utilizing the BERT model, our study showcases the potential of natural language processing techniques for predicting product ratings from textual data and enabling informed purchasing decisions. Venkata et al. [11] underline the significance of user feedback in the digital realm, the research accentuates how it can facilitate the service sector to cater to user requirements more efficiently. The bag-of-words approach was applied to extract hidden information from hotel reviews in Datafiniti's hotel reviews dataset, using natural language processing to predict user ratings. The methodology achieved an accuracy rate of around 60%.

Carneiro et al. [12] discuss the importance of argumentation-based dialogue models in decision contexts where decision-makers are dispersed or unknown. The study investigates the utilization of ML models and NLP to automatically anticipate the preferences of tourists by forecasting the evaluations of points of interest through other tourist reviews. The experiments manifest a remarkable accuracy in determining whether a review is favorable or unfavorable. However, the study struggles to foresee the rating allocated to the points of interest.

Saumya et al. [13] highlights the significance of online consumer reviews and ratings in e-commerce. It explains the inconsistency between the rating and the review text and proposes a machine learning-based model to identify this inconsistency. The model uses Long Short-Term Memory to classify reviews as positive or negative and checks their polarity with the user's rating for consistency.

Lu et al. [14] tackle the hurdle of predicting customized ratings of products that users haven't engaged with previously. A unique hybrid sequential model is proposed, that finds it's basis in the product reviews and is suggested to seize the sequence traits of users' reviews and rating sequences by exploiting both LSTM models and low-rank factorization. The results substantiate that the suggested framework surpasses the existing cutting-edge approaches (Table 1).

Table 1. Comparative analysis between prior research works

Paper	Description	Remarks
[15]	This research paper puts forth a novel approach to forecasting ratings in recommendation systems, which involves utilizing deep learning techniques known as EMRM which seek to counter the issue of data sparsity by acquiring auxiliary reviews from nearby and distant neighbors Moreover, EMRM employs an Ordered-Neurons Long Short-Term Memory (ONLSTM) mechanism to extract relevant semantic details from both user and auxiliary reviews. The study's empirical findings indicate that EMRM surpasses other models on three actual datasets when it comes to predicting rating accuracy	By employing the BERT model, our work emphasizes the efficacy of natural language processing techniques in predicting product ratings from textual data, empowering businesses to make well-informed decisions based on consumer feedback
[16]	Consumer reviews may contain sarcastic feedback resulting in inconsistent ratings, hindering valuable feedback for firms. This study proposes utilizing consumer review content to identify true ratings by applying multi-class classification methods. Results indicate that SVM with TF-IGM performed best, with average improved accuracies and F1 scores of 11.7% and 10.5%, respectively	Our research underscores the potential of natural language processing techniques, exemplified by the BERT model, to extract insights from copious amounts of unstructured data and improve the accuracy of product reviews by predicting ratings from textual data
[17]	The AspeRa model is used to estimate user ratings based on review texts and discovers coherent aspects to explain predictions and profile users. This advanced method has proven to surpass contemporary techniques such as DeepCoNN, HFT, NARRE, and TransRev on not one, but two genuine datasets. It also substantiates the effectiveness of aspect embeddings in fields other than recommendation systems	Our study explores the use of NLP in sentiment analysis by utilizing the BERT encoder model to predict numerical ratings from textual product reviews
[18]	This study identifies the most practical algorithms for sentiment analysis of product reviews, such as Random Forests and XGBoost. However, with the help of explainable AI techniques, the predictions of certain models were found to be faulty and biased. The study warns people against over-dependency on ML algorithms	By utilising the BERT model, our research highlights the potential of natural language processing techniques in the text classification and sentiment analysis by predicting accurate numerical ratings from product reviews

(*continued*)

Table 1. (*continued*)

Paper	Description	Remarks
[19]	This study used an unsupervised learning algorithm, Latent Dirichlet Allocation, to provide labels to product characteristics. These labels are later combined to predict the numerical ratings. The proposed model shows advantageous outcomes on a small review dataset	Our study demonstrates the effectiveness of natural language processing techniques, particularly the BERT model, in predicting product ratings from textual data and enhancing the accuracy of consumer feedback for enterprises
[20]	This paper addresses the need for text mining and analysis in the era of vast digital media. It proposes a unique approach for opinion mining of reviews given by users using n-gram classification and POS tagging, followed by machine learning algorithms. The proposed method shows improved accuracy and promising outcomes	Employing the BERT model, our research work emphasizes the potential of natural language processing techniques in extracting insights from textual data and enabling businesses to make datadriven decisions based on consumer feedback

3 Methodology

The approach included utilizing the "Flipkart Products Review" dataset, taken from Kaggle, as the source for reviews and ratings on which our encoder models could be tested. Data preprocessing involved removing redundant columns and missing values, eliminating punctuation marks and undefined characters, and filtering out common and redundant words using the NLTK library. A word cloud was generated from the final list of words contained in the 'Summary' reviews to select the most relevant features for the Encoder model. The data modeling stage of the research paper involved loading and preprocessing the Flipkart Reviews dataset, developing an artificial intelligence model utilizing a pre-existing BERT model from the Hugging Face Transformers repository, and assessing its efficacy in foretelling predicting on the test set.

3.1 Data Exploration

The "Flipkart Products Review" dataset on Kaggle, comprises 194276 rows of data and 5 columns, containing valuable information about 104 distinct product types available on Flipkart.com. The columns consist of "ProductName," a string datatype storing the product name, "Price," an integer datatype indicating the price in the designated currency, and "Rate," an integer datatype for customer rating on a scale of 1 to 5. The "Review" column is a string datatype with brief customer feedback, and the "Summary" column is another string datatype that provides a detailed review of the product review, encompassing specifications, features, and user experience.

3.2 Data Preprocessing

In the data preprocessing stage, four major stages were undertaken which included data reduction, removal of punctuation, and elimination of undefined characters and unnecessary words.

Data Reduction
Under data reduction, nonessential rows and columns were dropped from the dataset. These were rows and columns which did not provide any information which was meaningful for analysis. The columns dropped were – 'Price,' 'ProductName,' and 'Review.' This was done because the product's price, name, and short-worded review were not valuable to the task at hand. Our primary focus was on the column containing the actual product review, the 'Summary' column, and its allocated Rating.

Ensuring our working data set does not have incomplete data is essential. Hence, next, the dataset was evaluated for missing values using the 'isnull' function. Missing values in the 'Summary' and 'Rate' columns were eliminated from the data set using the 'dropna' columns. This helps remove any probable errors and inaccuracies in the modeling of data during mathematical modeling or statistical calculations.

Removing Punctuation
Next, all punctuation marks were removed from the 'Summary' column. This was done because punctuation marks, like commas, semicolons, and full stops, only serve as grammatical devices for better sentence structure. Even without them, the essence of the text is still preserved. Our data frame must only contain essential and relevant information. Hence, these extraneous symbols were removed.

Undefined Characters
Next, undefined values were removed from the dataset. This was done using the Regular Expressions library, which can differentiate between alphanumeric characters and whitespaces. Subsequently, a list of distinct constituents in the 'Rate' column of the dataset was obtained.

This step proved critical in ascertaining any undefined values that were not part of the pre-defined set of values, which included solely numeric ratings of integers ranging from 1 to 5. Thus, only those rows which met the set condition were retained. This was crucial in ensuring that the study did not encompass any invalid values that study could endanger the analysis's accuracy and integrity, thus preventing biased predictions.

Unnecessary Words
The Natural Language Toolkit (NLTK) library filters out redundant and commonly used words in the 'Summary' reviews data frame called Stopwords. Stopwords serve as supportive words in a sentence without carrying any distinct sentiment and, hence, can be removed. First, the most commonly used words in the dataset were identified, such as "bad," "good," "Flipkart," "also," "like," "item," "one," "much," "even," and "product." These words and the stopwords were then removed from the dataset using the NLTK library.

Subsequently, a word cloud was generated from the final list of words in the 'Summary' reviews. The resulting word cloud, as shown in Fig. 1 provided an exhaustive and

graphically clear representation of the words most commonly used by customers in their reviews. Words such as “good”, “products”, “cooler”, “nice”, “best”, “flipkart”, “water”, “air” were seen to be most frequent in reviews, as per the word cloud. These words were crucial in selecting the most relevant features to be used in the Encoder model.

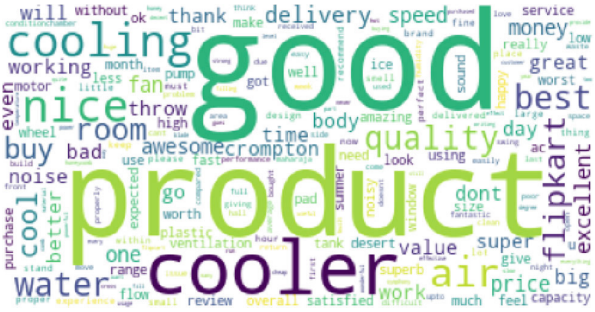


Fig. 1. Wordcloud of Content of the ‘Reviews’ Attribute

4 Modelling

The task at hand involves deciphering consumer sentiment from online product reviews. These reviews are generally heavily textual, unstructured, and varied. BERT, which stands for Bidirectional Encoder Representations from Transformers, is renowned for its proficiency in natural language understanding and processing. It can be successfully employed here due to its ability to capture the context from large textual corpora. Its bidirectional nature ensures that it considers both the previous and following word in a sentence while processing a word, thus fully understanding the contextual implications of the word in the given sentence. Hence it can be highly useful for vast unstructured data such as ours. Another benefit of using BERT here is, the presence of irony and sarcasm in reviews, which can be accurately handled by BERT’s sentiment analysis. The textual data was then prepared for analysis with BERT.

For better understanding, first, the columns in the dataset were renamed, with the ‘Summary’ and ‘Rate’ columns renamed as ‘text’ and ‘label,’ respectively. Next, an analysis of the dataset’s distribution was done, tabulating the number of reviews allocated to each rating. A pie chart was then generated using the Matplotlib library to illustrate the percentage distribution of reviews amongst the various rating categories. As seen in Fig. 2, it was observed that more than half of the reviews, 57.3% were of the rating 5. The next most frequent were reviews of rating 4, being 20.9% of the whole. The data was then prepared for analysis with BERT.

The data frame underwent a partition process, dividing it into two subsets for training and testing purposes. The testing set comprises 20% of the dataset, while the remaining 80% is allocated to the training set. These were then converted into data sets using the Hugging Face Datasets library. A pre-trained BERT model was downloaded, and the necessary functions for tokenization were installed. The text data was preprocessed

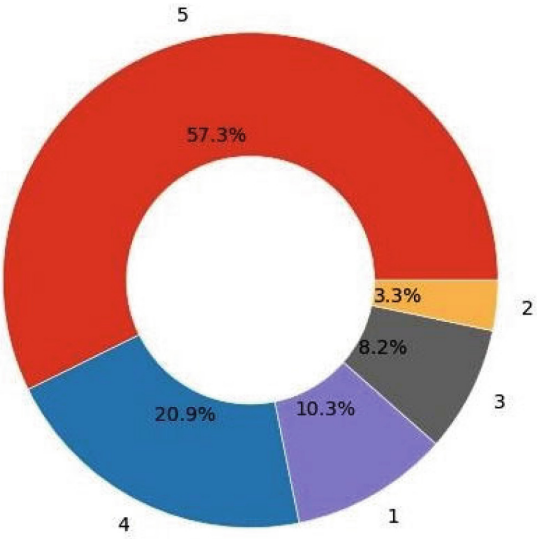


Fig. 2. Percentage Distribution of Reviews Amongst Ratings

and transformed into numeric data using the BERT tokenizer, which is a WordPiece tokenizer. Then, the tokens were encoded using the BERT vocabulary, using the BERT encode() function. The BERT model was then built using the build model() function, which is a TensorFlow 2.0 model.

This function added layers for classification, including a Dense layer with 64 units and a ReLU activation function, a Dropout layer with a rate of 0.2, another Dense layer with 32 units and a ReLU activation function, a second Dropout layer with a rate of 0.2, and a final Dense layer with 5 units and a softmax activation function.

The Adam Optimiser (of learning rate 2e−5) and a categorical cross-entropy function were executed to compile the model. The model's prediction accuracy was decided to be used as the comparison metric. The Hugging Face Transformers library's trainer() function enabled the model's training process. Training arguments, such as the duration of the training process in terms of epochs, the size of each batch, and the learning rate were selected. Finally, the model was evaluated against the testing data using the BERT tokeniser by encoding it using the BERT vocabulary. The results showed that the BERT had helped the composition of a vigorous and accurate text classification mode.

5 Results

The BERT model applied to predict ratings from Flipkart reviews demonstrated a commendable accuracy of 73%. This suggests that our model can extract relevant textual information from the Flipkart reviews and make accurate predictions about the corresponding ratings.

The success of our model can be attributed to its robustness in handling the vast varieties of reviews available within the dataset. Different people, coming from various

backgrounds have very different writing styles. Hence, our reviews are very diverse. There are variations in grammar, syntax, spelling, and language styles. There also might be ambiguous words, slang, sarcasm, or irony in reviews. Even with such diversity, our model's capability to navigate such varied language styles is a testament to its robustness and adaptability to make precise and accurate predictions. For better predictions, further study is required to improve the model's capacity to handle such diverse and context-specific usage of words, which would further improve accuracy. This level of precision has practical implementations. Businesses looking to understand their customer base and make improvements based on customer feedback can depend on this model's capabilities. Even customers looking to purchase high quality products can make informed decisions. They can rely on other customers' reviews and experiences to guide their purchasing decisions.

6 Conclusion and Future Scope

This study has delved deeper into the correlation of product reviews with allied ratings. The possibilities in this line of work are extensive and multifaceted. This study contributes to a more exhaustive understanding of consumer sentiment by converting complex textual reviews to numerical ratings. With an accuracy of 73%, the model's performance is satisfactory and provides a starting point for further expansion and investigation on this topic and the techniques utilized. By analyzing consumer reviews, businesses can better understand their customer base and, tailor their products and services accordingly, and make more informed decisions. The BERT model presented here demonstrated remarkable accuracy. However, there are still several areas that can be explored further:

- Integration of other features into the prediction system, such as product categories, prices, and images.
- Encompass the domain of multilingual sentiment analysis.
- Exploring and comparing the efficacy of novel deep learning models such as LSTM and CNN.

References

1. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018). arXiv preprint arXiv:1810.04805
2. Jhaveri, Y., Gandhi, T., Naik, T., Nisar, S., Sonawane, P.: Predicting doctor ratings from user reviews using deep learning. Int. Conf. Appl. Artif. Intel. Comp. (ICAAIC) **2022**, 641–648 (2022)
3. Akinlaja, O., Mosia, M.: Using deep learning and sentiment analysis to identify mismatches between online courses' reviews and ratings. Int. Multidisciplinary Info. Technol. Eng. Conf. (IMITEC) **2021**, 1–6 (2021)
4. Kusunoki, Z., Kamei, S., Morimoto, Y.: Overall rating prediction from review texts using category-oriented japanese sentiment polarity dictionary. Tenth Int. Sympos. Comp. Netw. (CANDAR) **2022**, 124–129 (2022)
5. Gezici, B., B¨olu¨cu¨, N., Tarhan, A., Can, B.: Neural sentiment analysis of user reviews to predict user ratings. 4th International Conference on Computer Science and Engineering (UBMK) 2019, pp. 629–634 (2019)

6. Liu, Y., Shen, Y.: Personal tastes vs. fashion trends: predicting ratings based on visual appearances and reviews. IEEE Access **6**, 16655–16664 (2018)
7. Xu, Y., et al.: Exploiting the sentimental bias between ratings and reviews for enhancing recommendation. ieee International Conference on Data Mining (icdm) 2018, pp. 1356–1361 (2018)
8. Shiju, A., He, Z.: Classifying drug ratings using user reviews with transformer based language models. IEEE 10th International Conference on Healthcare Informatics (ICHI) 2022, pp. 163–169 (2022)
9. Zhang, H., et al.: Dynamical rating prediction with topic words of reviews: a hierarchical analysis approach. Collaborative Computing: Networking, Applications and Worksharing: 15th EAI International Conference, Collaborate-Com 2019, London, UK, August 19–22, 2019, Proceedings 15, pp. 780–798 (2019)
10. Kavousi, M., Saadatmand, S.: Estimating the rating of the reviews based on the text. Data Analytics and Learning: Proceedings of DAL **2018**(2019), 257–267 (2019)
11. Raju, K.V., Sridhar, M.: Based sentiment prediction of rating using natural language processing sentence-level sentiment analysis with bag-of-words approach. First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI **2019**(2020), 807–821 (2020)
12. Carneiro, J., Meira, J., Novais, P., Marreiros, G.: Using machine learning to predict the users ratings on tripadvisor based on their reviews. Highlights in Practical Applications of Agents, Multi-Agent Systems, and Social Good. The PAAMS Collection: International Workshops of PAAMS 2021, Salamanca, Spain, October 6–9, 2021, Proceedings 19, pp. 127–138 (2021)
13. Saumya, S., Singh, J.P., Kumar, A.: A machine learning model for review rating inconsistency in e-commerce websites. Data Management, Analytics and Innovation: Proceedings of ICDMAI 2020, Vol. 1 2021, pp. 221–230 (2021)
14. Lu, Y., Zhang, W., Lu, P., Wang, J.: A novel hybrid sequential model for review based rating prediction. Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14–17, 2019, Proceedings, Part I 23, pp. 148–159 (2019)
15. Wang, X., Xiao, T., Shao, J.: Emrm: enhanced multi-source review-based model for rating prediction. Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part III 14, pp. 487–499 (2021)
16. Polpinij, J., Luaphol, B.: Comparing of multi-class text classification methods for automatic ratings of consumer reviews. Multi-disciplinary Trends in Artificial Intelligence: 14th International Conference, MIWAI 2021, Virtual Event, July 2–3, 2021, Proceedings, pp. 164–175 (2021)
17. Nikolenko, S.I., Tutubalina, E., Malykh, V., Shenbin, I., Alekseev, A.: Aspera: aspect-based rating prediction model. Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part II 41, pp. 163–171 (2019)
18. So, C.: What emotions make one or five stars? understanding ratings of online product reviews by sentiment analysis and xai. Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22, pp. 412–421 (2020)

19. VRPS, S.Y., Seshadri, K.: A framework for predicting item ratings based on aspect level sentiment analysis. International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) 2021, pp. 327–332 (2021)
20. Shah, A., Kothari, K., Thakkar, U., Khara, S.: User review classification and star rating prediction by sentimental analysis and machine learning classifiers. Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018 2020, pp. 279–288 (2020)

Diagnosis of Alzheimer Disease Using Machine Learning Algorithms

Tanya Kumari[1], Ritika Kumari[1,2(✉)], and Poonam Bansal[1]

[1] Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, India
{tanya147btcsai22,ritikakumari,poonambansal}@igdtuw.ac.in
[2] USICT, Guru Gobind Singh Indraprastha University, New Delhi, India

Abstract. Alzheimer's disease (AD) is a prominent nervous ailment. Even though its symptoms are mild at first, they get worse with time. To foretell the onset of AD utilizing psychological criteria such as age, gender, memory, Clinical Dementia Rating (CDR), and Mini-Mental State Examination (MMSE) scores, this study uses machine learning algorithms (MLAs): Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Support vector machine (SVM), and K closest neighbor (KNN) to identify AD in an Alzheimer's clinical dataset from the Kaggle Repository. We use five performance metrics for a comparative study: Accuracy, Precision, Recall, F1score, and AUC.

Keywords: Alzheimer's Disease · dementia · machine learning · performance metrics

1 Introduction

AD can be driven by both inherited and environmental triggers, both of which impact an individual's brain over time. It is an irreparable brain ailment that slowly deteriorates productivity, memory retention, and reasoning abilities [1]. AD is an illness that causes memory and other essential mental skills to deteriorate over the years. Initially, someone with AD may experience minor disorientation and difficulties remembering things. In today's world, AD is the most prevalent form of dementia [2]. AD can be identified primarily by memory loss as well as other cognitive dysfunctions which can deteriorate to the point of interfering with everyday functioning. Of dementia cases, it accounts for 60 to 80 % [3]. AD is growing more common in later life, with people over the age of 65 being particularly vulnerable. When AD advances, patients at various stages must be treated accordingly. To achieve this, early detection and classification of AD stages can be extremely beneficial in treating the disease's symptoms [4]. This disease gradually destroys brain tissue. Individuals with this disease survive for approximately 9 years, and one out of every eight individuals aged 65 and over possesses it. [5]. An increased risk of AD has been associated with diabetes, homocysteine, hypertension, heart disease, smoking, and obesity [6]. The primary goal of this study is to find the most effective way for detecting AD. In this work, we classify AD using a dataset from the Kaggle repository with the help of MLAs including NB, DT, RF, SVM, and KNN. The best performers are DT and RF, according to our findings.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 191–200, 2025.
https://doi.org/10.1007/978-3-031-75167-7_15

The following are the remainder sections of this paper: Sect. 2 includes a literature survey. Information about the classifiers under consideration are given in Sect. 3. Sections 4 and 5 go over the performance metrics and dataset description. The proposed methodology is explained in Sect. 6. Our experiment along with its findings are described in Sect. 7, while Sect. 8 concludes.

2 Literature Survey

In the literature, several studies conducted for Alzheimer's detection and other domains using machine learning (ML) are discussed below (Table 1):

Morshedul Bari Antor et al. [1] conducted the study using MLAs: SVM, Logistic regression (LR), DT and RF on "MRI and Alzheimer's" dataset (provided by the OASIS project) and it yielded the highest accuracy with SVM model i.e. 92%.

P. Lodha, A. Talele and K. Degaonkar [2] analysed the study using MLAs: SVM, Gradient Boosting Algorithm, Neural Network (NN), KNN, RF and achieved highest accuracy of 98.36% with NN.

Randa Ben Ammar and Yassine Ben Ayed [3] investigated speech processing using MLAs in order to identify AD beforehand. This work uses ML models to infer the presence of AD based on lexical indicators of language impairment that were collected from the spoken language of individuals with and without AD.

Muhammad Shahbaz et al. [4] investigated the classification of (AD) using MLAs and data mining and deep learning algorithms on the ADNI dataset. According to the analysis of the results, the GLM beats the remaining classifiers and achieves an accuracy of 88.24%.

J. Neelaveni and M.S. Geetha Devasana [5] used MLAs to predict AD and attained an accuracy of 85% using SVM.

Sandhya Joshi et al. [6] researched the identification of Parkinson's and AD with the help of ML and NN methods. They used test cases to validate their classification model, and on average, their model exhibited a 99.25% classification accuracy.

Ritika Kumari et al. [7] implemented ML approach for preliminary identification of PCOS in women using six MLAs: LR, RF, NB, SVM, DT, and Ada Boost. They were compared using five performance metrics. RF yields the best accuracy of 89.57%.

Suman Raj et al. [8] identified autistic spectrum disorder using various ML and deep learning methods. Convolutional NN-based prediction algorithms outperformed other models on every dataset, with a maximum accuracy of 99.53%, according to the results.

Deepika Bansal et al. [9] used four MLAs for determining dementia: J48, NB, RF, and Multilayer Perceptron, and the findings clearly showed that J48 is the best classifier to detect dementia among them all.

Table 1. Literature Survey.

Reference Number	Title of the paper	Methods used	Best Model
[1]	"A comparative analysis of machine learning algorithms to predict Alzheimer's disease"	SVM, LR, DT and RF	SVM with 92% accuracy
[2]	"Diagnosis of Alzheimer's Disease using Machine Learning"	SVM, Gradient Boosting Algorithm, NN, KNN, RF	Neurak Networks with 98.36% accuracy
[3]	"Speech Processing for Early Alzheimer Disease Diagnosis: Machine Learning Based Approach"	SVM, NN and DTusing feature selection methods where needed	SVM performed best with (79%) and without (60%) Feature selection methods
[4]	"Classification of Alzheimer's Disease using Machine Learning Techniques"	DT, KNN, NB, GLM	GLM performed the best with an accuracy of 88.24%
[5]	"Alzheimer Disease Prediction using Machine Learning Algorithms"	SVM and DT	SVM with 85% accuracy
[6]	"Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods"	NN and ML methods	99.25% classification accuracy
[7]	"SmS: SMOTE-Stacked Hybrid Model for Diagnosis of Polycystic Ovary Syndrome using Feature Selection Method"	LR, RF, NB, SVM, DT, and AdaBoost	RF with 89.57% accuracy
[8]	"Analysis and Detection of Autism Spectrum Disorder using Machine Learning Techniques"	LR, SVM, NB, KNN, ANN, CNN	CNN with 99.53% accuracy

(*continued*)

Table 1. (*continued*)

Reference Number	Title of the paper	Methods used	Best Model
[9]	"Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia"	J48, NB, RF, and Multilayer Perceptron	J48 with 99.52% accuracy
[10]	Optimum HMM combined with vector quantization for Hindi Speech Word Recognition	Hidden Markov Model, Vector Quantization, Linear Predictive Cepstral (LPC) coefficients, Mel-Frequency Cepstral (MFC) coefficients and delta MFCC	MFC coefficients outperforms others

3 Classification Models

We use five MLAs in this study which have been described below:

Naïve Bayes: It is a supervised MLA that calculates the conditional probability using the Bayes theorem. It discusses the likelihood of an event happening in light of an already completed occurrence. The likelihood is used to generate the prediction. [7, 9].

Decision Tree: It is one of the MLAs that makes predictions using a tree-like topology. Each branch represents a potential solution. In an effective DT, splits reduce the unpredictability of the data and increase the clarity of the information received. [1, 4, 5, 7].

Random Forest: This supervised MLA is the one that employs collaborative learning, which is the technique of merging various classifiers to take on a challenging task and increase the model performance. It is made up of several decision trees that collaborate to make a single prediction. Final model predictionis given to the class with the highest votes. By gathering several individual forecasts before generating a final prediction, this strategy reduces mistakes. [1, 2, 6, 7, 9].

KNN: This refers to a supervised, non-parameterized MLA that makes predictions according to neighbor familiarity. Becausethe principle of similarity relies on the value of K, cases are categorized based on the majority of votes cast by their neighbors. This happens when similar cases are close together. It is used when the dataset is small, marked, and noise-free [2, 4].

SVM: It is a guided MLA that can address classification and regression issues. After labelling the training data, the algorithm determines the best possible hyperplane for categorizing the unknown data [1–3, 5, 7].

4 Dataset

We use the dataset "Alzheimer's Clinical Data" [11] from the Kaggle repository. It consists of 1230 instances with 846 people detected with dementia, 367 people detected with uncertain dementia and 17 people detected with no dementia. It should be noted that this dataset has been simplified into a binary classification, where individuals are grouped into two categories: those with "AD dementia", designated as 1 and those with "No dementia" and "Uncertain dementia", designated as 0. It has 7 columns. Table 2 provides a thorough description for the dataset used.

Table 2. Description of the dataset.

Dataset Name	Number of instances	With AD (1)	Without AD (0)
Alzheimer's Clinical Data	1230	846	384

5 Performance Metrics

Here, we analyze the model efficiency using five metrics, which are explained below [7]:

Accuracy. It is among the simplest categorization metrics to use. It refers to the quantity of accurate forecasts made by the MLAs as indicated by formula (1):

$$\text{Accuracy} = \text{TP} + \text{NP}/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \tag{1}$$

Precision. It's use is to overcome the limitation of Accuracy (in circumstances of imbalanced datasets). It determines the fraction of positive predictions that were actually correct and is computed as indicated by formula (2).

$$\text{Precision} = \text{TP}/(\text{FP} + \text{TP}) \tag{2}$$

Recall. Comparable to Precision metric, it aims to ascertain the percentage of true positives that were mistakenly identified. It can be computed as True Positive (predictions that are truly true to the total number of positives) or False Negative (negative predictions that are mistakenly forecasted) and is computed as indicated by formula(3).

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{3}$$

AUC. This is obtained using formula (4) and is known as the region beneath the Receiver Operating Characteristic (ROC) curve.

$$\text{AUC} = (\text{TP rate} + \text{TN rate})/2 \tag{4}$$

F1 score. It combines precision and recall as indicated by formula (5).

$$\text{F1 score} = (\text{Precision} * \text{recall}) * 2/(\text{Precision} + \text{Recall}) \tag{5}$$

TP = true positives, TN = true negatives, FN = false negatives, FP = false positives.

6 Proposed Methodology

1. Feed the Dataset.
2. Preprocess the Dataset.
   a. Label Encoding.
   b. Addressing null values.
3. Train the dataset.
4. Apply the MLAs.
5. Evaluation of various Performance Metrics.
6. Best Model (Fig. 1).

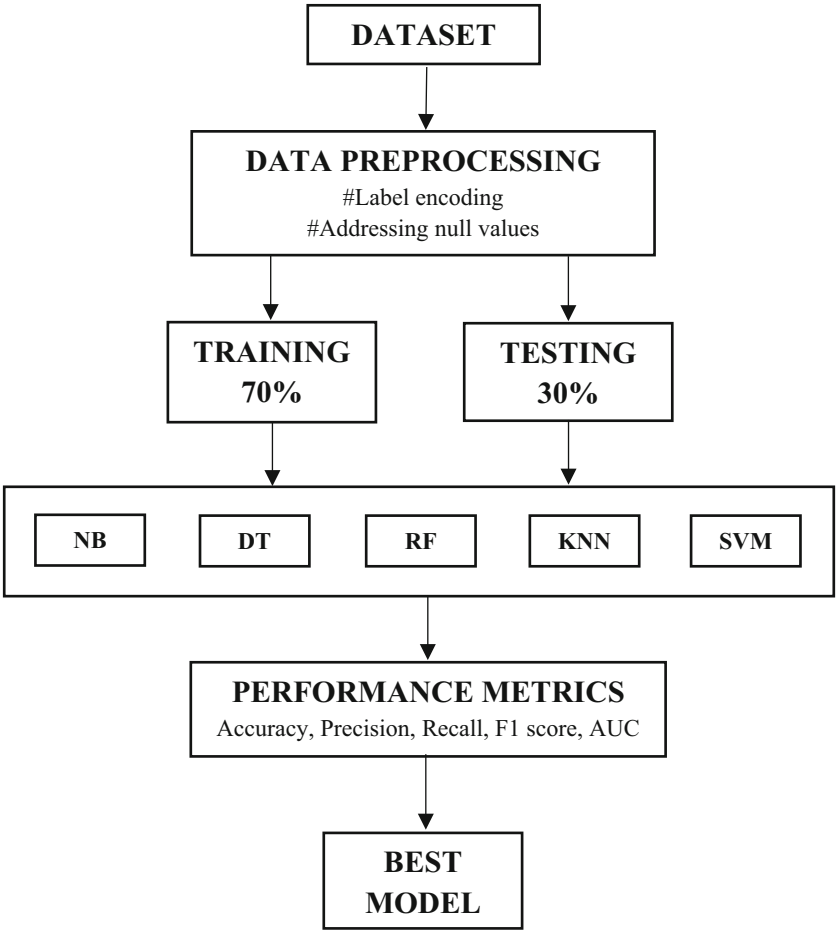


Fig. 1. Our Proposed Methodology

7 Experiments and Results

We are currently experimenting with five MLAs on the Alzheimer's clinical dataset obtained from the Kaggle repository: NB, DT, RF, KNN and SVM. Models are compared using different performance metrics as shown below in Table 3 and the pictorial representation of our results are expressed in Figs. 2, 3, 4, 5 and 6:

Table 3. The MLAs' performance.

Classifiers	Accuracy	Precision	Recall	F1 Score	AUC
NB	71.00%	73.52%	77.03%	70.50%	77.03%
DT	76.96%	73.51%	75.18%	74.14%	75.18%
RF	76.69%	72.81%	72.36%	72.58%	72.36%
KNN	70.73%	65.89%	65.89%	65.89%	65.89%
SVM	76.15%	72.15%	71.49%	71.80%	71.49%

Decision Tree and Random Forest are best performing models giving an accuracy of 76.96% and 76.69% respectively whereas KNN performed the least giving an accuracy of 70.73% as shown in table above. The interpretability of Decision Trees is well established. They provide simple guidelines for making forecasts, making it easy to comprehend why a specific forecast was made. This is especially important in situations where interpretation and explanation are crucial. RF is an ensemble-based learning method for improving accuracy through the integration of numerous Decision Trees. It often results in a more robust and accurate model by pooling the predictions of several

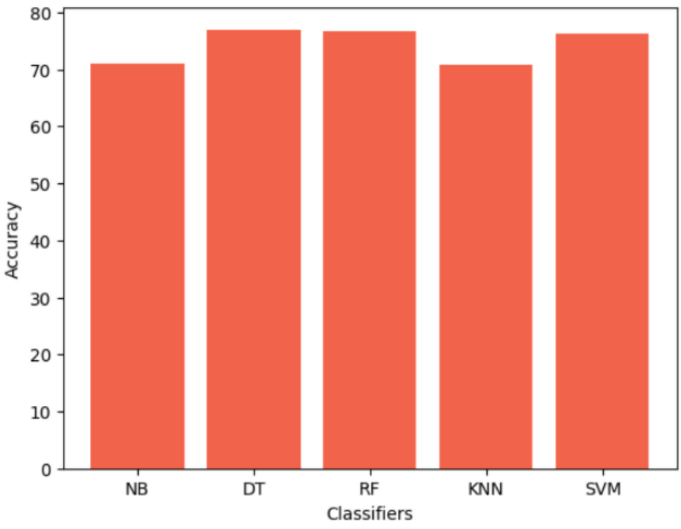


Fig. 2. Several MLAs' Accuracy performance

trees. Random Forest's combining nature aids in reducing overfitting, making it one of the best performers across this dataset.

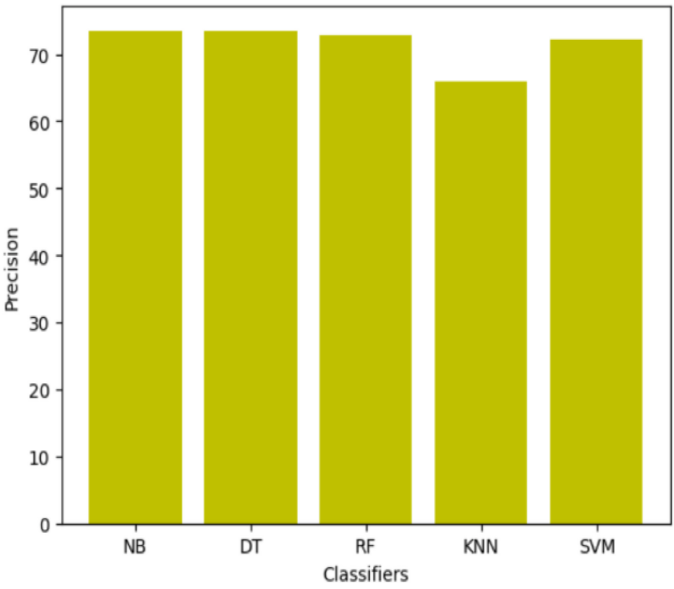


Fig. 3. Several MLAs' Precision performance

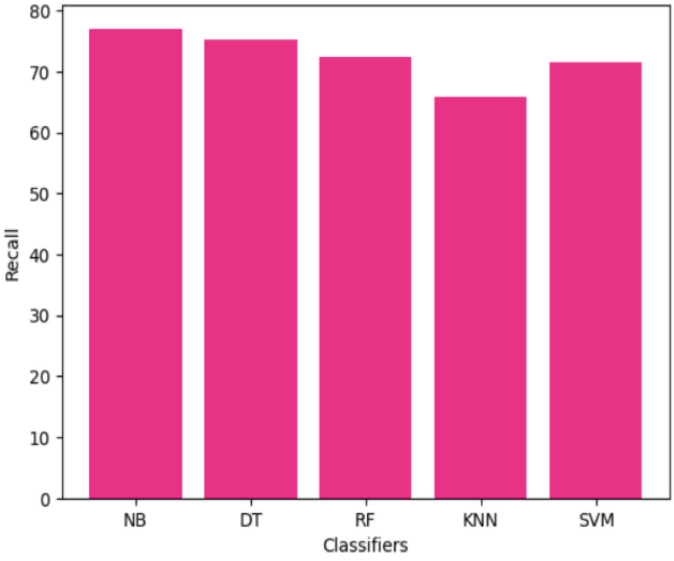


Fig. 4. Several MLAs' Recall performance

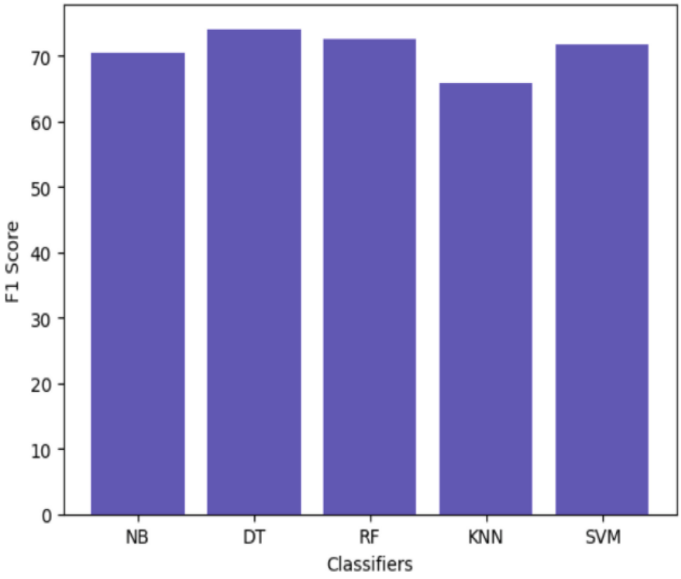


Fig. 5. Several MLAs' F1 Score performance

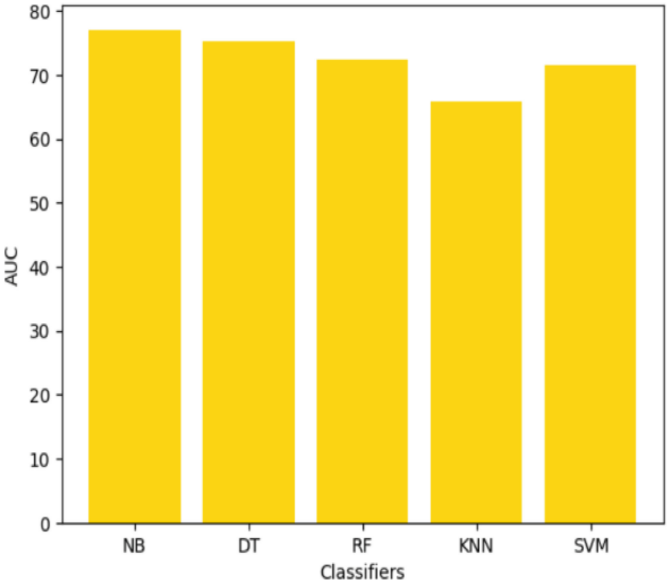


Fig. 6. Several MLAs' AUC performance

8 Conclusion

In this study, we used five MLAs (Naïve Bayes, Decision Tree, Random Forest, KNN, and SVM) to detect Alzheimer's on the "Alzheimer's Clinical Data" dataset. The following five performance metrics were applied: F1 Score, Accuracy, Precision, Recall, and AUC. We concluded that the best-performing algorithms were Decision Tree and Random Forest, exhibiting accuracy rates of 76.96% and 76.69%, respectively. In the future, we'll work with additional relevant datasets and carry out the investigations using various MLAs.

Acknowledgement. The authors are grateful to Indira Gandhi Delhi Technical University for Women for providing them with this opportunity to conduct this research.

References

1. Bari Antor, M., et al.: A comparative analysis of machine learning algorithms to predict alzheimer's disease. Journal of Healthcare Engineering (2021)
2. Lodha, P., Talele, A., Degaonkar, K.: Diagnosis of alzheimer's disease using machine learning. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA), pp. 1–4. IEEE (2018)
3. Ammar, R.B., Ayed, Y.B.: Speech processing for early Alzheimer disease diagnosis: machine learning based approach. In: 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), pp. 1–8. IEEE (2018)
4. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., Umer, A.: Classification of Alzheimer's Disease using Machine Learning Techniques. In: Data, pp. 296–303 (2019)
5. Neelaveni, J., Devasana, M.G.: Alzheimer disease prediction using machine learning algorithms. In: 2020 6th international conference on advanced computing and communication systems (ICACCS), pp. 101–104. IEEE (2020)
6. Joshi, S., et al.: Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. In: 2010 Second International Conference on Machine Learning and Computing, pp. 218–222. IEEE (2010)
7. Kumari, R., Singh, J., Gosain, A.: SmS: SMOTE-stacked hybrid model for diagnosis of polycystic ovary syndrome using feature selection method. Expert Syst. Appl. **225**, 120102 (2023)
8. Raj, S., Masood, S.: Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Comp. Sci. **167**, 994–1004 (2020)
9. Bansal, D., Chhikara, R., Khanna, K., Gupta, P.: Comparative analysis of various machine learning algorithms for detecting dementia. Procedia Comp. Sci. **132**, 1497–1502 (2018)
10. Bansal, P., Dev, A., Jain, S.B.: Optimum HMM combined with vector quantization for hindi speech word recognition. IETE J. Res. **54**(4), 239–324 (2008)
11. Kaggle Repository for Alzheimer's Clinical Dataset, https://www.kaggle.com/datasets/legendahmed/alzheimers-clinical-data?resource=download. Last accessed on 6 June 2023

Detecting Image-Based Fake News with Neural Sleuths

Nachiket Rathod(✉) and Prabhakar Ramteke

Department of Computer Science and Engineering, H.V.P. Mandal's College of Engineering and Technology, Amaravati, Maharashtra, India
rathod.nachiket@gmail.com

Abstract. In the period of extensive social media use, unconventional formats for news sharing often lack reliability, with misleading information disseminated through graphics emerging as a unique tactic. While the verification of the authenticity of news is a broad research domain, scholars have mainly focused on textual data. The misleading information detection model described in this paper uses neural sleuth-based networks to capture image data from the freely accessible Fakeddit dataset. The results are scrutinized using a confusion matrix, and a comprehensive evaluation of the model's performance metrics is conducted across six categories: "true," "satire," "false connection," "imposter content," "manipulated content," and "misleading content." This work contributes an inclusive analysis of the model's effectiveness as a problem posed by misleading information on social media.

Keywords: Deep Network · Image Fake News · Convolution · Neural Network

1 Introduction

In recent times, the wide adoption of the internet has led to an exponential increase in its use in the daily lives of humans. In India, the number of social media users reached 518 million in 2020, thanks to the country's widespread internet availability. Furthermore, it is expected that by 2040, the country will have about 1.5 billion users on social networks [1]. Social media platforms have become virtual hubs for connecting with friends, family, and society, offering an effective means of remote global communication. People are engaging these platforms for a wide range of purposes, including social work, political movements, public relations, dating, business, advertising, and news consumption. Traditional forms of news dissemination, such as newspapers and magazines, are rapidly being replaced by digital platforms and social media for news sharing.

The majority of individuals now turn to social media platforms like Twitter, Facebook, and others for their news consumption. To expand their audience, news agencies share news in the form of images on these platforms. Such image-based news spreads quickly on social media. In the age of technology, a significant number of people have become victims of photo fraud. Some individuals exploit software to create and circulate images containing misleading news captions to advance their agendas. Morphed

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 201–211, 2025.
https://doi.org/10.1007/978-3-031-75167-7_16

images, where one person's face is replaced with another's, are a prime example of fake images. The credibility of news is a pressing issue that needs to be addressed, as often images shared via social media are found to be fake or misleading. This can have serious consequences for individual integrity, organizations reputations, and public unrest. This anatomy is reflected in Fig. 1.

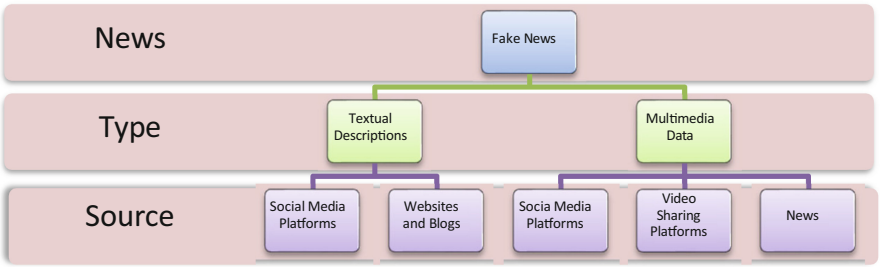


Fig. 1. Anatomy of Fake News

Human fact-checking is a labor-intensive and expensive process. Consequently, there is an urgent need to create technology that can autonomously distinguish and classify news as genuine or fraudulent. This can potentially be accomplished by utilizing artificial intelligence techniques, as conventional methods for detecting accuracy and extracting features have notable constraints as shown in Fig. 2.



Fig. 2. Solution for Problem of Fake News

Hence, this work proposes the implementation of NSNM: Neural Sleuth-based Network Model to automate the recognition and categorization of news articles and this is proposed to work with images. The model uses the openly accessible multimodal dataset named Fakeddit, which includes more than 1 million instances of different fake news types. Fakeddit stands out as a pioneering dataset for detecting fake news, and the examples are categorized through remote oversight, involving 2-way, 3-way, and 6-way classification categories following several rounds of evaluation [2]. The paper is structured into the mainly following sections: related work, proposed model, result, and conclusion.

2 Related Work

Within this section, an exploration of previous research endeavours in the domain of fake news detection is undertaken, placing particular emphasis on methodologies precisely adapted for the purpose of distinguishing fake news articles that rely on visual content as a pivotal component.

Rathod & Ramteke [3] explored the prevalence of multimodal data on social media platforms which is continually increasing. Consequently, noted that fake news detection models are rigorously evaluated against multiway classified multimodal datasets. The same researchers proposed neural network-based models and employed the distilBERT model for analyzing textual news obtained from multimodal datasets [4].

Masciari [5] has advocated for a sophisticated approach grounded in deep learning principles. His research not only propounds the utilization of Convolutional Neural Networks (CNNs), which are renowned for their prowess in image analysis, but also underscores the importance of a comprehensive strategy that transcends the boundaries between textual and visual content analysis. By embracing this multifaceted approach, Masciari's framework transcends conventional methodologies by seamlessly amalgamating natural language processing for textual data with multimedia analysis for visual content.

Uppada et al. [6] introduced a particularly intriguing approach to elevate the effectiveness of authenticity assessment. Their innovation resides in the advancement. of an ensemble framework that synergizes visual and textual features, thereby forging a comprehensive and robust system for discerning the authenticity of news content. Notably, the approach incorporates the remarkable accuracy achieved through fine-tuned BERT models, which excel in understanding and processing textual information, with the prowess of the Xception network, renowned for its capabilities in image analysis.

Shariah et al. [7] engineered a sophisticated model grounded in deep learning principles. This model, prominently featuring a Convolutional Neural Network (CNN), represents a cutting-edge approach in the sector of image analysis. With the CNN at its core, the model has the remarkable capability to process images as input and make accurate classifications, marking a significant stride in the development of effective systems for image-based tasks.

Steinebach et al. [8] conducted an extensive exploration into the development and practical application of the concept of image montage detection. Their study is driven by a paramount objective: to establish a mechanism that can swiftly, accurately, and with utmost precision identify instances of image montages. This innovative approach not only contributes to the area of feature detection but also serves as a critical tool in addressing the challenges posed by manipulated or misleading images frequently faced within the framework of image based fake news.

Meng et al. [9] enhanced the verification and source tracing of fake news detection, an image tagging scheme by proposed novel system. In this context, a neural network-based Dynamic Image Neural Network (DINN) has been designed to embed authentication tags and traceable tags. Additionally, models such as FAPM and DMGM have been developed to improve detection accuracy [9].

Singh et al. [10] introduced a cutting-edge deep learning model that operates under a multimodal framework, representing a significant advancement in the sphere of fake

news detection. This innovative approach leverages the strengths of two distinct but complementary technologies to comprehensively analyze both visual and textual components of news content. The utilization of the CNN model, specifically Efficient-NetB0, underscores the model's prowess in image analysis, allowing it to scrutinize images with precision and sensitivity.

Bondielli et al. [11] proposed the model as part of the campaign, focusing on the exploration of multimodality in the domain of fake news. This task had two distinct sub tasks: first sub task evaluated the effectiveness of multimodal fake news detection systems, while second sub task delved into the complex interplay between text and images, particularly how they mutually impact content interpretation when distinguishing fake news from real news. The study also suggests that the problem remains open and proposes potential directions for future research.

Zhang et al. [12] assessed four distinct measures of image-text similarity: text-based similarity, meaning based similarity, context-based similarity, and after training similarity. Text-based similarity, meaning based similarity gauge the inherent likeness between the original image and text content based on textual and image caption information. Context based similarity examines the alignment between image and text using meaningful named entities. In contrast, after training similarity investigates image-text similarity both before and after training a model.

Chen et al. [13] employed a two-step approach to enhance multi-modal fake news detection. Initially, they address psycholinguistic bias by employing causal intervention techniques to eliminate spurious correlations between text features and news labels. Subsequently, they apply counterfactual reasoning to imagine a scenario where news articles are exclusively represented by image features. This enables the estimation of the direct impact of images, thereby mitigating image-only bias. The comprehensive experiments carried out on two practical benchmark datasets within the study strongly validate the effectiveness of their framework in enhancing the detection of multi-modal fake news.

Karwa and Gupta [14] introduced a Hybrid Deep Neural Network Model that seamlessly integrates C-DSSM and Deep CNN models to address the challenge of identifying and classifying fake news. They utilized the LIAR dataset as the fundamental basis for their research. When compared to earlier studies that employed the LIAR dataset for fake news identification, the proposed model demonstrates notably superior performance. However, their work is not done on image data.

Many researchers are currently working on finding ways to detect fake news, especially on social media. Despite their advancements, there is still significant work ahead to truly enhance the capability of these methods in automatically detecting fake news. The problem is tricky because fake news is always changing, so researchers need to keep improving their techniques to stay ahead of it.

3 Neural Sleuths: Tackling Image-Based Fake News Using Deep Networks

This section gives the methodology used for classifying news articles into total six classes "true," "Satire," "false connection," "imposter content," "manipulated content," and "misleading content". Process is started with collecting a dataset for research, it is

acquiesced by the publicly available dataset. Next, a feature selection process is implemented to choose the most relevant characteristics for their model. Then, a pre-trained ResNet34 model is utilized to classify the articles within the dataset. The process involves training and testing the dataset to assess the model's accuracy and precision. Essentially, this methodology relies on conducting a series of experiments on the dataset to fine-tune the model and achieve the best possible results in identifying fake news articles that rely on images.

The main goal is to employ a set of classification algorithms to develop a robust classification model tailored for detecting fake news, with a specific emphasis on the intricacies of news identification. Additionally, Python code has been enhanced through suitable refactoring techniques to produce an optimized version. Fakeddit [2] is an innovative multimodal dataset designed for the identification of fake news, encompassing more than 1 million samples spanning various fake news categories. Following a meticulous review process, the samples are categorized to total three-way classification categories through remote supervision.

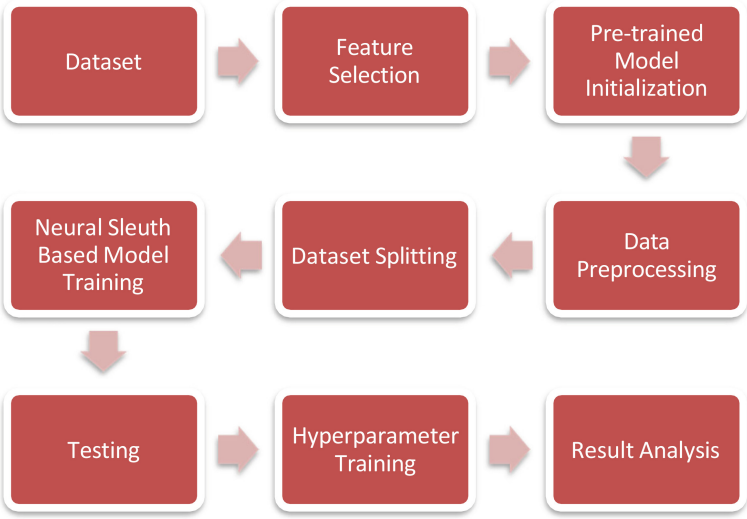


Fig. 3. Block Diagram of Tackling Image-Based Fake News Using Deep Networks

This model focuses on extracting meaningful visual features solely from Reddit-post images to facilitate the classification of Reddit-Posts using the Fakeddit benchmark dataset [2]. To achieve this, the Image-Feature Extractor leverages a pre-trained ResNet34 model [15], initially trained on the ImageNet 2012 benchmark dataset for image classification tasks. The primary objective is to attain high generalization capabilities by accurately categorizing sample posts into six predefined classes: authentic/true news content, satire/parody, content with false connections, imposter content, manipulated content, and misleading content. Among these classes, only the last four are considered as types of Fake News, while authentic/true content and satire/parody are categorized as Authentic Content in the context of a 2-way classification task.

To manage the large dataset, the train_test_split() function from Scikit-learn is utilized, aiming to create a more manageable dataset with over 700,000 samples. The entire dataset is divided into specialized training, validation, and test subsets, where only 20% of the initial data is reused, allocating 80% for training purposes and 10% each for both validation and testing.

The implementation is carried out using the PyTorch framework, where the Post-Dataset function encompasses all the pertinent information per batch, and the DataLoader function iterates over the entire dataset to provide batches of size 16 to the model. In the process, images are preprocessed and transformed to make them suitable for training. Additional data augmentation is performed by randomly cropping, resizing, and horizontally flipping the images to artificially diversify the training set, a common technique for regularization. Reddit-Post images undergo modifications, conversion into a PyTorch tensor representation, and final normalization to prepare them for model training. This is summarized in Fig. 3.

Algorithm 1: Neural Sleuth based Neural Network Model

Step 1: Loading Pre-trained ResNet34 Model

$$ResNet34 = models.resnet34(pretrained = "imagenet")$$

Step 2: Dropout Layer

$$dropout_output = dropout(input, p = 0.3)$$

Step 3: Fully Connected Layer (Linear Layer)

$$fc_image_output = Linear(in_features = 1000, out_features = num_classes, bias = True)$$

Step 4: SoftMax function

$$softmax\left(fc_{image_{output[i]}}\right) = \frac{e^{fc_{image_{output[i]}}}}{\sum_{j=1}^{num=classes} fc_{image_{output[i]}}}$$

---

4 Result and Discussion

This section contains the analysis of the results obtained from the constructed model. The dataset used in this research comprises news articles categorized into six distinct classes. The following parameters have been taken into account for the analysis of the classification model's results.

Figure 4 gives plot that explains how well an image classification model is learning during training. Training accuracy is how well the model is performing on the training data it has seen during each training cycle (epoch). Validation accuracy is how well

the model is performing on a separate set of data it has never seen before. This checks if the model is learning to generalize beyond the training data. The x-axis of the graph represents the number of training cycles, or "epochs." The y-axis represents the accuracy, which is a measure of how often the model makes correct predictions. Here, the model's performance was assessed over 20 training epochs, and the corresponding results are shown in the graph below. During this period, there was a noticeable but moderate increase in both training and validation accuracies. Specifically, by the end of the 20th epoch, the training accuracy reached 73%, while the validation accuracy stood at 71%. Figure 4 also shows that 20 epoch training history graphs is trimmed to 10 epochs for better visualization.

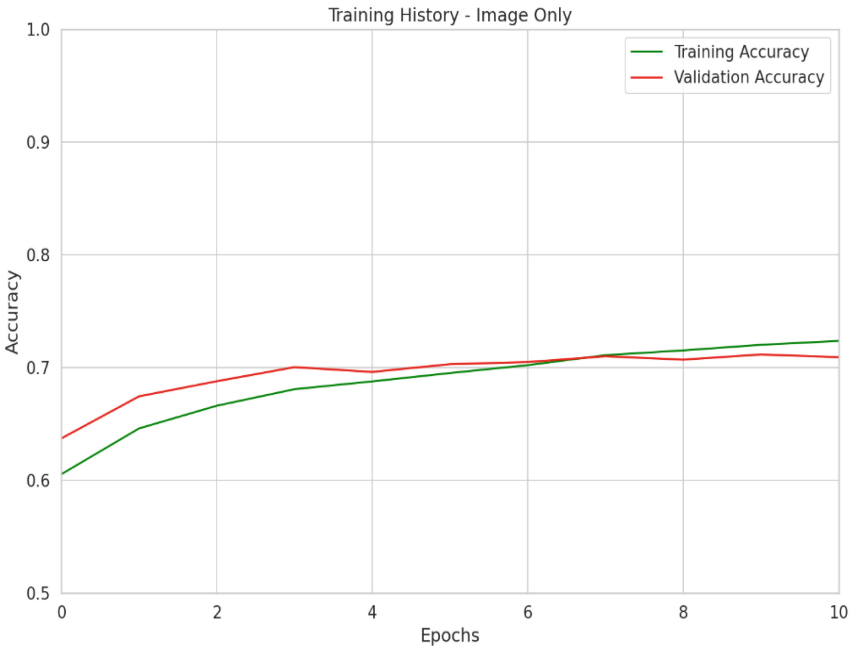


Fig. 4. Training History (Train accuracy vs Validation accuracy)

Figure 5 illustrates a graph depicting the evolution of an image classification model's performance during its training process. In simpler terms, this graph shows how well or poorly the model is doing. The Training Loss measures the model's performance on the training data throughout the training process. Similarly, the Validation Loss assesses performance but on a separate dataset known as the validation dataset. On the horizontal x-axis, you can observe the number of training cycles or "epochs," while the vertical y-axis represents the loss, specifically the "Cross Entropy Loss," which gauges how much the model's predictions deviate from the actual correct values. Over a span of 20 epochs, the training and validation losses were closely monitored. In the 20th epoch, the training loss was 1.475%, and the validation loss was 1.435%. To enhance clarity, the training history graph for 20 epochs is condensed to 10 epochs in Fig. 5.

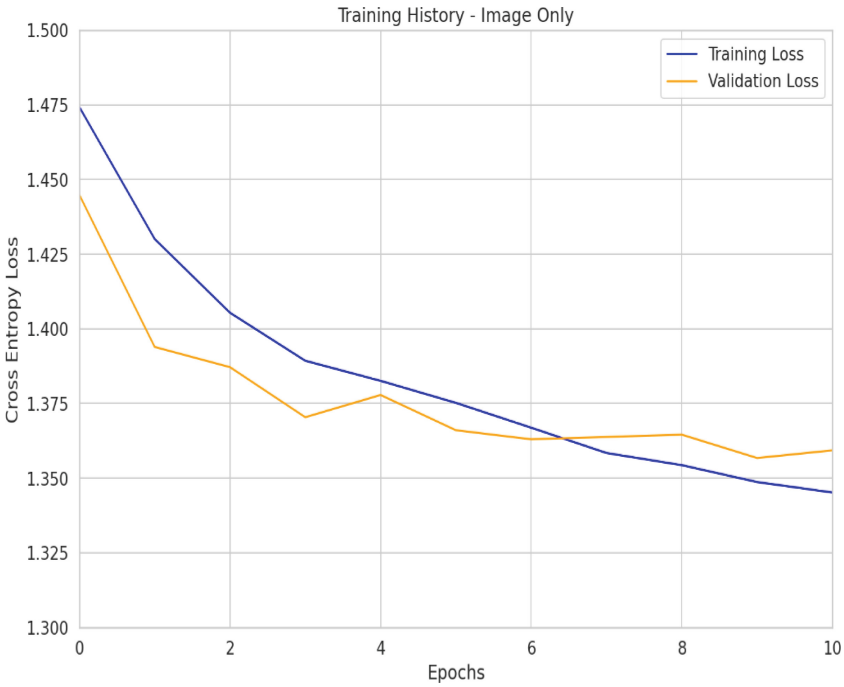


Fig. 5. Training History (Train Loss vs Validation Loss)

Classification Report provides a comprehensive assessment of the performance of a classification model. It offers a set of predefined evaluation metrics such as precision, recall, and F1-score, which are calculated for each class in the dataset. These metrics allow us to gauge the model's classification abilities not only across all classes as a whole but also on a class-by-class basis. In other words, it helps us understand how well the model is distinguishing between different categories or classes of data. Figure 6 report is a valuable tool for evaluating the model's overall effectiveness and its strengths and weaknesses in classifying specific groups within the dataset. Accuracy, Precision, Recall, and F1-Score metrics for a classification involving six distinct classes is examined. The corresponding table is displayed in the image below. The Accuracy was determined to be 0.71, while the weighted average values of Precision, Recall, and F1-Score were approximately 0.70. Figure 7 serves as a vital instrument in the model's efficacy in discerning fake news articles from image inputs. Each cell within the matrix captures critical information, detailing the count of accurately classified samples for each class while also shedding light on instances of misclassification, indicating the classes to which samples are mistakenly assigned.

The y-axis represents the actual ground truth labels for each Reddit-Post, specifically addressing research focus on identifying fake news articles through images, while the x-axis encapsulates the model's predictions. Research question revolves around the model's ability to accurately classify fake news articles from images, and the Confusion Matrix serves as a solution through which model's performance is dissected. The

Labels (6-way classification)	Precision	Recall	F1-score	Support
True	0.67	0.85	0.75	4442
Satire	0.53	0.28	0.36	670
False Connection	0.57	0.48	0.52	2144
Imposter Content	0.00	0.00	0.00	236
Manipulated Content	0.91	0.84	0.87	3357
Misleading Content	0.58	0.57	0.57	431
Accuracy			0.71	11280
Macro Average	0.54	0.50	0.51	11280
Weighted Average	0.70	0.71	0.70	11280

Fig. 6. Performance Report

central objective is to concentrate all numeric sample counts along the diagonal of the matrix, symbolizing precise classification. The intensity of colors within the matrix provides nuanced insights, with darker shades highlighting a higher percentage of correct classifications and thereby showcasing the model's proficiency in distinguishing

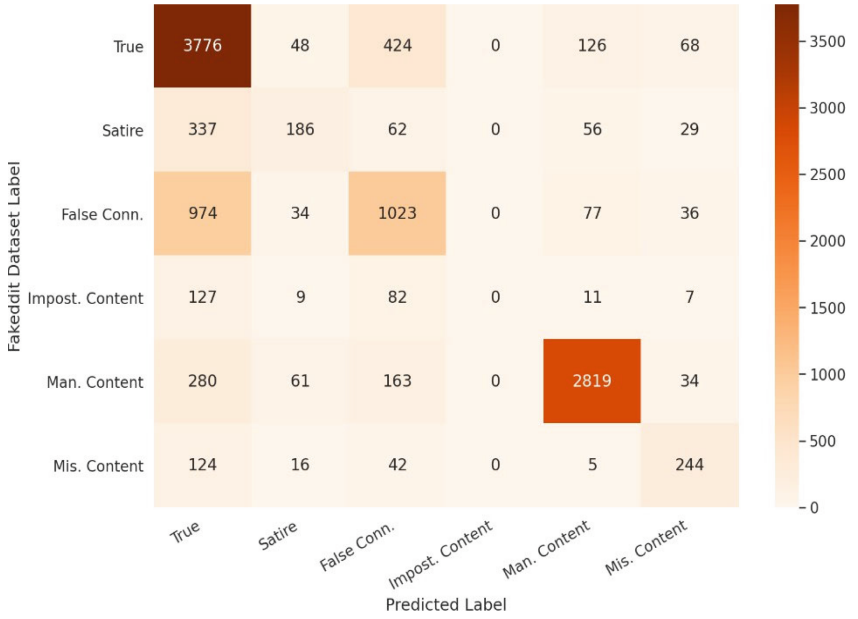


Fig. 7. Confusion Matrix

between authentic and misleading content. In the context of identifying fake news articles, this detailed interpretation of the confusion matrix is pivotal. It not only correlates with research question but also directly connects to hypotheses regarding the model's discriminatory power.

5 Conclusion and Future Scope

This paper introduced and successfully implemented a neural sleuth network-based deep learning model designed for the detection of fake news within images. Captivating the multimodal Fakeddit dataset, featuring a challenging 6-way classification task, the powerful ResNet34 model for image feature extraction is employed. Through rigorous training over 20 epochs, the model's performance is evaluated. Findings reveal that proposed model outperforms several contemporary models in this domain.

Looking ahead, there are several exciting avenues for future research in the field of authenticity assessment. Firstly, the exploration of more extensive and diverse datasets could further enhance the model's ability to detect increasingly sophisticated forms of fake news. Additionally, investigating the integration of other modalities, such as text and metadata, into a multimodal approach could yield more comprehensive and robust detection capabilities.

Furthermore, there is substantial potential for exploration in the area of constructing real-time detection systems capable of promptly recognizing and curbing the dissemination of fake news across social media platforms.

References

1. Basuroy, T.: Number of social network users in India from 2015 to 2020. https://www.statista.com/statistics/278407/number-of-social-network-usersin-india/#statisticContainer/. Access 31 July 2023
2. Nakamura, K., Yang Wang, S.L.W.: Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. arXiy, Computation and Language, 10 Nov 2019
3. Rathod, N.A., Ramteke, P.L.: Challenges and future scope in Fake news detection: a survey. AICTE Sponsored Springer International conference, REEDev (2022)
4. Rathod, N.A., Ramteke, P.L.: Results of Fake News Detection Model on Fakeddit multimodal dataset for textual data. GIS Science Journal (2023). ISSN NO: 1869-9391
5. Masciari, E., Moscato, V., Picariello, A., Sperli, G.: Detecting Fake News by Image Analysis. IDEAS 2020. Seoul, Republic of Korea (2020)
6. Uppada, S.K., Patel, P., Sivaselvan, B.: An image and text-based multimodal model for detecting fake news in OSN's. Journal of Intelligent Information Systems, part of Springer Nature (2022)
7. Al Shariah, N.M., Jilani Saudagar, A.K.: Detecting fake images on social media using machine learning. (IJACSA) International Journal of Advanced Computer Science and Applications **10**(12) (2019)
8. Steinebach, M., Gotkowski, K., Liu, H.: Fake news detection by image montage recognition. J. Cyber Sec. Mobi. **9**(2), 175–202
9. Meng, R., Zhou, Z., Cui, Q., Lam, K.-Y., Kot, A.: Traceable and Authenticable Image Tagging for Fake News Detection, arXiv:2211.10923

10. Singh, B., Sharma, D.K.: Predicting image credibility in fake news over social media using multi-modal approach. Neural Computing and Applications **34**, 21503–2151 (2022)
11. Bondielli, A., et al.: Multi-fake-detective at evalita 2023: Overview of the multimodal fake news detection and verification task. In: Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR. org, Parma, Italy (2023)
12. Zhang, X., Dadkhah, S., Weismann, A.G., Kanaani, M.A., Ghorbani, A.A.: Multimodal fake news analysis based on image–text similarity. IEEE Transactions on Computational Social Systems (2023)
13. Chen, Z., Hu, L., Li, W., Shao, Y., Nie, L.: Causal Intervention and Counterfactual Reasoning for Multi-modal Fake News Detection. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Vol 1, 627–638. Long Papers (2023)
14. Karwa, R.R., Gupta, S.R.: Automated hybrid deep neural network model for fake news identification and classification in social networks. Int. J. Auto. Smart Technol. **13**(1), 2409 (2023)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, arXiv: 1512.03385, https://doi.org/10.48550/arXiv.1512.03385



Background Subtraction Model with Advance GMM in Dynamic Background

Manisha[1](✉), Avadhesh Kumar[1], and Dileep Kumar Yadav[2]

[1] School of Computer Science Engineering, Galgotias, University, Greater Noida, India
manishacse2022@gmail.com
[2] SCSET, Bennett, University, Greater Noida, India

Abstract. Now a days to identify objects in video is a challenging task due to the real-time environment challenges. These challenges are illumination variation, cluttered background, rain, fog, dust, mist, *etc*. Such works are very helpful for security surveillance, transportation, agriculture, manufacturing, medical, healthcare, robotics, and multimedia applications. The *BGS* approach is an effective way of separating the foreground region from the background scene. It's often used in applications that operate in real-time for surveillance systems, motion detection, and recognition, or gestures-based pixel classification *BGS* that works by removing the static or dynamic background from the current frame to detect the foreground moving objects. So. This work focuses on video datasets to detect objects and remove the background scene by applying a background-subtraction approach. This work also focuses on the lack of a detailed description of the considered dataset. Here, a deep learning technique and *BGS* method are applied for object detection. This work enhances the existing GMM method and also compares it with advance GMM. The proposed advance GMM based method provides the better outcome as compared with existing method.

Keywords: *BGS* model · GMM · Deep learning Model · intelligent video analytics

1 Introduction

The computer vision is a branch of Machine Learning (*ML*) and Artificial Intelligence (*AI*), mainly concern on identifying and detecting objects in an image or scene. It tries to identify human vision, enabling robots to analyze, assess, and interpret images and videos, thereby extracting meaningful information. It's a very challenging task for video surveillance system to determine the motion of objects in frame sequences [1]. These approaches are pixel-based, block-based, and texture-based. Pixel based approaches are structures that model each pixel independently in a frame sequence of video using a pixel-based technique to describe each pixel as a Gaussian distribution. Researchers have keen interest in automatic video monitorization due to the combination of improved technology for cameras with intelligent video analytics. However, advancement in technology leads to increase the amount of complexity, and scale of monitoring videos, which help

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 212–222, 2025.
https://doi.org/10.1007/978-3-031-75167-7_17

to develop a novel approach for effectively handling the large amounts of video data. The effective Background Subtraction (*BGS*) approach provides better distinction method to distinguish foreground areas from the background of a video sequence. This type of method is required for many intelligent video surveillance applications for detection and extraction of moving objects. The main objective of *BGS* methods is to remove or detect foreground boundaries that are moving in video sequencing. For surveillance application *BGS* method is widely used for supervision applications for video analysis, such as optical motion scanning, computational models, target tracking, and automated video surveillance. In all these applications, first the background is analyzed, then a model is developed to extract the moving object from the frame sequences of the video. In this application, a fixed camera is used for video surveillance [2, 3]. This research in general mainly focuses on future trends in integrating complex mathematical models and machine learning models for detection of moving objects. The *BGS* method can be classified as: 1) statistical models: statistical approach-based models. 2) Single-Frame and Temporal Methods: 3) Pixel-Based and Region-Based 4) static and adaptive background 5) Foreground Detection Algorithms 6) Complexity and computational requirements 7) Environmental factors. Due to uncertainty and incompleteness in the data the desire outcome is not accurate, by using the concepts of *ML* the different issues like unclear data, varying illumination and weather condition can easily addressed.

2 Problem Statement

This paper describes various challenges related to background and foreground scene and camera related issues as shown in Fig. 1 for identifying moving objects. The *BGS* approach is completed through background initialization, modeling, foreground detection, pixel classification, and background maintenance [5–7]. There are various issues faced during the *BGS* techniques such as accuracy in detecting an object, foreground

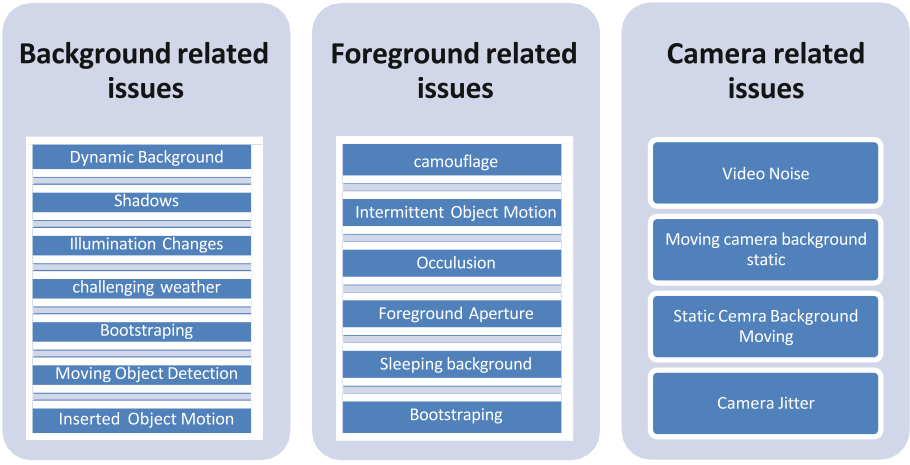


Fig. 1. Issues in development of *BGS* method for moving objects detection

challenges, illumination changes, shadows cast by moving objects, low-quality sensors, challenging weather, ghosting, and occlusions.

The detecting moving targets object in a video is major issue in the field of image processing. This is one of the most challenging problems to resolve. An object that is in motion within a video sequence is observed with the purpose of identifying and isolating it. *BGS* is a fundamental technique used for this purpose. Moving objects are identified by identifying pixels that deviate significantly from the static background model of the scene [4, 10]. The most widely used approach for *BGS* is Gaussian Mixture Model (*GMM*). It models each pixel's intensity distribution as a mixture of several Gaussian distributions. When a new frame is received, the GMM is updated in the maintenance step, and the pixel is considered foreground if its intensity is well-modelled by the current Gaussian components. This paper uses adaptive threshold values instead of fixed threshold values. This method can handle various weather conditions effectively. This paper proposes noise filtering techniques such as the Gaussian bullring method to reduce noise under different environmental conditions. This paper analyses the algorithm's performance quantitatively in terms of accuracy. Several datasets have been freely available for the analysis of *BGS* strategies. The primary use of background surveillance is smart vision monitoring. In this paper, we describe various video and image datasets for *BGS*. These datasets are used for such things as image classification, object detection, image segmentation, and more. The most popular image datasets are MNIST, CIFAR-10 and CIFAR-100, ImageNet, Pascal VOC, COCO, and Open Images. Video datasets are CD net, BMC 2012/2014, Vibe, SBI, SB Mnet, and LAP datasets.

3 Related Works

The GMM is an unsupervised analytical parametrical approach that employs three major factors: mean, covariance, and mixture weights from all component strengths. GMM is well-known for its tenacity in the face of progressive illumination shifts and dynamic background difficulties. GMM is a background-oriented illustration approach that generates a distribution model for individual pixels by analyzing color information. The video was created in order to meet the purpose of background modelling. It can characterize the multipeak pixels, among other things. GMM is appropriate for modelling complicated backgrounds such as swinging trees and light gradients. a GMM is used to build for each pixel in a sequence of frames. [12] In this model, small weight is related to the foreground and large weight is related to the background. Each pixel can be associated with a mixture of K Gaussian functions. it is the weighted sum of a finite number. as shown by Eq. 1:

$$f(x_t) = \sum_j^k w_{jt} * \emptyset(x_{tj} * \mu_{tj}, \sum_{jt}) \tag{1}$$

Here *wjt* is defined as j^{th} Gaussian's weight in the GM model at time 't', and ϕ (xjt * μtj, ∑jt) defined as Gaussian's probability density function having μtj as the mean value and ∑jt is the defined as variance at time 't'. The Gaussian's probability is shown in Eq. 2:

$$\emptyset(x_t; y_{jt}, \sum_{jt}) = \frac{1}{2\pi^{D/2} |\sum_{jt}|^{1/2}} \exp\{-\frac{1}{2}\Delta^2\} \tag{2}$$

Here, Δ is called Mahalanobis distance and is defined by using Eq. 3

$$\Delta = (x_{\mathrm{t}} - \mu_{\mathrm{jt}})^T \sum_{it}^{-1} (x_t - \mu_{jt})^{1/2} \tag{3}$$

Each new pixel compares with k Gaussian averages after image processing [14]. When the value of new pixel, represented as xt, is a multiple of the standard deviation from the mean, new pixel matches are identified.as shown by Eqs. 4 and 5:

$$x_t \varepsilon \, \emptyset(x_t; \mu_{jt}, \sum_{jt}) \quad if \, |x_t \, \mu_{jt}| < T\sigma_{jt} \tag{4}$$

Here The Constant multiplier of Standard derivation denoted as T, along with variance μjt, xt, σjt

$$\begin{aligned} w_{jt} &= (1-\alpha)w_{jt} + a \\ \mu_{jt} &= (1-\rho)\mu_{jt} - 1 + \rho_{xt} \\ \sigma_{jt}^2 &= (1-\rho)\sigma_{jt-1}^2 + \rho(\mathrm{xt} - \mu\mathrm{jt})\mathrm{T}(\mathrm{xt} - \mu\mathrm{jt}) \end{aligned} \tag{5}$$

Here ρ is represent learning factor and σ is represent learning rate, the background is comprised of one or more distributions characterized by height, weight, and minimal variances. Initially B distributions are chosen as the background hold processor. Threshold represent as *Th* used to determine the minimum data in continuing form as shown by Eq. 6:

$$B = \arg\min_{b}(\sum_{j=1}^{b} w_{jt} > Th) \tag{6}$$

4 Proposed Works

Here in this paper, noise is removed from the dynamic background during image processing for *BGS*. This process helps to remove the detection complexity in the dynamic back ground. Here in this study proposed advance dynamic *BGS* model with the GMM (Gaussian mixture model). The work is process into three steps. 1) background reconstruction; 2) moving target detection; 3) background updating. In these steps firstly detect moving object, then noise is removed by combining the wavelet half threshold method.

4.1 Background Reconstruction

In this phase, firstly select the block of images. If the selected image is large, then the number of blocks will be processed as small. The efficiency of image can be increases, but reduce the accuracy of the targeted object [13]. To restore the background, by eliminate the noise. In this phase, GM model is followed by employing K-GMM to access the alignment of current image frame matches with initial distribution or not.

If the GMM distribution matches the GMM distribution as shown by Eqs. 7 and 8:

$$|D_{ij}^{m,k}| \leq \mu \times M_{ij}^{m-1,k} \tag{7}$$

$$D_{ij}^{m,k} = G_{ij}^{m,k} - M_{ij}^{m-1,k} \tag{8}$$

Here accurate distance is represented by $D_{ij}^{m,k}$ between new pixel and mean of K^{th} Gaussian distribution. μ represent threshold derivation, the standard derivation of current image pixel in frame is represented by $M_{ij}^{m-1,k}$.

The standard deviation is represented by G^{m-1} and current frame is represented by kij and $G_{ij}^{m,k}$ represents the current image frame. The row number of the current image frame is represented by i, j is the column number of the current image frame, m represents the number of image frames, and K is the number of Gaussian distributions. if the new pixel value matches with k^{th} Gaussian distribution model, then update model paraments as shown by Eqs. 9 and 10:

$$M_{ij}^{m,k} = (1-p)M_{ij}^{m-1,k} + p \times G_{ij}^{m,k} \tag{9}$$

$$w_{ij}^{m,k} = (1-\alpha)w_{ij}^{m-1,k} + \alpha$$

$$\beta_{ij}^{m,k} = \sqrt{(1-p)(\beta_{ij}^{m-1,k})^2 + p \times (D_{ij}^{m,k})^2} \tag{10}$$

$$P = \frac{\alpha}{w_{ij}^{m,k}}$$

If the new pixel value does not align with any Gaussian distributions, then novel Gaussian distribution is created with the current minimum weight to replace the original Gaussian distribution value. The average value of the newly created Gaussian distribution is determined based on average value of the currently observed pixels. Shown by Eq. 11

$$w_{ij}^{m,k} = (1-\alpha)w_{ij}^{m-1,k} \tag{11}$$

4.2 Moving Target Detection

Noice is the major issue to detect object in dynamic background. Then we need to detect object in high accuracy.

$$|D_{ij}^{m,k}| > \delta * M_{ij}^{m,-1,k}$$

where $D_{ij}^{m,k}$ represent to detect absolute distance between two new pixels. δ represent threshold value. $M_{ij}^{m,-1,k}$ is standard derivation of current pixel [10, 15]. To remove noise here in this paper used mean filter denoise method combine median filter wavelet denoise method. The threshold function value selection is very complex and important for effective detection in image with noise.

$$f_{ij} = x_{ij} + n_{ij}$$

Here this paper both soft threshold denoising function and a hard threshold denoising function are introduced. That are shown by Eq. 12

$$\begin{aligned} \text{sgn}(t) &= \begin{cases} 1, & t > 0 \\ -1, & otherwise \end{cases} \\ f(t) &= \begin{cases} \text{sgn}(t), & (|t| - \tau), |t|0 \\ 0, & otherwise \end{cases} \\ f(t) &= \begin{cases} t, & t > 0 \\ 0, & otherwise \end{cases} \end{aligned} \tag{12}$$

The wavelet coefficients are represented by t, the threshold is represented by Th, and the wavelet coefficients following by threshold are represented by f(t).

4.3 Background Update

The background is an essential process to deliver a better result. A changing background affects target detection. A strong wind can cause the background trees' leaves to shake, making the test findings more audible. Using the conventional GMM The background cannot be removed using the *BGS* method. Updated in real time, which causes ghosting and major problems and has an impact on the identification of moving objects. In this study the adaptive background technique is updated [17]. This technique integrates current object identification frames with the existing background model. By upgrades the background model and retrieving the real-time frame. Ensuring that the adaptability of identified backgrounds can meet the needs of constantly changing actual time and recognizing more precise targets.as shown by Eq. 13:

$$B_{s+1}^{c} = h_s \times B_s^t + (1 + h_s) \times B_s^c \tag{13}$$

where hs denotes the update coefficient with [0,1] ranges; current target frame is represented by B**cs** and B^t denotes the instantaneous target frame.

5 Data sets

The paper utilizes a comprehensive approach that integrates these primary components to enhance the detection phase, incorporating techniques for noise reduction and handling dynamic backgrounds in the context of background modelling, target identification, denoising, and background updating. Table 1 shows the datasets used for the experimental analysis. The images and video datasets are chosen randomly and resize to 320 × 240 using OpenCV for training and testing of the model.

In the above table, describe the various datasets, such as image datasets, video datasets, pedestrians' datasets, overpass datasets, and so on. In this paper, we describe real-time video datasets using a background surface of water with moving objects and also take image datasets with static objects.

Table 1. Details of datasets

Datasets	Number of Frame	Size	Background	Moving Targets
Image Datasets	1800	320 × 240	People, cat, dog, tree, non-living object	People, Cat, Dog
Video Datasets	1750	320 × 240	Vehicle, tree, building,	Vehicle, boats
			pallor, moving boat	
Pedestrians	875	320 × 240	People, background	Pedestrians
Overpass	2100	320 × 240	Tree, Overpass	People
Caneo	950	320 × 240	surface of water	Boat

6 Experimental Result

To better analyses the performance of input data, this paper uses OpenCV library, which provides built-in functions to perform background subtraction using GM. The study uses various strategies for collecting data. This paper describes various datasets, like image datasets, video datasets, pedestrians' datasets, overpass datasets, and Caneo datasets. A description of datasets is mentioned in the above Table 1. *BGS* techniques mainly focus on the three phases: background initialization, detecting the object in the background, and background modelling. In this paper, we compare the Traditional BGS algorithm and advance *BGS* algorithm with GMM. To overcome the restraint noise in a dynamic background, during the detecting object denoising technique is performed. This work mainly focuses on background subtraction, object detection, noise removal, and background updating (Fig. 2).

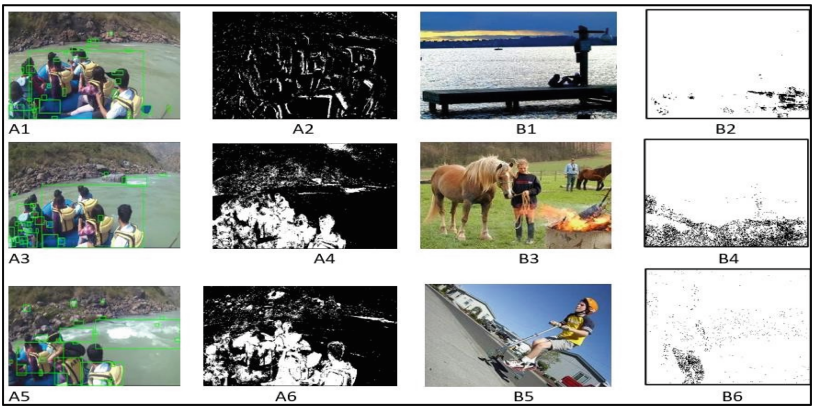


Fig. 2. *BGS* using GMM analysis result on dynamic background and static background Result group A show the result of video frame and group B show analysis result of static image

The conventional GMM technique's mean as well as standard deviation differ from the calculating approach of two pixels. The new method of standard derivation is more

accurate than traditional GMM. The GMM clustering procedure analysis the means and standard deviations of the resulting clusters. This approach is used to analyzing pixels in a video frame sequencing, including the calculate of average and standard deviation of selected pixels. Update your image folder for analyzing pixels in a series of video sequencing frame and calculating the average and standard deviation of selected pixels. Through enhanced background removal and the use of a Gaussian Mixture Model (GMM), this procedure assists in separating pictures that are more likely to be background pixels than foreground pixels (Fig. 3 and Table 2).

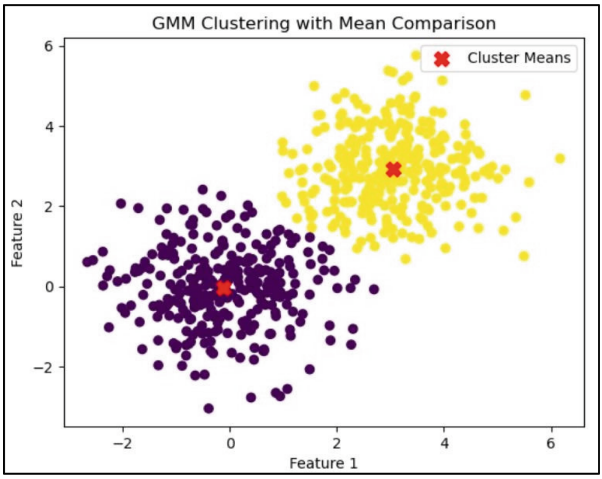


Fig. 3. GMM clustering with mean comparison using two pixels A1 and A2

This paper also compares the traditional GMM model with the modified GMM model. To analyse the mean and standard derivation of each frame in the image between two pixels. However, GMM is mainly used for the modelling and distribution of data. It is mainly focused on image segmentation, clustering data, and directly analysing the pixels. The traditional GMM model can be applied to clusters based on their intensity value. The mean value for every Gaussian element indicates the average brightness of the pixels within the cluster, while the standard deviation shows the dispersion or variety of intensity levels of pixels inside the cluster. GMMs may be used for picture segmentation, in which pixels are assigned to distinct clusters depending on their intensity levels. Advance GMM can be applied between two pixels for analysis of the mean and standard deviation. It can be applied in a small region to capture variation in detected pixels. Adaptive GMM can adjust the parameters based on detected pixels. In a review of moving target detection, the resultant pixel of the image is divided into two categories. Firstly, move the target object and background pixels. The records of these pixels store positive and negative values.

Table 2. Comparison Table of Traditional GMM model and Advance GMM model

Image	Tradional GMM model	Advance GMM model
	Results for pixel A1 (Background): Mean: 56.0 Standard Deviation: 0.001 Results for pixel A2 (Background): Mean: 103.0	Results for pixel A1: Average: 21.0 Standard Deviation: 0.0 Results for pixel A2: Average: 130.0 Standard Deviation: 0.0
	Standard Deviation: 0.001	
	Results for pixel A1 (Background): Mean: 66.0 Standard Deviation: 0.001 Results for pixel A2 (Background): Mean: 123.0 Standard Deviation: 0.001	Results for pixel A1: Average: 65.0 Standard Deviation: 0.0 Results for pixel A2: Average: 150.0 Standard Deviation: 0.0
	Results for pixel A1 (Background): Mean: 46.0 Standard Deviation: 0.001 Results for pixel A2 (Background): Mean: 99.0 Standard Deviation: 0.001	Results for pixel A1: Average: 18.0 Standard Deviation: 0.0 Results for pixel A2: Average: 120.0 Standard Deviation: 0.0
	Results for pixel A1 (Background): Mean: 26.0 Standard Deviation: 0.001 Results for pixel A2 (Background): Mean: 89.0 Standard Deviation: 0.001	Results for pixel A1: Average: 13.0 Standard Deviation: 0.0 Results for pixel A2: Average: 110.0 Standard Deviation: 0.0

7 Conclusion

This study mainly focuses on critical challenges to detection objects in various real time environments. The main contribution in the static and video datasets by using GMM model to detect object in background scene. The BGS method for object detection represent the significant the advanced GMM algorithm applied to the *BGS* method, outperforms traditional methods. The superiority is evaluated qualitative and quantities in a simulation experiment, demonstrating the effectiveness and flexibility of the enhanced approach. Overall, the research suggests that the proposed algorithm offers improvements in recognizing moving objects in dynamically changing backgrounds, contributing to the advancement of object recognition systems. This proposed work also help to develop foundation for future development in video analysis.

References

1. Rifat, R., Mou, J.R., Shahariar, R., Ahsan, A.: A New Approach of Moving Object Detection Using *BGS* Method. 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), pp. 256–259. Rajshahi, Bangladesh (2019)
2. Komagal, E., Vinodhini, A., Archana, Bricilla: Real time *BGS* techniques for detection of moving objects in video surveillance system. 2012 International Conference on Computing, pp. 1–5. Communication and Applications, Dindigul, India (2012)
3. Pandey, S., Jain, P., Patel, P.: Video *BGS* Algorithms for Object Tracking. 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), pp. 464–469. Kannur, India (2022)
4. Kim, Y.-C., Bae, T.-W., Ahn, S.-H.: *BGS* with shadow removal using hue and texture model for moving object detection. 2020 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1–2. Barcelona, Spain (2020)
5. Goyal, K., Singhai, J.: 'Review of B*GS* methods using Gaussian mixture model for video surveillance systems'. Artif. Intel. Rev. **50**(2), 241–259 (2018)
6. Devanne, M., et al.: Motion segment decomposition of rib-d sequences for human behavior understanding. Pattern Recognin. **61**, 222–233 (2017)
7. Zhang, X.Y., et al.: 'Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision. In: Proc. AAAI Conf. Artif. Intel. (AAAI), pp. 9227–9234 (2019)
8. Kaushal, M., Khehra, B.S.: 'BBBCO and fuzzy entropy based modified B*GS* algorithm for object detection in videos'. Appl. Intel. **47**(4), 1008–1021 (2017)
9. Akilan, T., Wu, Q.M.J., Yang, Y.: Fusion-based foreground enhancement for B*GS* using multivariate multi-model Gaussian distribution. Inf. Sci. **430–431**, 414–431 (2018)
10. Zhang, R., Liu, X., Hu, J., Chang, K., Liu, K.: A fast method for moving object detection in video surveillance image. Signal, Image Video Process. **11**(5), 841–848 (2016)
11. Ou, X., et al.: Adaptive GMM and BP neural network hybrid method for moving objects detection in complex scenes.' Int. J. Pattern Recognin. Artif. Intel. **33**(2), Art. no. 1950004 (2019)
12. Zhu, X., Zhang, X., Zhang, X.-Y., Xue, Z., Wang, L.: A novel framework for semantic segmentation with generative adversarial network. J. Vis. Communes. Image Represent. **58**, 532–543 (2019)
13. Aranda, L.A., Reverie, P., Maestro, J.A.: Error detection technique for a median filter. IEEE Trans. Nucl. Sci. **64**(8), 2219–2226 (2017)

14. Roy, J.S., Manam, L., Laskar, R.H.: Combination of adaptive vector median filter and weighted mean filter for removal of high-density impulse noise from colour images. IET Image Process **11**(6), 352–361 (2017)
15. Ou, X., et al.: Adaptive GMM and BP neural network hybrid method for moving objects detection in complex scenes. Int. J. Pattern Recognit. Artif. Intell. **33**(2), Art. no. 1950004 (2019)
16. Mirzaei, B., Nezam Abadi-Pour, H., Raoof, A., Derakhshani, R.: Small object detection and tracking: a comprehensive review. Sensors **23**(15), 6887 (2023)
17. Liu, L., et al.: Abnormal operation recognition based on a spatiotemporal residual network. Multimedia Tools and Applications, 1–13 (2023)



Video Summarization Using Firefly Algorithm

Shashank Pandey(✉), Sakshi Dwivedi, Vijay Bhan Singh, Neetu Verma, and Ranvijay

Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh, India
Shashankpandey388@gmail.com, {neetu.2022rcs01, ranvijay}@mnnit.ac.in

Abstract. In this research, Firefly algorithm to generate summary of a given video is proposed. The Firefly Algorithm is an optimization algorithm that draws inspiration from the behaviour of fireflies. There are many types of existing video summarization techniques. Here in this implementation, we focus on summarizing the video based on Object of Interest. Object of interest-based video summarization is a technique for identifying and choosing key portions of a video based on the existence of specified items or regions of interest within the frames. This method involves identifying and tracking relevant things in video frames, and then picking keyframes or short video clips containing those objects or regions of interest. In a surveillance video, for example, the item of interest could be a person or a vehicle entering a restricted area, and the summarization algorithm would select the frames or clips in which the object is present to provide a summary of the video's relevant events.

Keywords: Object of Interest · Firefly Algorithm · Video Summarization

1 Introduction

Video summarization plays a crucial role in the computer vision area and video processing, which proposes to design an incisive and meaningful summary of a given video [1]. The digital camera is extremely important in many different industries, such as advertising, news, entertainment, sports, event management, education, security surveillance, etc. Static and mobile security cameras are put in a variety of locations, including banks, businesses, airports, and residences, among others, and they continuously produce data. Analyzing video footage in order to extract interesting or valuable information from the video is a difficult and time-consuming operation. Users must view the entire film in order to get any useful information, which is the reason. Managing and protecting information faces significant issues as a result of the rapid expansion of data's digitization [2]. As day by day, video data is generating in huge amount, so the need for effective and efficient video summarization techniques has become more pressing than ever.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 223–234, 2025.
https://doi.org/10.1007/978-3-031-75167-7_18

The abundance of video data in recent years has raised interest in video summarization. Nowadays, video retrieval is undoubtedly accessible to both experts and consumers. A tool for quickly summarizing videos is video summarization. Either a sequence of motion pictures (video skims) or static pictures (keyframes) can be used [1]. Keyframes and video skims can both be used to sum up the video. In order to express significant information in videos, keyframes are utilized. They can also be referred to as representative frames, still-image synopses, R-frames, or a group of standout pictures gleaned from the video data [3]. By condensing the input video into a summary and only keeping the most important details, it provides quick browsing. Due to the exponential development in the volume of video data being daily collected by smartphones or caught by security cameras, video summarization is a hot topic in the study. Multiview Video Summarization (MVS) and Single view Video Summarization (SVS) are two general categories that might be used to classify it.

Using metaheuristic optimization algorithms like the Firefly Algorithm (FA) is one of the most promising methods for video summarization [4]. Firefly behaviour provides the inspiration for the FA, a relatively recent and potent optimization method. The FA has a good name for being effective at resolving challenging optimization issues. It is a popular choice for a variety of optimization problems, because of its speedy convergence to the best solution [5]. A wide variety of optimization problems, including those with nonlinear constraints, can be resolved by the highly adaptable Firefly Algorithm. A wide range of users, including those with no prior experience in optimization, may easily understand and use the FA because of its relative simplicity [6]. Because of its extreme resilience and ability to manage noisy and incomplete data, the Firefly Algorithm finds applicability in real-world scenarios where data quality may be a concern. The Firefly Algorithm enables the execution of numerous computations at once, because of its high parallelizability. It is hence appropriate for usage on parallel computing systems such as high-performance computing clusters.

Fireflies are a metaphor for various alternatives in this method, which are assessed and updated repeatedly in order to find the best one. The firefly algorithm is used in this study to provide an object-based method for video summarization. The proposed approach aims to identify the objects of interest in the video and create a summary that highlights these objects while minimizing the redundancy and preserving the coherence of the summary. The implementations show that the suggested method outperforms alternative video summarization techniques, and its outcomes are assessed on a set of benchmark datasets.

The paper is composed of 6 sections, in which Sect. 2 represents the motivation for initiating this research. The related works of video summarization are covered in Sect. 3. A concise introduction of used dataset is discussed in Sect. 4. As well as the proposed methodology for video summarization along with its mathematical model, and factors used in fitness function, fitness function is covered in Sect. 5. The results after applying the proposed method for video summarization are analyzed in Sect. 6.

2 Motivation

Video summarization (VS) is an emerging field in computer vision and multimedia research that has gained significant attention due to the increasing amount of digital video data on the worldwide network. With the available huge amount of data, it has become a challenging task to efficiently and effectively browse and search for relevant information within these large collections of videos. Apart from saving the valuable time of the viewer, VS contains the possibilities to decrease the storage and bandwidth needs of video data, leading to substantial economic and environmental effects. This has led to the development of video summarization techniques that aim to provide a condensed representation of the video content while preserving its essential information. Video summarization has many applications in various domains, such as video retrieval, surveillance, etc.

3 Related Work

The static video summarizing approach presented in this research by Bhattacharjee et. al. [2], is based on Artificial Bee Colony optimization and summarizes a video data by selecting its most important frames. Initially, the most important alterations in the content are determined by identifying the Regions of Interest or pixel clusters of the video frames. The Artificial Bee Colony optimization technique splits the video's frame sequence into pieces based on these characteristics. In [7], proposed a comprehensive survey of various nature-inspired optimization algorithms used for solving complex problems in various fields such as engineering, economics, and science. A thorough analysis of several nature-inspired optimization algorithms, such as GA, PSO, ACO, and others, is given in this study. The researchers delve over each algorithm's basic operation, benefits and drawbacks, and its uses in various domains. The efficiency, resilience, and complexity of the problem these algorithms can address are among the other aspects the study evaluates these algorithms based on. The authors also discuss the hybridization of nature-inspired optimization algorithms with other optimization techniques to achieve better performance. The researcher [8], introduced the Firefly Algorithm (FA). The primary objective behind developing the FA was to offer an optimization technique inspired by natural phenomena that can be applied to address intricate optimization problems in diverse fields, including but not limited to engineering, economics, and science. It is designed to mimic the flashing behavior of fireflies and their attraction toward brighter fireflies to find the optimal solution. Prashantha et. al [9], discussed about the segmentation and summarization of the videos. The process of video summarization is crucial for generating video thumbnails, video surveillance, and video downloading, by extracting key frames from each scene and creating a summary video that conveys the overall content in a short time. This research focuses on segmenting and summarizing video frames using a genetic algorithm (GA), which selects significant frames that highlight the event. The GA is modified with the purpose of selecting keyframes for the purpose of summarization, and a comparison is made between the modified GA and the original GA.

The researcher [10], proposed (i) Employing an image as a search criterion to determine user inclination, (ii) a mathematical model that uses a loss function and summary variance to minimize redundancy, and (iii) a similarity score to create a condensed video from the query image and video frames. The researchers also created an Object-based Query Image (OQI) dataset that contains the query images. The efficacy of the proposed method was evaluated on the UT Egocentric (UTE) dataset and demonstrated success in addressing the challenges of user preference and identifying important keyframes in daily life videos with diverse lighting conditions. Using a three-player generative adversarial network, query-conditioned summarization is described in [11]. Wherein the combined representation of the query and video is learned using a generator. By separating the true summary from the other two, randomly produced summaries, the discriminator uses three distinct summaries in input. The concept of key object-based video summarizing was introduced more recently by Meng et al. [12], who suggested condensing the video information into a set of key objects. They specifically stated that this reflective selection problem should be approached as a sparse dictionary selection problem, which would involve selecting a small number of representative object proposals in order to reconstruct the entire proposal pool

Song et al. [13], using multiple techniques like sampling, archetypal analysis, etc. applied to the TVSum dataset to generate a summary of videos based on the title of the video. Archetypal analysis is a multivariate statistical technique used for exploratory data analysis and pattern recognition. It aims to identify a small set of archetypes or prototypes that capture the essential features of a data set. These archetypes are defined as convex combinations of the data points and can be interpreted as extreme or representative cases of the data. The technique has applications in various fields, such as image processing, text mining, and social network analysis. Archetypal analysis is closely related to other methods such as principal component analysis, cluster analysis, and factor analysis.

Because of the increased information sharing in recent years, we have seen a rise in interest in copyright protection and multimedia data protection. The need to secure data is underscored by the fact that hackers are attempting to obtain private information from several sources. Several researchers have employed methods to conceal confidential information in order to preserve the confidentiality and integrity of data. The researcher [14], provided the secure digital video by inserting the encrypted watermark into the highlighted video frames, they were able to perform reversible data concealing based on histogram bit shifting. The approach is unusual in that it selects a small number of unique frames with the encrypted watermark symbol depending on a quality criterion. In order to eliminate the cumulative mistakes, A very effective watermarking method based on video coding was provided by Wang et al. [15]. The watermark would be included in the quantized coefficients of the video files. The results provided for a range of quantization parameter values showing that the suggested approach yielded satisfactory results even when the values were over 40. Similarly, a graph-based approach is used for text summarization [16] to reduce information loss and duplicacy of data. They proposed a textual graph-based extractive text summarization method which is based on single document (Table 1).

Table 1. Related work of Video Summarization

S. No.	Reference	Dataset	Method	Performances	Description
1	[2]	SumMe	Artificial Bee Colony optimization	CUSA = 0.71 and CUSE = 0.43	This technique analyses the frame sequence and highlights significant changes in the information to provide an efficient video synopsis
2	[7]	SumMe, TVSum, VSum, YouTube.8 M	GA, PSO, ACO	—	The study assesses these algorithms' effectiveness, adaptability to complicated issues, and investigation of hybrid techniques that combine these algorithms with others for improved performance
3	[8]	Random dataset	Firefly Algorithm	Calculate fitness function-43.5	The algorithm was created to discover the best answers to difficult optimization issues in science, engineering, and economics by utilizing the idea that fireflies are drawn to brighter objects
4	[9]	Yuv dataset	Genetic algorithm	Global optima-99%	They compare the performance of their modified genetic algorithm (GA) in video summary to the original GA, using it to identify important frames that emphasize events
5	[10]	UT Egocentric	Query-conditioned approach	Avg precision- 59.04, Avg recall- 62.04, Avg F1-score- 57.06	A mathematical approach that reduces duplication by utilizing a loss function and summary variance was presented by academics as a way to measure user interest using photographs as search criteria

(*continued*)

Table 1. (*continued*)

S. No.	Reference	Dataset	Method	Performances	Description
6	[11]	Public dataset	Query-conditioned approach	Avg F1- score- 46.05%	A generator learns the combined representation of the query and video. When three different summaries are used as input, a discriminator determines which of the two randomly produced summaries is the real summary
7	[12]	SumMe	Fast iterative shrinkage thresholding algorithm	—	Their goal was to pick a limited set of object suggestions that could reconstruct the complete pool, treating the selection process as a sparse dictionary issue
8	[13]	SumMe, TVSum50	Title-based approach	Avg F1- score- 48.06%	Principal component analysis, cluster analysis, factor analysis, and image processing are few of the domains in which this methodology finds use. It is closely related to these techniques
9	[14]	Own dataset	Firefly optimization	Precision- 94%, Recall- 96%, F1-measure- 95%	Specifically, they selected unique frames based on quality criteria to insert encrypted watermark symbols, achieving reversible data hiding by histogram bit shifting
10	[15]	Watermark dataset	Watermarking Algorithm	PSNR- 42.45	Their results over a range of parameter values showed that their approach was resilient and continued to be successful even when the quantization parameter value exceeded 40

4 Dataset

The 50 videos are taken from the TVSum as a dataset and each video has manual annotations by 20 judges which produces 20 reference summaries for each video. The 50 videos are divided into 10 categories, each category having 5 videos. The categories are getting Dog Show (DS), changing Vehicle Tire (VT), ParKour (PK), Making Sandwich

(MS), Flash Mob gathering (FM), PaRade (PR), BeeKeeping (BK), attempting Bike Tricks (BT), Grooming an Animal (GA), Vehicle Unstuck (VU),.

The provided reference summaries are not in the form of .mp4 files. Instead, the dataset includes the ratings of 20 judges for each of the 50 videos, where each rating corresponds to a 2-second long shot in the video. By using these ratings, can select the best frames based on a specified summary length (in this case, set to 15% of the video length). To achieve this, a MATLAB code has been provided to extract binary vectors from each reference summary. These binary vectors are the same size as the number of frames in the original video, with a value of 1 indicating that the frame should be included and a value of 0 indicating that the frame should not be included.

5 Proposed Methodology

The proposed methodology depends on 2 subsections mainly; factors used in fitness function and fitness function. The Fig. 1 represents how the algorithm is working to complete its purpose of summarizing input video.

Firefly algorithm is an optimization algorithm that employs metaheuristic techniques and takes inspiration from the flashing patterns of fireflies. It is commonly used for solving optimization problems in engineering, computer science, and other fields. In this algorithm, each firefly corresponds to a potential solution to the optimization problem, and the flashing behavior of fireflies is used to update the positions of the fireflies and search for the optimal solution. The concept of the algorithm relies on the attraction and movement of fireflies towards brighter fireflies, which represents better solutions. The Firefly algorithm has demonstrated its efficacy in resolving numerous optimization challenges and has been implemented in several practical issues. There are some properties of firefly algorithm:

Attraction. Fireflies are attracted to other fireflies that are brighter than themselves. In the algorithm, this attraction is modeled as a function of the distance between two fireflies and their relative brightness.

Randomness. The algorithm uses randomization to explore the search space and escape local optima.

Scaling. The algorithm scales the attractiveness of the fireflies with their brightness, so that brighter fireflies have a stronger influence on the search.

Diversity. The algorithm maintains a diverse population of fireflies to ensure that the search space is thoroughly explored.

Efficiency. The algorithm is computationally efficient and can handle a large number of variables and constraints.

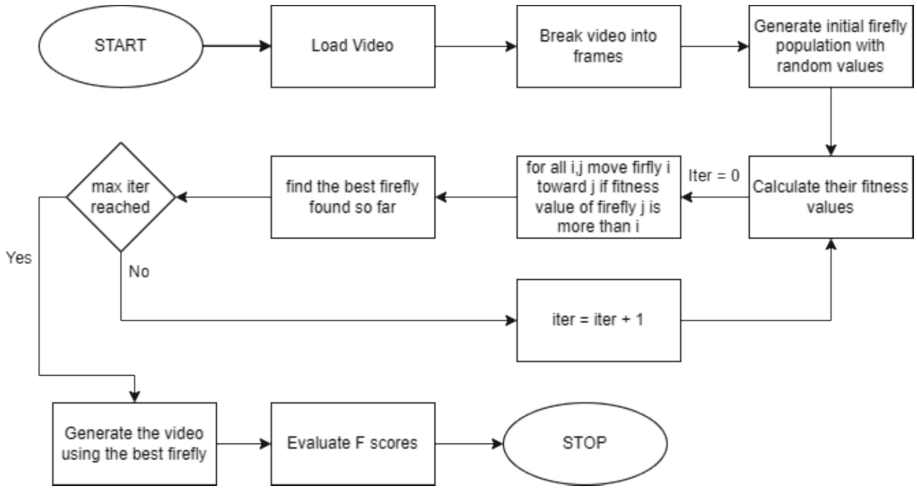


Fig. 1. Workflow diagram of proposed model

5.1 Mathematical Model of Firefly Algorithm

The algorithm operates under the presumption that all fireflies are non-gender specific, implying that a firefly can be enticed by another firefly; the degree of attraction is directly proportional to the brightness of the other firefly, which is determined by the objective function. The brighter the firefly, the more likely it is to be approached by another. Additionally, the inverse square law states that brightness decreases with distance.:

$$I\alpha\frac{1}{r^2} \tag{1}$$

In the event where light is moving through a medium whose light absorption coefficient is γ, then in that case intensity of the light at some distance r from the source can be shown as:

$$I = I_0 e^{-\gamma r^2} \tag{2}$$

The intensity of light at the source is represented by I0. Correspondingly, the level of brightness, denoted as β, can be expressed as:

$$\beta = \beta_0 e^{-\gamma r^2} \tag{3}$$

Algorithm 1

Parameter Initialization: $maxIter, maxIntensity, \alpha, \beta_O, \gamma$, etc
Initial population: Randomly initialize population of N fireflies X_i; $(i = 1, 2, 3, ..., n)$
Find Fitness value for fireflies
while iter < maxIter **do**
  for $i = 1$ to N **do**
    for $j = 1$ to N **do**
      if $F(X_i) < F(X_j)$ **then**
        Move X_i towards X_j
        Compute fitness of new X_i
        Move X_i towards X_j
      end if
    end for
  end for
  $iter = iter + 1$
  Find maximum intensity firefly
end while
Return the best solution

The algorithm generates viable solutions at random, referred to as fireflies, which are then given a fitness/objective function value or let us say light intensity. The firefly's light intensity and brightness are intimately correlated. For minimizing problems, the firefly with the lowest fitness value is given the highest light intensity. Once the intensity or brightness is assigned to the fireflies, each of them follows those with superior brightness. The brightest firefly moves randomly within its immediate surroundings. Consequently, if a given firefly j is brighter than another firefly i, firefly i will move toward firefly j, as per the following formula:

$$v = v_i + \beta_0 e - \gamma r_{ij}^2 (v_j - v_i) + \alpha(\varepsilon() - 0.5) \tag{4}$$

where β_0 is the brightnesss of vj at r = 0, in equation [4] the author recommended to use $\beta_0 = 1$ during practical implementation, γ determines the degree to which the firefly updating process is dependent on the distance between the two fireflies under consideration, The parameter α is used in the algorithm to determine the length of the random movement step, while $\epsilon()$ refers to a random vector obtained from a uniform distribution with values ranging from 0 to 1.

5.2 Factors Used in Fitness Function

The fitness function is the main fuel that drives the algorithm, it decides the direction in which the fireflies will move. It is thus extremely crucial to define a good fitness function according to the needs of the work required to be taken out from the algorithm.

Object of Interest (OoI). One of the important factors in the fitness function of the algorithm is Object of Interest. Similarity of the frames of video with the object of interest determines how important a frame is and how likely would it be present in the generated summary. This similarity factor is inversely proportional to sum of dissimilarity between

the frames that are selected for a candidate summary and the OoI image. The following equation defines this factor:

$$x = \frac{1}{\sum_{j=1}^{n} \text{dissimilarity}(F_j, I)} \tag{5}$$

where:

- X is similarity factor.
- n is number of frames under consideration.
- Fj is current frame.
- I Object of Interest image

Adjacent Dissimilarity Factor. While our goal is to choose frames that have a greater resemblance to the object of interest image, we must ensure that these frames are not solely from a particular section of the video and that they are evenly spread out throughout the video wherever similar frames occur.

Adjacent Dissimilarity Value is expressed as follows:

$$\text{ADV} = \sum_{j=1}^{n} \text{dissimilarity}(F_i, F_{i+1}) \tag{6}$$

where:

- ADV is cumulative adjacent dissimilarity value.
- n is number of frames under consideration.
- Fi is i th frame in the candiate summary. Adjacent Dissimilarity factor is given by:

$$Y = \frac{\text{ADV}}{\max \forall \text{ADVi}} \tag{7}$$

where:

- Y is adjacent dissimilarity factor.
- ADVi is is dissimilarity value between i th and i + 1th frame of candidate summary.

5.3 Fitness Function

The two factors we introduced earlier are utilized to create the fitness function, which can be customized by adjusting the values of its constants to meet the user's needs. The below equation illustrates how the fitness function can be modified to produce a summary that is more precise and comprehensible.

$$F = \frac{\alpha * X}{\alpha + \beta} \frac{\beta * y}{\alpha + \beta} \tag{8}$$

where:

- F is fitness function value.
- α user defined constant value.
- β user defined constant value.

6 Results Analysis

We ran the firefly algorithm with above proposed fitness function on our systems. Various details for the same are described in this section. We ran the algorithm on two physical systems (System A and System B) and a virtual machine based on Microsoft Azure(System C) with Python IDE; Spyder along openCV, pandas, sklearnmetrics, and functools library. The system's requirements are given Table 2.

Table 2. System requirements.

System	Hardware
A	6 gigabytes of RAM, 4 cores 8 threads Ryzen 5 3550H processor
B	8 gigabytes of RAM, 6 cores 12 threads Ryzen 5 5600H processor
C	Standard DS12 v2 (4 vcpus, 28 GiB memory) virtual machine on Microsoft Azure

Table 3. F1-score of proposed algorithm along with existing methods.

Cat name	SR's F1-score	AA1's F1-score	FA's F1-score
VT	0.29	0.33	0.33
VU	0.31	0.36	0.34
GA	0.36	0.28	0.34
MS	0.32	0.36	0.34
PK	0.32	0.35	0.34
PR	0.30	0.37	0.33
FM	0.30	0.41	0.35
BK	0.32	0.28	0.34
BT	0.29	0.24	0.33
DS	0.34	0.32	0.33
Avg	0.321	0.331	0.336

The F1 scores for two exactly same videos come out to be 1.0. As shown in Table 3, when comparing the binary vectors of the generated summary and the reference summary, if a reference summary video that is completely different is used as input, the resulting F score is significantly low. The F score can be as low as 0.03. The following table shows the output of our algorithm on the test dataset along with that of the two other techniques mentioned in the paper namely Random Sampling (SR) and Archetypal Analysis 1 (AA1).

7 Conclusion and Future Scope

The current fitness function uses Object of Interest (OoI) and Adjacent dissimilarity factor. Depending upon the requirements, it is possible to have multiple objects of interest in the same video. Moreover, it is possible that the importance that it is needed to give to different objects of interest is different. In such a case we can modify the fitness function and the basic structure of the algorithm to accommodate these requirements. Reducing the number of frames over which the algorithm runs could help us reduce the runtime. To achieve this, one possible way is to discard the irrelevant frames in the video and prevent them from being part of any firefly in the next iteration.

References

1. Haq, H.B.U., Asif, M., Ahmad, M.B.: Video summarization techniques: A review. International journal of Scientific and Technology Research (2020)
2. Bhattacharjee, T., Saha, S., Konar, A., Nagar, A.K.: Static video summarization using artificial bee colony optimization. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 777–784. IEEE (2018)
3. Garg, H.: Firefly algorithm step by step explanantion. YouTube link https://www.youtube.com/watch?v=QRttcJSMNK
4. Krishna, K.V.S.S.R., Chaitanya, K., Subhashini, P.P.S., Yamparala, R., Kanumalli, S.S.: Classification of glaucoma optical coherence tomography (oct) images based on blood vessel identification using cnn and firefly optimization. Traitement du Signal **38**(1) (2021)
5. Sennan, S., et al.: Energy efficient optimal parent selection based routing protocol for internet of things using firefly optimization algorithm. Trans. Emerg. Telecomm. Technol. **32**(8), 4171 (2021)
6. Amsaveni, A., Bharathi, M.: Use of firefly optimization algorithm for fractional fourier transform based reversible data hiding. J. Intel. Fuzzy Sys. **40**(1), 415–425 (2021)
7. Kushwaha, D.S., Sanchan, R.K.: Nature-inspired optimization algorithms: Research direction and survey. Neural and Evolutionary Computing (2020)
8. Yang, X.S.: Firefly algorithm for multimodal optimization. Fifth International Symposium SAGA (2009)
9. Prashantha, H.S.: Video segmentation summarization using modified genetic algorithm. International Journal on Computational Science Applications (IJCSA) (2018)
10. Baghel, N., Raikwar, S.C., Bhatnagar, C.: Image conditioned keyframe-based video summarization using object detection. Computer vision and pattern recognition (2020)
11. Zhang, Y., Kampffmeyer, M., Liang, X., Tan, M., Xing, E.P.: Query-conditioned three-player adversarial network for video summarization. arXiv preprint arXiv:1807.06677 (2018)
12. Meng, J., Wang, H., Yuan, J., Tan, Y.-P.: From keyframes to key objects: Video summarization by representative object proposal selection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1039–1048 (2016)
13. Song, Y., Vallmitjana, J., Stent, A., Jaimes, A.: Tvsum: Summarizing web videos using titles. IEEE explore 14 (2015)
14. Kumar, M., et al.: Secure video communication using firefly optimization and visual cryptography. Artificial Intelligence Review, 1–21 (2022)
15. Wang, C., Shan, R., Zhou, X.: Anti-hevc recompression video watermarking algorithm based on the all phase biorthogonal transform and svd. IETE Tech. Rev. **35**(sup1), 42–58 (2018)
16. Yadav, A.K., Ranvijay, Yadav, R.S., Maurya, A.K.: Graph-based extractive text summarization based on single document. Multimedia Tools and Applications, 1–27 (2023)

Exploring New Horizons in Computer Vision Research

An Image Processing-Based Tomato Leaf Disease Prediction Using Deep CNN

Kesana Vanisri(✉), K. Srujan Raju, and Bagam Laxmaiah

Department of CSE, CMR Technical Campus, Hyderabad, India
vanisri444@gmail.com

Abstract. Pest outbreaks have a significant impact on plants and crops, important to a reduction in national agricultural productivity. Conventionally, farmers and specialists rely on physical observation, and it can be timeconsuming, more expensive, and prone to errors, to classify diseases in plants. Though, by employing leading-edge image processing techniques, an instinctive detection system can be established, providing rapid and precise results. This paper aims to produce a Tomato Leaf Disease identification model with different categorizing leaf images and utilizing deep neural networks. The development of specific plant protection has the potential for substantial growth and improvement, with computer vision applications in the precision agriculture market also showing promising growth. This discussion about outlines all the critical stages for implementing the disease recognition model, initial from the image database group, assessment by agricultural specialists, and applying a deep learning framework to conduct the training of deep CNNs. By using this CNN, we will apply Max pooling layer, Multiple implementations are applied and finally we will get more accuracy.

Keywords: Diseases · Deep CNN · Dataset · neural network

1 Introduction

The primary approach for identifying plant diseases is through visual observation by a skilled plant pathologist [1]. However, detecting diseases at early stages or when symptoms are not apparent can be challenging for both amateur gardeners and experts. Detecting diseases at early stages means identifying and diagnosing health issues in plants before they become severe or widespread [2, 3]. Just like how doctors can catch illnesses in humans before they get really bad, early disease detection in plants helps catch problems while they're still small and manageable. This is crucial in agriculture because it allows farmers to take action promptly and prevent the disease from spreading and causing significant damage to their crops [4]. An automated system that uses computer vision [5], particularly Convolutional Neural Networks (CNN), could be beneficial for accurate and efficient plant disease diagnosis. CNN is inspired by how the brain processes information and is widely used in machine learning and cognitive research.

The connections between units can reinforce or inhibit activation states, and each unit may have a threshold or limiting function for signal propagation. These neural

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 237–245, 2025.
https://doi.org/10.1007/978-3-031-75167-7_19

networks are self-learning [6–8], which means they can be taught rather than explicitly programmed, making them particularly useful for tasks where conventional computer programs struggle with representation or feature recognition. The signal flow in neural networks generally moves from the front to the back, with multiple layers or a cube shape. Features are taken from the resized pictures using the convolution layers [9]. The dimension of the retrieved features is decreased by the pooling layer. The terms "max pooling" [10, 11] and "average pooling" refer to several forms of pooling. In order to down samples or reduce the spatial dimensions of feature maps work as a filter. During training with backpropagation, the correct outcomes are known, and forward stimulation helps adjust the weights on the front neural units to improve performance. Increasing the performance of disease detection techniques is the main objective of the research work.

2 Literature Review

Ms. Kiran R. Gavhale et al. [1], provided comprehensive research on leaf disease detection using IPT. This early work laid the foundation for subsequent studies in the field. Eknath Patil et al. [2] Infection causes significant crop damage, and in many situations, it is too late to identify the leaf disease. Offered in the realm of crop productivity prediction using hybrid ML models. By combining Decision Tree/Logistic Regression with Random Forest and employing essential climate and district-specific features, their research demonstrates the potential for improved agricultural forecast-ing and resource management.

K. Narsimha Reddy et al. [3], examined different image processing methods used for plant leaf disease detection. Their study highlighted the importance of accurate disease identification to enhance agricultural productivity. Riyao Chen et al. [4] Evaluated a DL-based approach for finding plant leaf diseases using channel attention and channel pruning. The authors demonstrated the significance of attention mechanisms in improving disease detection accuracy.

S. Bhuvaneswari et al. [5] presented an illness recognition model based on LearnNet, LILI LI and her co-authors [6] presented plant leaf diseases. Their approach highlighted the potential of identifying diseases for plants. M. T. M -Ap et al. [7] examined different leaf sickness. The authors discussed the advantages and challenges of various classification techniques [8], Prof. SUNITA KHEMAN el al [9] provided plant leaf diseases. M. Malathi, K. Aruli, and others, et al. [10] The presence of crop yield diseases can origin a significant decline in both the quantity and worth of farming yields.

Unfortunately, identifying signs of illness using just the naked eye can be difficult [11], especially for farmers managing large farms. C. Senthil Kumar et al. [12] mentioned that a plant disease leads to a reduction in agriculture production. They pointed out that the loss caused by various diseases each year remains a challenging aspect of agricultural production. They noted that although some work has been conducted on disease detection [13], proper segmentation of affected parts based on the type of plant family still remains an open problem in the field of research, G. Geetha et al. [14] presented complete research on plant disease leaf sicknesses using DLN techniques, A. Patil et al. [15] describes early-stage diseases.

3 Methodology

Proposed System: Fig. 1, describes as proposed system architecture. It takes the input image applies some CNN techniques and finally gets the healthy or diseased leaf description. The existing system, using a Support vector machine and HSP, gets low-accurate prediction. In this regard, I proposed a new methodology called Deep CNN, which may get a more accurate prediction.

This Architecture takes an input picture with a weight matrix (or filter) to create a stack of filtered images. Patches of the image matrix selected over a specific stride are multiplied by the filter.

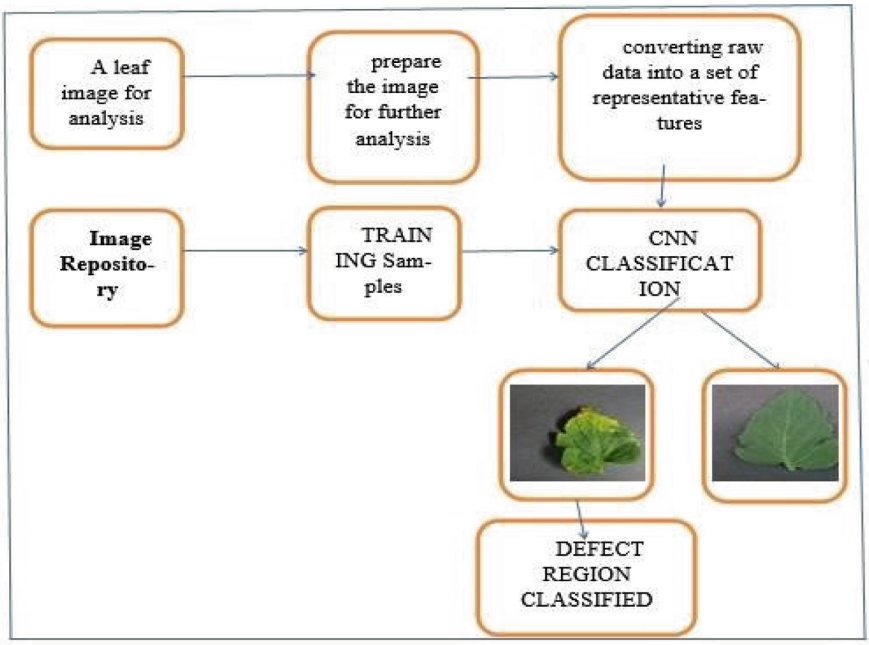


Fig. 1. It takes the input image applies some CNN techniques and finally gets the healthy or diseased leaf.

The input image is a grid of pixel values, typically represented as a matrix.

Convolution Operation: Convolution involves sliding a small filter (also called a kernel) over the input image. The filter is a smaller matrix with learnable weights. At each position of the filter, element-wise multiplication is performed between the filter and the portion of the input image it covers. The resulting values are summed up to produce a single value in the output feature map.

Stride and Padding: The filter moves across the input image in steps known as strides. Strides control how much the filter shifts at each step. Padding can be added around the input image to ensure that the output feature map has the same dimensions as the input. It helps preserve spatial information.

Feature Map: The output of the convolution operation is a feature map. It captures specific features that the filter is designed to detect.

Activation Function: After convolution, an activation function (e.g., ReLU) is applied element-wise to introduce non-linearity and enhance the network's ability to capture complex patterns.

Multiple Filters: A Convolutional Layer typically consists of multiple filters, each designed to detect different features (edges, textures, shapes) in the image. Each filter generates a separate feature map, and these maps are stacked to create the output volume of the layer.

4 Implementation

Table 1. The below table contains 18378 images for the preparation of the dataset

Plant	Disease Name	No.of images
	Healthy	1926
	bacterial pathogens	1702
	fungal pathogens	1920
	destructive and highly	1851
	contagious fungal disease	
	fungal disease that af-	
	fects the leaves of certain plants	1882
Tomato		
	fungal disease that af-	
	fects a wide range of	1745
	plants	
	damaging pest	1741
	fungus Corynespora cas-	1827
	siicola	
	viral disease	1961
	viral diseases in tomato	1790
	cultivation	

4.1 **Dataset Preparation:** The dataset may be manually collected using multiple digital cameras, enhanced, segmented as needed, and saved in distinct fold-ers to distinguish between different sick photos and healthy photographs for diverse plant leaves. In this preparation of the dataset, there are 9 catego-ries [Table 1, Website: 1] of

diseased images and healthy Images that are kept might be in color, grayscale, or in segments. The total images are 18378 for the preparation of the dataset. So based on the diseased image as input, it identifies the accurate image from the dataset.

4.2 **Preprocessing:** The below Fig. 2 [3], represented as Reducing the image's size and cropping it to a certain size and area of interest are both examples of preprocessing. The image is processed and improved to the necessary color scale. By determining the number of parameters in each layer based on the total of the weights and biases in each layer, each architecture's input resolution and the size of the model are determined. The input picture size in the proposed CNN architecture is 64×64, and the model size is 45 K. In comparison to processing RGB photos, a segmented and grayscale interpretation of the image does not offer a high level of performance. In order to facilitate subsequent processing, the proposed method first resizes color pictures to 64×64 dimensions.

4.3 **Attribute Extraction:** Features are taken from the resized pictures using the convolution layers. CNN learns the biases and weights of various feature maps to a particular process of capturing and selecting relevant information in comparison to generally designed feature extractors. The following convolution is the Rectified nonlinear activation function (ReLu). Because of this, CNN becomes nonlinear. The dimension of the retrieved features is decreased by the pooling layer. The terms "max pooling" and "average pooling" refer to several forms of pooling. In order to down samples or reduce the spatial dimensions of feature maps work as a filter.

4.4 **Attribute Extraction:** Features are taken from the resized pictures using the convolution layers. CNN learns the biases and weights of various feature maps to a particular process of capturing and selecting relevant information in comparison to generally designed feature extractors. The following convolution is the Rectified nonlinear activation function (ReLu). Because of this, CNN becomes nonlinear. The dimension of the retrieved features is decreased by the pooling layer. The terms "max pooling" and "average pooling" refer to several forms of pooling. In order to down samples or reduce the spatial dimensions of feature maps work as a filter.

Max Pooling Is a process in which the output value for a specific region, known as the pooling zone, is determined by selecting the highest value from the input values within that region. This pooling technique is frequently employed in Convolutional Neural Networks for entity identification process due to its ability to facilitate the recognition of an object's most distinctive features, such as edges and corners. By emphasizing the highest values within pooling regions, max pooling enhances the network's capability to identify crucial characteristics that aid in accurate object identification.

Average Pooling Is a pooling technique commonly used in Convolutional Neural Networks (CNNs) for tasks like picture segmentation and object recognition. In Average Pooling, the output value for each pooling region is always calculated as the average of its input values. This approach retains more information compared to Max Pooling, though it may make the most crucial features less prominent**.**

Fully Connected Layer (FCL): This layer finds extremely high-level properties that strongly correspond to an item or class. A set of characteristics for categorizing pictures without taking into account their spatial structure is the input to a fully linked layer. The

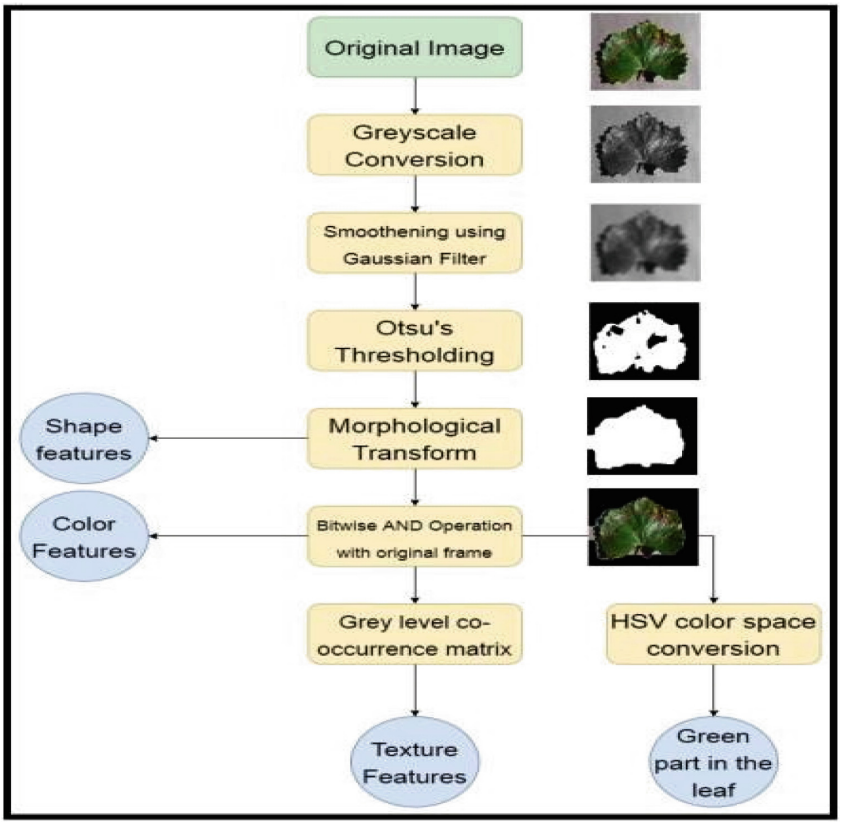


Fig. 2. The input picture size in the proposed CNN architecture is 64 × 64, and the model size is 45 K. In comparison to processing RGB photos, a segmented and grayscale interpretation of the image does not offer a high level of performance.

fully connected input layer, also known as the "flatten" layer, restructures the output from the previous layers into a single long vector. This transformation enables the data to be fed as input to the subsequent layer in the neural network.

The first fully connected layer: Uses weights to predict the proper label using inputs from the feature analysis.

The final probabilities: For each label are provided by the fully linked output layer.

4.5 **Classification:** The initial procedure performs Table 2, on the network's programmable arguments in order to enable the learning of a new class of classifiers from a constrained collection of samples. Learning to compare is the goal of metric learning methodologies. It implies that once a network masters class comparison, the capability to identify new image classes even when provided with only a limited number of labelled examples.

5 Result and Discussion

The training using the whole database of both original and enhanced photos is discussed in this part. The model's effectiveness was assessed using a collection of plant leaf images comprising both images (active and inactive). The dataset was carefully curated and labelled by agriculture specialists to ensure its reliability and accuracy. The image processing techniques used in the pre-processing stage effectively prepared the dataset, enhancing the high efficiency of the images and reducing complete data, and getting accurate results.

The model's confusion matrices provide valuable insights into the number of correct predictions made by the model, along with the occurrences of false positives and false negatives. These matrices allow for a comprehensive analysis of the model's high efficiency and accuracy measurements (Fig. 3).

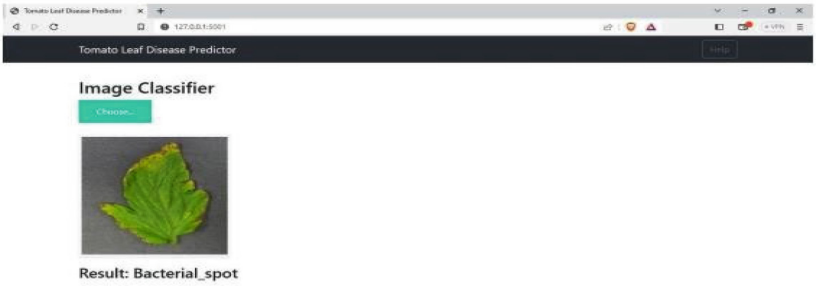


Fig. 3. Illustrates that the output image has Bacterial_spot disease caused by Tomato leaf.

Various Existing systems with proposed system: Our proposed system outperforms other existing systems in both accuracy and efficiency shown in Table 2.

Moreover, it offers the advantage of being more cost-effective as it does not require specialized hardware (Fig. 4).

Table 2. The below tables refers the Performance Analysis of Different Algorithms

Implementation	Backpropagation Neural Network	DIP and Support Vector Machine	HSP and Support Vector Machine	DIP	CNN
Correctness	-	83.34%	88.80%	93%	96.3%

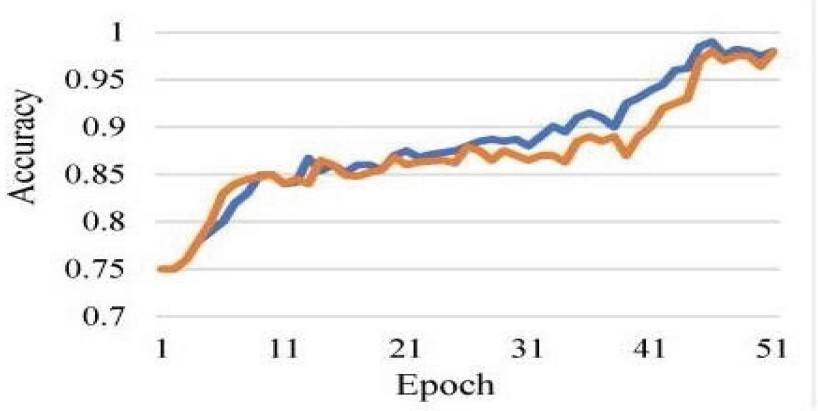


Fig. 4. Tomato leaf disease Prediction Analysis, its performance across various classification thresholds. It is a graphical representation that showcases how well the model distinguishes between different classes.

6 Conclusion

The dataset was carefully curated and labelled by agricultural specialists to ensure its reliability and accuracy. The deep CNN architecture, with its ability to learn hierarchical features from the images, played a critical role in capturing and distinguishing various disease patterns in the tomato leaves. By employing multiple convolutional and pooling layers, the model learned to extract relevant features and represent the input images in a manner that facilitated accurate disease classification. This model is more useful for the farmers, but at the same time, the combination of CNN and random forest algorithm may provide more accuracy and finding correct diseases.

The tomato leaf disease detection model presented in this study showed promising results and demonstrated the potential of image processing techniques and deep CNNs in automating and enhancing disease diagnosis in agriculture. Future work may focus on expanding the dataset with more samples of rare diseases and exploring transfer learning techniques to mitigate the need for extensive computational resources.

References

1. Gavhale, K.R., Gawande, U.: An Overview of the Research on Plant Leaves Disease Detection using Image Processing Techniques (2014)
2. Patil, M.E., et al.: A Hybrid Approach for Crop Yield Prediction using Supervised Machine Learning. In: 8th International Conference on Smart Structures and Systems, ICSSS, Institute of Electrical and Electronics Engineers Inc. (2022). https://doi.org/10.1109/ICSSS54381.2022.9782272
3. Reddy, K.N., Polaiah, B., Madhu, N.: A Literature Survey: Plant Leaf Diseases Detection Using Image Processing Techniques. IOSR Journal of Electronics and Communication Engineering **12**(03), 13–15 (2017). https://doi.org/10.9790/2834-1203021315

4. Chen, R., Qi, H., Liang, Y., Yang, M.: Identification of plant leaf diseases by deep learning based on channel attention and channel pruning. Front Plant Sci. **13** (2022). https://doi.org/10.3389/fpls.2022.1023515
5. Bhuvaneswari, S., Surendiran, R., Aarthi, R., Thangamani, M., Lingisetti, R.B.: Disease Detection in Plant Leaf using LNet Based on Deep Learning. International Journal of Engineering Trends and Technology **70**(9), 64–75 (2022). https://doi.org/10.14445/22315381/IJETT-V70I9P207
6. Li, L., Zhang, S., Wang, B.: Plant Disease Detection and Classification by Deep Learning - A Review. IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 56683–56698 (2021). https://doi.org/10.1109/ACCESS.2021.3069646
7. Map, M.T.: Plant Leaf Disease Detection using Deep Learning (2021). www.ijert.org
8. Bhattania, Y., Singhal, P., Agarwal, T.: Plant Leaf Disease Detection Using Deep Learning. Int. J. Res. Appl. Sci. Eng. Technol. **10**(5), 2518–1523 (2022). https://doi.org/10.22214/ijraset.2022.42892
9. Kheman, S.: Plant leaf disease detection using convolution neural network Poojya doddappa appa college of engineering (2022)
10. Kulkarni, P., et al.: Plant Disease Detection Using Image Processing and Machine Learning (2022)
11. Malathi, M., Aruli, K., Nizar, S.M., Selvaraj, A.S.: A Survey on Plant Leaf Disease Detection Using Image Processing Techniques. International Research Journal of Engineering and Technology (2015). www.irjet.net
12. Harakannanavar, S.S., Rudagi, J.M., Puranikmath, V.I., Siddiqua, A., Pramodhini, R.: Plant leaf disease detection using computer vision and machine learning algorithms. Global Transitions Proceedings **3**(1), 305–310 (2022). https://doi.org/10.1016/j.gltp.2022.03.016
13. Sahu, K., Tiwari, S., Mandal, S.: A review on leaf disease detection using image processing. International Research Journal of Engineering and Technology (2020). www.irjet.net
14. Geetha, G., Samundeswari, S., Saranya, G., Meenakshi, K., Nithya, M. : Plant leaf disease classification and detection system using machine learning. In: Journal of Physics: Conference Series, IOP Publishing Ltd. (2020). https://doi.org/10.1088/1742-6596/1712/1/012012
15. Patil, S., More, G., Gadilkar, S.A., Borade, A.N., Jagadale. S.S.: Leaf Disease Detection (2022). www.ijrpr.com
16. Website: https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf



Investigation on Existing Blockchain Based Architecture, AI/ML Driven for Boosting IoT Security and Privacy

Abdul Moid and Narendra Sharma(✉)

Department of Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences Sehore, Bhopal, MP, India
dr.narendrasharma88@gmail.com

Abstract. In the social world, the Internet of Things (IoT) can assimilate the physical as well as computational procedures as services. The quantity of administrations at the IoT edge is rising quickly because of the predominant purposes of smart gadgets and cyber-physical systems. One of the difficulties in exploring the encouraging applications of IoT services is making the services interoperable in a decentralized environment. It has been demonstrated that the Blockchain technology (BCT) is a promising option for calling for executions and establishing data trust. It could theoretically be utilized to aid in service interoperability. BCT records a verification of info or a procedure as a operation in a distributed ledger. Like the geography to IoT, considering BCT at the edges of the organization shows the dispersed trademark. Be that as it may, presently, BCT is as yet confronting the moves for interoperability because of various factors, for example, agreement conventions, block sizes, and time frame. The current BC-related security systems for the IoT environment are discussed in this paper. This paper further makes sense of the meaning of the provision of a safeguarded communication architecture for IoT devices utilizing a BC with Software defined network (SDN) empowered network. Further, this paper depicts the significance of the AI/ML driven techniques in tightening the security and confidentiality level of IoT frameworks.

Keywords: Internet of Things (IoT) · Blockchain (BC) Technology · smart frameworks · Artificial Intelligence · SDN · Security

1 Introduction

IoT has gained many attention in recent times. Consolidating with trend setting innovations, for example, service-oriented architecture (SoA) as well as block-chain technology (BCT), IoT is expected to build the interface amid people and PCs in the actual climate with an enormous scope. In addition, a broad applications range, including prevalent smart city as well as medical services, will be made possible by connecting physical entities in the real world [1]. IoT facilities have been expanded to the societal areas in household/human-oriented computing, like the persistent gathering, processing, as well

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 246–261, 2025.
https://doi.org/10.1007/978-3-031-75167-7_20

as separating of a huge quantity of info [2, 3]. With the rapid IoT development, both generated info as well as computing capabilities are exponentially growing at the network edge. Such applications display the dispersed trademark: information and registering assets are geologically conveyed as well as the controls are decentralized. Nonetheless, complicated errands require the joint efforts of numerous substances that need the interoperations of IoT administrations; whereas most Internet-based facilities must be international.

The IoT is becoming increasingly important to everyday life. The most popular gadgets of IoT innovation are cell phones, smartwatches, keen sensors, wellbeing checking gadgets, and numerous others. Some of these gadgets are just that— gadgets—and you can live without them, but many of them are very important to people's lives. This kind of innovation likewise has critical significance in the modern area. Robots, machines, automobiles, and airplanes, for instance, began to combine IoT components. The majority of IoT devices are small and limited, and they can also be considered embedded devices that are implemented as Systems on Chip (SoC). IoT gadgets are utilized generally, and the interest for these gadgets is high yet an opportunity to showcase is short. Various new applications require quicker execution, lower power utilization, and other explicit necessities. The new SoC devices' development phase begins to be divided into high-level modeling and low-level modeling phases in order to fulfill all of the application-specific necessities in a moderately smaller amount of time. Separating the demonstrating into two stages can speed up the turn of events and the engineering plan confirmation and approval process. Significant level displaying permits the equal improvement of an

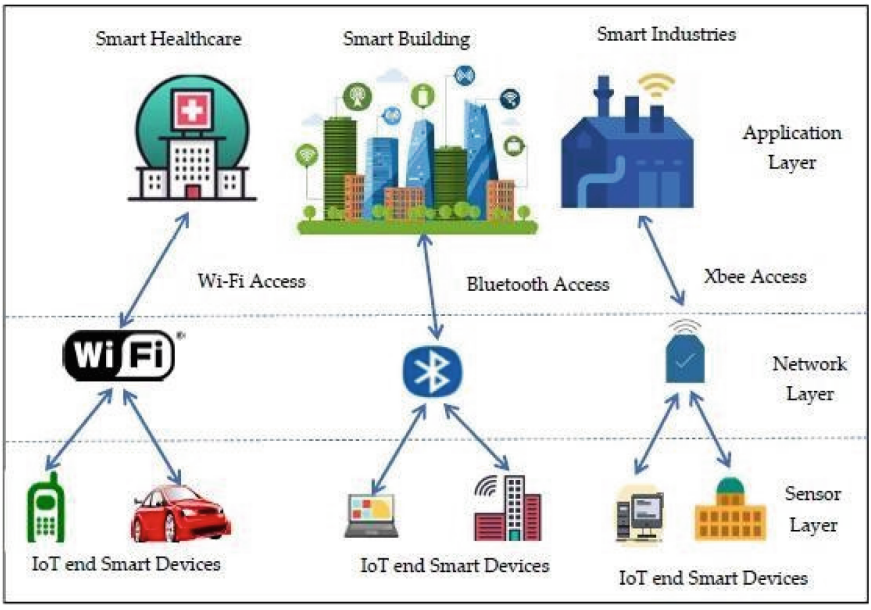


Fig. 1. Architecture of IoT [3]

essential useful or power controlled design and the application explicit programming that would run on the architecture.

High-level modeling can provide early results for the given architecture's power-controlled or functional model. These outcomes can likewise exhibit that the low level displayed plans should be changed to meet the particular necessities. The high level modeling, which is carried out by the SystemCTLM hardware description language, is the foundation for the majority of the contributions made in this thesis.

BC becomes a dispersed ledger innovation which likewise implies that the information contained by such a record is decentralized. Additionally, decentralization removes the requirement for an authorized third party to store the system's data; instead, it is distributed among the system's members (each participant owns a copy of the BC). It becomes substantial to outline that this technology is utilized not only for the exchange of cryptocurrencies but also for providing data traceability, integrity, and data immutability. On account of the highlights of BC innovation, the blend of block-chain with IoT frameworks can take care of numerous issues that are available in the present IoT based frameworks. The main issues with IoT structures are gadget confirmation, security of information sending and sharing, and confidence in unfalsifiability of gathered information.

2 Background

IoT proposes that physical gadgets will be able to participate and transmit data through established frameworks. This will induce a carried game plan of devices that can chat with the two people as well as one another. This will have an impact on the business and family streams of influence. One huge domain of utilization is in planning the supply's association. A's store will probably shift a thing or help from a maker to a client, as well as in an ideal world, this ought to be finished as capability as is sensible. An age game plan involves a business affiliation that engineers the development of individuals, exercises, data, and assets starting with one unit then onto the following [4]. Shifting IoT parts from the cloud to edge has lessens in general organization traffic and, thus, diminishes idleness. On the other hand, implementing IoT services on IoT edge gadgets presents new proposal and maintenance challenges. The operation of programming categorized IoT fragments as virtual IoT resources is one choice. Subsequently, the thing/gadget layer as well as the center IoT administration layer can be uncovered as assortments of microservices that can be circulated to a di-section hosts set [5].

Ways to deal with making consent BC circumstances in what character and quality association are chipped away at for the two clients and gadgets are being proposed by the creators [6]. The proposed worldview is charming; however, there is currently no implementation underway. Conquering the need of foremost BCs' tedious proof-of- work (PoW) engineering for molding as well as integrating information squares is a critical exploration issue not just for the IoT social request. Casper, a proof-of-stake (PoS) variant of Ethereum, will be distributed, encouraging to be all the more remarkable as well as safe after some duration [7]. To avoid key adjusting by the organization provider (which might be incorporated) [8], two affirmations were conducted prior to transmitting any message: When associating with the first client, various consumers verified the collector

open key, and this key did not vary frequently. Undefined solutions for BIoT devices have been illustrated, using BC advancement to arrive at leaders and develop their characters, showing that the BC is safeguarded from Internet protocol (IP) as well as mocking imitation assaults [9].

To address this issue, a software-defined BC architecture has been proposed for the implementation of complicated BC frameworks. After that, a consensus feature virtualization method for application-aware workflow is suggested. This technique has the ability to abstract and handle multiple consensus tools. The application-layer packet study is then applied to a smart transfer-learning-based network for effective management of virtualized consensus facilities. The aftereffects of the investigation recommend the suitability of the proposed system. This study is vital for working on the adaptability as well as expandability of BCs in the IoT. On the off chance that BC usefulness can be adjusted to suit it, the IoT can give a few benefits. As a result, a number of current IoT issues can be resolved. Nonetheless, a few issues may likewise be forced by the presentation of BCs for the IoT. A review distinguished five significant parts that ought to be tended to while planning the BC design for the IoT. This research zeroed in on the potential outcomes of coordinating BC as well as IoT advancements, bringing about the idea of BIoT, and the advantages as well as hardships that such solidification could offer. Contemporary smart home IoTs automation models examine the viability of block-chain technologies in this scenario and provide a valid environmental assessment. Another option anticipates the implementation of private local BCs in the home area network because of the substantial amount of usage data, event and status though not essentially for each appliance-generated transaction. Table 1 provides a comparison of BIoT technology options.

Table 1. Comparative analysis of BC based IoT techniques for data privacy and security.

References	Implications	Affected Layers	IoT Level	BC Technology Used	Propose Solutions
Hammi et al. [4]	Disruption and denial-of- service	Adaptation layer, network layer	Intermediate level	Private BC	As the basic security priorities, like availability integrity and secrecy the suggested IoT-BC framework is adequately protected.

(*continued*)

Table 1. (*continued*)

References	Implications	Affected Layers	IoT Level	BC Technology Used	Propose Solutions
D.W. Kravitz and J. Cooper [5]	Privacy violation	Network layer	Intermediate level	Private BC	Authorized BC has been suggested as an effective method to save and handle such systems.
Wee et al. [6]	Disruption and denial-of-service	Network layer	Low level	BC Consortium	For such condition, the data is available; though, the obtainability for conducting exchanges might be paused by the invader that manipulates the BC.
Nissar et al. [7]	Privacy violation	6LoWPAN Network/adaptation layer	Low level	Hybrid BC	Various performances have been reviewed for several kinds of aspects and give an effective answer for malware identification on Android devices.

(*continued*)

Table 1. (*continued*)

References	Implications	Affected Layers	IoT Level	BC Technology Used	Propose Solutions
Ahmad et al. [8]	Privacy violation	Network layer level	Low level	BC Consortium	Modum.io, a beginning up that utilizes IoT sensor to take advantage of block-chain innovations to guarantee temperature record information permanence and public perceivability while lessening working costs in the production network of drugs, is introduced.
Kaplan et al. [9]	Denial-of-service and privacy violation	Adaptation layer, network layer	Low level	Private BC	Assurance was continuously perplexed in BIoT conditions, in light of the fact that BIoT devices may discover phenomenal client data that can be saved in a BC whose security provisions are remarkable about a country to another country.

3 Internet of Things with BC

To resolve the issues with the centralized IoT framework and to exploit the different merits of BC innovation, it has become fundamental to consolidate BC with IoT. The basic layered BC with IoT design is comprised of four levels. In the IoT design, another layer called the BC is set between the network and application layers. IoT network issues might be better tended to by BC applications. In situations where the Internet of Things (IoT) network is growing, there are more chances that more devices will connect to it. Every one of these gadgets will be utilized as a device to endeavor an association across the web.

Since the obtained information has been all kept up with in the central repository, this could represent a great deal of issues. Be that as it may, the IoT has been addressed as an enormous scope gadget with cutting edge innovative coordination as the need might arise for IoT and its executions. Such extensive IoT frameworks wouldn't be best served by the brought together server. The idea of a centralized server underpins every traditional IoT solution currently in use. In the Web of Things frameworks, the sensor gadgets accumulate information from the designated things and empower information transmission across wired and remote net-works, alluding to the principal server as the web. Examination from a solitary system were completed rapidly and as indicated by client particulars. The computational force of the contemporary web is adequate to reconfigure the web framework in the event that the exploration were to be led by an enormous scope IoT framework. Giving the highlights "Autonomous Machine Coordination (ADC), Distributed File Storage (DFS) and Peer-to-Peer Networking (PPN)" for disseminated/decentralized networks is one of the most amazing ways of dealing with this issue. BC can do these three undertakings, permitting IoT organizations to observe the gigantic quantity of associated gadgets as well as organizations. BC upholds the IoT frameworks' cooperative treatment of gadget to-gadget communications.

3.1 Framework Design

This part gives an outline of the suggested design as well as portrays its primary components. Additionally, it provides a functional outline of the framework and elaborates on the workflow.

Architecture. The various entities in the architecture offer a BC-based IoT info communication system. IoT devices, device discovery, clients, the BC network, gateways, and the registry are the primary architectural entities depicted in Figure 1. Additionally, the system has a one-of-a-kind smart contract that is embedded in the BC framework. There are two primary forms of users for the network user as well as administrator of the system The framework administrator is liable for dealing with the framework, approving client access, keeping up with the BC network, as well as positioning/refreshing the savvy contract. As long as the system can write/read the program in a BC network, the client can be any remote user device. An operational explanation of every constraint is delivered in this subsection.

IoT Device. IoT gadgets are commonly little estimated and obliged actual gadgets with restricted handling, stockpiling, as well as power assets. They are inserted with detecting and correspondence modules to give the usefulness of IoT information assortment

as well as transmission. They structure the primary wellspring of information which can be gotten through various sorts of sensors for empowering specific IoT applications like natural checking, shrewd industry, and e-wellbeing. Any physical IoT-enabled or IoT-customized device can be an IoT device, including smart wearables, wireless cameras, microcontrollers with actuators/sensors, and others. As IoT gadgets are resourced-restricted gadgets, the framework keeps them from being conveyed as a component of the BC framework or in any event, having straight collaboration with the BC system. All things considered, they just speak with the entryway element, which regularly has more strong assets. Moreover, shifting scale arrangement of IoT gadgets is upheld by the framework. IoT frameworks are laid out amid the IoT gadgets to keep up with their interconnectivity utilizing extensive choices of correspondence advances.

Such incorporate 4G as well as WiFi and IoT-situated ones like LoRaWAN and NB-IoT. In addition, the design needs the IoT gadgets to be remarkably and all around the world distinguished in the BC framework. A special identifier is relegated to every gadget to work on personality the board transversely the framework.

Gateway. It is in computer science engineering is a multioerational gadget that is non-obliged with regards to computing, storage, and power capacities. The usefulness of the door incorporates gadget enlistment, personality the board, confirming IoT gadget exchanges, savvy contract summon, and information encryption/unscrambling.

In addition, IoT data communication via the BC is made possible by the gateway entity, which manages device–BC interface. It interconnects IoT gadgets as well as furnishes them with protected admittance to the BC framework. Every one of the correspondences with the BC are taken care of by the door in the interest of the IoT gadgets. In order to manage protected information communication between IoT devices as well as clients, it raises and listens to smart contract measures. Each gateway must be given a unique ID because the architecture allows for multiple deployments. It likewise empowers a gateway to interface a fluctuating quantity of IoT gadgets and permits a BC hub to interface with one or numerous passages.

BC Network. The architecture's key component for ensuring data integrity and decentralized communication is the BC network. A peer-to-peer network of multiple interconnected nodes makes up its physical structure. It gives the primary usefulness of miners, smart contracts and distributed ledgers. The BC system in framework can be public or private. It is used for overseeing data communication, client access and resource sharing (Fig. 2).

One of the key design elements of the architecture is the need for the BC system to support the development of smart contracts. In our engineering, the BC involves the accompanying primary modules:

- Registration module: To initiate integrating an IoT gadget with an exceptional identifier that refers the addressing the BC.
- Verification module: To initiate confirming client authorization to the various framework resources.
- Mining module: To initiate handling of the entire BC exchanges by the miner hubs.
- Smart contract module: To initiate implementing a smart agreement as well as processing of the triggered actions.

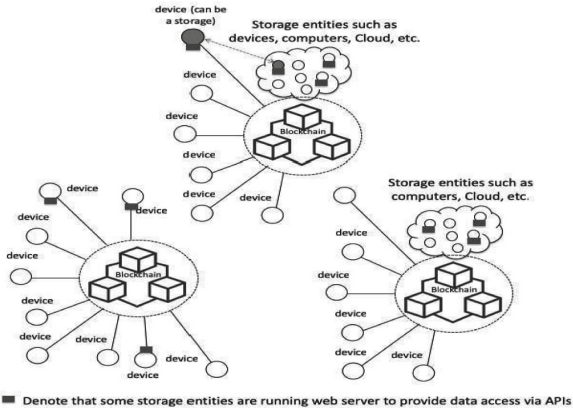


Fig. 2. Three storage schemes of distributed BC systems [4].

Device Detection. The retrieval of specific client-requested resource information is the responsibility of the device discovery module. Resource discovery is a critical cycle in keeping up with accessibility and usage at a lesser level of complexity. It depends on mentioning data about specific accessible assets in the framework and giving adequately point by point reactions to make the query activity compelling. It makes it simple for customers to look for and locate the system's required and available resources. This is a significant plan thought to work on asset revelation and usage, especially for enormous scope organizations.

It is necessary to form the system resources recognizable with sufficient detail to improve system utilization in order to enable this functionality. Thusly, at a beginning phase, another asset is at first handled in the vault to add its data in an enlightening way. For instance, any pertinent client request for information regarding the Internet of Things (IoT) instruments that support ecological monitoring in a particular position must be made accessible. The device discovery module responds to resource requests with the IDs of the entire accessible gadgets that can complete the request.

Registry. Registry becomes the integration of nearby stockpiling module for putting away the data of IoT gadgets and empowering basic administration of the actual assets of the framework. It can be executed using any storage capability, including standalone files and database servers of any size. The registry is an external module in the architecture to keep the level of intricacy of BC interaction low. The entryway element regularly collaborates with the vault for adding, refreshing, and eliminating the resources data. For better resource allocation as well as device characterization, information can be systematized in a variety of methods, like grouping particular devices into particular groups or classifying resources into various categories.

Clients. Given that it is capability to write/read info within the BC framework, any remote user device can be authorized as a client. Clients connect with the framework to get to its assets and lay out information interchanges with IoT gadgets. For instance, an industry client can utilize a PC to screen the situation with different IoT- enabled equipment by getting to the pertinent IoT information in the BC at a specific timeframe. Any

IoT-enabled object, remote server, or remote gateway, can serve as a client. The client gets to the framework through the BC to guarantee solid, decentralized and secure information correspondence. Client correspondences are taken care of through the brilliant agreement module of the BC.

3.2 Overview of Conceptual Model

The framework of IoT system depends on a progressive plan of various conceptual layers. It gives a deliberate and basic plan that can work with compelling and secure information correspondence. The design was created with just four various leveled layers: IoT gadgets, edge, BC, and application layers. Every layer is intended to give specific unmistakable operations of the IoT framework. The following provides a general overview of these layers:

- **IoT-Device Layer:** IoT gadgets are conveyed at the physical layer situated at the lower part of the framework. Moreover to data collection, transmission and acquisition, for the Internet of Things (IoT), this layer also offers network connectivity and device management. It also includes using the gateway to manage communication with IoT devices. Due to generally having gadgets of restricted capacities, this layer is kept up with in a straightforward structure with fundamental usefulness.
- **Edge Layer:** Numerous principal parts of the framework are situated at the edge layer to keep up with various key functional aspects. One is controlling how IoT devices connect to the network through the gateway entity. This needs laying out IoT-based wired/wireless network amid the IoT gadgets as well as the gateways. The gateways can then receive and send IoT data collected at the bottom layer to the appropriate clients thanks to the established connections. In order to facilitate efficient resource allocation and device characterization, this layer performs identity management as well as device registration and. Also, this layer keeps up with the gadget revelation interaction to give powerful control of client admittance to the various assets of the framework. It also sustains access authorization, user authentication, and data encryption, among other features. Furthermore, only this layer manages the system's interaction with the BC network. This makes the system modular adequate to be implemented with more progressive proficiencies and allows it to abstract its complexity at this fundamental layer.
- **BC Layer:** The BC layer involves every one of the activities that give the various decentralized operations of the BC framework. A smart contract, event management, miners, a distributed ledger, consensus, access control and identity management, are all examples of these. Event-triggered smart contract-based IoT data interface between data producers as well as IoT clients in a unreliable atmosphere is made possible by this layer. Activities conducted on the IoT information are approved as well as handled as exchanges in the BC framework at this layer. The system can encompass and abstract the BC modules operation thanks to this layer's design. In order to make it easier for the BC system to be effectively integrated into the architecture, this is an essential design feature.
- **Application Layer:** At the application layer, various application-related aspects are managed. Client access definition, service subscription, data request and resource

lookup, are all examples of these. This layer considers the high level of the engineering to keep up with direct connection with IoT application clients. IoT information are requested as well as gotten at this layer for a particular IoT application like natural observing or distant observation. The information are then handled to give brilliant application situated administrations.

3.3 Smart Contract

The framework includes the shrewd agreement module as a component of the BC system. The smart contract is used to carry out various system processes, allowing protected/controlled authorization to IoT devices. Data publishing, device discovery and client subscription/unsubscription, are examples of these. The system necessitates the implementation of a one-of-a-kind integrated smart contract for data sharing on the BC. As a result, the BC network's processes would be made simpler and communication costs between nodes would be reduced. Being coordinated into the BC would keep it carefully designed and vaccinated from being eliminated. The admin user must insert the smart contract into the BC framework. Its address is back to the admin user after it has been deployed, where it can be switched with any client/gateway that has successfully linked to the framework. The gadget disclosure module is likewise designed with the location of the savvy contract. As a result, such entities are able to connect with the smart contract in order to trigger and detect the system's block-chain events.

The smart contract of the framework is a portion of code sent in a decentralized organization of BC hubs. Any solicitation or implementation of a savvy contract is converted into an exchange on the BC. The system's smart contract defines a set of functions that initiate its various smart contract-based procedures. These oversee the IoT asset access and information sharing that are set off by BC exchanges. The framework administrator is the main element that has the option to deal with the brilliant agreement. An event-driven smart contract serves as the foundation of the system, making it easier to interact with the BC. Any activity performed through the shrewd agreement causes a significant BC occasion to be delivered. The brilliant agreement characterizes the accompanying fundamental activities: Subscribe, Unsubscribe, UnsubscribeAck ResListRes, GetResList publish, and PublishAck In like manner, the occasions that can be set off are characterized in the framework as follows: GetResListEvent, ResListResEvent, subscribeEvent, subscribeAckEvent, UnSubscribeEvent, UnSub-scribeAckEvent, publishEvent, and PublishAckEvent.

4 Internet of Things with BC -Functional Requirements

The technical difficulties associated with meeting the functional requirements (FRs) of IoT facilities that are crucial to interoperability are examined in this section. The accompanying FRs are normal to IoT administrations.

1. **Distribution:** The things over the Web like sensors, portable gadgets, or other relevant facilities are geographically dispersed. Moreover, IoT administrations should uphold the dispersed clients to authorized as well as utilize administrations over IoT.

2. **Big Data Processing:** Any smart device connected to an Internet of Things service generates real-time data; as the many IoT service users expands, so will the amount of info that needs to be processed.
3. **Heterogeneity:** Context-aware systems in a variety of data formats, operating systems, and protocols are typically included in IoT services. Additionally, local laws and regulations may impose restrictions on a portion of the network.
4. **Dynamicity:** Clients who authorize an IoT administration and the things the assistance rely upon are extremely powerful. For instance, of the arranged gadgets or different administrations, they might join, interruption, or leave the organization, or they might become failed or negotiated out of the blue. Subsequently, administration arrangements are problematic and difficult to be relied upon.
5. **Mobility:** Smartphones, body sensors, and automobiles are just a few of the Internet's mobile devices. They change areas as well as the organization spaces over the long run. IoT services must be able to deal with service usage uncertainties and the trustworthiness of produced info from various domains.
6. **Scalability:** The quantity of organized things and administrations remember for the myriad of an IoT administration might increment quickly with the commonness of administration. Scalability is essential for the developed IoT service; particularly, to the interoperations with different administrations in the myriad (Figs. 3 and 4).

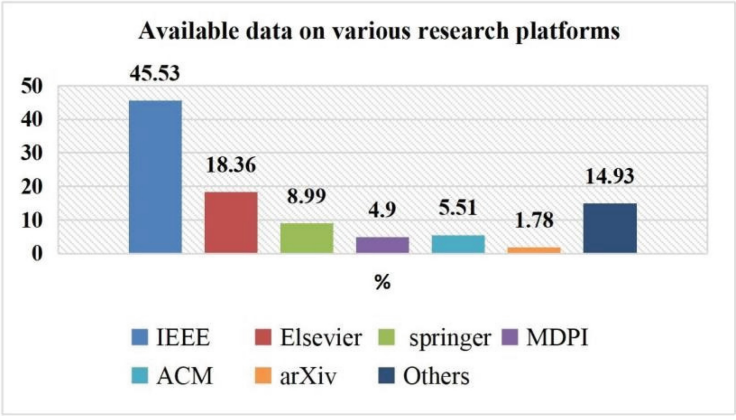


Fig. 3. The IoT statistics with advanced works in BC with respect to various publishers (2018–2022)

A lot of examination endeavors has been accounted for to manage a bunch of previously mentioned FRs. A few necessities, like information use and information security, might be clashed with one another, and no arrangement is yet accessible to sufficiently meet these FRs. Specifically, these FRs are pretty much applicable to the interoperations of the administrations, prevailing exploration research are confronting the accompanying difficulties in supporting help interoperations.

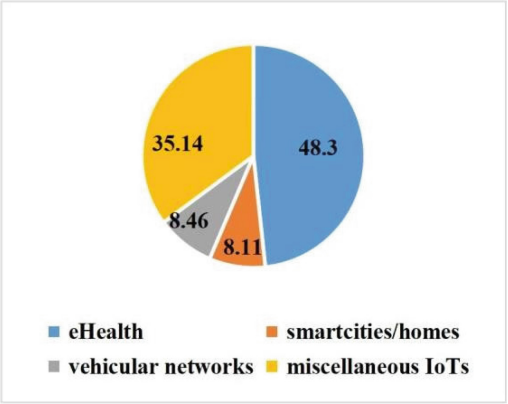


Fig. 4. Significant domain here IoT technique with BC network are used

- **Data Validity:** It is fundamental for empower interoperations in a distributed environment, yet steadily expanding sum and the dynamicity of the created information carry trouble to guarantee data validity.
- **Heterogeneity:** The mainstream of IoT services are not application-agnostic because of private-by-design as well as the requirement to upsurge system autonomy. Interoperations should be empowered among heterogeneous administrations with various proprietary protocols, interfaces, sizes and operating systems.
- **Uncertainties:** The things in an IoT might be portable as well as shifting the long haul; interoperations amid the things should be flexible, protected, and adjustable to vulnerabilities related with portability and dynamicity.
- **Trust:** The trust represents the test because of the dynamicity of independent and different gadgets and cycles in IoT administrations. It is unreasonable to utilize a typical safety control or cycle for each IoT administration; although the reliability of an IoT administration depends on the trusts on the entirety of the arranged gadgets and different administrations.
- **QoS:** An IoT administration includes in the assistance choices progressively when different service providers with comparable capabilities are accessible. One must trust on the QoSs data for service assortments as the number of accessible service providers grows, as well as how to assure the trustworthiness of QoS data becomes a concern.

As an outline, most of FRs of an IoT administration depends on the interoperations amid the organized things and different administrations in the heap of the IoT administration. No arrangement is accessible to fulfill the entire of FRs agreeably because of the difficulties of information legitimacy, heterogeneity, vulnerabilities, as well as the trust confirmations of information, procedure implementations, as well as QoS data of accessible administrations. There is an arising requirement to coordinate empowering answers for all of the previously mentioned difficulties in the advancement of IoT administrations.

5 Enhancing IoT Privacy and Security Using AI/ML Techniques

The Internet of Things (IoT) is reforming present day innovation by interfacing various gadgets and applications. In any case, its absence of normalization has prompted security and protection challenges. Man-made consciousness (computer based intelligence) procedures, for example, AI, profound learning, and support learning, show the potential in resolving these issues by empowering the execution of interruption location frameworks, verification systems, and security conservation strategies.

Simulated intelligence based innovations for confirmation utilize assorted strategies for various applications. Of all the verification frameworks, the proposition that are standing out as far as references are the ones that integrate systems for united learning and move learning, and furthermore BC as the innovation that will safeguard the data exchanges between end gadgets and the servers that carry out the learning calculations. For example, Maurya et al. [10] propose a methodology called federated transfer learning for validation and protection conservation utilizing novel strong twin deferred DDPG (S-TD3) calculation for IIoT. The methodology involves BCs to save security/protection for a broad modern applications range. It likewise presents a validation component in light of move figuring out how to upgrade the conservation and security norms for modern applications. The novel steady twin delayed DDPG trains the client model to validate explicit locales. The method facilitates the positive transfer of various types of info among devices in industry-wide and regional functions.

One more model is an investigation of Wang et al. [11], which proposes an authentication mechanism relied on transfer learning empowered BC (ATLB) that influences BCs for protection conservation in modern applications. ATLB presents an exchange learning-based verification system and diminishes model preparation time by moving locally or cross-provincially. In addition, Salim et al. [12] propose a handover authentication (HO-Auth) plot utilizing DL to verify gadgets and construct a client profile-based framework for sure fire approval. The plan guarantees that the BC decentralized networks get information from substantial gadgets, guarding cloud applications from degenerate information. To distinguish between an honest user and an attacker, the proposed scheme looks at the CSI movement pattern collected by multiple receivers.

AI based advancements have been applied to get to control (approval) in different spaces. In the medical services area, Lin et al. [13] propose a property dependent secure access control mechanism (SACM) for IoT-Wellbeing utilizing federated deep learning (FDL) to oversee delicate clinical information in IoT-empowered medical care. The proposed instrument finds the connection between clients' social ascribes and their trust attributes. By applying graph convolutional network (GCN) to the social chart with the defenseless contaminated recuperated model-dependent misfortune capability, clients' persuasions are acquired and afterward changed into their trust qualities. For every occupation, clients' trusts permit them to get to explicit clinical information provided that their reliability are greater than the relating edge. The FDL is applicable to get the ideal limit and important access control boundaries to improve access control exactness and the upgrade of security conservation.

Mrabet and others [14] propose a safe engineering for Industrial Internet-of-Things (IioT) in savvy producing applications. BC technology (BCT) as well as machine learning are incorporated into the architecture in five layers: sensing, network, transport,

application, and advanced services. BCT empowers assembling sensor access control data, whereas ML acquires its efficiently assault recognition like DoS, DdoS, infusion, man-in-the-middle (MitM), beast force, cross-site sequence (XSS), and filtering assaults by utilizing classifiers separating among ordinary and pernicious movement. The model can decrease essentially the quantity of DdoS, infusion, beast force, and XSS assaults inside a high level structure for sensor access control in IioT connectivity in view of a brilliant agreement alongside ML classifiers.

6 Conclusions

The incorporation of BC innovation into IoT frameworks provides a promising answer for address critical difficulties, for example, IoT information security and protection that would restrict the broad reception of IoT. The suggested design gives a straightforward as well as secure BC-based IoT information correspondence framework. It has a good scheme which mostly uses an event-driven smart contract to make data interface in a trustless/distributed atmosphere easy and safe. The brilliant agreement characterizes every one of the tasks and occasions to empower a sensible and consistent trade of IoT information in light of a BC driven publish/subscribe in model. The safe sharing of data and increased security of BIoT networks were discussed in this paper. IoT devices' current capabilities are insufficient to defend themselves against attackers. In an IoT ecosystem, device security, communication protocols, as well as various layers must all be taken into account. Tragically, countless IoT gadgets contain security defects and are helpless, permitting programmers and awful people to hurt and disturb the activity of these gadgets, as well as compromise the protection of their clients. This is because of asset limitations in IoT gadgets, as well as juvenile guidelines, an absence of secure equipment, and programming structures.

In future work, coordinating IoT-situated information examination utilizing AI calculations into the design will be contemplated. Executing and testing the architecture on a more accurate physical testbed is another consideration. Additionally, future research should look into the performance of the suggested design in light of additional real-world situations for various IoT applications, for example smart industries and smart cities.

References

1. Sahay, R., Geethakumari, G., Mitra, B.: A novel BC based framework to secure IoT- LLNs against routing attacks. Computing **102**, 2445–2470 (2020)
2. Patel, C., Vyas, S., Saikia, P.: A Futuristic Survey on Learning Techniques for Internet of Things (IoT) Security: Developments, Applications, and Challenges (2022)
3. Abdi, A.I., et al.: Hierarchical BC-Based Multi-Chaincode Access Control for Securing IoT Systems. Electron **11**, 711 (2022).
4. Hammi, M.T., Hammi, B., Bellot, P., Serhrouchni, A.: Bubbles of trust: a decentralized BC-based authentication system for IoT. Comput. Secur. **78**, 126–142 (2018)
5. Kravitz, D.W., Cooper, J.: Securing user identity and transactions symbiotically: IoT meets BC. In: 2017 Global Internet of Things Summit (GIoTS), pp. 1– 6. IEEE (2017)

6. Xian Wee, L., et al.: Rom-p: route optimization management of producer mobility in information-centric networking. In: Intelligent Transport Systems, From Research and Development to the Market Uptake: Second EAI International Conference, INTSYS 2018, Guimarães, Portugal, November 21–23, 2018, Proceedings 2, pp. 81–91. Springer International Publishing (2019)
7. Nisar, K., Amphawan, A., Hassan, S.B.: Comprehensive structure of novel voice priority queue scheduling system model for VoIP over WLANs. Int. J. Adv. Pervasive and Ubiquitous Comp. (IJAPUC) **3**(4), 50–70 (2011)
8. Ahmad, A.: Integration of IoT Devices via a BC-Based Decentralized Application. Master's Thesis. University of Stuttgart, Stuttgart, Germany (2018)
9. Kaplan, A.: How Alibaba Is Championing the Application of BC Technology in China and Beyond (2018). https://smartereum.com/7630/how-alibaba-is-championing-the-application-of-BC-technology-in-china-andbeyond-sun-jun-09/. Accessed on 22 June 2018
10. Maurya, S., Joseph, S., Asokan, A., Algethami, A.A., Hamdi, M., Rauf, H.T.: Federated transfer learning for authentication and privacy preservation using novel supportive twin delayed DDPG (S-TD3) algorithm for IIoT. Sensors **21**(23), 7793 (2021)
11. Wang, X., Garg, S., Lin, H., Piran, M.J., Hu, J., Hossain, M.S.: Enabling secure authentication in industrial iot with transfer learning empowered BC. IEEE Trans. Industr. Inf. **17**(11), 7725–7733 (2021)
12. Salim, M.M., Shanmuganathan, V., Loia, V., Park, J.H.: Deep learning enabled secure IoT handover authentication for BC networks. Hum. Cent. Comput. Inf. Sci. **11**, 21 (2021)
13. Lin, H., et al.: Privacy- aware access control in IoT-enabled healthcare: A federated deep learning approach. IEEE Internet of Things Journal (2021)
14. Mrabet, H., Alhomoud, A., Jemai, A., Trentesaux, D.: A secured industrial Internet-of-things architecture based on BC technology and machine learning for sensor access control systems in smart manufacturing. Appl. Sci. **12**(9), 4641 (2022)



Machine Learning-Based Estimation of Caloric Expenditure

Shivani Trivedi, Jahanvi Gupta(✉), and Himanshu Singh

ABES Engineering College, Ghaziabad, India
jahanvigupta1775@gmail.com

Abstract. In today's fast-paced technological age, people tend to neglect the health and opt for unhealthy fast food due to busy schedules. This leads to consuming more calories, contributing to obesity. This paper gives a thorough analysis of five different machine learning algorithms—Linear Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, and XGBoost—to predict the number of human calories burned during physical activity. This study aims to evaluate the predictive Mean Absolute Error of different algorithms and pinpoint the most precise human calorie burn calculation model. To train and test the models on a diverse dataset including details on age, gender, amount of activity, heart rate, and other pertinent variables. Following feature engineering and preprocessing, we assessed the algorithms' capacity to forecast calorie burn. The findings of this study have practical ramifications for several puroses, such as fitness observing, individualized health advice, and the creation of wearable technology. People can improve their general health, fitness, and well-being by making educated judgments about their physical activities and diet by precisely forecasting their calorie burn. Additionally, this work highlights the importance of choosing the right algorithms for certain predictive tasks and contributes to the body of knowledge in machine learning for wellness-related predictions.

Keywords: Machine learning · Kaggle · XGBoost · Decision tree regression · Support Vector Machine · Linear regression · Random Forest

1 Introduction

Many factors can contribute to obesity. Consuming excessive amounts of calories is one of these causes. A person who consumes too many calories has a calorie intake that is greater than their calorie expenditure. The extra calories are turned into body fat by the body [1]. People must monitor their calorie intake in order to lose weight or retain their weight at a healthy level. However, this procedure can be challenging and draining. People frequently fail to keep track of their food intake because they prefer to avoid challenging or taxing activities. And obesity could result from this [2, 3].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 262–271, 2025.
https://doi.org/10.1007/978-3-031-75167-7_21

Calories burn prediction is a branch of health and fitness analytics that seeks to predict how many calories a person will probably burn when engaging in physical activity, exercise, or daily activities. It is essential for managing weight, planning meals, and making training programs more effective, among other areas of health. For weight maintenance, the average adult needs to expend between 2,000 and 2,500 cal per day, as per the Centers for Disease Control and Prevention (CDC) [4]. A person expend more calories than he consume in order to decrease weight. A healthy weight reduction rate is 1–2 lb each week. A person's weight, the intensity of the activity, and the length of the activity all affect how many calories you burn throughout each activity.

This paper depicts the approach of identifying and categorizing the top algorithms for estimating the number of calories burned in order to obtain the best outcome. We also provide a brief overview of the xgboost, linear regression, random forest, decision tree, and support vector machine algorithms (see Sect. 3.4).

2 Literature Survey

According to the Nipas M. et al. regular physical conditioning is necessary to remain fit and healthy. The evaluation of calories expended by individualities is grounded on a formula and through the values of their metabolic equivalent of task using map [5]. This study used a retrogression model as one of the machine learning methods to estimate the amount of calories expended in order to obtain more accurate results. The main steps before data may be fed to retrogression models are data medication, cleansing, and analysis. The popular model for the study was selected by model training and evaluation using K-fold confirmation. The performance and vaticination delicacy of regression models were estimated grounded on the outcome of the model's testing after replications. The average delicacy was reckoned according to the outcome, Random Forest regression is the best model for the research due to delicacy of 95.77.

Handling is the winner for the highest calories burned in a single hour, according to Manjunathan N. et al. [6]. Running, swimming, and stationary bicycling are also fantastic options. HIIT workouts are simply not realistic for calorie burning. Following an HIIT exercise, your physique will continue calorie-burning for over 24 h. Vaticinating the calories burned by the drill is still a challenge due to natural disasters and changes in physical health. The calories burnt out during the course of an individual's exercise can be predicted via machine literacy techniques. To estimate the caloric expenditure of a drill, the exercise dataset was removed from the machine learning library at UCI. Four methods are available to attain the smooth the burnt calorie rate. The data set underwent point scaling and missing value reprocessing initially. Second, the target highlight's scattering is imaged and an exploratory point investigation is conducted.

Kalpesh J. et al. recognizes the mortal physical exertion with the modern technical developments. Among the detectors integrated into mobile phones are accelerometers, gyroscopes, and magnetometers. In moment's world powered by data, mortal exertion recognition is vital in many different domains, including medical operations, health shadowing, mortal check systems, and many more [7]. This exploration study examines the information attained from mobile detectors similar as gyroscopes, accelerometers, direct accelerometers, and magnetometers. Additionally, the suggested system forecasts mortal exertion by utilising the information collected from a mobile detector.

According to Hazifa N. et al. the notion of incorporating a bus-recognition point for Malay gleeful Seasons Food grounded on pictures is a veritably grueling project in computer vision as it's a commodity fresh and unexplored until lately. Still, acknowledgment is crucial for Malaysian druggies to control calorie input, particularly around Hari Raya, one of Malaysia's largest gleeful seasons and utmost famed carnivals [8]. Given that colour is significant part in securing the type of food, this exploration targets to apply the Colourpoint birth system after using segmentation ways during the preprocessing stage, where every hue found in the pictures is uprooted collectively. Also, outcome of the Colorpoint birth system is employed to determine the kind of food by using the Error-Correcting Affair (ECOC) bracket, which is included in the Support Vector Machine (SVM) algorithm.

Graduate scholars, like Acharya M. et al. always struggle to decide which universities to apply to while applying to master's programs. Despite the large amount of consultants and forecasters that provide assistance to a pupil, they are not invariably dependable because the decisions are made based on a limited number of prior enrollments [9]. We describe a machine lireracy system in this research when several regression techniques are compared, like Linear Regression, Support Vector Regression, Decision Trees, and Random Forest, considering the pupil's information.

As more health problems related to rotundity and gluttony have been documented, Naryana D. et al. and the team came to the conclusion that people have become more cautious about their dietary choices in an effort to protect themselves from situations like hypertension, diabetes, and other heart-related problems which is brought on by rotundity. 2.8 million individuals pass away every year resulting from being overweight, based on data collected by the WHO. Any healthy diet plan must have a certain number of calories. The method is deeply rooted in literacy to calculate the caloric intake in the food items visible in the stoner's photo. A subcaste grounded approach is used to prognosticate food's calorie content which includes Acquiring Pictures, Food product bracket, Surface area discovery and calorie vaticination [10].

3 Methodology

The method used to calculate and assess a person's calorie expenditure is described in this section. Choosing the appropriate attributes and filling in the gaps in the values are the two main aspects of the initial step of data cleaning, which is crucial. The process of attribute selection is determining and keeping only the most important variables, such as age, weight, height, gender, and degree of activity, which influence calorie expenditure. The relationships between the selected attributes are then clarified using a variety of graphical techniques. Different graphical depictions help to find patterns, correlations, and dependencies in the data, providing insightful information about how different elements interact and affect calorie expenditure. This method offers thorough explanations of the dataset's origins, make-up, and applicability to the task at hand. Finally, the machine learning techniques used for calorie expenditure prediction, laying

the groundwork for precise and individualized predictions based on the carefully chosen variables and enhanced dataset. The approach includes information about the dataset used, data collecting, data preprocessing, and machine learning methods for prediction as shown in Fig. 1.

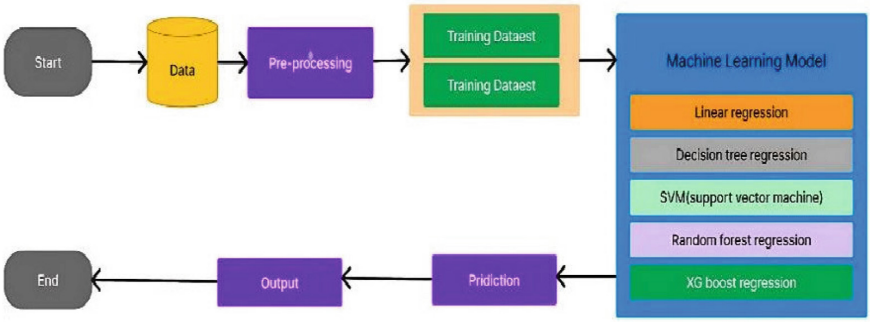


Fig. 1. Flow Diagram of Proposed Methodology

3.1 Data Collection

Gathering the dataset is the initial step, and Kaggle was our primary source of information."Exercise.csv" and "Calories.csv" are two datasets. Exercise.csv has 15000 instances and eight distinct attributes shown in Table 2, whereas calories.csv has the same amount of rows and number of columns are two that is "User ID" and "Calories" shown in Table 1. The data was then uploaded to the Collab platform. This dataset contains both categorical and numerical data.

Table 1. Description of Primary Dataset (exercise.csv)

User ID	Gender	Age	Height	Weight	Duration	Heart Rate	Body Temp
14733363	male	68	190	94	29	105	40.8
14861698	female	20	166	60	14	94	40.3
11179863	male	69	179	79	5	88	38.7
16180408	female	34	179	71	13	100	40.5
17771927	female	27	154	58	10	81	39.8

3.2 Data Preprocessing

The data must be prepared before being fed into the model. This stage includes managing null and missing values because data's quality greatly affects how well the model can

Table 2. Description of Secondary Dataset (calories.csv)

User_ID	Calories
14733363	231
14861698	66
11179863	26
16180408	71
17771927	35
15130815	123
19602372	112
11117088	143
12132339	134

learn. Taking care of the superfluous columns will help the model be more accurate and efficient. Merge the two datasets together and deal with missing and null values once more.

The Table 3. Given below shows number of null values with respect to different attributes.

Table 3. Attributes having null values

Attributes of Merge Dataset	Number of Null Values
Gender	0
Age	0
Height	0
Weight	0
Duration	0
Heart Rate	0
Body Temperature	0
Calories	0

3.3 Data Analysis

Data visualisation approaches are used using a variety of graphs and charts. First, analyze the data on gender distribution, then plot the distribution graph between count and gender (see Fig. 2).

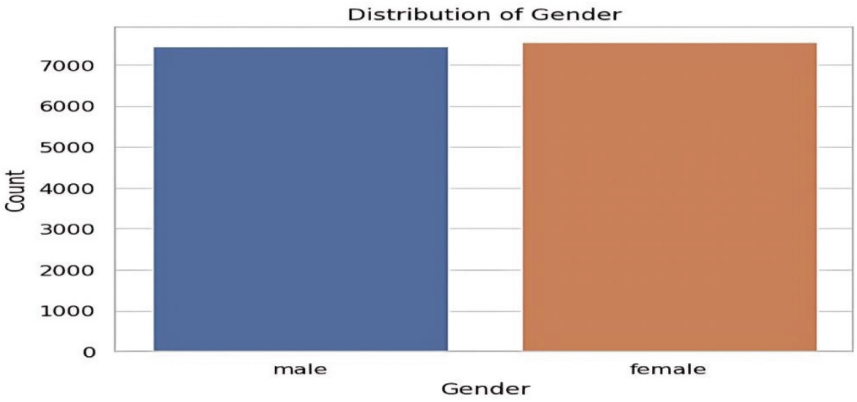


Fig. 2. Graph showing Distribution of Gender

Figure 3 shows the visual representation of the positive and negative relationships between different qualities. Regression models are then implemented after creating training and test sets from the data. The accuracy of the forecasts is then assessed using test data.

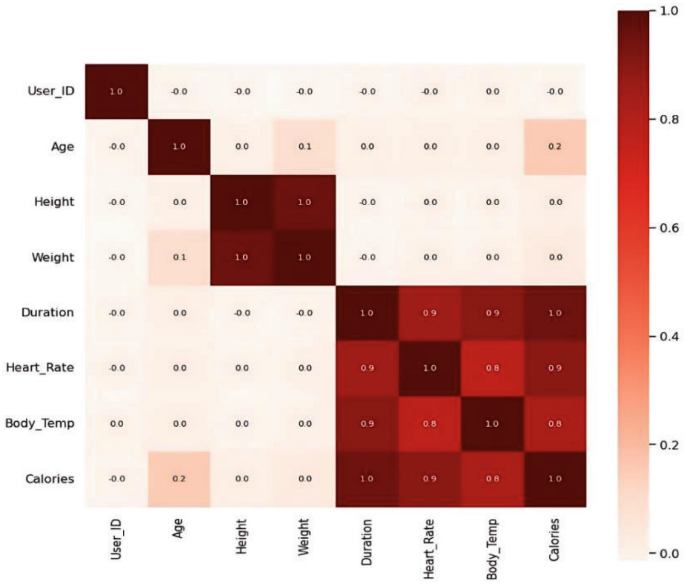


Fig. 3. Correlation Heatmap between different attributes of data.

3.4 Machine Learning Models

Five machine learning algorithms—linear regression, SVM, random forest, XGBoost, and decision tree regression—are used in this study. This stage includes the application of certain techniques to compute the R-squared value and absolute mean error value. A statistical technique called regression is employed in predictive modeling and is particularly helpful for forecasting continuous outcomes by looking at the correlations between variables.

Linear Regression. A predictor variable and a dependent variable are connected to one another sequentially in linear regression, a statistical method. Its goal is to reveal the connection or relationship between two variables. It can be used, for instance, to ascertain the relationship between the amount of exercise performed and temperature fluctuations [11].

Extreme Gradient Boosting Regression. XGBoost is short for this type of regression. By pooling the results of several weak models, this ensemble learning technique produces predictions that are reliable. It works especially well for managing large datasets and for addressing missing data [12, 13].

Decision Tree Regression. Regression models are built using the decision tree regression method, which resembles a tree. Recursively, it chooses the feature with the least impurity at each level after choosing a feature as the root node in the beginning. When considering impurity measurements, it sorts the data and computes the averages of nearby values to determine node splits.

Support Vector Machine (SVM). Depending on the class to be predicted, SVM is able to categorise new data points based on their position relative to this hyperplane. It is designe determine a hyperplane that effectivel divides two classes [14].

Random Forest Regression. The supervised learning approach known as "random forest regression" combines expectations from various machine learning models to increase accuracy. It is a potent and accurate technique for a variety of prediction applications [15].

4 Results and Discussions

4.1 Implementation

This model is analysed to determine the optimal algorithm for estimating caloric intake burned while working out based on characteristics like age, height, weight, body temperature, gender, heart rate, and activity duration. Evaluate the results of all five models after analysing them on the dataset to see which model predicts best. The optimal algorithm is identified by determining the one with the lowest Mean Absolute Error (MAE). For each model, MAE are tabulated as shown in Table 4.

As per Table 4, the smallest prediction errors and the highest predictive performance are shown by XGBoost, which has the lowest MAE. Linear regression has the highest MAE value of 8.38.

Table 4. Mean absolute error of different algorithms.

Regression Models	Mean Absolute Errors (MAE)
XGBoost Regression	1.48
Random Forest Regression	1.67
Decision Tree Regression	3.39
Support Vector Machine (SVM)	8.10
Linear Regression	8.38

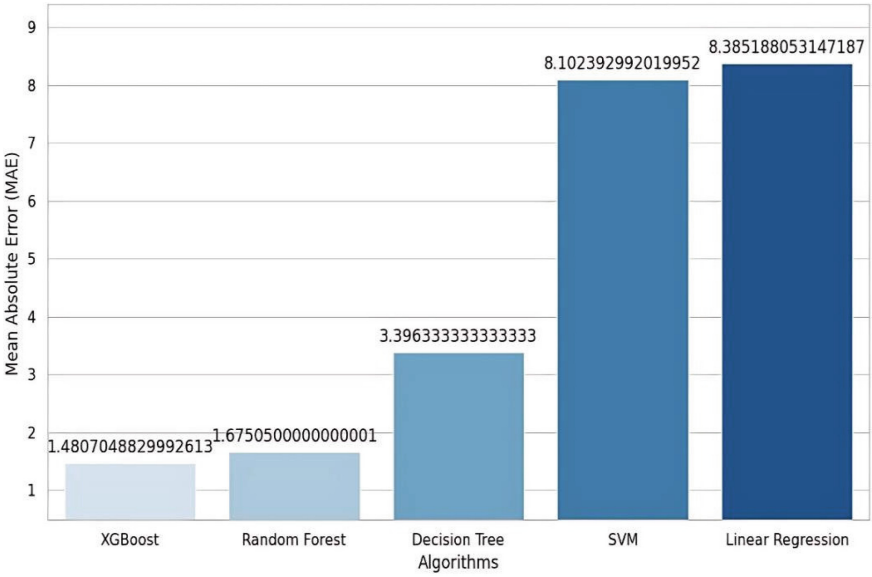


Fig. 4. MAE Comparison of Calorie Burnt Prediction

Figure 4 shows the outcomes across several models using a visualisation approach. The largest Mean Absolute Error (MAE) value is shown by the tallest bar in this graph, while the lowest MAE value—the desired outcome of the study is represented by the shortest bar.

4.2 Prediction

Finally, calorie burn is predicted using the data set to validate the outcome (see Fig. 5).

The performance of a trained model on random data sets is shown in Fig. 5 above. The model can predict the proper data, as seen in the figure.

```
missing_feature_value = 0
X_array = np.array([[1, 25, 160, 24, 102, 40.3, missing_feature_value]])
y_pred = model1.predict(X_array)
print("Prediction:", round(y_pred[0], 2))


     Prediction: 132.61
```

Fig. 5. Predicted Data

5 Conclusion

This study reveals that XGBoost emerged as the standout performer among the five machine learning algorithms we tested for calorie prediction based on dietary and activity data. XGBoost consistently delivered the most accurate predictions, outshining Random Forest, Linear Regression, Decision Trees, and Support Vector Machine. These findings provide clear guidance for individuals seeking precise calorie estimations, highlighting XGBoost as the algorithm of choice. Future research endeavors may explore further refinements and enhancements to harness the full potential of XGBoost in calorie prediction, ultimately advancing our capabilities in health and nutrition monitoring.

6 Future Scope

Further more improvement in calorie burn prediction system using machine learning is promising and holds potential for various applications in the fields of healthcare. Machine learning algorithms can be used to create personalized fitness programs. By constantly monitoring a person's work, diet, and physical data, the body can adjust and optimize exercise and nutrition recommendations to increase calorie count and overall health. The calorie burn estimation will be embedded in medical devices and medical facilities to help control diseases such as obesity, diabetes and heart disease. Doctors can use this technology to create better treatment plans and remotely monitor patients' progress. Machine learning algorithms can analyze user behavior and provide longterm behavioral analytics. This can help users identify patterns that affect calorie burn and make sustainable lifestyle changes.

Acknowledgment. We express our gratitude to everyone who assisted us to successfully complete this research paper. We would like to extend our appreciation to Ms. Shivani Trivedi for her invaluable guidance, unwavering support, and insightful advice. Her counsel significantly contributed to our project's leadership, research analysis, and overall project improvement. We want to express our gratitude to the whole team of ABESEC Department of Computer Science and Engineering for generously granting us the necessary materials and information. Lastly, we would like to convey our heartfelt thanks to our family members and our circle of colleagues for the consistent and unwavering guidance throughout the endeavour. We extend our heartfelt gratitude to the inside evaluators who contributed insightful remarks on a previous version of the paper.

References

1. Munguia Tapia, E.: Using machine learning for real-time activity recognition and estimation of energy expenditure. Massachusetts Institute of Technology (2008)
2. Liang, Y., Li, J.: Computer vision-based food calorie estimation: dataset, method, and experiment (2017)
3. Reichherzer, T., Timm, M., Earley, N., Reyes, N., Kumar, V.: Using machine learning techniques to track individuals & their fitness activities. In: CATA 2017, pp. 119–124. ISCA (2017)
4. Falsetti, H., Blau, S., Burke, E., Smith, K.: Heart rate response and calories burned in a spinning® workout. Spinning® Instructor Manual. Mad Dogg Athletics, Inc, Venice, Italy B1 (2010)
5. Nipas, M., et al.: Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm. IEEE (2022)
6. Manjunathan, N.: Feature selection intent machine learning based conjecturing workout burnt calories. Turkish J. Comp. Math. Educ. **12**(9), 1729–1742 (2021)
7. Kalpesh, J., Rushikesh, J., Swaraj, K., Rohan, K., Bharadwaj, R.: Human Physical Activities Based Calorie Burn Calculator Using LSTM. Springer International Publishing (2022)
8. Basiruddin, N.H., Zulkifli, Z., Ahmad, S.: Malay festive seasons food recognition for calorie detection using SVM and ECOC approaches/Nurul Hafiza Binti Basiruddin, Zalikha Zulkifli and Samsiah Ahmad. Math. Sci. Informat. J. (MIJ) **3**(2), 55–64 (2022)
9. Acharya, M.S., Armaan, A., Antony, A.S.: A comparison of regression models for prediction of graduate admissions. In: 2019 international conference on computational intelligence in data science, pp. 1–5. IEEE (2019)
10. Darapaneni, N., et al.: Food Image Recognition and Calorie Prediction. In: 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–6. IEEE (2021)
11. Su, X., Yan, X., Tsai, C.L.: Linear regression. Wiley Interdisciplinary Reviews: Computational Statistics **4**(3), 275–294 (2012)
12. Gaikwad, V., Khodke, M., Shahare, S., Terkar, P., Talmale, R.: Implementation of XGBoost Regression for Calories Burnt Prediction Using R. In: International Conference on Advances in Computing and Data Sciences, pp. 377–390. Springer Nature Switzerland (2023)
13. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794 (2016)
14. Wang, H., Hu, D.: Comparison of SVM and LS-SVM for regression. In: 2005 International conference on neural networks and brain, vol. 1, pp. 279–283. IEEE (2005)
15. Biau, G., Scornet, E.: A random forest guided tour. TEST **25**, 197–227 (2016)



FarmScan: Smart Rotten Fruit and Vegetable Detector

Shivani Trivedi, Ankit Yadav, Suraj Kumar Gupta, and Shivam Gupta(✉)

Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, India
shivam216702@gmail.com

Abstract. Quality evaluation of fruits and vegetables is a vital factor in maintaining qualities of food and ensures consumer satisfaction. Freshness is an essential asset of fruit and vegetable quality that impacts the physical health and motivation of consumers. This paper presents a transformative approach to revolutionizing quality evaluation in fruits and vegetables by applying image processing and machine learning algorithms, specifically utilizing the GrabCut algorithm for background subtraction, The traditional subjective methods of assessing product quality are replaced with objective and data-driven techniques. Through case studies, the efficacy of this approach is demonstrated, showcasing its potential to enhance quality assessment precision and efficiency. This innovative fusion of image processing and machine learning presented a promising avenue to redefine industry standards, ensuring higher consumer satisfaction and reduced wastage through consistent and reliable quality evaluation.

Keywords: Image processing · image segmentation · background subtraction · grab cut algorithm

1 Introduction

In the recent past, the quality assurance of food products has achieved notable attention due to its critical significance in affirming consumer safety and fulfillment. The modern food industry is described by its global reach and varying range of products, making traditional primer inspection procedures both time-saving and error-prone. Food contamination is a part of significant buyer concerns in a growing nation due to its direct punch on individual person health. Food contamination appears due to falling-off production values and more significant market appeal than supply, especially for latter-day COVID-19 pandemic circumstances [1]. As a resolution, the incorporation of image segmentation methods with advanced machine learning algorithms has evolved as a promising approach to automating and enhancing food quality inspection procedures. Assessment of fruits can be carried out by manual operation Owing to inconsistent outcomes and substantial time requirements, it is essential to comply with the quality tests with the automated systems [6]. Customer fulfillment of needs and internal satisfaction towards the foods are essential for the growth of the market and quality assurance of the product [2].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 272–280, 2025.
https://doi.org/10.1007/978-3-031-75167-7_22

The determined system's justification lies in the achievement of high-quality images under controlled conditions. Images serve as the raw input that fuels subsequent methods, making their purity and representational accuracy supreme. To accomplish this, a precisely labeled dataset encompassing discrete quality variations is collected, forming the bedrock of supervised learning. This dataset not only facilitates algorithm training but also enables the system to conclude its understanding of quality essences across different food items. The problem experienced in the human review is minor in reliability and correctness as carried out by manual operation by the human eye will take more time and energy [7].

Image segmentation, the subsequent step, involves the separation of acquired images into definite regions. In summary, this research paper received a comprehensive passage to food quality inspection by synergizing image segmentation and machine learning techniques. The amalgamation of these components promises a prototype shift in the efficiency and accuracy of food quality estimation. As consumer demands for higher quality standards and increased product variety persist, the proposed framework offers a timely and innovative clarification to meet these challenges head-on. Through empirical evaluation and real-world application, the effectiveness and versatility of the proposed system are authenticated, highlighting its potential to revolutionize the field of food quality inspection.

The findings in this document, as we first showed a review of related work in quality evaluation of fruits and vegetables, assessing the qualities of the foods and how they complement and ensure the consumers' satisfaction. Then we discussed the methodology of our system using different algorithms and performed implementation on datasets of different fruits and vegetables and obtained results with an accuracy of 95%. We also provided the novelty we proposed in our system. We concluded the paper with the success of our result in depicting the freshness of fruits and vegetables with its future scope.

2 Literature Review

Tan et al. [1] discussed a method that will examine and depict the features currently applied to the image, Odor, electromagnetic, and taste, like to the food and beverages evaluation and it will further analyze and examine every technique and add suggestions related to recent technology. This method will consider technology alliance and the entanglement of deep learning to enable numerous variety of recent technologies, like electromagnetic sensing, odor, taste, and imaging, which help in the inspection of food quality.

Azgomi et al. [3] presented a method that focused on detecting apple diseases by image processing and an artificial neural network and focused on providing a sound and accurate method for the diagnosis of three fungal diseases as it can help in disease diagnosis and prevent crop losses and also uses color and texture features. This provides a low-cost method of apple disease diagnosis using a neural network and fruit categorization into four classes scab, bitter rot, black rot, and healthy fruits.

Kyaw et al. [4] proposed a method that examines the potential of artificial intelligence in food processing and also provides the innovations and ways with probable appliances in adopting automated processing with technological assistance to enhance productivity.

Engaging in the conceptualization of an automated food manufacturing line consists of many operational layers and processes aimed at increasing the accuracy of microbial safety and quality assurance measures for liquid foods, including milk and beverages, is ensured.

Mukhiddinov et al. [13] emphasizes the critical role of classifying the freshness of fruits and vegetables in the food industry. This system involves optimizing the YOLOv4 model, creating an extensive image dataset, employing data augmentation techniques, and enhancing the model's backbone with the Mish activation function for more precise detection. The experimental results reveal that the proposed method outperforms previous YOLO models, achieving an average precision rate of 50.4 percent. This system holds great promise for automating real-time fruit and vegetable classification in the food industry and marketplaces. Furthermore, it can serve as a valuable tool for visually impaired individuals, enabling them to select fresh produce and avoid potential health risks associated with spoiled food.

Anwar et al. [5] explored a powerful non-destructive method for food quality assessment by combining electronic nose systems with machine learning algorithms. This method examined various food items such as meat, dairy, edible oil, fish, tea, and coffee products. They also reviewed the effects of the performance of various combinations of machine learning models and gas sensors improved the accuracy and minimized the bias associated with the manual evaluation and human error by using advanced machine learning techniques.

3 Methodology

The proposed methodology of our system as per Figure 1, for quality evaluation of fruits and vegetables is structured with several stages as shown in the figure. These stages can be described as the following steps:

- Firstly, we input the dataset of several images of fruits and vegetables.
- Gaussian elimination is employed for preprocessing the image dataset. This method facilitates noise removal through processes such as image restoration, color space transformation, image filtering, and enhancement. These enhancements collectively contribute to improved image quality, thereby enhancing the accuracy of image classification.
- Perform image segmentation and background subtraction using the method of the GrabCut Algorithm.
- We extract various features from the images, including color histograms, texture patterns, edge, shape, and size descriptors. The categorization of fruits and vegetables is accomplished through image processing methods.
- The preprocessed image dataset undergoes classification using Convolutional Neural Network (CNN). Subsequently, a quality analysis is performed on the images to determine whether they are rotten or not.

Why GrabCut Algorithm:

- It excels in accurately segmenting objects even in the presence of complex backgrounds..

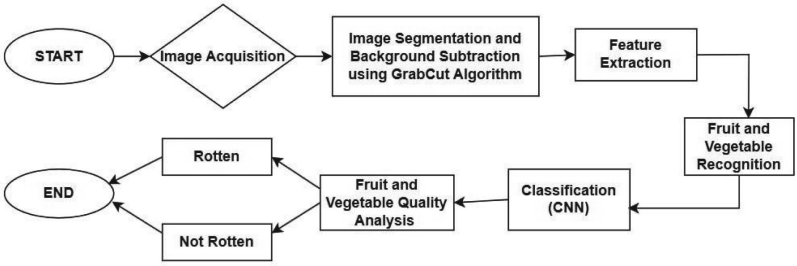


Fig. 1. Process flow diagram of our system

- Its computational efficiency makes it a practical choice for real-time or largescale quality evaluation processes.
- It develops a more robust and accurate quality evaluation system for the higherquality segmented images.
- It reduces the time and effort required for manual inspection, enhancing the efficiency of the overall evaluation process.

3.1 Image Acquisition

Image acquisition is a fundamental step in the quality evaluation of fruits and vegetables using image assessment methods. Image acquisition is a technique by which we achieve a digital form of an image [8]. It uses Gaussian elimination to preprocess the dataset of the images. This Gaussian elimination is used to improve the quality of the images which helps in better classification of images. It ensures that the image data is enhanced for accurate analysis. Firstly, color correction and normalization techniques are applied to make uniformity in color representation across all images. Contrast adjustment techniques are being applied to emphasize finer details, especially in regions of changing textures or shades. Moreover, techniques like histogram equalization can be applied to improve the overall distribution of pixel values, improving the graphic depiction of the result. After applying these preprocessing steps, the images are enhanced and streamlined, setting the stage for more accurate and reliable quality evaluation of fruits and vegetables using subsequent machine learning and image processing techniques. This ensures that the analysis has been done based on advanced high-quality image data, ultimately leading to more accurate evaluations of the output.

3.2 Image Segmentation and Background Subtraction

Image segmentation and background subtraction, enabled by algorithms like GrabCut, play a crucial role in the quality evaluation of fruits and vegetables. The GrabCut algorithm uses an iterative optimization process, outset directed by user input, to accurately segment objects within an image. In terms of fruits and vegetables evaluation, GrabCut can precisely separate similar components from the background. This is particularly crucial when performing analysis on items like packaged foods, where perceptive fine details are necessary, and by iteratively refining the segmentation boundaries, GrabCut provides an elaborated and accurate representation of the food or beverage. In tandem

with background subtraction, this algorithm further improves quality assessment [9]. By removing the background, the focus is completely on the food product, eliminating potential distractions. This is especially valuable when evaluating items with complex textures or fine details. Together, image segmentation and background subtraction using the GrabCut algorithm allow a more accurate and efficient quality evaluation process, providing detailed insights into the attributes and features of fruits and vegetables.

3.3 Feature Extraction

Feature extraction is an important process in the quality evaluation of fruits and vegetables using machine learning and image processing techniques. It involves the recognition and extraction of essential attributes or features from the image data of food items. These features deliver the basis for subsequent analysis and decision-making. In terms of food quality assessment, different types of features can be extracted like Color-based features which tell us about the hues, saturation, and intensity levels, through which it helps in evaluating the characteristics like ripeness or freshness of the foods. Texture features, which improve the surface qualities like smoothness, roughness, or graininess. Shape features that tell us about the geometrical properties of the food items, aiding in tasks like portion size estimation. Additionally, statistical features like mean, standard deviation, and skewness provide a numerical summary of pixel intensity distributions, contributing to a comprehensive evaluation. By extracting a combination of these features, a detailed and multi-dimensional representation of the food and beverage images is obtained [10]. This feature set serves as a basis for machine learning models to learn and make accurate quality assessments. Through this process, feature extraction links the gap between raw visual data and actionable insights, enabling a robust and automated approach to food and beverage quality evaluation.

3.4 Convolutional Neural Network (CNN)

Merging attributes class and learning, the convolutional neural network (CNN) is a feed-forward neural network with convolution estimation and deep structure [11]. It is one of the representative algorithms of deep learning. The network consists of four convolutional layers to acquire various levels of characteristics from the images [12]. The utilization of Convolutional Neural Networks (CNNs) in the quality evaluation of fruits and vegetables performs a significant advancement in automated assessment processes. CNNs are adept at learning complex patterns and features from images, making them particularly well-suited for this task. In this term, after pre-processing and feature extraction, the CNN is processed on a dataset of labeled fruits and vegetables images, where every class belongs to a particular quality category. Through successive layers of convolutions, pooling, and fully connected nodes, the CNN learns to discern subtle visual cues indicative of quality attributes. It can recognize color variations, texture patterns, and shape characteristics that are linked with freshness, ripeness, or other quality properties. The trained CNN can then discover its learning to give predictions on new, unseen images, effectively automating the quality evaluation process. It provides a more objective and consistent assessment and minimizes the need for manual inspection. Overall, the integration of CNNs in fruits and vegetables quality evaluation shows significant

potential for improving efficiency, accuracy, and reliability in providing high-quality products that reach consumers.

3.5 Food Quality Analysis

Food quality analysis, an important part of ensuring consumer satisfaction and safety of foods, involves a complete evaluation of various attributes and properties of fruits and vegetables. It checks the quality of the food whether it is rotten or nonrotten. This process involves a combination of techniques, including sensory evaluation, chemical analysis, and some advanced technologies like image processing and machine learning. Image processing techniques give information on the extraction of objects from images of food items, obtaining details related to color, texture, shape, and overall appearance. These features present important characteristics of freshness, ripeness, and potential defects. Machine learning models, mainly Convolutional Neural Networks (CNNs), excel at learning and detecting these complex visual patterns, enabling automated quality evaluations. Also, chemical analysis techniques judge the composition and nutritional content of food, ensuring compliance with safety and labeling standards. Sensory evaluation, involving human sensory opinions like taste, smell, and texture, gives invaluable vision into the overall eating experience. With the combination of both approaches, food quality analysis provides a holistic and scientifically grounded assessment, which guarantees that customers receive high-standard products. This process secures the reputation and success of food producers and distributors in a competitive market and also benefits consumers by ensuring safe and enjoyable food experiences.

4 Results and Implementation

The proposed system was designed on the platform Jupyter for training the dataset and modeling our data to identify rotten fruits and vegetables. We modeled our data for the fruits and vegetables that keep us healthy and fresh. The dataset was firstly classified on the basic scale such as its size, image quality, etc. Then, Image acquisition is performed for the improvement of quality of the images.

As per Figure 2, we can see some of the images of our datasets that have 14 different fruits and vegetables that have more than 12000 images. We trained the dataset on sets of different images that have several images of rotten and fresh fruits and vegetables. Now, we applied the Grab Cut algorithm, an iterative optimization process that is used for quality evaluation from similar components from fruits and vegetables. The feature is being extracted that recognizes the freshness of fruits and vegetables.

Classification of the images for the rotten and fresh fruits and vegetables is being performed using CNN (Convolutional Neural Network). Figure 3, depicts the classified fruits and vegetables on our dataset which we trained to classify as if they are rotten or fresh. We can see that the results show some of the fruits (Apple, Banana, etc.) and vegetables (Tomato, etc.).

After the training, we obtained the results classifying the freshness of the fruits and vegetables with an accuracy of 95%.

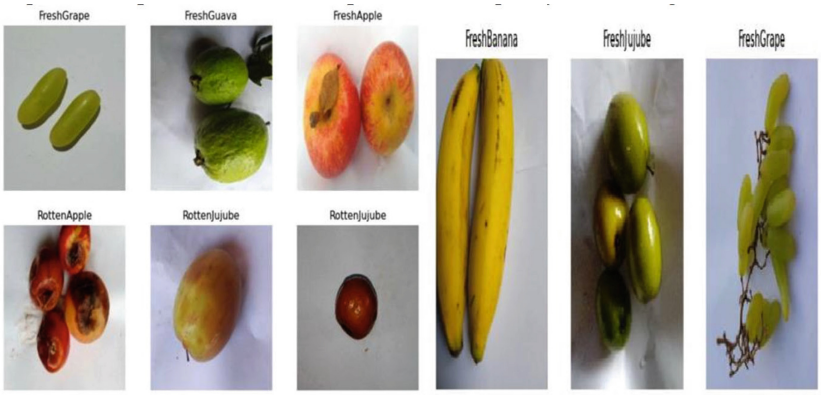


Fig. 2. Some related images of the fruits and vegetables from the dataset

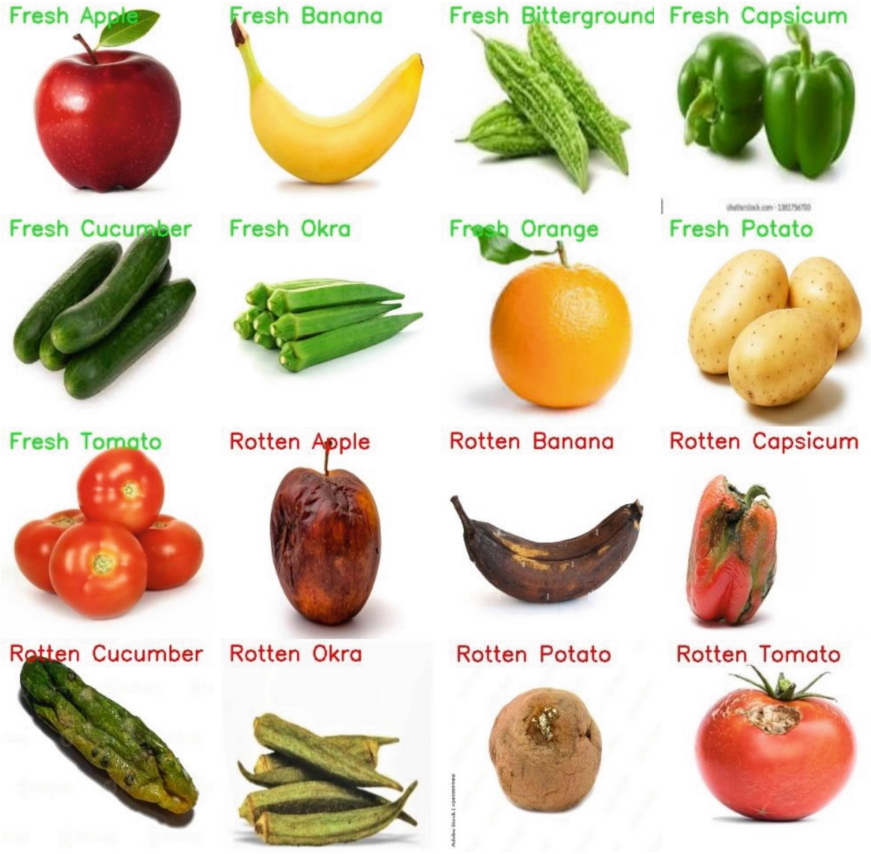


Fig. 3. Some classified rotten and fresh fruits and vegetables

These results depicted in Figure 3, showcase a commendable level of accuracy. This achievement was realized through a rigorous process involving numerous classification iterations with diverse sets of images of each fruit and vegetable. The reported accuracy is a reliable measure derived from the careful averaging of multiple accuracy values obtained in multiple classification rounds. The final calculated accuracy stands at an impressive approximately 95%.

5 Novelty

- More number of fruits and vegetables: Our project detected 16 different types of fruits and vegetables and detected their freshness. The existing model can detect only one or two fruits or vegetables.
- Multi-Task Learning: We have implemented multi-task learning in our model. In addition to detecting whether a fruit or vegetable is rotten or fresh, incorporate additional tasks such as identifying the specific type of fruit or vegetable.
- This shows that the results of our project are more accurate than the existing projects.

6 Conclusion

This proposed method presents a way for the monitoring and analysis of fruit and vegetable spoilage using a Grab cut algorithm, an adequate preprocessing step to section and isolates the fields of interest in the images, specifically the fraction of fruits and vegetables that exhibit signs of rotten, and CNN for extracting significant features from images and making defined classifications. The analysis determined in this paper is able to detect the spoilage of fruits and vegetables. Moreover, this method will help us to prevent fruits and vegetables from rotten by suspecting the analysis of their lifespan. This will analyze the fruits and vegetables and detect whether it is rotten or not. The model is trained well to detect its spoilage.

7 Future Scope

The future of smart rotten fruit and vegetable detectors is flashing, offering a field of interesting possibilities to direct critical points such as food waste reduction, food quality assurance, and consumer health. As technology continues to forward, many methods have more enhanced algorithms and classifications with good accuracy will have the potential to become indispensable appliances at different phases of the food supply chain, from production and transportation to consumption in households. With the potential for better accuracy, integration with AI and machine learning, and broader acceptance, smart detectors can add to a more feasible and efficient food system while advantaging both businesses and consumers. As society grows increasingly aware of the priority of reducing food waste and establishing food safety, the role of smart detectors in achieving these goals is likely to expand and advance.

Acknowledgement. I am thankful to all the respective people who supported me on the way to the conclusion of the research and heartfelt gratitude to my dedicated supervisor, Ms. Shivani Trivedi, for her invaluable guidance, and to my team members. I also want to express our gratitude to the entire staff at ABESEC's Department of Computer Science and Engineering for giving us the resources and information I needed. My family and friends provided unwavering support during this research journey. This work is a collective effort, and I am immensely thankful to all who played a vital role in its completion.

References

1. Tan, W.K., Husin, Z., Yasruddin, M.L., Ismail, M.A.H.: Recent technology for food and beverage quality assessment: a review. J. Food Sci. Technol. **60**(6), 1681–1694 (2023)
2. Nizam, N.Z., Alkaabi, M.R.A.Z., Husseini, S.A.: Customer service quality assessment and customers' satisfaction in food and beverage in mcdonald's restaurant in united arab emirates. J. Posit. School Psychol. **6**(3), 4684–4693 (2022)
3. Azgomi, H., Haredasht, F.R., Motlagh, M.R.S.: Diagnosis of some apple fruit diseases by using image processing and artificial neural network. Food Control **145**, 109484 (2023)
4. Kyaw, K.S., Adegoke, S.C., Ajani, C.K., Nwabor, O.F., Onyeaka, H.: Toward in-process technology-aided automation for enhanced microbial food safety and quality assurance in milk and beverages processing. Critical Reviews in Food Science and Nutrition, 1–21 (2022)
5. Anwar, H., Anwar, T., Murtaza, S.: Review on food quality assessment using machine learning and electronic nose system. Biosensors and Bioelectronics: X, 100365 (2023)
6. Mehta, D., Choudhury, T., Sehgal, S., Sarkar, T.: Fruit quality analysis using modern computer vision methodologies. In: 2021 IEEE Madras Section Conference (MASCON), pp. 1–6. IEEE (2021)
7. Foong, C.C., Meng, G.K., Tze, L.L.: Convolutional neural network based rotten fruit detection using resnet50. In: 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC), pp. 75–80. IEEE (2021)
8. Perez-Sanz, F., Navarro, P.J., Egea-Cortines, M.: Plant phenomics: an overview of image acquisition technologies and image data analysis algorithms. GigaScience **6**(11), gix092 (2017)
9. Arulnathan, D.N., Koay, B.C. W., Lai, W.K., Ong, T.K., Lim, L.L.: Background subtraction for accurate palm oil fruitlet ripeness detection. In: 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), pp. 48–53. IEEE (2022)
10. Bonilla, J.C., Sørensen, J.L., Warming, A.S., Clausen, M.P.: Quantitative image analysis of protein foam microstructure and its correlation with rheological properties: Egg white foam. Food Hydrocolloids **133**(108010), 2022 (2022)
11. Hao, Y., Li, X., Zhang, C., Lei, Z.: Online inspection of browning in Yali pears using visible-near infrared spectroscopy and interpretable spectrogram-based CNN modeling. Biosensors **13**(2), 203 (2023)
12. Jana, S., Parekh, R., Sarkar, B.: Detection of rotten fruits and vegetables using deep learning. Computer Vision and Machine Learning in Agriculture 31–49 (2021)
13. Mukhiddinov, M., Muminov, A., Cho, J.: Improved classification approach for fruits and vegetables freshness based on deep learning. Sensors **22**(21), 8192 (2022)



Closing the Data Divide in COVID-19 X-ray Datasets: Advancing Diagnosis Through Stable Diffusion-Based Synthetic Image Generation

Saumya Mehra(✉) and Bhawna Jain(✉)

Computer Science Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW), Madrasa Road, Delhi 110006, India
saumya.mehra23@gmail.com, bhawnajain@igdtuw.ac.in

Abstract. Addressing the limited availability of COVID-19 patient X-ray images for precise diagnostic tools, this research harnesses Stable Diffusion, a robust generative modelling technique. Deep neural networks are employed to learn the underlying data distribution and generate synthetic X-ray images. By employing the Stable Diffusion model, the study aims to overcome the scarcity of annotated COVID-19 patient X-ray datasets. Through a process of fine-tuning with a Kaggle dataset, synthetic X-ray images closely resembling authentic patient X-rays are created. This synthetic data generation technique effectively addresses the data scarcity challenge, facilitating more robust model training and improving the accuracy of COVID-19 detection and categorization. The methodology for utilizing Stable Diffusion for data augmentation involves several crucial steps and processes. Initially, the model is fine-tuned to capture essential characteristics and patterns present in real patient X-rays. Following fine-tuning, the model generates synthetic X-ray images that closely mimic genuine patient X-rays by drawing from the learned data distribution. These synthetic images are then seamlessly integrated into the original dataset, thereby expanding and diversifying the collection of X-ray images. In the final stage, machine learning models, such as deep neural networks, are trained using the augmented dataset. This approach leverages information from both real and synthetic images, significantly enhancing the accuracy and robustness of COVID-19 case detection and classification.

Keywords: Stable Diffusion · COVID 19 · Generative AI

1 Introduction

The COVID-19 pandemic has posed unprecedented challenges to healthcare systems globally, necessitating the development of accurate and efficient diagnostic tools for the identification and monitoring of infected individuals. The limited availability of annotated X-ray images from COVID-19 patients presents a significant barrier in the development of reliable machine learning models for automated diagnosis. In this context, Stable diffusion, a robust generative modelling technique, offers a promising solution as stated in [1]. Stable diffusion refers to a computational approach that allows for the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 281–291, 2025.
https://doi.org/10.1007/978-3-031-75167-7_23

generation of high-quality synthetic images by fine-tuning a diffusion model with an existing dataset. The technique has garnered attention for its ability to generate realistic and diverse synthetic images across different domains. Researchers can solve the shortage of annotated X-ray pictures from COVID-19 patients by using Stable diffusion. The procedure entails fine-tuning the Stable diffusion model with a dataset gathered from sites such as Kaggle, which comprises X-ray pictures of COVID-19 patients. The model learns the underlying patterns and traits inherent in genuine patient X-rays through this fine-tuning process, allowing it to create synthetic pictures that closely match actual X-rays [2]. These simulated X-ray pictures are a great addition to the current collection, considerably increasing its size and diversity. The enhanced dataset, which includes both real patient X-rays and synthetic pictures, is subsequently used to train machine learning algorithms. Researchers can increase their accuracy and resilience in detecting and categorising COVID-19 instances by training these models on the expanded dataset [3]. To address the aforementioned issue, researchers investigated the use of Stable diffusion, a strong generative modelling tool, to produce synthetic X-ray pictures that can supplement the restricted dataset. The goal is to generate a larger and more diversified dataset that properly represents the main properties of genuine patient X-rays by fine-tuning the Stable diffusion model using a dataset supplied from Kaggle that includes X-ray pictures from COVID-19 patients. This enriched dataset may then be used to train and improve machine learning models built for automated COVID-19 diagnosis [4].

The use of stable diffusion to generate synthetic X-ray pictures has various advantages. For starters, it tackles the issue of data scarcity, as having a bigger dataset allows for more robust and accurate model training. Second, it enables the investigation of a broader spectrum of illness presentations and changes, allowing for a better knowledge of the disease's radiological properties. Furthermore, the produced dataset allows you to test the generalizability of machine learning models by introducing variables that were not included in the original dataset [5]. The quality and realism of the generated synthetic X-ray images are evaluated through visual inspection and quantitative metrics. Furthermore, the impact of augmenting the dataset with synthetic images on the performance of machine-learning models for COVID-19 diagnosis is investigated [6] . The study aims to contribute to the field of medical imaging and AI-assisted diagnosis by showcasing the potential of Stable diffusion in addressing data scarcity and enhancing the accuracy and reliability of automated COVID-19 detection systems. Clinical Importance: The development of accurate and efficient diagnostic tools for COVID-19 is of paramount importance in effectively managing the ongoing pandemic. However, the limited availability of annotated X-ray images from COVID-19 patients hampers the development of reliable AI-based diagnostic systems. The clinical importance of this research lies in its potential to address the scarcity of annotated data by leveraging Stable diffusion, a generative modeling technique [7]. By generating synthetic X-ray images that closely resemble real patient X-rays, we can significantly augment the available dataset. This augmented dataset enables the training of machine learning models on a larger and more diverse set of COVID-19 cases, thereby improving their accuracy and reliability in automated diagnosis. The use of Stable diffusion to generate synthetic X-ray images offers valuable clinical implications [8]. Firstly, it enhances the robustness and generalizability

of AI models by providing a more comprehensive representation of COVID-19 manifestations. Moreover, the availability of a larger dataset through synthetic image generation allows for the exploration of rare or subtle radiological features that may be challenging to encounter in the original dataset. This expanded understanding of disease variations contributes to the refinement of diagnostic criteria [9].

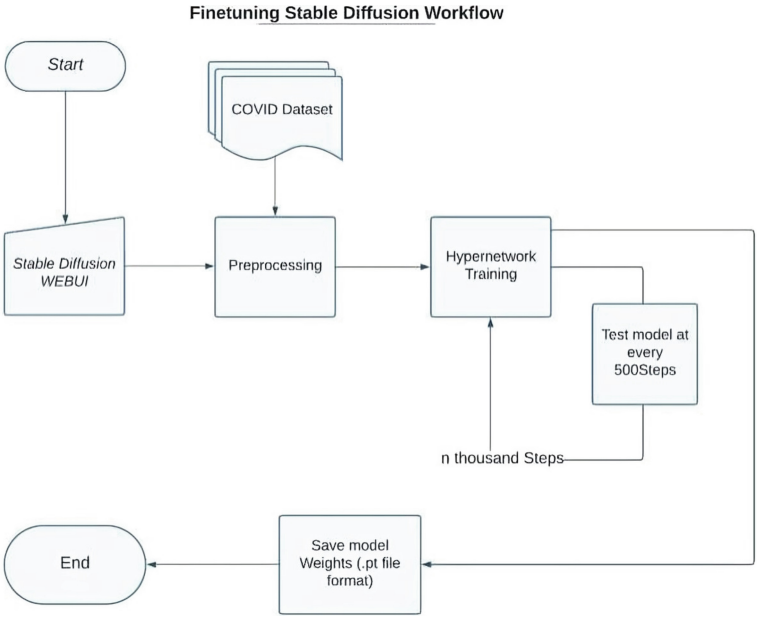


Fig. 1. Flowchart for Stable Diffusion Model of this Study

The remainder of the paper is organised as follows: Section II delves into the existing literature and previous studies on the use of medical imaging and the challenges associated. Section III is Implementation which contains the details, methodology and steps involved in utilizing Stable diffusion for generating synthetic X-ray images. Section IV presents the findings of the research. It includes the evaluation of the quality and realism of the generated synthetic X-ray images. Section V discusses the limitations and obstacles encountered during the research process. It addresses the potential difficulties and complexities in implementing Stable diffusion for data augmentation. Section VI concludes the paper by summarizing the findings on stable diffusion in generating synthetic X-ray and its effect on detecting COVID-19 cases.

2 Background and Related Work

Stable diffusion operates by iteratively applying a series of diffusion steps to an initial noise vector, gradually transforming it into a realistic image. Each diffusion step involves a controlled blending of the current image with a noise signal, guided by a diffusion model. This diffusion model, typically based on deep neural networks, captures

the intricate mapping between the noisy image and the target distribution, effectively learning the underlying data distribution from the available training data. It discusses the opportunities for extending the application of Stable diffusion beyond COVID-19 to other respiratory diseases and imaging modalities [10]. Unlike traditional generative models, Stable diffusion avoids the mode collapse issue and produces diverse and high-fidelity samples. It achieves this by employing a diffusion process that enables fine-grained control over the level of noise in the generated images. By gradually reducing the noise level during the diffusion process, Stable diffusion ensures that the generated images exhibit increasingly realistic and precise features. The original stable diffusion model is shown in Figure 1. The Stable diffusion framework is employed in this research to generate synthetic X-ray images that closely resemble real patient X-rays. By fine-tuning the Stable diffusion model on a dataset of X-ray images from COVID-19 patients, the model captures the underlying patterns and characteristics of the disease [10]. This allows for the generation of synthetic images that exhibit realistic radiological features specific to COVID-19, including opacities, consolidations, and groundglass opacities. By utilizing the Stable diffusion framework, the researchers effectively expand the dataset and create a more comprehensive representation of COVID-19 manifestations, thereby enhancing the training of machine learning models for automated diagnosis [11] . By leveraging the capabilities of Stable diffusion, the limitations imposed by the available dataset are effectively overcome, and a significant number of additional images capturing the diverse variations of COVID-19 in X-ray imaging are generated. This innovative approach proves instrumental inenhancing the performance and robustness of AI-based diagnostic systems, thereby enabling more accurate and reliable detection of COVID-19 cases.

3 Implementation

The implementation of the research involved utilizing the Stable diffusion repository available on GitHub as the foundation. This repository provided a well-established framework for generative modeling using Stable diffusion [12]. To train the model on a dataset of COVID-19 X-ray images, a curated dataset from Kaggle, consisting of images from COVID-19 patients, was obtained. The dataset was preprocessed to ensure uniform image sizes and incorporate appropriate data augmentation techniques, enhancing the model's robustness and generalization capabilities. Subsequently, the Stable diffusion model was fine-tuned by adjusting its hyperparameters and optimizing its architecture to align with the specific characteristics of the dataset [13]. To maximize the model's ability to create realistic synthetic X-ray pictures, an iterative fine-tuning approach was used. After fine-tuning the model, it was used to produce synthetic pictures. The learnt weights and latent space representations were used to generate sets of 100 synthetic pictures at a time using a batch generation technique. This iterative generation approach allowed for the rapid production of many different and realistic synthetic X-ray pictures [14]. The technique flowchart used in this investigation is shown in Figure 2. The quality and authenticity of the produced photographs were constantly monitored throughout the installation. To confirm that the synthetic pictures properly reflected the major radiological aspects of COVID-19, visual inspections, quantitative analyses, and

comparisons with genuine patient X-rays were performed. The goal of this rigorous assessment method was to confirm the realism and authenticity of the synthetic X-ray pictures produced by the Stable diffusion model.

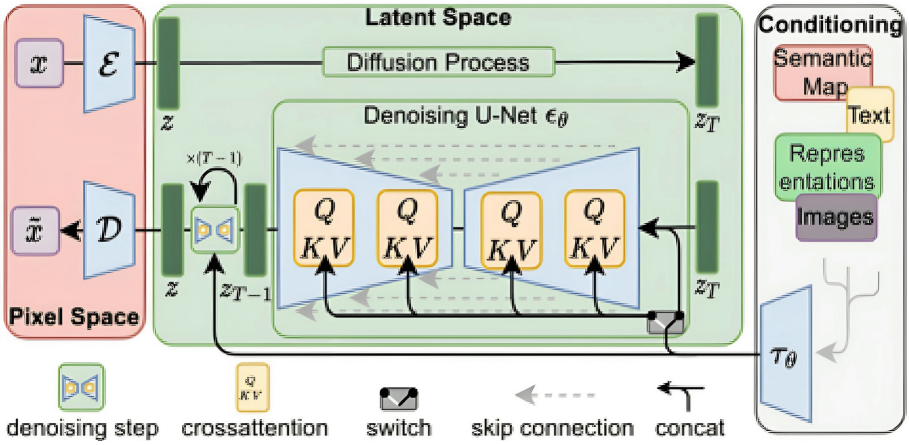


Fig. 2. Original Stable Diffusion Model

Powerful computational resources, including GPUs, were used to facilitate implementation and expedite the generating process. This enabled efficient and effective model inference and picture production, allowing for the creation of many synthetic images. The Stable diffusion model was fine-tuned using a COVID-19 X-ray dataset, followed by batch-wise production of realistic synthetic pictures [15]. The successful augmentation of the dataset revealed Stable diffusion's potential for creating useful synthetic data for enhancing COVID-19 diagnosis. The use of modern computer resources was critical in assuring the implementation's efficiency and scalability [16].

4 Results

A cooperation with a medical practitioner with experience in diagnosing COVID-19 was created to test the correctness of the generated synthetic X-ray pictures. The generated images, produced using the Stable diffusion model, were presented to two professional doctors for their evaluation. With meticulous attention, the doctors thoroughly examined the images, conducting a comparative analysis with real patient X-rays. These doctors were given two parameters - Good and Bad for determining the quality of the X-ray images. Their professional expertise and insights proved invaluable in assessing the quality and fidelity of synthetic images. Examples of a good and bad augmented image are shown in Figure 3. Based on the doctor's assessment, it was determined that approximately 68 percent of the generated images accurately represented the characteristic features of COVID-19 patient lungs in X-ray imaging. This result shown in Figure 4 demonstrates the potential of the Stable diffusion approach in generating synthetic X-ray images that closely resemble real cases. All the augmented images by stable diffusion

model are collectively shown in Figure 5. It is important to note that the evaluation conducted in this research involved subjective assessment by two medical professionals. While the doctor's expertise adds credibility to the evaluation, a larger and more diverse panel of medical experts would be necessary for a comprehensive and objective evaluation of the accuracy and reliability of the generated images. generating synthetic X-ray images representative of COVID-19 cases. This result underscores the potential of generative modeling techniques to address data scarcity and augment the training datasets for improved COVID-19 diagnosis. Further research and collaborations with a larger medical expert panel, along with quantitative evaluation metrics, are needed to validate the diagnostic accuracy and clinical utility of the synthetic X-ray images generated by the Stable diffusion approach. These efforts will contribute to enhancing two parameters - Good and Bad for determining the quality of the X-ray images. Their professional expertise and insights proved invaluable in assessing the quality and fidelity of synthetic images. Examples of a good and bad augmented image are shown in Figure 3. Based on the doctor's assessment, it was determined that approximately 68 percent of the generated images accurately represented the characteristic features of COVID-19 patient lungs in X-ray imaging. This result shown in Figure 4 demonstrates the potential of the Stable diffusion approach in generating synthetic X-ray images that closely resemble real cases. All the augmented images by stable diffusion model are collectively shown in Figure 5. It is important to note that the evaluation conducted in this research involved subjective assessment by two medical professionals. While the doctor's expertise adds credibility to the evaluation, a larger and more diverse panel of medical experts would be necessary for a comprehensive and objective evaluation of the accuracy and reliability of the generated images. Nevertheless, these initial findings provide a promising indication of the effectiveness of the Stable diffusion model in the reliability and trustworthiness of AI-assisted diagnostic systems and their potential impact on patient care in the context of COVID-19.

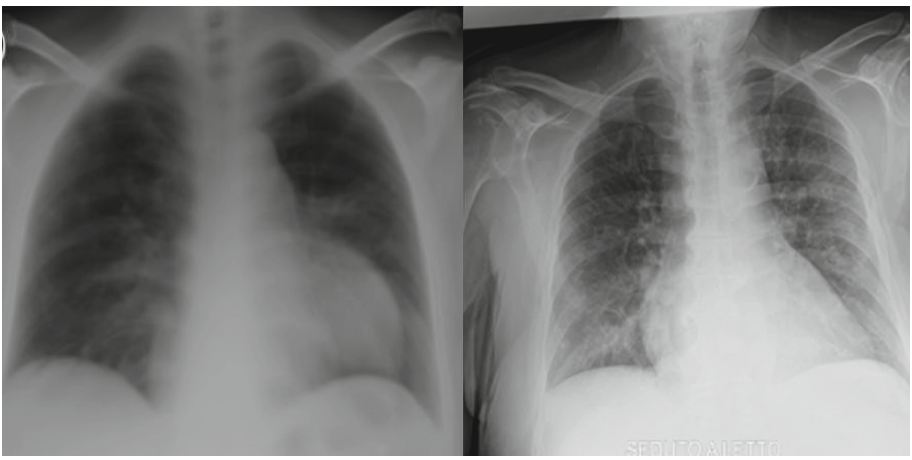


Fig. 3. Generated image(Left) and Original Image(Right)

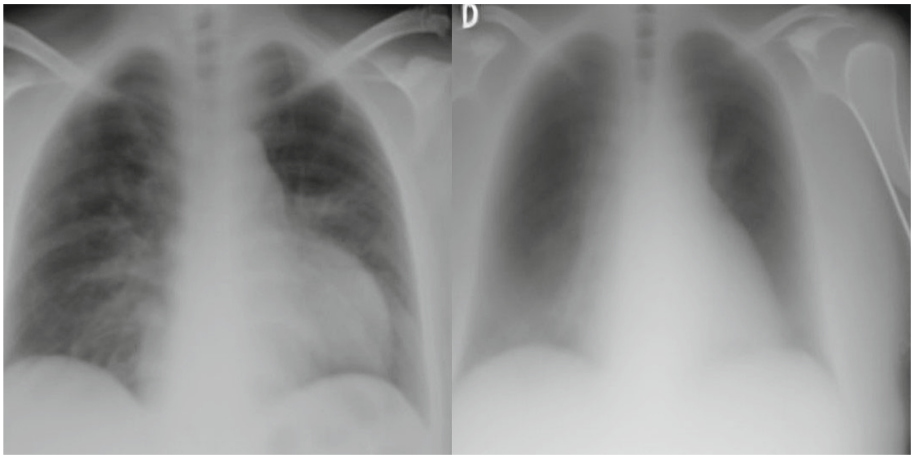


Fig. 4. Good augmented image(Left) and Badly augmented Image(Right)

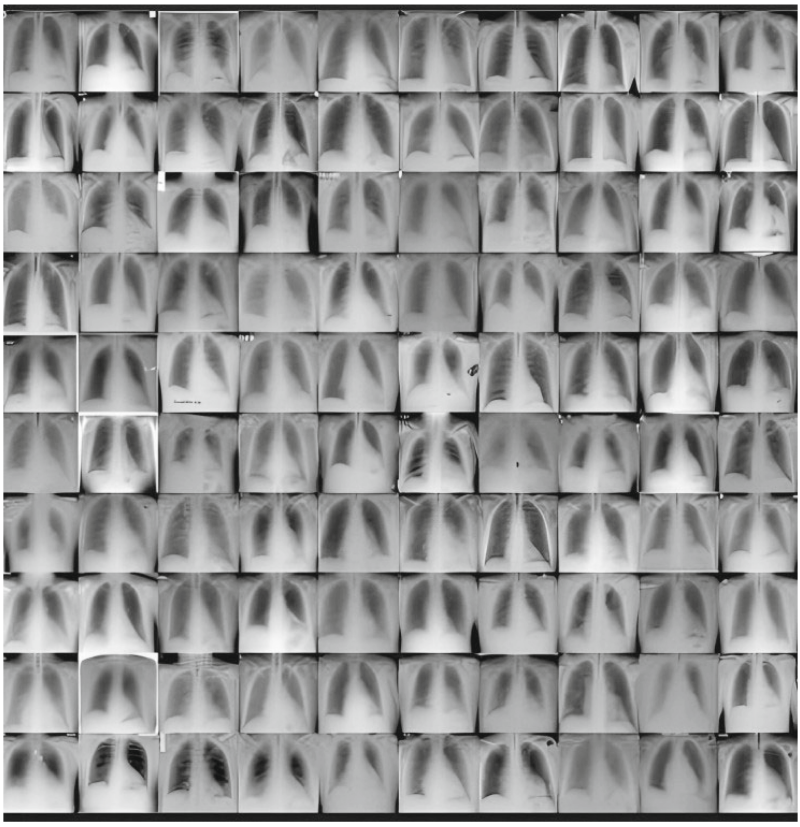


Fig. 5. All Augmented images

5 Challenges

While the application of Stable diffusion for generating synthetic X-ray images in COVID-19 diagnosis shows promising results, there are several avenues for challenges that should be addressed. These include:

1. Incorporating Clinical Metadata: Future work can involve incorporating clinical metadata, such as patient demographics, comorbidities, and laboratory results, into the Stable diffusion framework. By incorporating this additional information, the generated synthetic images can better capture the complex relationships between imaging features and patient characteristics, leading to more accurate and personalized diagnostic models.
2. Uncertainty Estimation: Estimating and quantifying uncertainty in the generated synthetic images is an essential aspect to consider. Future research can focus on developing techniques to provide reliable uncertainty estimates in the Stable diffusion framework. This will enhance the interpretability and trustworthiness of the generated images, allowing clinicians to make more informed decisions based on the AI predictions.
3. Evaluation on Diverse Data Sources: The generalizability of the Stable diffusion model and the effectiveness of the generated synthetic images can be further evaluated on diverse data sources and external validation datasets. This will help assess the robustness and reliability of the approach across different healthcare systems, imaging equipment, and populations.
4. Ethical Considerations and Bias Mitigation: As with any AI-based technology, ethical considerations and bias mitigation are crucial aspects to address. Future work should focus on investigating potential biases in the generated synthetic im-ages and developing strategies to ensure fairness, transparency, and equity in the diagnostic models trained on these images.
5. Integration into Clinical Workflow: Successfully integrating AI-generated synthetic images into the clinical workflow poses practical challenges. Future research can explore strategies for seamless integration, considering factors such as data privacy, regulatory compliance, and clinician acceptance. This will facilitate the translation of Stable diffusion-generated images into real-world clinical acceptance.
6. Data Privacy and Security: The privacy of patient data and the security of the created synthetic pictures are critical. Future research should focus on developing effective data anonymization and privacy preservation strategies to ensure patient confidentiality while using the benefits of generative modelling.
7. Reliability of the evaluation: A medical practitioner is used to verify results. It would have been much more efficient and reliable if a matrix was available as a credible benchmark to provide a more impartial evaluation of the correctness of the generated synthetic X-ray pictures. This matrix, which includes proven COVID-19 instances and X-ray pictures, would be gathered from a variety of sources, including reliable medical databases and expert comments.

The amount of authenticity and realism attained by comparing the synthetic pictures created by the Stable diffusion model can be assessed reference matrix. This method allows us to objectively examine the correctness of the produced pictures, minimizing

potential biases and subjective interpretations. The solicited advice and experience of medical authorities that specialize in COVID-19 diagnosis was used. These specialists analyzed the synthetic photos, offering useful insights and opinions. Their experience contributed to the review process by validating the veracity of the synthetic photographs. A thorough assessment system that can strengthen the reliability and correctness of the generated synthetic X-ray pictures by can be guaranteed by combining objective comparison against the reference matrix with the experience of medical experts. This comprehensive assessment technique will increase the validity of our findings and conclusions on the Stable diffusion model's effectiveness and promise in resolving data shortage and enhancing COVID-19 diagnosis. Addressing these future work areas and problems will progress the field of AI-based medical imaging, improve diagnostic model reliability, and make it easier to integrate Stable diffusion-generated synthetic pictures into clinical practice. By addressing these issues, we will be able to fully utilize the promise of generative modelling approaches in enhancing healthcare outcomes and patient care.

6 Conclusion

The use of stable diffusion in the generation of synthetic X-ray pictures for COVID-19 diagnosis is a significant strategy for overcoming data shortages and improving the performance of AI-based diagnostic systems. We were able to develop realistic synthetic pictures that closely mimic genuine patient X-rays by finetuning the Stable diffusion model using a small dataset. These synthetic photos are a useful addition to the collection, allowing for more robust training and enhancing COVID-19 detection accuracy. Stable diffusion's promise goes beyond COVID-19, providing chances for multi-modal data synthesis, modelling of uncommon illnesses, and data augmentation in a variety of respiratory conditions. However, obstacles remain, including as adding clinical metadata, calculating uncertainty, correcting biases, and assuring ethical concerns and data protection. Despite these obstacles, the incorporation of Stable diffusion-generated synthetic pictures into clinical practise holds significant potential for future advancements in AI-assisted medical imaging and patient care. We may improve the capabilities of AI-based diagnostic systems by exploiting the benefits of stable diffusion, allowing for more accurate and efficient identification of respiratory disorders, eventually leading to improved patient outcomes.

7 Future Work

The Stable diffusion technique, in addition to solving the lack of annotated X-ray images for COVID-19 diagnosis, provides up prospects for additional study and applications in the medical imaging sector. In this section, we outline potential future projects that may be investigated using the same approach:

1. Multimodal Data Synthesis: Stable diffusion may be extended to provide synthetic data in a variety of modalities, including computed tomography (CT) scans and magnetic resonance imaging (MRI). This extension enables the synthesis of various medical pictures, allowing the construction of complete diagnostic models that can use several imaging modalities.

2. Rare Disease Simulation: Stable diffusion can be employed to generate synthetic images for a multitude of rare illnesses and diseases with very little data. The resultant diverse data could therefore be utilised effectively in the development and training of machine learning models for a more accurate diagnosis and rigorous treatment planning.
3. Data Augmentation for related Respiratory Diseases: this technique is not restricted to COVID-19 data sets only. It may be replicated in other cases of related respiratory disorders such as pneumonia, TB, and lung cancer. The data generated by the method can be used in the creation of more reliable and precise diagnostic tool for treatment of a variety of respiratory disorders.
4. Detection of aberrations and Outlier Identification: Stable diffusion may be used to generate quality images for both diseased and normal state. Therefore it will allow us to discover anomalies or outliers in medical imaging,. This opens a wider field for wider and more detailed inspection of a disease and would warrant a more precise line of treatment that could be formulated in the face of that anomaly.
5. Transfer Learning and Domain Adaptation:Stable diffusion may be used for domain adaptation tasks, where models trained on synthetic pictures created by the method can be fine-tuned on real-world data. When applied to diverse imaging datasets or healthcare situations, this transfer learning process can improve the generalizability and performance of AI models.

The Stable diffusion methodology can contribute to the development of robust AI-based diagnostic tools, data augmentation approaches, and anomaly detection methods in the field of medical imaging by investigating these connected study fields. Beyond COVID-19 diagnosis, the potential applications are numerous, addressing different healthcare difficulties and improving patient care across a wide spectrum of illnesses and ailments.

References

1. Kazerouni, A., et al.: Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis **88**, 102846 (2023). ISSN: 1361-8415. https://doi.org/10.1016/j.media.2023.102846. URL: https://www.sciencedirect.com/science/article/pii/S1361841523001068
2. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis **58**, 101552 (2019)
3. Jiang, Y., et al.: Covid-19 CT image synthesis with a conditional generative adversarial network. IEEE J. Biomed. Health Informat. **25**(2), 441–452 (2020)
4. Chen, M., et al.: Deep feature learning for medical image analysis with convolutional autoencoder neural network. IEEE Transactions on Big Data **7**(4), 750–758 (2017)
5. Ali, H., et al.: The role of generative adversarial networks in brain MRI: a scoping review. Insights into Imaging **13**(1), 1–15 (2022)
6. Ahmad, W., et al.: A new generative adversarial network for medical images super resolution. Sci Rep. **12**(1), 1–20 (2022)
7. Munawar, F., et al.: Segmentation of lungs in chest x-ray image using generative adversarial networks. IEEE Access **8**, 153535–153545 (2020)
8. Antun, V., et al.: On instabilities of deep learning in image reconstruction and the potential costs of AI. Proceedings of the National Academy of Sciences **117**(48), 30088–30095 (2020)

9. Armato, S.G., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics **38**(2), 915–931 (2011)
10. Bakas, S., et al.: Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data **4**(1), 1–13 (2017)
11. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated MRI (2022)
12. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis **34**, 8780–8794 (2021)
13. Kazerouni, A., et al.: Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis **88**, 102846 (2023)
14. Mikolajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 117–122 (2018)
15. Razavi, A., van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with VQ-VAE-2. In: Advances in Neural Information Processing Systems, Vol. 32. Curran Associates, Inc. (2019)
16. Song, Y., et al.: Solving inverse problems in medical imaging with score-based generative models (2022)



Machine-Learning Algorithms for Predicting Cardiac Arrhythmias

Sakshi Shete, Suvarna Udgire(✉), Anuradha Joshi, Shravani Desai, and Ronit Murpani

Vidyalankar Institute of Technology, Mumbai 400037, India
{sakshi.shete,suvarna.udgire,anuradha.joshi,shravani.desai,
ronit.murpani}@vit.edu.in

Abstract. Cardiac arrhythmia, a frequently encountered cardiac disorder characterized by irregular heart rhythms, poses a substantial risk to a patient's well-being. This research explored various algorithms for forecasting cardiac arrhythmias derived from ECG datasets available on Kaggle. Employing machine learning techniques, the study achieved significant accuracy rates: Logistic Regression (71.43%), Random Forest Classifier (58.24%), and Decision Tree Classifier (65.93%). Extracting the different characteristics of cardiac electrograms (ECG), focusing on R peak, QRS duration, and T interval, forms the foundation for arrhythmia categorization. The results are employed to transform arrhythmia management, emphasizing accuracy and key features, with Logistic Regression leading at 73.28%, Random Forest Classifier at 76.82%, and Decision Tree Classifier at 61.22% in AUC-ROC scores. This research contributes to advancing datadriven methodologies in cardiac care, enhancing patient outcomes by facilitating early detection and personalized treatment plans.

Keywords: Electrocardiography (ECG) · R-R Interval (Peak-to-Peak Interval) · Heart Rate (Beats per Minute) · P-Wave to QRS Peak Interval · Convolutional-Neural-Networks (CNN) · Artificial-Neural-Network (ANN)

1 Introduction

Cardiac arrhythmia is a fatal cardiac condition marked by aberrant electrical impulses that interfere with the heart's regular rhythm. This circumstance of an electrocardiogram (ECG or EKG), Holter monitoring, stress testing, echocardiogram, or electrophysiological investigations are commonly utilized in diagnosis to evaluate the cardiac electrograms for determining the rhythm of the heart. One can show up as an abnormally rapid, slow, or irregular heartbeat (tachycardia, bradycardia). Cardiac arrhythmias can happen to people of all ages and, if addressed, can have major health repercussions. Heart irregularities affect millions of individuals globally and are a serious health concern. Atrial fibrillation (AFib), a condition impacting one to two percent of the global population, is known for its potential to disrupt the normal functioning within the heart. It is expected that roughly 2–4% of humans on earth encounter atrial fibrillation (AFib), which represents among the most prevalent forms of arrhythmias. Cardiac arrhythmias encompass a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 292–302, 2025.
https://doi.org/10.1007/978-3-031-75167-7_24

wide spectrum of abnormal heart rhythms, including ventricular tachycardia (VT), atrial flutter, atrial fibrillation (AFib), and bradycardia, among others. Each variety has distinct traits and possible health hazards. Heart palpitations, chest discomfort, exhaustion, dizziness, and syncope are a few of the symptoms of cardiac arrhythmia.

However, some individuals with arrhythmia may remain asymptomatic. The occurrence of developing cardiac arrhythmias is influenced by age, family history, cardiovascular disease, smoking, hypertension, and severe alcohol consumption and substance abuse. If left untreated, cardiac arrhythmias can lead to significant issues like heart failure, stroke, and sudden cardiac arrest. Early identification and suitable medical intervention are pivotal in mitigating these risks. Treatment options for heart disturbances encompass lifestyle modifications, medication, electrical cardioversion, catheter ablation, and the implantation of pacemakers or defibrillators, considering the types of the arrhythmia. Research on cardiac arrhythmia underscores its common occurrence, particularly pertaining to age, genetic factors, and risk factors like hypertension. Atrial fibrillation (AFib), because of its association with serious cardiovascular events, is the subject of extensive investigation. Arrhythmias can manifest with a wide array of symptoms, including palpitations and chest discomfort, emphasizing the significance of modern diagnostic approaches such as wearable technology and remote monitoring for early detection.

Treatment of arrhythmias includes medication, catheter ablation, and implanted devices. Medical intervention is complemented by lifestyle changes. Electrocardiogram (ECG) data has been analyzed to create machine learning models, such as Deep-Learning Methods like CNNs (convolutional neural networks) and RNNs. Arrhythmia care is being transformed through telemedicine and remote monitoring, which improves prompt treatments. Studies conducted over an extended period to evaluate patient outcomes and the effects on public health, recognizing arrhythmia as a major healthcare burden. To improve management techniques for cardiac arrhythmia, ongoing research combines data-driven methodologies and technology.

2 Literature Survey

The data source is taken from Kaggle (https://www.kaggle.com/datasets/bulentesen/cardiac-arrhythmia-database) and is attributed to MIT-BIH. Various Machine-Learning Algorithms are implemented using Electrocardiogram (ECG) dataset of five sets of 100,022 pulses provided by MIT-BIH rhythmic database [1]. SVM and logistic regression are methods utilized to categorize cardiac arrhythmias. The UCI machine repository is where the dataset is located. Systems with two stages of serial fusion are employed [2]. Models are now a common method for classifying ECG data. The author of a survey article, Hong illustrates how to make use of Deep-Learning (DL) to analyze ECG data techniques [3, 4]. Models for Deep-Learning are the topic of more research since Murat et al.'s survey of the subject, and they are now widely employed. to classify ECG data [5].

Features like R-R interval, QRS interval, and heart rates are generally utilized to categorize the various forms of heartbeat irregularity in the ECG [6]. These characteristics may be extracted using multiple techniques [7], including Discrete Wavelet Transform (DWT) [8], Principal Component Analysis (PCA) [9, 10], and others. Machine learning is employed to categorize the characteristics into the proper groups following their extraction [11]. The utilization of first level of SVM classifier to link the distance outputs to the confidence measure. The database was classified using classifiers such SVM, CNN, Genetic Algorithm (GA) [12], and Recurring-Neural-Network (RNN), ANN [13]. While Artificial-Neural-Networks (ANN) incorporate training data to understand the features of inputs that could later be utilized in classification, genetic algorithms categorize features using processes like natural selection. Each characteristic may be given weight to devise a quick optimization method [14]. Arrhythmia identification is accomplished using an anomaly recognition approach [15]. Prior to the anomaly recognition procedure, the most crucial pre-processing stage is carefully considered. The arrhythmia dataset is pre-processed using the random forest ensemble techniques and the recursive feature removal approach [16]. Anomaly detection comes when the pre-processing operation is complete. Anomaly detection is accomplished using the clustering based oversampling PCA approach. SVM and logistic regression are methods utilized to categorize cardiac arrhythmias. The UCI machine repository is where the dataset is located. Systems with two stages of serial fusion are employed. For both positive and negative ECG samples, rejection levels are applied [17]. The logistic regression classifier moves upon the next stage with the samples that are rejected. The classifiers' outcomes are combined to offer the conclusion. And appropriate for issues involving linearly separable data.

3 Dataset

The dataset and references were acquired from Kaggle and MIT-BIH, respectively. The dataset has multiple features, but we eliminated a few of the features to acquire better accuracy so that applied model learns the data effectively. It has various features like height, weight, p-r interval, age, sex, qrs_duration, q-t interval, t_interval, p_interval, heart rate etc. The data significantly resembles the cardiac electrograms which gives a clear idea about the parameters regarding cardiac arrhythmia. Also, there are 16 classes of arrhythmia, in which only the first one is normal kind of arrhythmia where the other classes can be super classed as risky arrhythmia cases. All the cases are classified as either normal or risky. Risky being the positive (True) case and normal being the negative (False) case. Class 1 indicates a "normal" ECG, classes 2 through 15 indicate several

arrhythmia classifications, and class 16 indicates the remaining unclassified ones. There are 245 instances of 'normal' and the rest 207 instances belong to the other classes. There are 279 features in the database overall, 206 of which have linear values while the remaining ones are nominal.

4 Proposed Work

The dataset and references were acquired from Kaggle and MIT-BIH, respectively. The initial data was taken which had many anomalies, which needed to be sorted i.e., preprocessing was necessary. The noise removal process involved eliminating the null values along getting rid of unnecessary columns, which are insignificant in the prediction. After data processing, the dependence between remaining attributes and target variable is analyzed so that it helps in creating the model to improve its accuracy and efficiency.

The process to gain insights regarding the connection between attributes is known as Exploratory Data Analysis (EDA). This relation can be best understood by plotting various graphs between the attributes, which assists in increasing the understandability about the attributes. Various graphs like Count, Bar, Box Plots are as per Figs. 2, 3 and 4 (see Figs. 2, 3 and 4) For plotting different libraries like Matplotlib, Seaborn etc. are used. After gaining insights from EDA appropriate machine-learning algorithms need to be selected for building predictive models. The algorithms selected for this project are Logistic-Regression, which is a linear classification algorithm used to predict binary outcomes. Decision-Tree-Classifier, which splits the data based on the feature thresholds to create tree like structure. And Random-Forest-Classifier is an ensemble method based on decision trees. Each selected algorithm was trained using subsets of the data. The dataset was split into training and testing sets to evaluate the model performance effectively. The performance of each model was accessed using the standard evaluation metrics like accuracy, precision, F1 score, recall and AUC-ROC Score. The complete workflow is represented in the form of a flowchart as shown in Fig. 1. (see Fig. 1) Within the dataset there are 16 classes of diagnosis, which are helpful in finding the degree and kind of arrhythmia, which is given below:

Class code	Class	Number of instances:
01	Normal	245
02	Ischemic changes (Coronary Artery Disease)	44
03	Old Anterior Myocardial Infarction	15
04	Old Inferior Myocardial Infarction	15
05	Sinus tachycardia	13
06	Sinus bradycardia	25
07	Ventricular Premature Contraction (PVC)	03
08	Supraventricular Premature Contraction	02
09	Left bundle branch block	09
10	Right bundle branch block	50

(*continued*)

(*continued*)

Class code	Class	Number of instances:
11	1. Degree Atrioventricular block	00
12	2. Degree AV block	00
13	3. Degree AV block	00
14	Left ventricle hypertrophy	04
15	Atrial Fibrillation or Flutter	05
16	Others	22

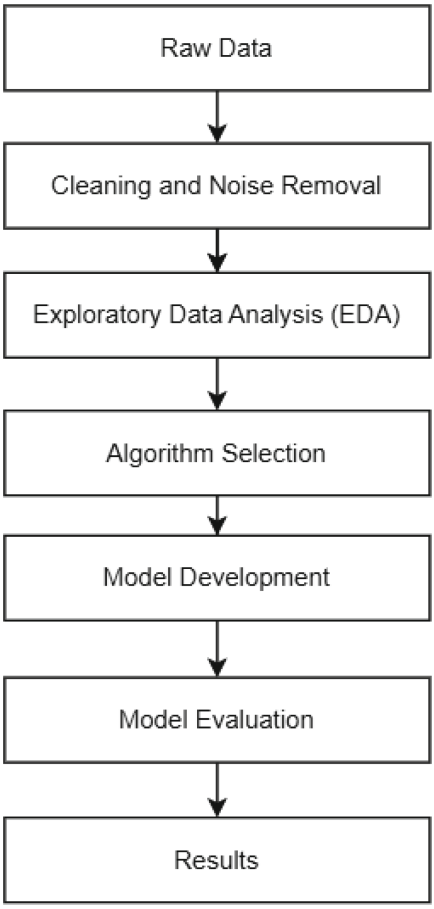


Fig. 1. Flowchart for Arrhythmia Prediction

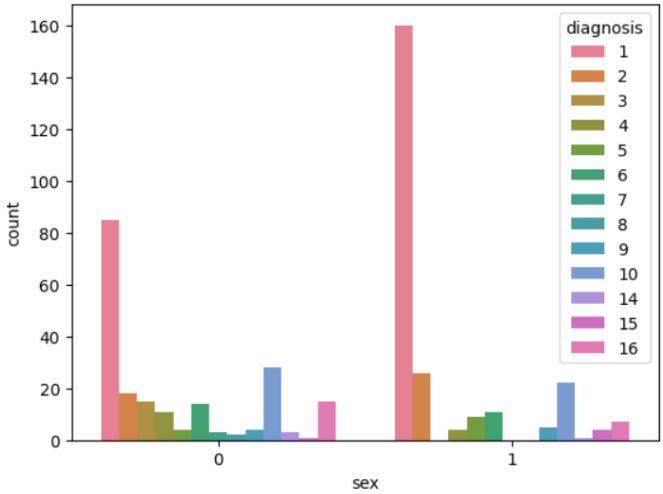


Fig. 2. Count Plot (count Vs sex)

The above fig represents the count plot between the 16 class of diagnosis and sex. It notifies that which gender is more prone to with type of arrhythmia, where 1 represents male and 0 as female. So as mentioned earlier the most number of instances are of normal type of arrhythmia. And the count plot clearly states the same. As a result, we get to know that around 150 male and 85 female suffer from normal arrhythmia.

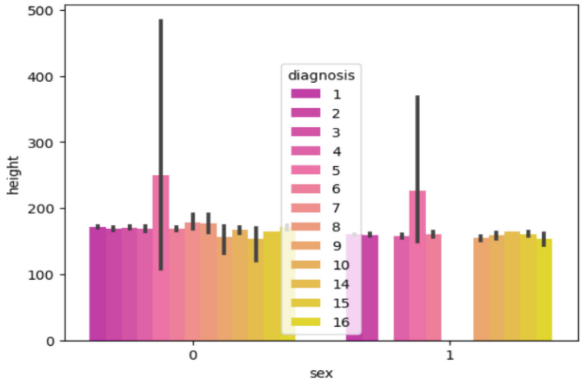


Fig. 3. Bar Plot (height Vs Sex)

Above fig represents the bar plot plotted between height and sex. This bar plot notifies that what number of male or female having a specific height may suffer from which type of arrhythmia among the 16 types mentioned earlier, where 0 represents female and 1 represents male.

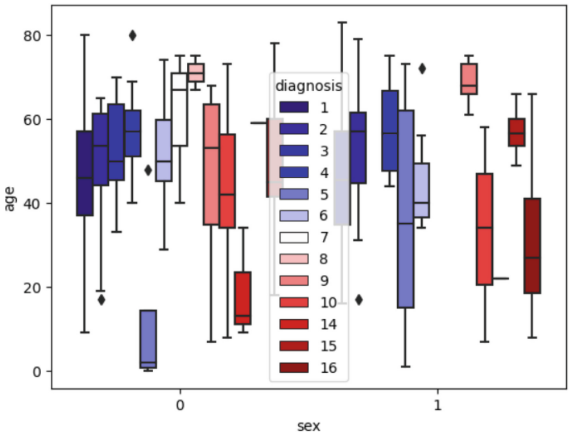


Fig. 4. Box Plot (age Vs sex)

Above figure represents a box plot plotted against age and sex to recognize that what range of age is more inclined towards which specific class of arrhythmia. So, the graph clearly states that in female the age group between 40 to 60 has a probability to suffer from class 1 type having the minimum age of 10 to maximum of 80 and that in male it is from 50 to 60 suffering from class 1 arrhythmia where the minimum age is 30 and maximum 75–79.

4.1 Algorithm for Arrhythmia Classification

A labeled dataset is utilized to implement supervised learning algorithms which includes regression and classification. We shall utilize Logistic-Regression (LR), Random - Forest-Classifiers (RFC) and decision tree classifiers more precisely in regression and classification. 20% is the percentage utilized to test the data. The remaining data is utilized for training the model, which aids in efficient training. Logistic Regression is basically a classification algorithm.

Logistic Regression predicts whether an instance belongs to a specific class or not. As we have multiple classes here, logistic regression fits perfectly for the same. Decision Tree is known as tree-structured classifier as internal nodes represent the features of a dataset which here are age, sex, height etc. and branches represent the decision rules those are qrs_duration, p-r interval, q-t interval etc. along with every leaf node indicating the outcome i.e., diagnosis variable or target variable here.

The Random-Forest-Classifier, which is also utilized by multiple decision trees, is adopted due to a substantial quantity of trees in the forest results in higher accuracy and prevents the overfitting problem. The primary benefit of utilizing a decision tree classifier is that, contrary to other algorithms, it requires less training time and produces accurate and highly precise output predictions.

Measures like accuracy and the AUC_ROC score may serve as a basis for evaluation and the effectiveness of classification algorithms.

4.2 Performance Analysis of Algorithm

To estimate the classifier's performance, we can calculate accuracy and the AUC-ROC score using the subsequent formulas:

$$Accuracy = \frac{Number\ of\ Correct\ \text{Pr}edictions}{Total\ Number\ of\ \text{Pr}edictions} \tag{1}$$

This accuracy metric is determined by correlating the count of accurate predictions per total number of forecasts. To calculate accuracy, we typically iterate through the ground truth and predicted values in a loop. It is essential to emphasize that accuracy is most reliable when there is an equivalent of instances within each category.

The AUC-ROC curve is a crucial tool for visualizing the output of classification model. Its performance at various threshold levels is shown graphically by the receiver-operating-characteristic curve (ROC) where the True Positive Rate is plotted against the False Positive Rate.

A simple statistic to evaluate the validity of classifier is accuracy, while the AUCROC curve provides a visual representation of its performance across various threshold levels, helping us understand how well the classification model performs.

True Positive Rate is computed as:

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

False Positive Rate is computed as:

$$TPR = \frac{FP}{FP + TN} \tag{3}$$

We may test a logistic regression model repeatedly with varying classification criteria to determine value at every point on a ROC curve, however this method is not highly effective. Thus, one effective method—the AUC—is employed for this. AUC is the term to denote the underside of the ROC curve. Total area underneath, the whole ROC curve is in two dimensions, or AUC, is as its name implies. The AUC value is between zero and one.

5 Results

While examining the arrhythmia database data, several key features were extracted to better understand and categorize the information. These features included the R peak, QRS duration, and T interval, which are significant components of an electrocardiogram (ECG) signal.

R Peak: The ECG waveform's peak point, the Q-R-S pattern, is symbolized by peak of R wave. It indicates the starting point of ventricular contraction and the gradual depolarization of the lower chambers. The R peak's exact timing and magnitude helps in determining cardiac electrical activity.

QRS Duration: The QRS duration indicates the time duration required for the electrical impulses to navigate through the ventricles of inner heart. It encompasses the Q

wave, R wave, and S wave of the complex of the QRS. The QRS complex's duration might change Depending upon presence of any irregularities in the heart's conduction system.

T Interval: It represents the repolarization of the ventricles as they prepare the following heartbeat. Variations in T interval is related to certain cardiac conditions.

Following the characteristics extraction were then used as a foundation for categorizing the data into various groups or categories of arrhythmia. By analyzing these features and their variations within the dataset, researchers and medical professionals could discover more about the types and to what extent arrhythmias are present in the patients' ECG recordings. This categorization is crucial for diagnosing and treating arrhythmias effectively (Table 1).

Table 1. Classification of accuracy for implemented algorithms.

Type of Algorithm	Accuracy
Logistic Regression	71.43%
Random Forest	58.24%
Decision Tree	65.93%

The ROCAUC or AUROC score is important to evaluate the correctness of model. The higher the AUC, the better is the model's performance at distinguishing between the positive and negative classes. As it is common knowledge, the acronym AUC is Area beneath the ROC (Receiver Operating Characteristic) Curve, as mentioned in the table below (Table 2).

Table 2. Classification of AUROC Score.

Type of Algorithm	AUROC Score
Logistic Regression	73.28%
Random Forest	76.82%
Decision Tree	61.22%

6 Conclusion

The research proposes the implementation of three Machine-Learning (ML) techniques to predict cardiac arrhythmias. Through meticulous data analysis and preparation, we were able to develop highly accurate prediction models. Python was utilized to extract features. And the creation of classification models. These models play a crucial part in aiding healthcare professionals in making critical decisions, facilitating early detection,

and tailoring treatment plans to individual patients. Training of dataset is important to be both precise and diverse since even minor variations in peak-to-peak values could impact heart health significantly. This study places emphasis on harnessing data-driven methodologies to revolutionize cardiac care and enhance patient outcomes.

References

1. Subramanian, K., Prakash, N.K.: Machine Learning based Cardiac Arrhythmia detection from ECG signal. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1137–1141. Tirunelveli, India (2020)
2. Shalini, B.N., Nandini, V., Sandhya, M., Bharathi, R.: Prediction and Classification of Cardiac Arrhythmia. Int. Res. J. Eng. Technol. (IRJET) **6**(6) (2019)
3. Sharma, T., Verma, S., Kavita: Intelligent heart disease prediction using Machine Learning: A review. Int. J. Recent Res. Aspects **4**(2) (2017)
4. Hong, S., Zhou, Y., Shang, J., Xiao, C., Sun, J.: Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Compute Biol Med. **122**, 103801 (2020). https://doi.org/10.1016/j.compbiomed.2020.103801. Epub 2020 Jun 7. PMID: 32658725
5. Murat, F., et al.: Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. Comput Biol Med. **120**, 103726 (2020)
6. Siva, A., Siddharth, S., Hari, S.M., Rajesh, C., Nithin, M.: B, Classification of arrhythmia using Wavelet Transform and Neural Network model. Journal of Biomedical Science & Bioengineering **8**(1), 244 (2018)
7. Chan, K.L., So, H.H.: Development of QRS detection method for real-time ambulatory cardiac monitor. Proceedings of the 19th Annual International Conference of the IEEE (Institute of Electrical and Electronics Engineers) Engineering in Medicine and Biology Society. (Cat. No.97CH36136), pp. 289–292. Chicago, IL, USA (1997)
8. Beniczky, S., Fuglsang Frederiksen, A., Jeppesen, J., Johansen, P., Sidenius, P.: Modified automatic R-peak detection algorithm for patients with epilepsy using a portable electrocardiogram recorder. Conf. Proc. IEEE Eng. Med. Biol. Soc. (2017)
9. Cui, D., Yuan, D., Li, H., Wang, Y., Cao, L.: Arrhythmia classification based on Multi-Domain Feature extraction for an ECG recognition system. Sensors **16**, 1744 (2016)
10. Hernández, A.I.: (Artificial Intelligence), Carrault G and Portet F, Evaluation of real-time QRS detection algorithms in variable contexts. Med. Biol. Eng. Comput. **43**(3), 379–385 (2005)
11. Gurgen, F., Uyar, A.: Arrhythmia classification using serial fusion of support vector machine and logistic regression. IEEE International Workshop on Intelligent data Acquisition and Advanced Computing Systems (2007)
12. Lee, D., Chen, C., Tseng, K.-K.: ECG identification system using neural networks with global and local features. International Association for Development of the Information Society (2016)
13. Vinayakumar, R., Swapna, G., Soman, K.P.: Automated detection of cardiac arrhythmia using deep learning techniques. Procedia Computer Science **132**, 1192–1201 (2018)
14. Cheng, Q., Gutta, S.: Joint feature extraction and classifier design for ECG-based biometric recognition. IEEE J Biomed HealthInform. **20**(2), 460–468 (2016)
15. Ashok, A., Smitha, S., Kavya Krishna, M.H.: Attribute reduction based anomaly detection scheme by clustering dependent oversampling PCA. In: IEEE 2016 International Conference on Advances in Computing, Communications and Informatics (2016)

16. Goldberger, A.L., et al.: PhysioBank, Toolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23
17. Marchesi, C., Paoletti, M.: Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis, Vol. 82 (2006)
18. Kaggle DataSet - https://www.kaggle.com/datasets/bulentesen/cardiac-arrhythmia-database. Last accessed on 06 October 2023

Natural Language Generation (NLG) with Reinforcement Learning (RL)

Shakir Khan[1,6(✉)], Mohammad Serajuddin[2], Zafarul Hasan[3], Syed Ahad Murtaza Alvi[4], Rashid Ayub[5], and Arun Sharma[7]

[1] Information Technology Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
sgkhan@imamu.edu.sa
[2] College of Dentistry, King Saud University, Riyadh, Saudi Arabia
mserajuddin@ksu.edu.sa
[3] College of Nursing, King Saud University, Riyadh, Saudi Arabia
zhasan@ksu.edu.sa
[4] College of Applied Computer Sciences, King Saud University, Riyadh, Saudi Arabia
salvi@ksu.edu.sa
[5] Department of Science Technology and Innovation Unit, College of Computer and Information Sciences, King Saud University, P. O. Box-2454, Riyadh 11451, Saudi Arabia
rayub@ksu.edu.sa
[6] University Centre for Research and Development, Chandigarh University, Mohali, India
[7] Department of Information Technology, Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi, India
arunsharma@igdtuw.ac.in

Abstract. Finally, utilizing an interpretative framework as well as a deductive method this study studied the use of reinforcement learning strategies in the development of natural languages. Although secondary data collection provided a solid foundation, the descriptive method allowed for a complete study. The technical solutions used included the incorporation of RL computer programs, advanced text processing, linguistic analysis, contextual adaptability, and moral concerns. As the consequence of our research, we now have a better understanding of RL-NLG systems, which opens the way to more sophisticated and environmentally conscious text production. This research not only enhances the field of NLG, although it also emphasizes the importance of ethical as well as objective applications for artificial intelligence. The insights have application that encompass intelligent machines to the generation of customized content.

Keywords: Reinforcement Learning · Natural Language generating · RL-NLG systems · mechanism · disinformation penalized · Python modules · RL engine

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 303–318, 2025.
https://doi.org/10.1007/978-3-031-75167-7_25

1 Introduction

1.1 Research Background

"Natural Language Generation (NLG)" is a key area of synthetic intellect inside which machines create human-like readers from underlying data. While NLG maintains created great progress, tailoring its creation for special purposes resumes to be a tribulation. Reinforcement Learning (RL) constitutes a promising specific approach to consider improving NLG systems. RL including operating training utilizes a reward-based Function which teaches optimal patterns of action [1]. RL could be utilize to refine published text in NLG through delivering rewards depending on metrics particularly such as coherence, effortlessness, as well as relevance. This method differs from typical supervised learning in that models learn from specified input-output pairings, allowing for greater depth and contextually relevant text production. Furthermore, RL-NLG systems have the ability to adapt to a wide range of domains and customer preferences, making them useful tools for tasks ranging from chatbots including digital assistants to content generation and data summarization. Despite the potential it holds, RL-NLG is a developing discipline that faces difficulties in harmonizing exploration-exploitation compromise, establishing appropriate reward functions, and assuring ethical overall unbiased text production [2]. Taking care of these challenges will pave a path for more advanced and adaptive NLG systems.

1.2 Aim and Objectives

Aim:
The goal of this study is to improve the generation of natural languages (NLG) by using Reinforcement Learning (RL) techniques, which will improve the quality and responsiveness of generated text.

Objectives:
To look into incorporating RL algorithms with existing NLG platforms for improved text generation.

To create and improve reward functions that accurately assess the quality, coherence, as well as relevance of created text.

To evaluate the adaptability along with specific to a domain utility of RL-NLG models in a variety of applications, such as virtual assistants and document creation.

To assure ethical as well as unbiased text production by putting in place mechanisms to avoid the spread of damaging or biased literature in RL-NLG systems.

1.3 Research Rationale

The combination of Reinforcement Learning (RL) and Natural Language generating (NLG) holds great promise for improving the capabilities of computerized text generating systems.

Traditional NLG models frequently rely on the use of supervised learning, which necessitates big, annotated datasets [3]. RL provides an alternate way by allowing models to acquire information directly via interaction with their surroundings, resulting

in more complex and contextually relevant output. This is especially important when explicit inputs and outputs are few or difficult to get. Furthermore, RL-NLG systems have the ability to adapt to different domains and customer preferences, making them adaptable tools for a variety of applications ranging from chatbots and digital assistants to content generation as well as data summarization. However, this field is still in its infancy, and issues like as establishing appropriate incentive functions, dealing with exploration-exploitation sacrifices, and assuring ethical as well as unbiased production of texts remain [4]. Addressing these difficulties will not only help to progress NLG science and technology, but it will additionally open up new paths for applications in a variety of areas.

2 Literature Review

2.1 Integration of Reinforcement Learning in Natural Language Generation

The incorporation of Reinforcement Learning (RL) within Natural Language generating (NLG) is a significant step forward in automated text-generating systems. RL, a subset of machine learning, allows models to acquire information from their experiences in their surroundings, allowing them to draw sequential judgments in dynamic environments. This means that, rather than relying entirely on pre-defined training information, models may develop the ability to generate text through experimentation as well as receiving feedback within the form of payments [5]. A model outputs a sequence of characters in RL-NLG systems, as well as those patterns undergo assessment based on predetermined criteria such as coherence, fluency, and relevancy. The model is rewarded positively or negatively based on how effectively it meets these requirements. The model continues to generate ever more precise and contextually suitable content over time through an approach of exploration and exploitation (Fig. 1).

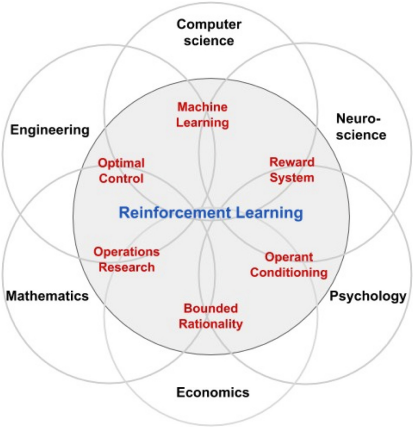


Fig. 1. Reinforcement Learning

This integration addresses a basic problem in NLG: the lack of large-scale annotated datasets. Because RL-NLG models may learn directly through their interactions with the environment, they are very adaptable to a wide range of domains and activities. Furthermore, RL-NLG systems have demonstrated potential in applications like as chatbots, digital assistants, and content production, where versatility and context awareness are critical [6]. However, issues such as designing effective reward mechanisms and managing exploration-exploitation alternatives remain.

2.2 Reward Function Design for Evaluating Text Generation Quality

A fundamental part of Reinforcement Learning (RL) utilized for the generation of natural languages (NLG) is the design of an appropriate reward function. The incentive function guides the RL agent in learning and improving its text creation capabilities [7]. It delivers a quantitative assessment of how well the created text matches predetermined standards such as coherence, relevancy, and overall quality. The reward mechanism in NLG must be carefully customized to the task's specific objectives. In a summary assignment, for example, awards could be based on the manner in which the resulting text conveys the essential points of the originating material. Incentives could be awarded in a conversational situation depending on qualities such as transparency, fluency, and involvement. Creating a successful reward function necessitates a thorough understanding of the intended outcome, as well as the hazards and problems unique to the NLG assignment at hand. It could entail experimenting with various incentive systems and fine-tuning them via repeated feedback cycles (Fig. 2).

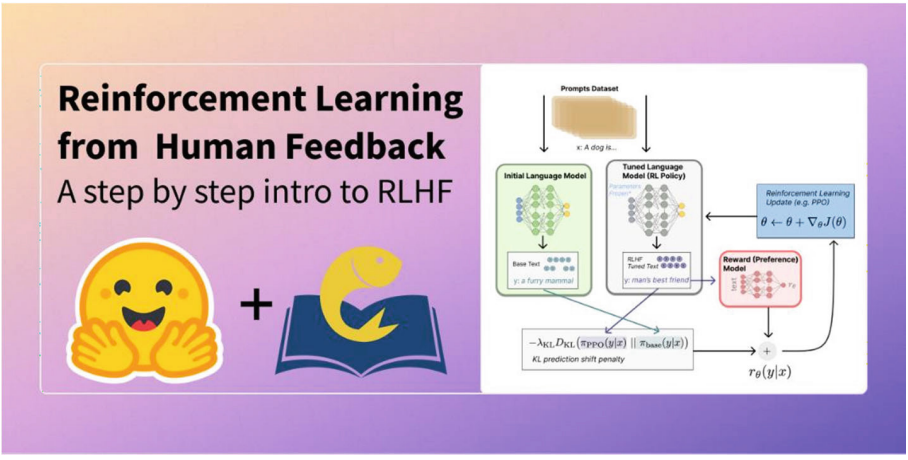


Fig. 2. Evaluating Learning

Furthermore, the system of rewards should be structured to provide unambiguous direction to the RL agent while avoiding excessively simple as well as misleading signals [8]. This is critical for teaching the RL-NLG engine to generate content that is additionally technically correct but also contextually suitable and cohesive. Developing an effective reward function is a critical step in improving the overall caliber of generated text.

2.3 Domain Adaptability and Performance Assessment

In the environment of applying reinforced learning (RL) to the generation of natural language (NLG), domain adaptation as well as achievement assessment are critical issues. NLG systems must be capable of producing coherent and contextually appropriate text in a variety of domains including applications [9]. The capacity of an RL-NLG modeling to effectively migrate while functioning well in diverse topics or specialized domains is referred to as sector adaptability. This feature is especially crucial given the breadth of NLG applications, which include healthcare, banking, customer service, and others. Studying how RL-NLG programs adapt and succeed in different circumstances is critical for their wider use in reality. Assessing achievement in different areas entails comparing the RL- NLG technique's performance to specific to the domain benchmark along with metrics. Tasks like as summation accuracy, assessment of sentiment, and domain-specific proficiency in languages may fall under this category. Furthermore, the evaluation should take into account factors such as coherence, effortlessness, and knowledge, which are crucial for producing high-quality writing (Fig. 3).

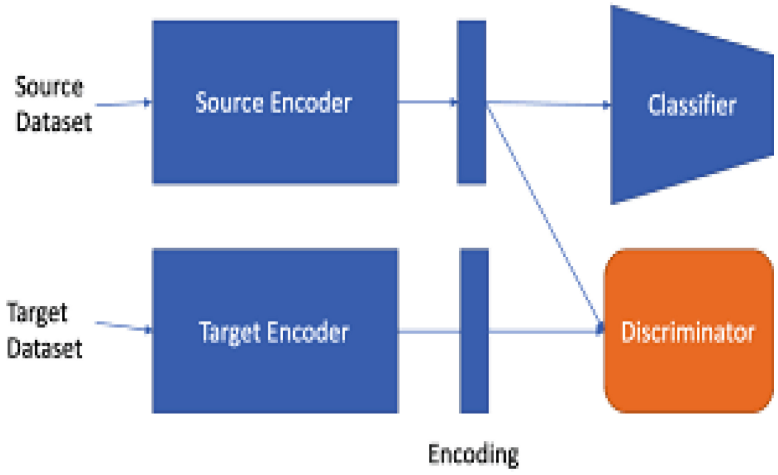


Fig. 3. Domain Adaptability

Furthermore, evaluations of performance ought to consider account of user preferences as well as requirements [10]. A virtual assistant created for health care providers, for example, may have different performance objectives than one built for general usage

by consumers. Understanding and assessing domain adaptation and performance are critical aspects in determining the efficacy and suppleness of RL-NLG systems.

2.4 Ethical and Unbiased Text Generation in RL-NLG

Ensuring ethical and objective text creation in Reinforcement Learning The development as well as deployment of Learning-based natural language creation (RL-NLG) systems is crucial. The potential for generating text to contain rude, discriminatory, or harmful information raises ethical considerations. Addressing these concerns is critical to developing responsible along with trustworthy NLG applications [11]. One critical part is the development of reward functions. Ethical issues should be incorporated into the reward structure, with outputs that reflect bias, vulgar speech, or disinformation penalized. Furthermore, models have to be trained on broad and comprehensive datasets to reduce biases in information used for training. Continuous evaluation and monitoring of the results of the model is required to mitigate bias in RL-NLG. This involves putting in place controls to prevent the creation of content that promotes stereotypes or differentiates against specific groups. Furthermore, fine-tuning models on certain elements of fairness, for example gender, ethnicity, or socioeconomic position, may be required. Another important factor is transparency (Fig. 4).

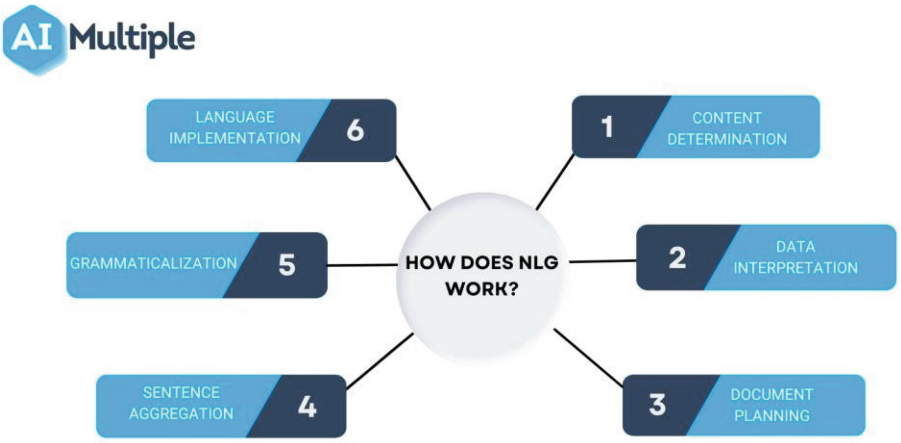


Fig. 4. Natural Language Generation

Users should be able to understand how RL-NLG models work, including the disclosure associated with training data, compensation functions, and any biases. Furthermore, explicit norms for dealing with delicate themes and potentially dangerous content should be created [12]. RL-NLG algorithms can improve communication, information distribution, and user experience by addressing unethical and unbiased text production, thereby building a more inclusive and environmentally conscious AI environment.

2.5 Literature Gap

The present literature on RL-NLG focuses mostly on generic applications. Specialized domains, such as legal or medical situations, where nuanced language creation is critical, have received less attention. Furthermore, there is a significant gap in the literature addressing real-time adaptations of RL-NLG modeling to dynamic settings, which is critical for applications that require instantaneous and contextually exact responses.

3 Methodology

This study, which is based on interpretivism, recognizes the complexities of life experiences and highlights the necessity of comprehending the subjective interpretations that individuals give to things. This concept is useful in investigating the complex behaviors of RL-NLG networks under specific circumstances [13]. A deductive method will be used, beginning with a framework of theory developed from a thorough survey of the RL-NLG literature. This methodical methodology enables the formation of predictions based on existing ideas, which are subsequently evaluated against empirical facts. A descriptive study design will be used to extensively document and examine RL-NLG systems. This approach is ideal for investigating the properties, behaviors, as well as patterns of these networks across a wide range of applications other domains. It allows for a thorough examination of the phenomenon that's under investigation. Secondary data will be gathered from trustworthy sources such as academic journals, proceedings for conferences, technical reports, and commercial publications. This data set includes RL computer programs, NLG applications, metrics of performance, as well as case studies. The careful categorization and investigation of the acquired material will be part of the complete review [14]. The Natural Language Toolbox (NLTK), spaCy, including Gensim Python modules will be used for advanced speech processing tasks which include tokenization, and the lemmatization recognition of names of entities, and evaluation of sentiment. These tools allow for detailed linguistic examination of created text. For quantitative analysis, statistical programming packages such as R, as well as specific libraries such as pandas as well as NumPy in the case of Python, will be used. Examining RL engine performance measures, convergence interest rates, as well as NLG high-quality output scores is part of this. Git will be used for version control, enabling for code modification tracking and sharing of code [15]. Technologies like GitHub will make it easier for teams to collaborate while also ensuring repeatability and accountability. RL algorithms will be built with widely used frameworks like OpenAI Gym as well as TensorFlow. This connection allows for the seamless experimentation and assessment of RL-NLG models in a variety of contexts. Mechanisms will be implemented to anonymize and secure sensitive data, as well as to ensure compliance with ethical principles for the use of secondary data [16]. Precautions will also be taken to address any possible prejudices in the datasets used.

4 Results

4.1 RL Algorithm Integration and Optimization

The subject of RL Algorithm Implementation and Optimization is to ensure facilitating the seamless incorporation of Reinforcement Learning (RL) methodologies into NLG, or natural language generation, frameworks, as well as improving the effectiveness of these algorithms for creating texts tasks [17].

Integration with NLG Frameworks:
This element entails carefully integrating RL algorithms alongside NLG systems. It will be necessary to modify RL techniques to function in tandem with NLG designs, ensuring that the output text meets the desired standards [18]. Existing neural networks pipelines may need to be modified to enable RL-based decisions to be made as part of this integration.

Algorithm Selection and Customization:
This theme is concerned with the selection of suitable algorithms for RL for NLG problems. Proximal Regulation Optimization (PPO), Underground Deterministic models Policy Gradient (DDPG), and other algorithms can be modified to meet the specific needs of text generation [19]. To maximize performance, personalization may entail tweaking hyperparameters, developing reward functions, and adopting exploration techniques (Fig. 5).

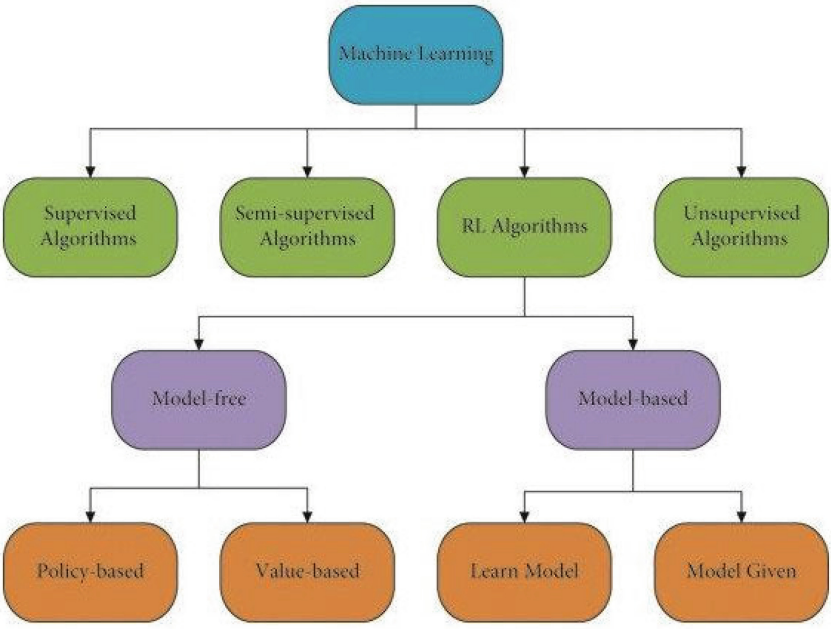


Fig. 5. RL Algorithm

Incorporating State-of-the-Art Techniques:
It is critical to stay up to date on the newest advances in RL in order to achieve optimum efficiency in RL-NLG networks. This may entail utilizing cutting-edge algorithms such as Soft Actor-Critic (SAC), Dual Delayed DDPG as (TD3), or other advanced methods [20]. The goal is to use the most effective strategies for the NLG scenario at hand.

Performance Metrics and Evaluation:
A rigorous evaluation method is required while optimizing the RL algorithms for NLG. Defining appropriate performance measurements, such as the acronym BLEU scores, ROUGE metrics, or customized NLG evaluation criteria, is required. The metrics chosen should be in line with the specific goals of the NLG implementation, ensuring that the output text satisfies the appropriate quality criteria. Finally, the goal of RL Algorithm Integration including Optimization is to improve the synergy underlying RL and NLG, resulting in the establishment of resilient text generation algorithms capable of providing high-quality, contextually pertinent output across a variety of applications including domains [21, 22]. This theme is critical to the advancement of NLG technology.

4.2 Text Processing and Linguistic Analysis

Text Processing as well as Grammatical Analysis is a theme that includes the complex management of text data as well as the implementation of linguistic approaches to improve the quality as well as contextuality that generated language in Positive reinforcement Learning-based natural language synthesis (RL-NLG) systems [23].

Advanced Text Processing:
This aspect entails utilizing specialist tools such as the Natural Language Toolkit. (NLTK), spacey, as well as Genism for complex text analysis. Tokenization, lemmatization as recognized named entities (NER), and evaluation of sentiment are examples of such procedures [24]. These techniques transform unstructured text into structured information that RL-NLG machines can use efficiently.

Linguistic Analysis for Quality Assessment:
It is critical to incorporate linguistic analysis when evaluating the level of quality of published text. Syntactic analyzing, semantic role categorizing, and discussion analysis allow for a more in-depth comprehension of the supplied content [25]. Furthermore, analysis of language aids in finding and correcting faults with coherence, fluency, as well as context relevance (Fig. 6).

Named Entity Recognition (NER) for Contextual Precision:
The ability to recognize and classify items inside text is critical for contextual correctness. Names of individuals, places, organizations, and other entities are identified and classified using NER algorithms. This data enhances the output text, guaranteeing it is relevant to specific areas and applications.

Sentiment Analysis for Contextual Appropriateness:
Using sentiment analysis techniques helps evaluate the emotional tone as well as sentiment communicated in the created text. This is especially important for applications requiring comprehension and matching of emotional context, such as interactions with

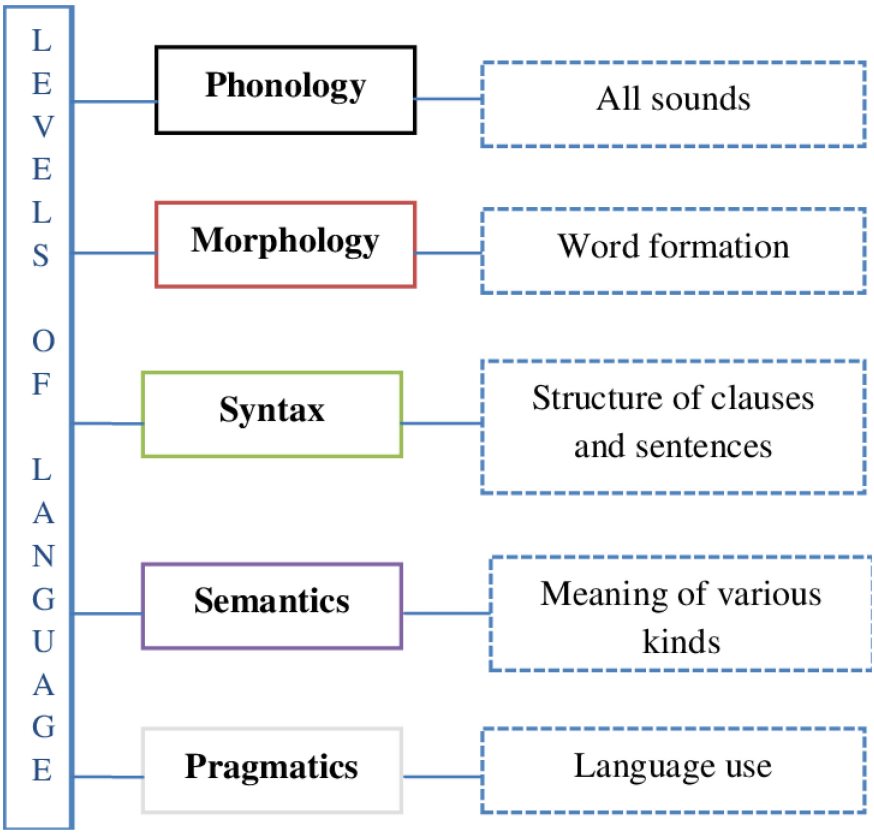


Fig. 6. Levels of Natural Language

customers with sentiment-specific creation of material. RL-NLG methods can produce material that is not solely technically correct but also appropriate for the situation and culturally refined by adding advanced language processing along with linguistic analysis [26]. This subject is critical in improving the quality including relevance of created text, hence adding to the general efficiency and usefulness of RL-NLG algorithms across a wide range of applications and areas.

4.3 NLG Contextual Adaptability and Domain-Specific Performance

The issue of NLG Contextual Being adaptable and Domain-specific Success focuses on the potential of Positive reinforcement Learning-based naturally language generation (RL- NLG) systems to adjust to different situations and excel in specialized fields.

Contextual Adaptability Across Domains:
This aspect entails assessing the RL-NLG system's ability to generate contextually suitable text across a broad spectrum of applications as well as subject areas [27]. It evaluates the model's capacity to recognize and respond to various clues from the environment, ensuring that the output text is relevant and meaningful regardless of domain.

Performance in Specialized Domains:
This component focuses on how effectively the RL-NLG algorithm performs in specific to the domain circumstances such as law, medicine, or technology. It assesses the system's capacity to generate sophisticated and accurate speech customized to the unique requirements as well as terminology of these specialist fields (Fig. 7).



Fig. 7. Domain-specific Performance

Adaptation to Dynamic Environments:
In dynamic situations, RL-NLG systems must adapt in real time to changing contexts. This includes situations where circumstances rapidly changes, for example customer support contacts or live broadcast coverage [28]. The flexibility of the RL-NLG technology to dynamically modify its output under various conditions is critical for its usefulness.

User-specific Customization and Preferences:
For individualized NLG interactions, it is critical to understand and accommodate user preferences. This entails adapting the resulting text to the different users' needs, tastes, and modes of communication. Customizing RL-NLG devices to user-specific requirements improves the overall satisfaction of users and utility [29]. This theme addresses the practical usability of RL-NLG structures across an extensive variety of everyday situations by analyzing NLG Contextual Being adaptable and Domain-specific Effectiveness. It guarantees these systems may produce high-quality, appropriate for the situation text in a variety of settings and specialized areas, hence increasing their adaptability and efficacy in a variety of applications.

4.4 Ethical Considerations and Bias Mitigation

The topic of moral considerations along with Bias Mitigation through Reinforcement is discussed. Learning-based organic language production (RL-NLG) addresses the important need for conscientious and unbiased text production, ensuring that RL-NLG systems' output adheres to ethical principles and avoids reproducing harmful biases [30].

Formulating Ethical Reward Functions:
One critical issue is to create incentive systems that incorporate ethical considerations. Outputs containing offensive, discriminatory, or hurtful content should be penalized by these functions. It is critical for responsible authors of texts to ensure that the real-time agent receives precise instruction on ethical bounds (Fig. 8).

Data Preprocessing for Bias Reduction:
Training data preprocessing is critical for decreasing biases within RL-NLG models. Debiasing technologies and training methods that are adversarial can be used to detect and correct biased tendencies in data, increasing justice and equity in output text [31].

Regular Bias Auditing and Monitoring:
Constant oversight and tracking of RL-NLG outputs is required to detect and correct any biases. Periodic assessments, utilizing techniques such as fairness indications and unfairness-aware evaluation metrics, aid in the identification and mitigation of biases that may occur during RL training.

Transparency and Explain ability:
The importance of transparency regarding the RL-NLG process cannot be overstated. Users should understand how RL models work, including information on the data used for training, compensation functions, and any biases. Furthermore, the algorithm should be built to provide justifications for its created outputs, providing readers with clarity concerning the text's sources and reason [32–36].

Guidelines for Handling Sensitive Topics:
It is critical for responsible creators of texts to establish clear criteria for dealing with sensitive themes such as belief systems, politics, or controversial matters. In delicate subject conversations, RL-NLG solutions should include methods to ensure that developed content is courteous and considerate. RL-NLG systems can improve interactions, knowledge dissemination, including user experience by adding these ethical issues and

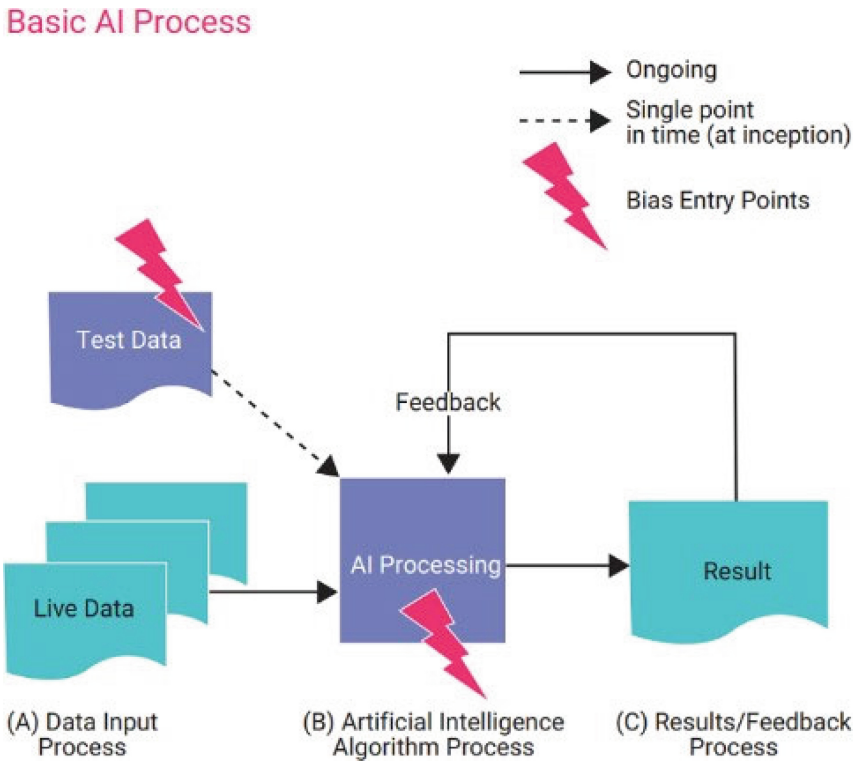


Fig. 8. Ethical Considerations in Machine Learning

bias mitigation measures [37–40]. This subject emphasizes the necessity of responsibly AI development, with the goal of creating an additional equitable and welcoming AI environment. Moreover, the proposed model can be efficiently used for other applications of deep learning and machine learning other than RG-NLG [41, 42].

5 Evaluation and Conclusion

5.1 Conclusion

Finally, this study investigated the incorporation of reinforcement learning techniques in developing natural languages using an interpretive framework and a deductive technique. The descriptive design allowed for a thorough investigation, although secondary data collecting offered a solid foundation. The technical techniques included the integration of RL algorithms, advanced processing of text, analysis of language, contextual versatility, and moral considerations. We've gained a better knowledge of RL-NLG systems as a result of our efforts, paving the door for greater sophistication and environment aware text production. This study not only advances the subject of NLG, but it also underscores the significance of ethical and objective AI applications. The discoveries have applications ranging from artificial intelligence to specific content development.

5.2 Recommendations

Several significant recommendations come from the research. For starters, more research into improved RL algorithms optimized for NLG purposes is needed. Furthermore, combining sentiment analysis with identification of named entities can improve contextuality. Regarding specialized NLG obligations, domain-specific modifications should be considered. RL-NLG systems must incorporate ethical principles and bias detection techniques. It is critical to conduct regular audits for fairness and openness. Finally, to enable ethical AI applications, collaboration between AI programmers, linguists, and philosophers of ethics should be promoted. These suggestions collectively push RL-NLG forward more adaptive, contextually sensitive, as well as unethical systems.

5.3 Future Work

Future study should concentrate on improving RL-NLG systems by investigating novel RL techniques and strategies for text production. Contextuality will be improved by investigating new linguistic assessment techniques and implementing real-time adaptation features. For customized NLG applications, domain- particular customizations should be fine-tuned. Furthermore, research initiatives should be focused on the development of comprehensive guidelines for ethics and prejudice mitigation measures. Further research could focus on user-specific modification to improve personalization. Finally, multidisciplinary collaboration among AI researchers, ethicists, as well as domain specialists will be critical in developing the subsequent generations of RL-NLG devices that are both technologically advanced and ethically accountable.

References

1. Scialom, T.: Natural Language Generation with Reinforcement Learning. Doctoral dissertation, Sorbonne Universite (2022)
2. Gao, J., He, D., Tan, X., Qin, T., Wang, L., Liu, T.Y.: Representation degeneration problem in training natural language generation models (2019). arXiv preprint arXiv:1907.12009
3. Wolf, T., et al.: Huggingface's transformers: State-of-the-art natural language processing (2019). arXiv preprint arXiv:1910.03771
4. Martin, A., Quispe, G., Ollion, C., Le Corff, S., Strub, F., Pietquin, O.: July. Learning natural language generation with truncated reinforcement learning. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 12–37 (2022)
5. Wolf, T., et al.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45 (2020)
6. Morris, J.X., Lifland, E., Yoo, J.Y., Qi, Y.: Textattack: A framework for adversarial attacks in natural language processing. In: Proceedings of the 2020 EMNLP, Arvix (2020)
7. Ohashi, A., Higashinaka, R.: Adaptive Natural Language Generation for Task- oriented Dialogue via Reinforcement Learning (2022). arXiv preprint arXiv:2209.07873
8. Ritschel, H., André, E.: Shaping a social robot's humor with natural language generation and socially-aware reinforcement learning (2018)
9. Ammanabrolu, P., Hausknecht, M.: Graph constrained reinforcement learning for natural language action spaces (2020). arXiv preprint arXiv:2001.08837

10. Maulud, D.H., Zeebaree, S.R., Jacksi, K., Sadeeq, M.A.M., Sharif, K.H.: State of art for semantic analysis of natural language processing. Qubahan Acad. J. **1**(2), 21–28 (2021)
11. Lu, S., et al.: Codexglue: A machine learning benchmark dataset for code understanding and generation (2021). arXiv preprint arXiv:2102.04664
12. Liu, G., Hsu, T.M.H., McDermott, M., Boag, W., Weng, W.H., Szolovits, P., Ghassemi, M.: Clinically accurate chest x-ray report generation. In: Machine Learning for Healthcare Conference, pp. 249–269. PMLR (2019)
13. He, D., Zhao, X., Huang, J., Li, F., Liu, X., Wen, S.: Read, watch, and move: Reinforcement learning for temporally grounding natural language descriptions in videos. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, No. 01, pp. 8393–8400 (2019)
14. Torfi, A., Shirvani, R.A., Keneshloo, Y., Tavaf, N., Fox, E.A.: Natural language processing advancements by deep learning: A survey (2020). arXiv preprint arXiv:2003.01200
15. Janner, M., Li, Q., Levine, S.: Offline reinforcement learning as one big sequence modeling problem. Adv. Neural. Inf. Process. Syst. **34**, 1273–1286 (2021)
16. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L.: A survey of human-in-the- loop for machine learning. Futur. Gener. Comput. Syst. **135**, 364–381 (2022)
17. Rasley, J., Rajbhandari, S., Ruwase, O., He, Y.: Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3505–3506 (2020)
18. Hashimoto, T.B., Zhang, H., Liang, P.: Unifying human and statistical evaluation for natural language generation (2019). arXiv preprint arXiv:1904.02792
19. Chernyavskiy, A., Ilvovsky, D., Nakov, P.: Transformers: "the end of history" for natural language processing? In: Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part III 21, pp. 677–693. Springer International Publishing (2021)
20. Goldberg, Y.: Neural Network Methods for Natural Language Processing. Springer Nature (2022)
21. Zhao, T., Xie, K., Eskenazi, M.: Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models (2019). arXiv preprint arXiv:1902.08858
22. Lauriola, I., Lavelli, A., Aiolli, F.: An introduction to deep learning in natural language processing: models, techniques, and tools. Neurocomputing **470**, 443–456 (2022)
23. Chen, Y., Wu, L., Zaki, M.J.: Reinforcement learning based graph-to-sequence model for natural question generation (2019). arXiv preprint arXiv:1908.04942
24. Canhoto, A.I., Clear, F.: Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential. Bus. Horiz. **63**(2), 183–193 (2020)
25. Song, K., Tan, X., Qin, T., Lu, J., Liu, T.Y.: Mass: Masked sequence to sequence pre-training for language generation (2019). arXiv preprint arXiv:1905.02450
26. Jin, D., Jin, Z., Zhou, J.T., Szolovits, P.: Is bert really robust? a strong baseline for natural language attack on text classification and entailment. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, No. 05, pp. 8018–8025 (2020)
27. Strieth-Kalthoff, F., Sandfort, F., Segler, M.H., Glorius, F.: Machine learning the ropes: principles, applications and directions in synthetic chemistry. Chem. Soc. Rev. **49**(17), 6154–6168 (2020)
28. Ekman, M.: Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow. Addison-Wesley Professional (2021)
29. Batra, R., Song, L., Ramprasad, R.: Emerging materials intelligence ecosystems propelled by machine learning. Nat. Rev. Mater. **6**(8), 655–678 (2021)
30. Peng, B., Zhu, C., Li, C., Li, X., Li, J., Zeng, M., Gao, J.: Few-shot natural language generation for task-oriented dialog (2020). arXiv preprint arXiv:2002.12328

31. Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., Sen, P.: A survey of the state of explainable AI for natural language processing (2020). arXiv preprint arXiv:2010.00711
32. Paullada, A., Raji, I.D., Bender, E.M., Denton, E., Hanna, A.: Data and its (dis) contents: a survey of dataset development and use in machine learning research. Patterns, **2**(11) (2021)
33. Liang, P.P., Zadeh, A., Morency, L.P.: Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions (2022). arXiv preprint arXiv:2209.03430
34. Radford, A., et al.: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning, pp. 8748–8763. PMLR (2021)
35. Khan, S., et al.: HCovBi-caps: hate speech detection using convolutional and Bi-directional gated recurrent unit with Capsule network. IEEE Access **10**, 7881–7894 (2022)
36. Khan, S., et al.: BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection. J. King Saud Univ.-Comput. Inf. Sci. **34**(7), 4335–4344 (2022)
37. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput. Surv. **55**(9), 1–35 (2023)
38. Fazil, M., Khan, S., Albahlal, B.M., Alotaibi, R.M., Siddiqui, T., Shah, M.A.: Attentional multi-channel convolution with bidirectional LSTM cell toward hate speech prediction. IEEE Access **11**, 16801–16811 (2023)
39. Khan, S.: Artificial intelligence virtual assistants (Chatbots) are innovative investigators. Int. J. Comput. Sci. Netw. Secur. **20**(2), 93–98 (2020)
40. Khan, S., et al.: Transformer architecture-based transfer learning for politeness prediction in conversation. Sustainability **15**(14), 10828 (2023)
41. Gupta, G., et al.: DDPM: a dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms. Diagnostics **13**(6), 1093 (2023)
42. Khan, S.: Business Intelligence Aspect for Emotions and Sentiments Analysis, pp. 1–5: IEEE



i-BMD: AI-Based Opportunistic Screening for Osteoporosis on Abdominal CT Using Deep Learning

Soumya Jindal[1](✉) and Meemansa Jindal[2]

[1] Department of IT, IGDTUW, Delhi, India
soumyajindal0707@gmail.com
[2] Department of Radio-Diagnosis, MAMC, Delhi, India

Abstract. Osteoporosis is a common curable ailment, more commonly seen in geriatric population, particularly prevalent in post-menopausal women. It often presents with generalized bone pain, and fragility fractures, most common in the distal radius, vertebrae, and pelvic bones. When it comes to diagnosing osteoporosis, the best and widely accepted golden standard is the DEXA scan, often interpreted in the form of a T-score and a Z-score of less than −2.5. Though, even in the modern era of evolved imaging, DEXA machines continue to be scarce and are not so profitable to be installed, leading to a massive population of the elderly having untreated and under-diagnosed osteoporosis. CT scan is one of the most common imaging techniques utilized in India for a variety of indications. There has been a lot of buzz regarding the discussions over – "What if one could utilize HU values for diagnosing osteoporosis?" "What if one could predict the risk of fragility fractures without letting one happen?". A lot of research lately has been focused on employing ML algorithms for achieving this, by using texture analysis of the vertebral body, cortex and trabecular pattern, HU values, and also by using Genant classification of vertebral body fractures. This article aims to review the current situation of opportunistic screening of osteoporosis worldwide and the implications of ML in achieving the same.

Keywords: Osteoporosis · Deep learning · CT · DEXA

1 Introduction

Pathologic, traumatic, or atraumatic reasons can all result in vertebral compression fractures. The latter might happen following low stress as a result of bone mineralization loss, which would lower mechanical bone strength. The two chronic skeletal diseases osteopenia and osteoporosis, which primarily afflict the elderly and chronically unwell, are the most frequent underlying causes. Osteoporosis-related fractures play a major contribution in disability and demise. Within the United States, there are more than 1 million fractures characterized by fragility every year (injuries resulting from trauma no more severe than a fall from standing height), and as more "baby boomers" reach their seventies, this number will rise [1]. Despite a substantial risk of more fractures,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 319–331, 2025.
https://doi.org/10.1007/978-3-031-75167-7_26

less than one-third of individuals who have experienced fragility fractures receive the proper evaluation and treatment for osteoporosis. The ultimate outcome of an osteoporosis screening would be to estimate every possibility of any fragility fracture or any risk of the same to occur at any time over the next five to ten years. Such projections still need to be improved, although some estimations are already available. If a T score indicative of low bone mineral density (−2.5 or lower) is observed in a postmenopausal woman with no additional risk factors are present, her cumulative 10-year risk of sustaining fragility fractures is below 5% at age 50, but exceeds 20% when she reaches age 65 [2]. Additional risk factors, especially a history of fragility fracture, substantially raise absolute risk.

The need for classifiers in disease screening is gradually growing due to the vast amount of data that routine clinical examinations and research studies provide. The understanding of illness patterns and the ability to make useful forecasts are both aided by pattern recognition techniques. SVM produces the best results in noisy and complicated environments when compared to other classification approaches. There have been studies using ANN, and CNN for the screening of osteoporosis on non-contrast, contrast-enhanced CT scans and even on radiographs. The existing ML-based approaches for the screening of osteoporosis on routine imaging scans have been properly reviewed and listed in this article.

2 Methodology

In July 2023, a literature search on the PubMed repository was conducted by employing specific keywords related to the topic of interest, such as ((“Artificial Intelligence” [All Fields] OR “Machine learning” [MeSH Terms]) AND (“Osteoporosis” [All Fields]) AND “Opportunistic screening” [All Fields]). This search, followed by a free full-text filter, resulted in a total of 73 English-language articles. Further, in this initial screening, articles lacking machine learning-based methodologies were systematically removed in accordance with PRISMA guidelines, as visually presented in (Fig. 1).

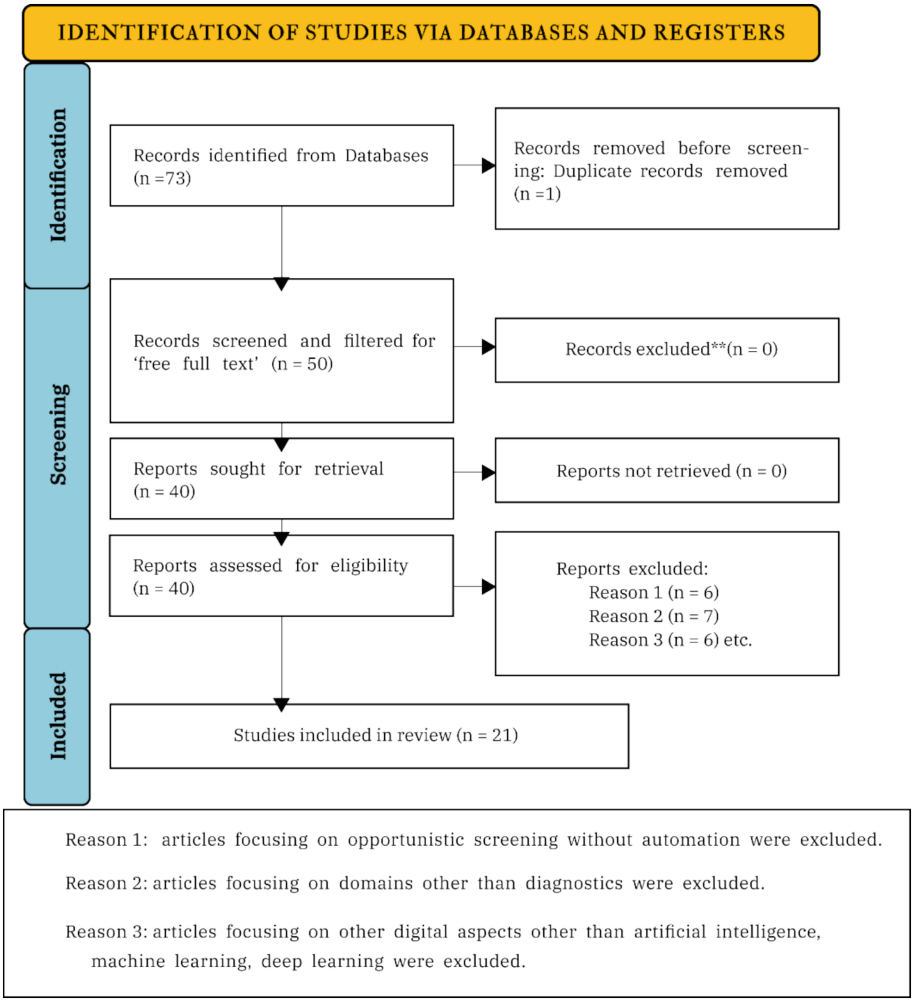


Fig. 1. Procedure followed in adherence with the PRISMA 2020 guidelines for identifying the relevant articles/studies for the research.

3 Literature Review

3.1 Public Burden of Osteoporosis

Although osteoporosis is common, curable, and carries a high lifelong fracture risk, it is nevertheless gravely underdiagnosed and undertreated [1]. In the US, more than two-thirds of people with a significant osteoporosis-related fracture do not receive pharmacologic medications to lower fracture risk, and approximately 50% of Medicare beneficiaries, especially females, have never undergone bone mineral density (BMD) testing performed [3]. Clear and prompt diagnosis of a fragility fracture followed by timely management can reduce the evident risk of fracture. It is said that after sustaining fragility

fractures, the risk of sustaining another one is increased by many folds, in both men and women [4].

3.2 Current Screening and Diagnostic Modalities

Currently, two methods, namely DEXA scan and QCT, are available for assessing bone density. QCT provides direct density measurements in Hounsfield Units (HUs), allowing for distinct measurements of trabecular and cortical bone density as well as precise 3-dimensional anatomical localization [5].

These techniques facilitate early intervention and the potential to enhance the accuracy of non-invasive fracture risk prediction. Both CT scan and QCT employ a similar method to measure bone mineral density, but QCT involves higher radiation exposure. The Hounsfield Unit (HU) is the standardized unit for representing values from reconstructed CT scans. This unit system transforms initial measurements of the linear attenuation coefficient into values where water and air are set at 0 and ± 1000, respectively. Cortical bone typically has values ranging from $+1000$ to $+1600$ HU, whereas trabecular bone has lower HU values. Osteoporotic bone exhibits a reduced overall HU value and a coarser trabecular pattern on CT scans [5].

DEXA scanning is a widely adopted practice for the evaluation of bone health, focusing on the radius, proximal femur and lumbar spine, adhering to well-established protocols. To meet the inclusion criteria, patients are required to have a valid T-score from any of these specific anatomical sites. The classification of patients is based on their minimum recorded T-scores, leading to categorization into three groups: those with normal bone density (T-score > -1.0), individuals with osteopenia ($-2.5 <$ T-score ≤ -1.0), and those diagnosed with osteoporosis (T-score ≤ -2.5), aligning with recognized clinical guidelines [6]. While dual-energy X-ray absorptiometry (DXA) of the lumbar spine and hips stands as the universally acknowledged gold standard for osteoporosis diagnosis, its utilization remains infrequent [2, 7]. The imperative challenge is the quest for reliable, cost-effective solutions to advance the detection of osteoporosis. The Fracture Risk Assessment Tool (FRAX®) represents a pivotal development in the field of osteoporosis care. This sophisticated tool, which calculates an individual's 10-year probability of experiencing a hip fracture, was introduced in 2008 as a collaborative effort involving the World Health Organization Centre in Sheffield, UK [8, 9]. The inclusion of factors such as age, sex, BMI provide for a more tailored and comprehensive risk assessment through FRAX® [8]. Apart from these, there is a provision of including 7 varied clinical risk factors in FRAX® such as past history of fragility fractures, parental history of hip fractures, smoking, alcohol, use of corticosteroids, and other osteoporosis-related risk factors [10]. Since its introduction, FRAX® has been widely accepted into clinical practice globally. So far, 86 discrete models of FRAX® have been developed and are in implementation phase in more than 70 different countries, for risk assessment of osteoporosis [11]. This has only provided a higher accuracy in fragility fracture risk predictions but has also empowered clinicians with valuable information for facilitating preventive measures for guiding treatment protocols, ultimately improving the healthcare quality for such patients.

3.3 Utility of CT in Detecting Osteoporosis

In 2011, there were more than 75 million computed tomography (CT) scans performed in the US, the majority of which contain potentially helpful BMD data [12]. No additional money, patient time, resources, technology, or radiation exposure are needed for the retrieval of BMD data from body CT scans ordered for other purposes. These data can also be obtained retroactively. Therefore, it might increase efforts to test the general population for osteoporosis.

It has been shown that the AUC curves for attenuation value thresholds do not significantly differ in contrast-enhanced and non-contrast-enhanced scans.

The ability of CT scans to detect the presence of cortical thinning, morphological abnormalities, and osteoporotic compression fractures, gives a significant edge over DXA. The widely used Genant semi-quantitative screening (Fig. 2) provides an easy visual grading on sagittal CT cuts for wedge deformity, biconcave osteoporotic compression, and crush deformity [13]. This reduces the number of false negatives which can easily occur in DXA if the compression fracture is missed, besides the skewing of measurements by sclerotic bone lesions, vascular calcifications, adipose tissue, etc.

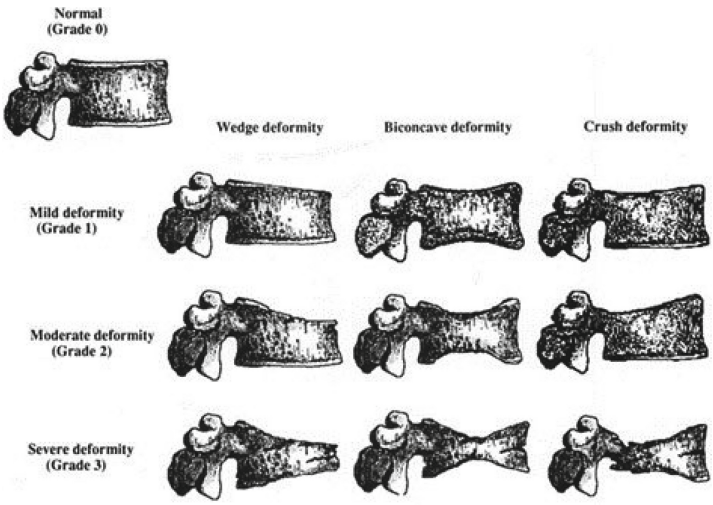


Fig. 2. Genant classification [12]

Despite the absence of a bone density diagnostic, a vertebral fracture is consistent with the diagnosis of osteoporosis; it is a sign that pharmacologic therapy with osteoporosis medicine is necessary to lower the risk of further fractures [2]. DXA has a limited application in patients with spinal deformities or those who have undergone prior spinal surgery.

Scalability is another benefit that CT screening may have over DXA screening. Results at L1 are equally as accurate as results at subsequent levels, including outcomes from multilevel assessment, if not more so. The ability to quickly identify the L1 level boosts efficiency and reproducibility. It is incorporated into all common chest and abdominal CT scans, significantly raising the possibility of a higher screening yield overall. It has been shown that mean attenuation values of less than 150 Hounsfield Units (HU) corroborate with occurrence of osteopenia and osteoporosis with a sensitivity of over 90%, when compared with DXA [14, 15]. Broadly speaking, when L1 trabecular attenuation registers at less than 100 HU at 120 kVp, it demonstrates relatively high specificity, approximately 90%, in identifying individuals with osteoporosis based on DXA results. This revelation highlights the potential for improving the classification of individuals at risk of fractures through the application of CT examination, especially considering the limitations of planar DXA as a reference standard [14].

Furthermore, the value of 90 HU emerges as an intriguing threshold for estimating the risk of osteoporotic fractures [16]. The results of Lee et al.'s Receiver Operating Characteristic (ROC) analysis suggest that CT-derived femoral neck T-scores might marginally outperform L1 attenuation values in identifying individuals susceptible to subsequent hip fractures. By adopting a higher threshold of sensitivity for example through a lower attenuation value, one can assume that the trabecular attenuation value of L1 can serve as an initial screening technique, offering valuable insights for excluding elevated hip fracture risk [15].

Some studies have proposed the use of facial CTs for estimating the risk of maxillofacial fractures in the elderly in case of low-impact traumas [17].

Pisu et al. have proposed an intriguing strategy for individuals aged 65 years and older, both women and men, who are already undergoing abdominal CT scans and lack recent Dual-Energy X-ray Absorptiometry (DXA) assessments. Their suggestion entails the implementation of CT screening for osteoporosis, a measure that seems not only cost-effective but, in many scenarios, potentially cost-saving [18]. This innovative approach not only optimizes the utilization of existing medical procedures but also underscores the importance of proactive osteoporosis screening, particularly in older demographics, where the risk of fractures is notably heightened.

3.4 AI in Opportunistic Screening

When utilized in everyday practice, abdominal CT scans typically yield useful data on body composition that can prove pivotal in the patient's management. The use of additional imaging data from CT-based opportunistic screening, which can be used to assess patient risk and forecast future negative clinical outcomes, is gaining popularity. These pertinent biometric measurements may now be evaluated quickly and accurately using artificial intelligence algorithms, which allows for full automation. Examples of potential scope for metabolic and cardiovascular screening on CT include assessment of

aortic calcium for cardiovascular risk assessment, measuring subcutaneous and splanchnic fat for assessment of metabolic syndrome, measuring muscle density and bulk for sarcopenia diagnosis, and measuring visceral and subcutaneous fat. Table 1 shows the critical analysis of various automated approaches toward opportunistic screening of osteoporosis.

Table 1. Comparative analysis between prior research works.

	Aims & Objectives	Type	Place of Study	Modality Chosen	Results	Limitations
[18]	To compare automated Bone Mineral Density (BMD) measurements between cardiac CT scans used for Coronary Artery Calcium (CAC) scoring and CT chest scans employed for carcinoma lung screening	Cross-sectional | Harbor UCLA Lundquist Institute	CT	Significant correlation (value of R2:0.95) was observed between Auto BMD measurements (obtained for the respective vertebral bodies) in both scans		• The fact that AutoBMD's ability to detect vertebrae in both the CT scans depends on the patient's particular field of view is a possible drawback • TheAutoBMD readings may be impacted if any of these vertebrae are brokenorobstructed by artifacts
[19]	To correlate between Hounsfield Unit (HU) values obtained from the proximal femur in dualenergy X-ray absorptiometry (DEXA) and CT colonoscopy measurements	Retrospective study | Bethesda USA	CT colonoscopy	To identify individuals with low bone mineral density (BMD), specific HU thresholds were determined. An HU value of 214 exhibited 100% sensitivity, while 59 HUs showed 100% specificity. An HU threshold of 113 resulted in 73% sensitivity and 76% specificity		• The main drawback of the study was its retrospective design, which resulted in potential selection bias when recruiting patients who had undergone both a DEXA scan and a CT colonoscopy
[20]	Examining the correlation between bone mineral density (BMD) and dual-energy X-ray absorptiometry (DEXA) T scores alongside computed tomography (CT) Hounsfield Unit (HU) values, with the aim of establishing reference values for CT HU measurements	Retrospective study | Saudi Arabia	CT Lumbar spine	The mean attenuation values at L1-L4 were 174.05, 120.45, and 115 HU for normal, osteopenia, and osteoporosis		• Calculationsand measurements done on different CT and DEXA scanners;therefore, reproducibility of results is questionable • Limitedsample size • 2-year gap between CT and DEXA imaging • CT HU values were derived solely from trabeculatedbone, whereasDEXA
						encompassed both trabeculated and cortical bone, potentially impacting comparability

(*continued*)

Table 1. (*continued*)

	Aims & Objectives	Type | Place of Study	Modality Chosen	Results	Limitations
[21]	Texture analysis of craniofacial bones in non-contrast CT scans of the head for osteoporosis detection	Cross-sectional | Japan	Non-contrast headCT scans	• Non-contrast head CT scans involved manual segmentation of bilateral sphenoid triangles, clivus, and mandibular condyles • 41 distinct textural features were extracted • Patient groups: 29 with normal bone density and 29 with osteoporosis • In the left sphenoid region: 24 texture characteristics displayed significant differences between osteoporosis and normal BMD • Both sides of mandibular condyles: 22 texture characteristics exhibited significant variations This set encompassed: o 6 histogram features o 3 GLCM features (Gray Level Co-occurrence Matrix) o 9 GLRL features (Gray Level Run Length) o 2 Law's features o 2 GLGM features (Gray Level Gradient Matrix)	• Limited sample size • 1-year gap between CT and DEXA imaging • CT HU values represented trabeculated bone only, while DEXA included both trabeculated and cortical bone
[22]	Evaluation of the capacity of ML model (RF classifier) to identify normal vs osteoporotic fracture patients by regional analysis of volumetric BMD and 3D texture data	Cross-sectionall Muchen Germany	Contrast-enhancedCT scans	High discriminating power was demonstrated by the random forest classifier (AUC = 0.88). The suggested algorithm's AUC (AUC = 0.64) was significantly greater than the AUC of volumetric global BMD alone	• Clinical parameters were excluded • Unavailability of DEXA • Excluded biomechanical parameters based on cortical bone or finite element analysis (FEA)

(*continued*)

Table 1. (*continued*)

	Aims & Objectives	Type	Place of Study	Modality Chosen	Results	Limitations
[23]	A new, reliable method for measuring marrow spacing and trabecular bone thickness in vivo at a low resolution. Utilizing voxel sizes similar to TB thickness, the methodology efficiently addresses the partial volumizing effects of in vivo imaging	Cross-sectional | Iowa city, USA	CT	The novel method's measurements of marrow spacing and trabecular thickness showed significant correlations(R2 [0.87,0.83]) with bone strength. Additionally, both the measurements permitted distinction between male and female volunteers as well as between athletes and non-athletes (p [0.01, 0.04] and [0.005, 0.03] respectively)	• None mentioned
[24]	To analyze the effects of aging on trabecular texture, bone mineral density (BMD), and the value of texture analysis in predicting the likelihood of fracture	Cross-sectional | Japan	CT	It was determined the breadth of the trabecula (T-texture) and the intertrabecular space (Itexture). Women who sustained fractures had BMD that was noticeably low and high I-texture. BMD and I-texture showed the highest sensitivity and highest specificity, respectively for predicting fracture risk in females aged 60 to 69	• None mentioned
[25]	To identify patients at risk for vertebral fragility fractures, by using ML based bone texture analysis (TA)	Cross-sectional | Zurich, Switzerland	CT	Combining machine learning and bone texture analysis enables highly accurate identification of patients at high risk for insufficiency fractures on CT	• Small dataset • Excluding cortical thickness, vertebral size etc • TA features limited to MaZda software and newer approaches were not used
[26]	Correlating Xray parameters, biochemical parameters and patient's symptoms for detecting osteoporosis and comparing performance of ANN with logistic regression	Cross-sectionall China	Xray	The diagnostic outcomes of the artificial neural network model and those of logistic regression are contrasted. The sensitivities are, respectively, 94.5% and 63.6%. The specificities are, respectively, 96.9% and 87.5%	• Small dataset
[27]	To compare machine learning algorithms based on support vector machines (SVM), random forests, artificial neural networks (ANN), and logistic regression (LR) based on clinical decision-making tools	Cross-sectionall Seoul, South Korea	Clinical data	SVM considerably outperformed ANN, LR in terms of area under the curve (AUC) of the receiver operating characteristic for the training set	• Model was prepared for Korean women and therefore reproducibility is limited • Imbalanced class distribution

(*continued*)

Table 1. (*continued*)

	Aims & Objectives	Type | Place of Study	Modality Chosen	Results	Limitations
[28]	To test the viability of utilizing texture analysis in routine contrast-enhanced MDCT scans for opportunistic osteoporosis screening	Cross-sectional | Munich	CT	Using certain textural metrics (energy, entropy, and homogeneity) and SVM, opportunistic osteoporosis screening at the spine had an accuracy of 83%	• Texture analysis and BMD were not correlated
[29]	To compare ML models based on parameters pertaining to the microstructure parameters of the trabecular bone	Case control study | New York, USA	MRI	The classifiers having highest performance for predicting osteoporotic fracture are RUSboosted trees, logistic regression, and linear discriminant. In their respective ways, MRI and FRAX® are both useful for spotting osteoporotic fractures. For the top three classifiers, the femoral head, greater trochanter, and inter-trochanter anatomical regions of the proximal femur produced stronger F1-scores	• Small sample size • Imbalanceddata cohort
[30]	Employing low-dose chest computed tomography (LDCT) scans taken during screening of carcinoma lung to do oppor-tunistic osteoporosis Screening	Cross-sectionall China	CT chest	The area under the curve (AUC) was 0.942 for low BMD differentiation and 0.927 for osteoporosis detection	• Paired LDCT and QCT measurements could have led to selection bias
[31]	To compare the existing ML techniques for opportunistic screening	Reviewl Germany	CT	When compared to traditionalradiological semi-quantitative Genant scoring, automated fracture identification has an accuracy of roughly 80%. When compared to simultaneous phantom-based calibration utilized in conventional quantitative CT (QCT), alternative BMD calibration methods' accuracy faults range from under 5% to over 10%	• None mentioned
[32]	To evaluate the performance of CNN in opportunistic detection of osteoporosis on DPRs	Cross-sectionall Seoul, South Korea	Dental panoramic radiographs (DPRs)	Overall area under the receiver operating characteristic for the topperforming model was 0.858	• Small number of convolutional layers

4 Conclusion

The foundation of our study's inspiration is in the current osteoporosis diagnosing scenarios. The number of studies on the osteoporosis diagnostic system is few, but we also see that the findings of the present studies are occasionally in agreement and occasionally even in conflict. The most crucial element is to look for an appropriate mathematical strategy to combine all the clinical data and X-ray features extracted from osteoporosis patients since too much complex clinical data appears to be another issue for doctors and takes too much time to analyze.

References

1. Raisz, L.G.: Clinical practice. Screening for osteoporosis. N. Engl. J. Med. **353**(2), 164–171 (2005)
2. Marshall., D., Johnell, O., Wedel, H.: Meta-analysis of how well measures of bone mineral density predict the occurrence of osteoporotic fractures. BMJ: Br. Med. J. **312**(7041), 1254 (1996)
3. King, A.B., Fiorentino, D.M.: Medicare payment cuts for osteoporosis testing reduced use despite tests' benefit in reducing fractures. Health Aff (Millwood). **30**(12), 2362–2370 (2011)
4. Elliot-Gibson, V., Bogoch, E.R., Jamal, S.A., Beaton, D.E.: Practice patterns in the diagnosis and treatment of osteoporosis after a fragility fracture: a systematic review. Osteoporos. Int. **15**(10), 767–778 (2004)
5. Lee, I.J., Lee, J.J., Bae, J.H., Hwang, E., Lee, S., Cho, M., et al.: Significance of osteoporosis in facial bone density using computed tomography. J. Craniofac. Surg. **24**(2), 428–431 (2013)
6. Cosman, F., de Beur, S.J., LeBoff, M.S., Lewiecki, E.M., Tanner, B., Randall, S., et al.: Clinician's guide to prevention and treatment of osteoporosis. Osteoporos. Int. **25**(10), 2359–2381 (2014)
7. Baim, S., Binkley, N., Bilezikian, J.P., Kendler, D.L., Hans, D.B., Lewiecki, E.M., et al.: Official positions of the international society for clinical densitometry and executive summary of the 2007 ISCD position development conference. J. Clin. Densitom. **11**(1), 75–91 (2008)
8. Bonjour, P., Compston, J., Dawson-Hughes, B., Delmas, P., Lyon, H.E.H., Ms, F., et al.: Members, observers and secretariat
9. Kanis, J.A., Johnell, O., Oden, A., Johansson, H., McCloskey, E.: FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos. Int. **19**(4), 385–397 (2008)
10. Vandenput, L., Johansson, H., Mccloskey, E.V., Liu, E., Åkesson, K.E., Anderson, F.A., et al.: Update of the fracture risk prediction tool FRAX: a systematic review of potential cohorts and analysis plan. Osteoporos. Int. **17**, 106 (2022)
11. Kanis, J.A., Harvey, N.C., Johansson, H., Liu, E., Vandenput, L., Lorentzon, M., et al.: A decade of FRAX: how has it changed the management of osteoporosis? Aging Clin. Exp. Res. **32**(2), 187–196 (2020)
12. IMV Medical Information Division: IMV 2012 CT market
13. Genant, H.K., Wu, C.Y., van Kuijk, C., Nevitt, M.C.: Vertebral fracture assessment using a semi-quantitative technique. J. Bone Miner. Res. **8**(9), 1137–1148 (1993)
14. Pickhardt, P.J., Pooler, B.D., Lauder, T., del Rio, A.M., Bruce, R.J., Binkley, N.: Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Ann. Intern. Med. **158**(8), 588 (2013)

15. Lee, S.J., Anderson, P.A., Pickhardt, P.J.: Predicting future hip fractures on routine abdominal CT using opportunistic osteoporosis screening measures: a matched case-control study. AJR Am. J. Roentgenol. **209**(2), 395–402 (2017)
16. Lee, S.J., Graffy, P.M., Zea, R.D., Ziemlewicz, T.J., Pickhardt, P.J.: Future osteoporotic fracture risk related to lumbar vertebral trabecular attenuation measured at routine body CT. J. Bone Miner. Res. **33**(5), 860–867 (2018)
17. Pisu, M., Kopperdahl, D.L., Lewis, C.E., Saag, K.G., Keaveny, T.M.: Cost-effectiveness of osteoporosis screening using biomechanical computed tomography for patients with a previous abdominal CT. J. Bone Miner. Res. **34**(7), 1229 (2019)
18. Naghavi, M., De Oliveira, I., Mao, S.S., Jaberzadeh, A., Montoya, J., Zhang, C., et al.: Opportunistic AI-enabled automated bone mineral density measurements in lung cancer screening and coronary calcium scoring CT scans are equivalent. Eur. J. Radiol. Open **10** (2023)
19. Christensen, D.L., Nappo, K.E., Wolfe, J.A., Wade, S.M., Brooks, D.I., Potter, B.K., et al.: Proximal femur hounsfield units on CT colonoscopy correlate with dual-energy X-ray absorptiometry. Clin. Orthop. Relat. Res. **477**(4), 850–860 (2019)
20. Alawi, M., Begum, A., Harraz, M., Alawi, H., Bamagos, S., Yaghmour, A., et al.: Dual-energy X-ray absorptiometry (DEXA) scan versus computed tomography for bone density assessment. Cureus **13**(2) (2021)
21. Kawashima, Y., Fujita, A., Buch, K., Li, B., Qureshi, M.M., Chapman, M.N., et al.: Using texture analysis of head CT images to differentiate osteoporosis from normal bone density. Eur. J. Radiol. **116**, 212–218 (2019)
22. Valentinitsch, A., Trebeschi, S., Kaesmacher, J., Lorenz, C., Löffler, M.T., Zimmer, C., et al.: Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures. Osteoporos. Int. **30**(6), 1275 (2019)
23. Liu, Y., Jin, D., Li, C., Janz, K.F., Burns, T.L., Torner, J.C., et al.: A robust algorithm for thickness computation at low resolution and its application to in vivo trabecular bone CT imaging. IEEE Trans. Biomed. Eng. **61**(7), 2057 (2014)
24. Ito, M., Ohki, M., Hayashi, K., Yamada, M., Uetani, M., Nakamura, T.: Trabecular texture analysis of CT images in the relationship with spinal fracture. Radiology. **194**(1), 55–59 (1995)
25. Muehlematter, U.J., Mannil, M., Becker, A.S., Vokinger, K.N., Finkenstaedt, T., Osterhoff, G., et al.: Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. Eur. Radiol. **29**(5), 2207–2217 (2019)
26. Yu, X., Ye, C., Xiang, L.: Application of artificial neural network in the diagnostic system of osteoporosis. Neurocomputing **19**(214), 376–381 (2016)
27. Yoo, T.K., Kim, S.K., Kim, D.W., Choi, J.Y., Lee, W.H., Oh, E., et al.: Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. Yonsei Med. J. **54**(6), 1321 (2013)
28. Mookiah, M.R.K., Rohrmeier, A., Dieckmeyer, M., Mei, K., Kopp, F.K., Noel, P.B., et al.: Feasibility of opportunistic osteoporosis screening in routine contrast-enhanced multi-detector computed tomography (MDCT) using texture analysis. Osteoporos Int. **29**(4), 825–835 (2018)
29. Ferizi, U., Besser, H., Hysi, P., Jacobs, J., Rajapakse, C.S., Chen, C., et al.: Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data. J. Magn. Reson. Imaging **49**(4), 1029–1038 (2019)
30. Pan, Y., Shi, D., Wang, H., Chen, T., Cui, D., Cheng, X., et al.: Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. Eur. Radiol. **30**(7), 4107–4116 (2020)

31. Engelke, K., Chaudry, O., Bartenschlager, S.: Opportunistic screening techniques for analysis of CT scans. Curr. Osteoporos. Rep. **21**(1), 65 (2023)
32. Lee, K.S., Jung, S.K., Ryu, J.J., Shin, S.W., Choi, J.: Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. J. Clin. Med. **9**(2) (2020)



Machine Learning in IoT: An In-Depth Dataset Analysis Based on Attack Detection

Kajal Tyagi(✉), Anil Ahlawat, and Himanshi Chaudhary

Department of Computer Science, Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India
kajal9013570940@gmail.com, anil.ahlawat@kiet.edu

Abstract. Internet of Things (IoT) has transformed how people interact with the physical environment, generating vast amounts of data from interconnected devices and sensors. This study offers a comprehensive survey of IoT network traffic datasets. The efficient analysis of these massive datasets has become a challenging task, leading researchers and practitioners to harness the power of machine learning algorithms. This paper delves into the exciting field of IoT dataset analysis using machine learning techniques. We explore the different IoT datasets on which ML is implemented, offering insights into the future prospects of this rapidly evolving domain. Overall, we examine seventeen datasets that we found through the analysis of various scientific publications, which we have covered in this study. Additionally, we examined which ML approach is more effective at classifying attacks and how we can implement Feature Engineering or Data Pre-processing before applying ML techniques on these datasets. Many researchers get some limitations while applying the ML techniques on different datasets which we have discussed in our analysis, also they highlights emerging trends and future directions on these datasets. Atlast, we have discussed five well known datasets which has been created and given labels corresponding to their different attack categories.

Keywords: IoT · Machine Learning (ML) · Datasets · Data Preprocessing · Attack samples

1 Introduction

IoT refers to networked systems that enable smooth information flow between physical devices. Remote monitoring and regulation of devices such as medical, healthcare, have self-driving cars, manufacturing robots, connected smart televisions & smart city networks is possible. Over time, IoT devices will outnumber devices having mobility & have accessibility to sensitive data including individual details and information. Increasing attacking surface area leads to higher attack probabilities. To protect communications provided by IoT technologies, IDS must be made as security is a key component of most IoT applications [1]. IoT and the proliferation of connected devices have resulted in the generation of diverse datasets in various domains. Researchers need data that is freely accessible to test new algorithms across numerous configurations and make scientific investigations verifiable and reliable.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 332–348, 2025.
https://doi.org/10.1007/978-3-031-75167-7_27

The application of machine learning algorithms in processing and extracting meaningful insights from IoT data has gained significant attention [2]. Effective ML-based classifier training and validation require datasets with a relevant subset of data. Without sufficient actual evidence, defining a classifier as "good" is sometimes meaningless. For Network IDS, the classifier must be trained and demonstrated on a large, representative, and accurately labeled dataset which consists both benign and harmful internet traffic [3].

This section, provide an introduction of IoT dataset analysis landscape & importance of employing ML techniques. Three essential elements need to be looked at in order to get research results of the highest caliber. The first one is Datasets' Quality which contributes to ensuring the validity and generalizability of the results. Scientists require data that is accurately categorized, comes from the realistic environment, after that the Quantity(amount) of datasets that guarantees the sufficient data need for analysis, which is crucial when using ML techniques, to train and test approaches and tools & the third one is Data accessibility. This is essential because it starts the research process and guarantees repeatable outcomes that advance the state of the art [4]. We have different methods to collect datasets or it is possible to portray real-world traffic using a variety of different approaches, each of which has a distinct level of precision [3] and these are shown in Fig. 1.

These datasets enable ML algorithms to learn patterns, correlations, and insights from vast and diverse IoT-generated data, enhancing predictive accuracy, anomaly detection, and decision-making in various applications. Moreover, IoT datasets facilitate the development and benchmarking of innovative ML techniques specific to IoT, driving advancements in optimizing resource utilization, improving operational efficiency, and fostering smarter and more sustainable IoT-driven systems.

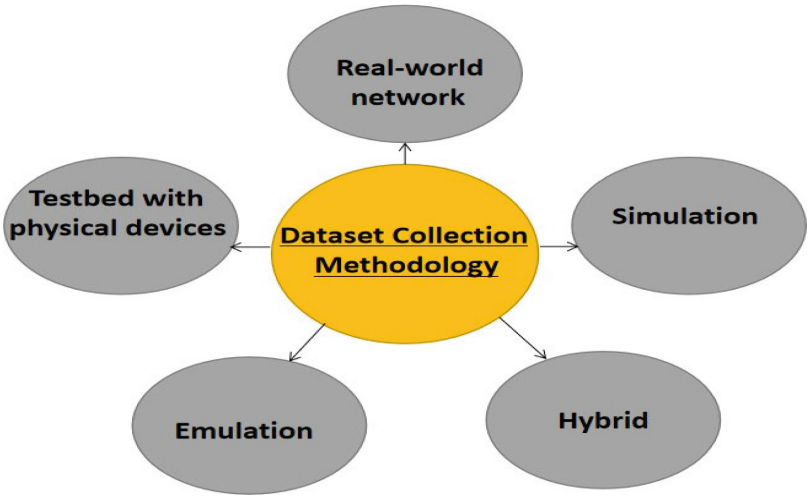


Fig. 1. Techniques employed by the researchers in gathering their dataset.

This paper makes the following contributions:

1. We give a brief summary of the various ML/DL/FL models proposed by various researches which uses different IoT datasets in the section mentioned below.
2. After that, the next section provides a summary of seventeen different IoT datasets upon which the researchers has worked upon. These datasets are related to healthcare, Industrial IoT, network traffic from IoT devices and sensor, etc.
3. Then, we demonstrate the different ML methods implemented on these datasets, with their results & future scopes.
4. And then, analysis of the dataset's attack labels with their count is presented and atlast we concluded the paper.

2 Related Work

Many authors has performed various ML, DL, and federated learning techniques on various IoT datasets, among which some are mentioned here. Marwa Keshk et al. [5] presented that framework which evaluate DL algorithms for IoT IDS implementations. The SPIP framework uses UNSW-NB15, NSL-KDD, TON_IoT datasets & XAI and DL methods like LSTM, SHAP, PFI, ICE, and PDP. Future investigations can create a framework by examining the characteristics of network traffic produced during intrusions, which may indicate possible vulnerabilities.

A clustered FL architecture was developed by Xabier Sáez-de-Cámara et al. [6] that enable unsupervised training of intrusion detection and anomaly models across sizable networks of diverse Internet of Things devices. Since, supervised data labeling is not required for the proposed FL architecture. Datasets N-BaIoT, NSL-KDD, Bot-IoT, merlin-cnc-dos, masscan, and mirai-cnc-dos were analyzed by the authors.

According to Deepak Kumar et al. [7], a DL model, which is probably more effective than a ML model can be used to categorize DDoS attacks on the CICDDoS2019 dataset.

By using a set of metrics taken from packet captures, Olivier Brun et al. [8] are able to forecast the likelihood that a network attack is still in progress. Additionally, they have noted that the outcomes of this approach are similar to those of a basic threshold detector.

An effective deep learning model named MM-WMVEDL is presented by P. Sanju [9] for the effective identification of IDS in an IoT setting on IoT-23, UNSW-NB15, and CICIDS2017. He used RNN and hybrid models. The suggested model use a multimodal architecture in order to efficiently represent the connections among various kinds of data present in network traffic. The accuracy they obtained was 98.12% and 99.98%. Future improvements will include adding network topology and device details to the proposed model. The concept is adaptable to dynamic and changing IoT networks.

Nuno Prazeres et al. [10] used the MQTT-IoT-IDS2020 dataset, to capture IoT traffic data. ExtraTreesClassifier model is used to classify them. Feature engineering included the use of LabelEncoder, SimpleImputer, and the StandardScaler algorithm for normalization (for IoT-23), Z-Score (for MQTT-IoT-IDS2020), and the other tools. In subsequent work, they plan to assess the robustness of their counter measure against adversarial machine learning and broaden the examination of deep models using fresh datasets.

Abderrahmane Ed-daoudy et al. [11] use Kafka streaming, decision tree(C4.5) algorithm, MapReduce, Hadoop, and Apache Spark to analyze diabetes datasets from diverse

diseases achieving results for the binary classification scheme(from 88.13 to 90.85%). Integrating real data sources is the next goal for addressing consumer requests.

HONG QING YU et al. [12] suggested a IoT-Fog-Cloud & dynamic AI-model architecture for remote skin disease diagnosis. They introduced an improved classification approach for skin disease identification. They tested existing ML models in a controlled environment to achieve it. Unlike prior evaluations, their setting uses a well known HAM10000 dataset instead of the customized dataset, obtaining 86.5% accuracy (DenseNet). The future scope can be the testing the two-phase classification procedure as a skin-related transfer model disorders like rash, allergy, and bone.

AROOSA HAMEED et al. [13], Investigated IoT traffic classification. They offer a two-stage composite learning approach to overcome this challenge & used Dataset 1 (network traffic traces from 28 smart devices) & Dataset 2 (Traffic from 81 IoT devices in the US and UK) for this approach. The ML techniques are applied which gives accuracy of 99.9% (first dataset) and 99.8% (second dataset). The future step is to integrate their suggested model with a resource allocation mechanism that may use workload estimation to dynamically adjust allocation strategies at access and Edge networks.

A FUG resource allocation strategy based on LSTM & SVM was proposed by Eslam Eldeeb et al. [14]. They used an SVM system to categorize devices into various priority groups. Afterwards, they created an LSTM architecture with a 98% accuracy rate to forecast the traffic of real-time dataset(MTD). Selecting a proper exploration rate through the application of learning-based methods is an open research challenge for further development.

To reduce data volume and maintain accuracy in real-time spatial data prediction, YUICHI INAGAKI et al. [15] proposes an IoT device control system that capitalizes on data from the Random forest ML model. The dataset used here, contains 25-day location logs of 536 taxies. Additional machine learning models and feature selection methods should be studied in future research.

3 IoT Datasets Overview

Understanding the unique characteristics of IoT datasets is vital to developing effective ML solutions. This section discuss the typical features of IoT data, such as variety basis on their attack types, protocols, and variability, and their implications for data analysis. Effective preprocessing and data cleaning are critical steps in the IoT dataset analysis. We explore various techniques and methodologies for dealing with missing values, noise, outliers, data balancing and data normalization to ensure the quality of the analysis. There are different IoT datasets explained below in Table 1 which are widely used as experimental purpose.

Table 1. A summary of IoT datasets showcases various attack kinds and ML applications in various cyber applications.

Dataset	Description
NSL-KDD	This dataset updates the KDD Cup 1999 dataset [5]. It tests intrusion detection systems (IDS) at discriminating between regular network traffic and DoS, probing, U2R, and R2L assaults. The dataset contains labeled network traffic data for training and testing ML models to detect intrusions and malicious activity
UNSW-NB15	University of New South Wales (UNSW) researchers generated the UNSW-NB15 dataset for security in intrusion detection and networks research. To fix outdated attack characteristics and unrealistic traffic distributions in the NSL-KDD & KDD Cup 1999 datasets, the dataset was created. UNSW-NB15 simulates modern network traffic and attacks. DoS, R2L, U2R, and probing attacks are among its nine varieties. The data includes source and destination IP addresses, ports, protocol types, packet payloads, & connection durations [5, 9, 16]
TON_IoT	Moustafa created the ToN_IoT dataset [35] to validate IIoT systems. Telemetry data from IoT devices is used to detect intrusions that alter these devices. IoT telemetry data was collected in an experiment environment with three layers (Edge, Fog, and Cloud) to simulate real-time data from current production networks [5, 18]
N-BaIoT	N-BaIoT stands for "Network-Based Behavioral Analysis for IoT," and the dataset was created by researchers to address the increasing security concerns in IoT ecosystems. It focuses on capturing network traffic from IoT devices and sensors to study their behavior and detect potential security threats or anomalies [6]
Bot-IoT	Bot-IoT stands for "Botnet of Things," and it was created to address the growing concern of botnet threats targeting IoT devices. Botnets are networks of compromised devices that are managed by malicious actors to carry out distinct tasks, including launching distributed denial of service attacks, spreading malware, and participating in other illegal behaviors [6, 19]
mirai-cnc-dos	Mirai, a notable malware, exploits weak default passwords or known flaws to infect internet-connected cameras, routers, and DVRs. A device hacked joins the Mirai botnet, a malicious network of compromised devices. Data about Mirai botnet CNC (Command and Control) communication during DDoS assaults is in the Mirai CNC DoS dataset. Botnet operators command and coordinate infected IoT devices via CNC servers. DNS queries, HTTP requests, and other botnet communication protocols used to obtain orders and launch attacks may be in the dataset. The Mirai CNC DoS dataset helps researchers and security professionals understand the Mirai botnet, its communication patterns, and how to protect against such attacks [6]
merlin-cnc-dos	"Merlin CNC DoS dataset" captures CNC transmission data from another IoT botnet or malware. The collection may contain network traffic data from the botnet's command and control servers and compromised IoT devices used to launch DDoS attacks [6]
CICDDoS2019	The public CICDDoS2019 dataset studies DDoS assaults in IoT contexts. The Canadian Institute for Cybersecurity's Cybersecurity Research Lab provided the dataset. This dataset captures IoT-targeted DDoS network traffic. Researchers and practitioners in IoT security, network intrusion detection, and DDoS defense will benefit from the dataset [7]

(*continued*)

Table 1. (*continued*)

Dataset	Description
AWID	The AWID (AARL Wireless IDS) dataset is freely available for Intrusion Detection System (IDS) research in wireless communication and IoT security. Advanced Aerospace Research Laboratories at Georgia Tech created the AWID dataset. Its wireless communications data makes it useful for analyzing security vulnerabilities in IoT-adopted Wi-Fi and other wireless networks [17]
IoT-23	IoT-23 is a public dataset for IoT security research. To meet the requirement for IoT-specific security analysis and intrusion detection datasets, University of New Brunswick and University of California, Irvine researchers created it. To research IoT device activity and identify security threats and abnormalities, this dataset collects network traffic data. Traffic from a variety of IoT devices makes the dataset typical of different deployments [9, 10]
CICIDS2017	The CICIDS2017 dataset was created by the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick. It is a comprehensive dataset that encompasses distinct types of network traffic, including both normal traffic and attacks [9, 19]
CSE-CICIDS2018	The CSE-CIC-IDS2018 is an updated version of CSE-CICIDS2017 dataset, was created by the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick. They extracted 75 network data features using their own tool, CICFlowMeterV3. It is a comprehensive dataset that includes a wide range of network traffic, both benign and malicious, to simulate real-world scenarios as well as make possible to develop and assess the intrusion detection systems [18]
MQTT-IoTIDS2020	The MQTT-IoT-IDS2020 dataset uses simulated MQTT network data. Raw pcap files and unidirectional & bidirectional flow features are in the dataset. Data includes MQTT brute-force attack, Sparta SSH brute-force attack, regular operation, UDP scan traffic, etc. It has 32 bidirectional flow features: protocol, addresses, packet length, ports, flags, etc. We integrated bidirecional flow data sets to create an 18,000-flow dataset [10]
KDD Cup 99	The KDD Cup 1999 dataset (KDD99) is a benchmark dataset for NID and ML. This was for the Third International Knowledge Discovery and Data Mining Tools Competition at KDD-99. In 1999, the KDD Cup sought to create models that could discern between normal network traffic and various attacks in network communication data. The most popular anomaly detection data collection has 41 features and categorizes assaults as remote-to-local (R2L), denial of service (DoS), user-to-remote (U2R), and probing [20]
HAM10000	HAM10000 stands for "Human Against Machine with 10000 training images." The dataset was introduced to measure the efficiency of ML algorithms distinguishing between various skin conditions, including benign and malignant skin lesions. It was created to aid in the development of computer-aided diagnosis (CAD) systems for dermatological applications [12]
DS2OS	The DS2OS dataset was developed by recording traces in an IoT surroundings utilizing four simulated IoT sites that provided services. The dataset includes hazardous and benign (normal) traffic. The "DS2OS" dataset catalogs novel IoT assaults from real-world traffic. Common network traffic attacks shape attacks. Network traffic samples are labelled by behavior [21]
Med-295 BIoT	MedBIoT integrates IoT technology with healthcare. Connecting medical devices, wearables, and sensors to the internet allows healthcare practitioners to remotely monitor, analyze, and make decisions. In MedBIoT, datasets are essential for creating and testing IoT applications, machine learning algorithms, and healthcare data analytics. These datasets contain real-world medical device and sensor data [22]

4 Machine Learning Algorithms for IoT Dataset Analysis

This section provides an extensive survey of popular ML algorithms used in IoT dataset analysis. These algorithms can be supervised, unsupervised, or reinforcement learning, and we discuss their suitability for numerous types of IoT data and tasks. Before applying any ML classifier we perform Data preprocessing and Feature engineering which is so important for improving the performance of ML models. We investigate feature extraction and selection techniques tailored for IoT datasets, considering the specific challenges posed by high-dimensional and heterogeneous data sources and this is shown in Fig. 2.

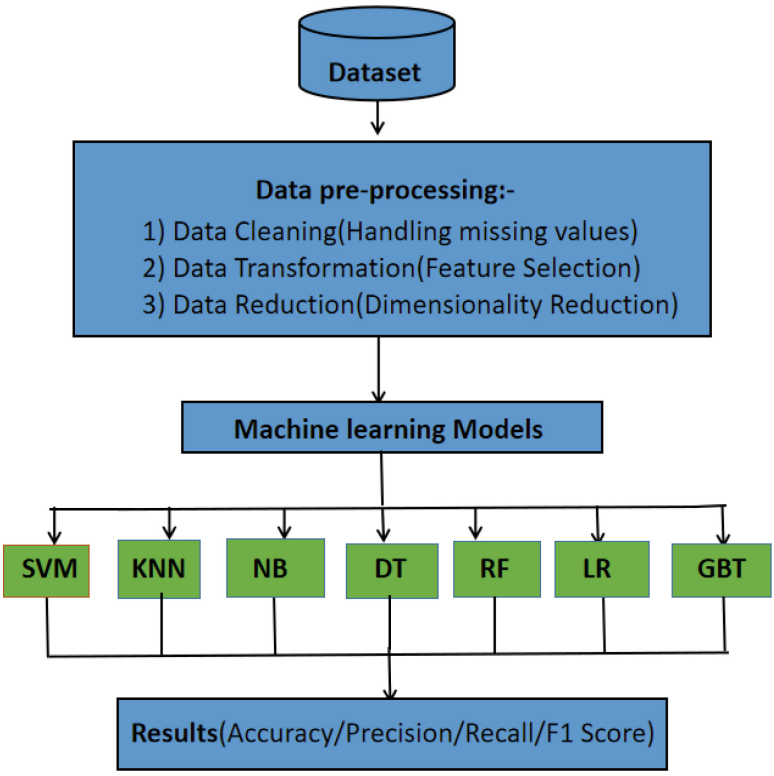


Fig. 2. Steps to be performed while implementing various ML Algorithms on IoT datasets.

Different ML approaches can be applied to different IoT datasets to create effective models [23]. Table 2 summarizes diverse machine learning strategies for building hybrid models for varied objectives. In this paper, we review numerous datasets using ML approaches and discuss pertinent concerns to improve future scope.

Table 2. An overview of the ML algorithms performed on the various IoT datasets.

Ref.	Dataset Used	Data Preprocessing	ML Technique	Results	Future scope
[16]	UNSWNB15	min–max normalization, Principal Component Analysis(PCA)	Extreme gradient boosting, Cat Boost, KNN, SVM, QDA, NB	Accuracy = 99.99%	Experimentally compare BoT-IoT to UNSWNB-15 using a deep learning traffic classification model
[17]	AWID	constant removal, recursive feature elimination & variance threshold feature selection	SVM, Ran dom Forest, Decision Tree	Accuracy = 99%	Future research will compare the suggested feature selection strategy to other well-known methods, examining convergence time, iteration, and detection accuracy
[18]	CSECICIDS2018, UNSWNB15, and ToNIoT	PCA, LDA, and Auto-encoder (AE)	CNN, RNN, Deep Feed Forward (DFF), Lo-gistic Regression (LR), DT and NB	various accuracy of datasets with different ML algorithms	Combining various feature engineering method and ML classifier for optimal performance across diverse datasets and practical applications is challenging and requires further research
[19]	CICIDS2 017 dataset and BoT-IoT	Removed all rows that have the feature "Flow Packets/s" equal to 'Infinity' or 'NaN'& also removed identical rows of the da-Taste	DT and rules-based concepts (JRip algorithm, REP Tree, and Forest PA)	Accuracy = 96.665% & 96.995%	Not provided
[20]	KDD Cup 99	centralization and normalization	C5.0 decision tree and time-series analysis	Accuracy = 96%	Not provided

(*continued*)

Table 2. (*continued*)

Ref.	Dataset Used	Data Preprocessing	ML Technique	Results	Future scope
[21]	DS2OS	Univariate selection, correlation matrix with a heat map	LR, RF, Extreme Gradient Boosting, and Light Gradient Boosting	Accuracy = 99.92%	The model's low time, memory, accuracy, and mistake rate may aid IoT-enabled criminal investigations
[22]	N-BaIoT and MedBIoT	filter model & wrapper model	KNN, DT, RF, extremely randomized trees classifier (ET), LR, SVM, Ada-Boost	Accuracy(XGB oost) = 99.9 6%	Not provided
[24]	UNSWNB15	Undersampling balancing method	Random Forest (RF), SVM & ANN	Accuracy(RF) = 98.67% & 97.37%	To improve the accuracy of harmful traffic patterns, future efforts will focus on collecting features from various IoT protocols
[25]	UNSW-NB15 and ToN_IoT	filter methods, wrapper methods, min-max normalization	Genetic algorithms Back Propagation Neural Networks (BPNN)	upto 37.07% reduction in execution time	Suggested testing the model with multiclass classification on varied datasets in the future
[26]	UNSWNB15	Min–Max scaling, XGBoost	SVM, KNN, LR, ANN & DT	ANN(highest performance) = 86.71%	Plan to test the XGBoost-based feature selection algorithm on the NSLKDD dataset and compare its performance to current methods

(*continued*)

Table 2. (*continued*)

Ref.	Dataset Used	Data Preprocessing	ML Technique	Results	Future scope
[27]	Bot-IoT	Information gain, information gain ratio, and Chi squared	LR, DT, NB, a multi-layer perceptron (MLP), RF, CatBoost, XGBoost, and LightGBM	CatBoost = 99.987%	Evaluate alternative classifiers trained on the same datasets. Additionally, classifier performance for information theft detection can be assessed on different IOT identification datasets
[28]	UNSWNB 15 and TON_IoT	Gini Impuritybased Weighted Random Forest (GIWRF) model and Min–Max scaling	DT, Gradient Boosting Tree (GBT), MLP, AdaBoost, LSTM, and GRU	AdaBoost(for UNSWNB) = 90.51% & GBT(TON_IoT) = 99.98%	Future work could include a temporal complexity-based multiclass IDS categorization scheme
[29]	IoT-23	MinMax Scaling & batch and stream processing	LSTM, Convolutional Neural Network and ANN	Accuracy = 99.93 %	The ensemble model can be improved by adding deep-learning approaches to improve detection ac-curacy
[30]	TON_IoT	LSTM, min-max normalisation	SVM, KNN, NB, RF, Classification and Regression Trees (CART)), LR, and LDA	various accuracy with different ML algorithms	Bayesian and evolutionary methods can optimize model hyperparameters and improve results

5 Datasets Analysis

Despite the numerous advantages of employing machine learning in IoT dataset analysis, several challenges remain. The integration of ML in IoT faces significant challenges pertaining to attack categories within datasets. Adversarial attacks, including poisoning, evasion, and model inversion, present substantial threats to the security and accuracy of ML models. In IoT datasets, attack categories encompass various malicious activities targeting Internet of Things devices. These can be broadly classified into physical attacks,

such as tampering or theft; eavesdropping and man-in-the-middle attacks; software-based attacks like malware & denial of service; and data privacy breaches, where sensitive information is compromised. Understanding and categorizing these attacks are crucial for developing robust security measures and strategies to safeguard IoT ecosystems, ensuring their resilience against evolving threat landscapes. This section will provide the analysis of those attacks in five well known datasets.

Firstly, we have NetFlow-based format dataset called NF-UNSW-NB15 which is obtained by the UNSW-NB15 dataset, and each attack category has been labeled accordingly in it. There are total 1,623,118 data flows, of which 1,550,712 (95.54%) are benign and 72,406 (4.46%) are attack samples [31]. Nine subcategories used to further classify the attack samples shown in Table 3.

After that, we created the NetFlow records for the ToN-IoT dataset [32] using the publicly accessible pcaps, resulting in the creation of the NF-ToN-IoT NetFlow-based IoT network dataset. The distribution of the NF-ToN-IoT dataset is shown in Table 4, with 1,379,274 total data flows, of which attack samples are 1,108,995 (80.4%) and benign are 270,279 (19.6%).

Table 3. The distribution of all attack flows in the NF-UNSW-NB15 dataset is shown in the table below.

Class	Count
Benign	1550712
Fuzzers	19463
Analysis	1995
Backdoor	1782
DoS	5051
Exploits	24736
Generic	5570
Reconnaissance	12291
Shellcode	1365
Worms	153

Nextly, the BoT-IoT dataset was used to create the NF-BoT-IoT dataset, which is an IoT NetFlow-based dataset [33]. The features were taken from pcap data that were made available to the general public, and the flows were labeled with the suitable attack types. There are total 600,100 data flows, of which 13,859 (2.31%) are benign and 586,241 (97.69%) are attack samples. The dataset comprises four distinct attack categories. The NF-BoT-IoT distribution of all flows is depicted in Table 5 below.

The CSE-CIC-IDS2018 dataset's original pcap files were used to create the NF-CSE-CIC-IDS2018 NetFlow-based dataset [31]. The distribution of the dataset is shown in Table 6, with 8,392,401 total flows, of which attack samples are 1,019,203 (12.14%) and 7,373,198 (87.86%) are benign ones.

Table 4. The distribution of attack categories of the NF-ToN-IoT dataset is presented in the table below.

Class	Count
Benign	270279
Backdoor	17247
DoS	17717
DDos	326345
Injection	468539
MITM	1295
Password	156299
Ransomware	142
Scanning	21467
XSS	99944

Lastly, we have CIC-BoT-IoT dataset with eighty-three features extracted from the pcap files of the BoT-IoT dataset [34] using the CICFlowMeter-v4. The total no. of records in the dataset is 13,428,602, out of which attack samples are 13,339,356 (99.34%) and benign samples are 89,246 (0.66%). The attack categories are shown in Table 7.

Table 5. The NF-BoT-IoT's distribution of all attack flows is shown in the table below.

Class	Benign	Reconnaissance	DDoS	DoS	Theft
Count	13859	470655	56844	56833	1909

Table 6. The distribution of all attack flows in NF-CSE-CIC-IDS2018 dataset is shown in the table below.

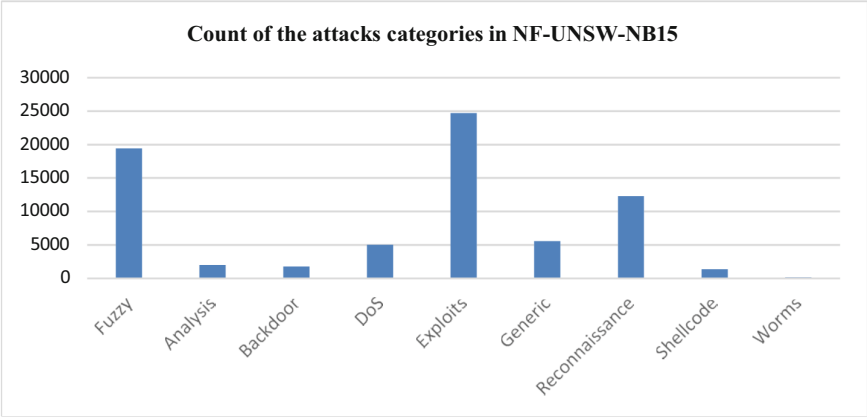
Class	Benign	BruteForce	Bot	DoS	DDoS	Infiltration	Web Attacks
Count	7373198	287597	15683	269361	380096	62072	4394

Table 7. The attack distribution of CIC-BoT-IoT dataset is shown in the table below.

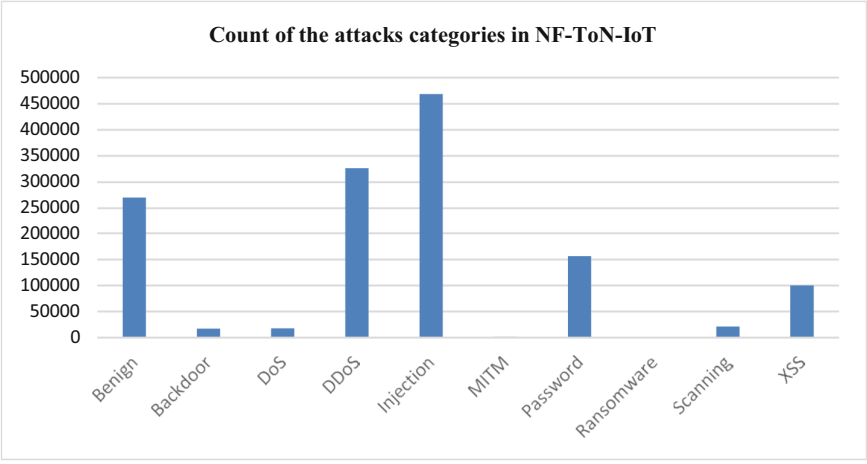
Class	Benign	Reconnaissance	DDoS	DoS	Theft
Count	89246	3514330	4913920	4909405	1701

As, there are numerous attack categories shown in Fig. 3 of the different datasets among which some is discussed as follows. Note that, we have not shown the count of NF- UNSW-NB15's and NF-CSE-CIC-IDS2018's benign in Fig. 3(a) & Fig. 3(d) respectively due to it's large count in both the datasets.

In the NF-UNSW-NB15 dataset [32] "Fuzzers" refer to attacks generated by fuzz testing tools. Fuzzers attacks often involve injecting malformed or unexpected data into network traffic to exploit weaknesses in network protocols, applications, or devices. Then, backdoor attack involves maliciously altering or inserting patterns into the network traffic data to create a hidden vulnerability or a "backdoor" within the dataset.

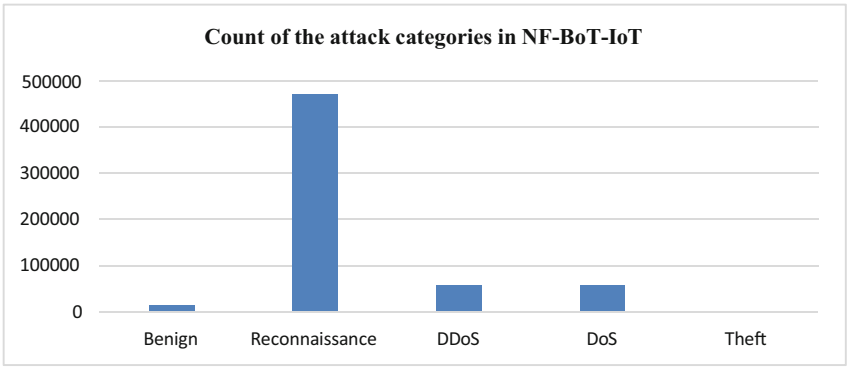


(a)

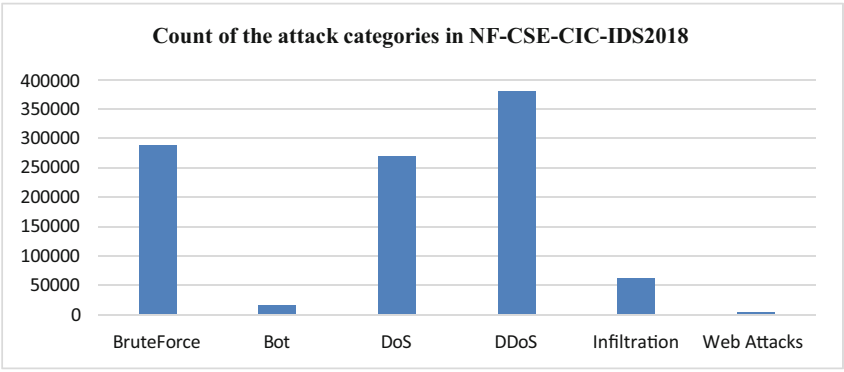


(b)

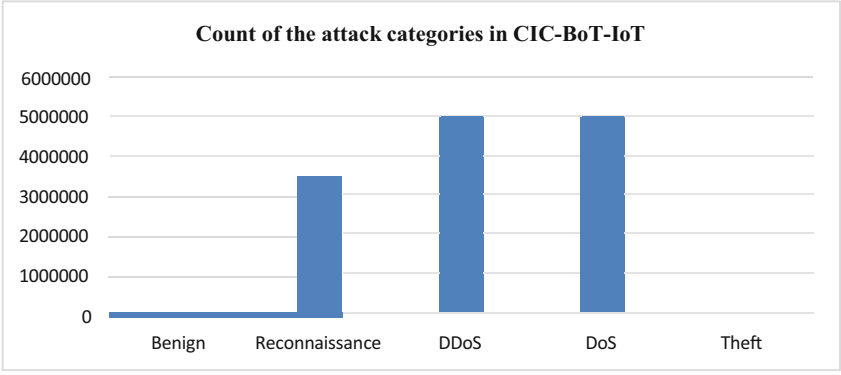
Fig. 3. The distribution of all attack flows of the datasets are shown in the figure. **(a)** Presents NF-UNSW-NB15's distribution (excluding benign). **(b)** Presents NF-ToN-IoT's distribution. **(c)** Presents NF-BoT-IoT's distribution. **(d)** Presents NF-CSE-CIC-IDS2018's distribution (excluding benign). **(e)** Presents CIC-BoT-IoT's distribution.



(c)



(d)



(e)

Fig. 3. (*continued*)

This alteration may occur in a subtle and inconspicuous manner, making it difficult to detect during normal analysis. It's goal is to compromise the effectiveness of IDS or other ML models trained on this data, allowing an attacker to bypass detection or gain

unauthorized access to the IoT system. In most of these datasets, Reconnaissance attack involves probing and information gathering by potential adversaries. It includes actions like network scanning and host enumeration, aiming to identify vulnerabilities and assess the IoT system's structure, critical for planning subsequent targeted attacks.

Worms attack represents malicious software propagating through a network, exploiting vulnerabilities in IoT devices. This attack type involves self-replication, enabling rapid infection and significant disruption to IoT networks and systems. Studying Worms attacks in the dataset is crucial for IoT security assessment and defense development. Then, there are other several attacks such as DDoS floods a system with traffic to disrupt its services. DoS (Denial of Service) similarly disrupts services but is executed from a single source. Theft attacks involve unauthorized access to steal sensitive data or resources, compromising system integrity and security.

6 Conclusion

Finally, this section summarize insights presented in the paper, reiterating the significance of machine learning in IoT dataset analysis and its potential to drive innovations in the IoT ecosystem. The versatility of machine learning enables its application in various IoT domains. In this paper, we explore real-world case studies and applications of machine learning in diverse IoT domains. We get to know that the datasets needs feature engineering approaches as they're highly imbalanced and many other issues resides on them & only after that ML techniques are implemented to get the better performance on classifying the different features or attacks in the dataset. After that, we have shown the labels of the various datasets with their count to get familiar with the attacks which a dataset consists. To conclude the paper, we present potential research directions and emerging trends in IoT dataset analysis with machine learning. We highlight areas for improvement and emphasize the need for interdisciplinary collaboration to address the challenges in this domain.

References

1. Khraisat, A., Alazab, A.: A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity **4**, 1–27 (2021)
2. Boutaba, R., et al.: A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J. Internet Serv. Appl. **9**(1), 1–99 (2018)
3. De Keersmaeker, F., Cao, Y., Ndonda, G.K., Sadre, R.: A Survey of public IoT datasets for network security research. IEEE Commun. Surv. Tutor. (2023)
4. Grajeda, C., Breitinger, F., Baggili, I.: Availability of datasets for digital forensics– and what is missing. Digit. Investig. **22**, S94–S105 (2017)
5. Keshk, M., Koroniotis, N., Pham, N., Moustafa, N., Turnbull, B., Zomaya, A.Y.: An explainable deep learning-enabled intrusion detection framework in IoT networks. Inf. Sci. **639**, 119000 (2023)
6. Sáez-de-Cámara, X., Flores, J.L., Arellano, C., Urbieta, A., Zurutuza, U.: Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks. Comput. Secur. **131**, 103299 (2023)

7. Kumar, D., Pateriya, R.K., Gupta, R.K., Dehalwar, V., Sharma, A.: DDoS Detection using deep learning. Procedia Comput. Sci. **218**, 2420–2429 (2023)
8. Brun, O., Yin, Y., Gelenbe, E., Kadioglu, Y.M., Augusto-Gonzalez, J., Ramos, M.: Deep learning with dense random neural networks for detecting attacks against IoT-connected home environments. In: Security in Computer and Information Sciences: First International ISCIS Security Workshop 2018, Euro-CYBERSEC 2018, London, UK, February 26–27, 2018, Revised Selected Papers 1, pp. 79–89. Springer International Publishing (2018)
9. Sanju, P.: Enhancing intrusion detection in IoT systems: a hybrid metaheuristics- deep learning approach with ensemble of recurrent neural networks. J. Eng. Res. 100122 (2023)
10. Prazeres, N., Costa, R.L.D.C., Santos, L., Rabadão, C.: Engineering the application of machine learning in an IDS based on IoT traffic flow. Intell. Syst. Appl. **17**, 200189 (2023)
11. Ed-daoudy, A., Maalmi, K.: A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. J. Big Data **6**, 1–25 (2019)
12. Yu, H.Q., Reiff-Marganiec, S.: Targeted ensemble machine classification approach for supporting IoT enabled skin disease detection. IEEE Access **9**, 50244–50252 (2021)
13. Hameed, A., Violos, J., Leivadeas, A.: A deep learning approach for IoT traffic multi-classification in a smart-city scenario. IEEE Access **10**, 21193–21210 (2022)
14. Eldeeb, E., Shehab, M., Alves, H.: A learning-based fast uplink grant for massive IoT via support vector machines and long short-term memory. IEEE Internet Things J. **9**(5), 3889–3898 (2021)
15. Inagaki, Y., Shinkuma, R., Sato, T., Oki, E.: Prioritization of mobile IoT data transmission based on data importance extracted from machine learning model. IEEE Access **7**, 93611–93620 (2019)
16. Saheed, Y.K., Abiodun, A.I., Misra, S., Holone, M.K., Colomo-Palacios, R.: A machine learning-based intrusion detection for detecting internet of things network attacks. Alex. Eng. J. **61**(12), 9395–9409 (2022)
17. Gaber, T., El-Ghamry, A., Hassanien, A.E.: Injection attack detection using machine learning for smart IoT applications. Phys. Commun. **52**, 101685 (2022)
18. Sarhan, M., Layeghy, S., Moustafa, N., Gallagher, M., Portmann, M.: Feature extraction for machine learning-based intrusion detection in IoT networks. Digital Commun. Netw. (2022)
19. Ferrag, M.A., Maglaras, L., Ahmim, A., Derdour, M., Janicke, H.: Rdtids: rules and decision tree-based intrusion detection system for internet-of-things networks. Fut. Internet **12**(3), 44 (2020)
20. Zhu, B., et al.: IoT equipment monitoring system based on C5. 0 decision tree and time-series analysis. IEEE Access **10**, 36637–36648 (2021)
21. Rani, D., Gill, N.S., Gulia, P., Arena, F., Pau, G.: Design of an intrusion detection model for IoT-enabled smart home. IEEE Access (2023)
22. Kalakoti, R., Nõmm, S., Bahsi, H.: In-depth feature selection for the statistical machine learning-based botnet detection in iot networks. IEEE Access **10**, 94518–94535 (2022)
23. Sarker, I.H., Kayes, A.S.M., Badsha, S., Alqahtani, H., Watters, P., Ng, A.: Cybersecurity data science: an overview from machine learning perspective. J. Big data **7**, 1–29 (2020)
24. Ahmad, M., Riaz, Q., Zeeshan, M., Tahir, H., Haider, S.A., Khan, M.S.: Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set. EURASIP J. Wirel. Commun. Netw. **2021**(1), 1–23 (2021)
25. Mohamed, D., Ismael, O.: Enhancement of an IoT hybrid intrusion detection system based on fog-to-cloud computing. J. Cloud Comput. **12**(1), 1–13 (2023)
26. Kasongo, S.M., Sun, Y.: Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. J. Big Data **7**, 1–20 (2020)
27. Leevy, J.L., Hancock, J., Khoshgoftaar, T.M., Peterson, J.M.: IoT information theft prediction using ensemble feature selection. J. Big Data **9**(1), 1–48 (2022)

28. Disha, R.A., Waheed, S.: Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity **5**(1), 1 (2022)
29. Alghamdi, R., Bellaiche, M.: An ensemble deep learning based IDS for IoT using Lambda architecture. Cybersecurity **6**(1), 5 (2023)
30. Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., Anwar, A.: TON_IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven intrusion detection systems. Ieee Access **8**, 165130–165150 (2020)
31. Sarhan, M., Layeghy, S., Moustafa, N., Portmann, M.: Netflow datasets for machine learning-based network intrusion detection systems. In: Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings 10, pp. 117–135. Springer International Publishing (2021)
32. Sarhan, M., Layeghy, S., Portmann, M.: Feature analysis for machine learning- based IoT intrusion detection. arXiv preprint arXiv:2108.12732 (2021)
33. Jayalaxmi, P.L.S., Kumar, G., Saha, R., Conti, M., Kim, T.H., Thomas, R.: DeBot: a deep learning-based model for bot detection in industrial internet-of- things. Comput. Electr. Eng. **102**, 108214 (2022)
34. Rubab, J., Afzal, H., Shahid, W.B.: A survey of network features for machine learning algorithms to detect network attacks. In: Asian Conference on Intelligent Information and Database Systems, pp. 77–88. Springer Nature Switzerland, Cham (2022, November)
35. Moustafa, N.: A new distributed architecture for evaluating AI-based security systems at the edge: network TON_IoT datasets. Sustain. Cities Soc. **72**, 102994 (2021)

Applications of Machine Learning and Deep Learning

Investigation of Satellite Data for Monitoring Air Quality Over Remote Sensing Technology

Panimalar Kathiroli[1](✉), V. Vijayalakshmi[1], Piyush Gopal[1], and Sivananda Panda[2]

[1] Department of Data Science and Business Systems, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu District, Chennai, Tamilnadu 603203, India
{panimalk,vijayalv3,pg5838}@srmist.edu.in

[2] Pricewaterhouse Coopers Service, Bengaluru, Karnataka 560071, India

Abstract. This paper presents a comprehensive Investigation of Satellite data for Monitoring Air quality through Remote sensing Technology tool (ISMART tool). Specifically, it utilizes data from the Sentinel-5 precursor (Sentinel-5p) Tropospheric Monitoring Instrument (TROPOMI) datasets to monitor the columnar surface of co gas across India, with a focus on high-pollution regions. After rigorous data processing, essential attributes were extracted to monitor the columnar CO level. Haversine formula is employed for specific area monitoring, Basemap is used to visualize the pollution levels and machine learning models (ARIMA, LSTM) for precise monitoring and forecasting. This research culminated in a user-friendly website, built with the Django framework, where users can easily identify pollution hotspots by inputting specific parameters. Ultimately, this paper underscores the cost-effectiveness and vast coverage advantages of remote sensing over traditional ground-level instruments, urging governments to harness this technology for timely interventions against pollution.

Keywords: Tropomi · Sentinel-5p · Air pollution · Long Short-Term Memory · Auto Regressive Integrated Moving Average · Carbon Monoxide

1 Introduction

Clean air is fundamental to the sustenance of all terrestrial life forms. Despite its importance, air quality, particularly in densely populated regions like India, has deteriorated significantly. As the world's second-most populous nation, India grapples with the gravity of air pollution, as evidenced by 21 of its cities being among the world's 30 most polluted. Alarmingly, over 140 million people in India are subjected to subpar air quality daily, resulting in approximately 2 million fatalities annually. Notably, atmospheric Carbon Monoxide (CO) [1] has been identified as a significant pollutant. Produced predominantly by industrial activities, CO undergoes oxidation to form Carbon Dioxide (CO_2), intensifying global warming by trapping solar heat.

Traditionally, air pollution monitoring has heavily relied on sensors and manual data collection.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 351–366, 2025.
https://doi.org/10.1007/978-3-031-75167-7_28

However, these methodologies have intrinsic limitations like limited spatial coverage, recurrent maintenance needs, declining accuracy over time, inability to trace pollution sources, insufficient tools for large-scale future predictions. Moreover, while air quality monitoring networks are predominantly urban-centric, the data they provide is spatially limited. Such punctual data fails to offer a holistic view, underscoring a significant research gap. Harnessing advancements in technology, satellites have emerged as potent tools in gauging atmospheric pollution levels. Space agencies globally have recognized this potential, deploying satellites to monitor harmful gases. Considering escalating air pollution and its adverse impacts on health and the environment, air quality monitoring has gained pivotal importance. The evolution of satellite remote sensing technology [2] enables comprehensive, high- resolution monitoring of air quality across vast areas. The main contributions of this research are detailed as:

- The paper introduces the ISMART tool, a pioneering approach to monitor air quality using remote sensing technology. This tool leverages data from the Sentinel-5P TROPOMI datasets to observe and analyze the columnar surface of CO gas, especially in high-pollution regions of India.
- The research employs cutting-edge techniques, such as the Haversine formula, for precise area monitoring. It also integrates advanced predictive models like ARIMA and LSTM for forecasting, showcasing the potential of machine learning in environmental studies.
- The paper offers a robust data visualization approach using tools like Basemap, facilitating an intuitive understanding of pollution levels across India. This aids both the public and policymakers in identifying pollution hotspots and understanding trends.
- While emphasizing the advantages of satellite-driven insights, the research underscores the potential benefits of integrating these insights with traditional ground-level instruments. This collaborative approach can lead to even more accurate and granular air quality assessments.

The remainder of the paper is organized as Sect. 2 offers an in-depth literature survey about the European Space Agency's Sentinel 5P satellite and its onboard TROPOMI instrument. The capabilities of this satellite, including its spectral resolution, operational mode, and the various atmospheric gases it can monitor, are discussed in detail. Section 3 elaborates various methodologies used in the research, the importance of monitoring specific gases like CO, NO_2, and the process of extracting and analyzing data from the Sentinel 5P TROPOMI datasets for specific regions. Section 4 gives a comprehensive overview of the ISMART tool, explaining how users can interact with it to get insights into air quality based on specific parameters. The advantages of using satellite-driven insights combined with machine learning models are highlighted. Section 5 presents the findings of the research, the insights derived from the various datasets, and the performance of the predictive models. The results from ARIMA and LSTM models, as well as other comparative models, are discussed in detail. Section 6 culminates with a conclusion that encapsulates the primary findings and their implications. Additionally, potential future directions for this research are proposed.

The results from ARIMA and LSTM models, as well as other comparative models, are discussed in detail. Section 8 culminates with a conclusion that encapsulates the

primary findings and their implications. Additionally, potential future directions for this research are proposed.

2 Literature Survey

The study [3] delves into the impacts of the Ukraine–Russia war on anthropogenic activities in Kiev which employs Sentinel-5P imagery coupled with an artificial intelligence model to predict air pollution levels in the metropolis. Leveraging a multilayer perceptron neural network (MLPNN) optimized with the Electromagnetic Field Optimization (EFO) algorithm, the research predicts particulate matter 2.5 (PM2.5) concentrations. Notably, the hybrid EFO-MLPNN model, when compared to conventional models, showcases superior performance, thus emphasizing the potential of hybrid models in air quality prediction. Also [4] addresses the challenges posed by military operations in Ukraine, particularly in terms of monitoring air quality in conflict zones. Using satellite data, the research investigates changes in the concentration levels of crucial pollutants like sulfur dioxide, nitrogen dioxide, and carbon monoxide in both Poland and Ukraine from 2018–2022. The study highlights the indispensability of geospatial data during emergencies and provides vital insights into the ramifications of war on environmental quality. The role of greenhouse gases (GHGs) are focussed in [5] influencing local air pollution and global climate. The study uses a ground based EM27/SUN Fourier transform infrared spectrometer (FTIR) to retrieve data on Carbon Dioxide (CO_2), Methane (CH_4), and Carbon Monoxide (CO). The research underscores the importance of continuous monitoring of GHGs and validates the accuracy of Sentinel-5P TROPOMI satellite data against ground-based measurements. The challenges posed by regional ozone pollution are highlighted in [6] emphasizing the insufficiency of ground-level data due to the exorbitant costs associated with establishing and maintaining monitoring sites. Addressing the shortcomings of traditional models, especially in regions with complex terrains like Chongqing, the study proposes an optimization approach integrating ResNet (Residual Network). This novel approach successfully addresses the "nonstill image" problem associated with ozone distribution and provides a robust and cost-effective solution for business operation. Study in [7] assesses the potential of Sentinel-5 air pollution products using the Google Earth Engine (GEE) platform[8] to monitor pollutants in Arak city, Iran. By leveraging Sentinel-5 data and automated approaches like GEE, the study demonstrates superior performance compared to traditional monitoring methods, offering accurate spatially distributed data.

The pressing issue of air pollution resulting from rapid urbanization, transportation, and industrial activities has been addressed in [9]. Nitrogen dioxide (NO2), primarily released from the combustion of fossil fuels in vehicular emissions, is highlighted as a major pollutant. The study focuses on harnessing the capabilities of the Sentinel-5 Precursor, a dedicated satellite mission for monitoring air pollution, equipped with the Tropospheric Monitoring Instrument.

By analyzing the mean NO2 values retrieved from this satellite from July 2018 to January 2019 over Turkey and juxtaposing it with the 2017 population statistics, Kaplan and team discern a significant correlation (greater than 0.72) between population density and the peak NO2 values. This study not only underscores the direct relationship

between human settlements and air pollution but also paves the way for future research on correlations between various pollutants, population dynamics, and other environmental factors. Also [10] accentuates the criticality of air pollution as a global environmental and health challenge. The study emphasizes the difficulties in obtaining detailed spatial-temporal data on pollutants due to the sparse distribution of ground-level monitoring stations. To counter these limitations, the author introduces three groundbreaking techniques: A localized gradient boosting approach that simulates conventional land-use regression, An OpenStreetMap (OSM) integrated gradient boosting technique, which, despite a slight reduction in performance, enables global-scale NO2 estimations and a deep learning model utilizing remote sensing data that combines high accuracy with global coverage, focusing on estimating nitrogen dioxide (NO2) levels. By leveraging these methods, the study achieves impressive results, with the deep learning model. The presented techniques underscore the potential of using advanced computational methods combined with remote sensing data to estimate high-resolution, surface-level NO2 pollution globally.

The existing literature offers valuable insights into the utilization of remote sensing technology, particularly Sentinel-5P TROPOMI [5], for air quality monitoring and assessment. However, several research gaps can be identified. Firstly, many studies concentrate on specific regions, leaving a gap in understanding air quality dynamics globally. Secondly, while NO2 [11] monitoring is crucial, there's a need to broaden the scope to encompass a wider range of pollutants for a more holistic assessment. Thirdly, literature presents historical data; hence, research focusing on real-time monitoring is needed. Furthermore, advanced predictive models for air quality forecasting should be explored to enhance preparedness. Moreover, studies often emphasize technical aspects without bridging the gap to practical implications or policy recommendations. Additionally, simplifying the fusion of remote sensing data with other sources and exploring standardized protocols for data sharing and analysis could enhance global collaboration and consistency in air quality monitoring. Addressing these gaps will lead to a more comprehensive and accessible approach to air quality assessment with broader coverage, enhanced data integration, and increased relevance for policy and decision-making. The surveyed literature accentuates the transformative potential of advanced computational methods and satellite data in revolutionizing air quality monitoring. The amalgamation of satellite imagery, deep learning techniques, and data assimilation approaches promises enhanced accuracy, making these tools indispensable for researchers, and environmentalists.

3 Methodology

3.1 Data Collection

The European Space Agency (ESA) launched the Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel 5P satellite in October 2017, with its primary mission being the capture of high-resolution observations to monitor air quality and air pollution. Sentinel 5P is the latest addition to the Copernicus Earth Observation Satellite (EOS) sentinel series, launched by ESA in 2015, and designed to operate for seven years. This satellite follows a sun-synchronous orbit at an altitude of 824 kilometers (Km) with an inclination of 98.742°. It boasts a spectral resolution ranging from 0.25 to 0.55 nm

and operates in nadir observation mode, providing global daily coverage with a ground resolution of 7 × 7 square Km. It boasts a swath width of 2600 km, encompassing several bands, ensuring comprehensive data capture. The radiometric accuracy of the satellite is approximately 2%, and it can capture observations of the entire As in Fig. 1, earth's surface within 24 h, offering a significant opportunity for global daily monitoring. TROPOMI, the onboard instrument, is capable of monitoring various gases in the atmosphere, cloud distribution, and aerosols to detect the UV index. This multispectral imaging instrument, located on the Sentinel 5P satellite, detects sunlight reflected from the Earth's surface and employs algorithms to analyze the electromagnetic spectrum in which the reflected radiation falls. It specifically detects gases like methane (CH4), nitrogen dioxide (NO2), ozone (O3), sulfur dioxide (SO2) [12], and carbon monoxide (CO) using visible (VIS), near-infrared (NIR), and shortwave infrared bands (SWIR). The data collected by the imager provides measurements of the total column density of these gases, effectively measuring the entire depth of the atmosphere.

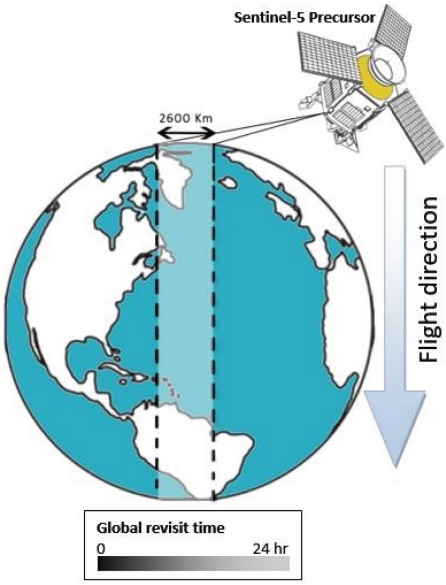


Fig. 1. Sentinel 5P monitoring earth's surface.

3.2 Data Extraction

This paper aims to scrutinize recent progress in this field, with a specific focus on research utilizing the Sentinel-5P TROPOMI satellite. Hence the foundation of this research rests on data sourced from the "Sentinel 5P" satellite, specifically from the TROPOMI, an integral component of Sentinel 5P. While a myriad of instruments is available, the choice of TROPOMI was driven by its superior spatial resolution, which ranges between 7.2 and 5.6. The data, updated daily, is available for download in the

Network Common Data Form (NetCDF) format. To comprehend the critical attributes of these datasets, Panoply software is used. As in Fig. 2, the subsequent step involved scripting in Python to transform these datasets into a more accessible .CSV format, facilitating its integration into the predictive model. A significant challenge encountered during this phase was the conversion of the extensive NetCDF files into the .CSV format, coupled with the need to filter data based on coordinates. The system is designed to accept user input for a specific region. Leveraging third-party libraries, establishment of precise coordinates of the input region is done, filtering the requisite data accordingly. Once the coordinates are pinpointed, the system provides insights into the air pollution levels for that region, offering metrics like average pollution levels, median, and standard deviation. In instances where exact coordinates are not available, the system defaults to data from the nearest available coordinates, determined by a grid system spanning three to five levels. This proximity-based selection employs the Haversine Formula to ensure accuracy. It's worth noting that certain values, represented as QA-value or NaN, indicate either a zero-concentration level or instances where recordings were interrupted, potentially due to atmospheric conditions like cloud cover or thunderstorms. With these data preprocessing steps complete, the machine learning model stands primed for deployment, offering predictive insights into air quality.

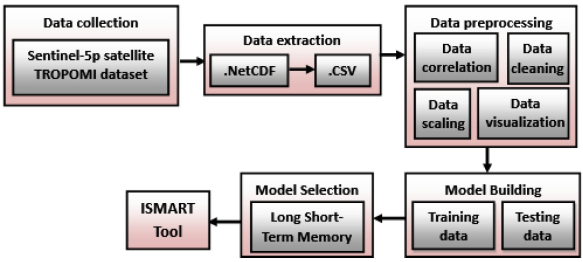


Fig. 2. Architecture of the ISMART tool

3.3 Haversine Formula

The Haversine Formula is a dependable method for determining the shortest route, often referred to as the "great circle distance," between two specific points on a sphere using their longitude and latitude coordinates. This calculation is especially relevant for locations on our spherical Earth.

Using the formula ensures a high degree of accuracy, which is particularly crucial when correlating geographical data, such as pinpointing pollutant sources. In vast regions, even minute location inaccuracies can produce significant variances in data readings. This formula is rooted in the recognition of Earth's curved shape, closely approximated as a sphere. For effective application, longitude (lambda (λ)), and latitude (phi (φ)) need to be converted to radians, followed by specific computational steps to derive the distance.

3.4 Dataset Consideration

The assessment of air quality, especially the concentration of gases like Carbon Monoxide (CO), Nitrogen Dioxide (NO2) and Sulfur Dioxide (SO2) are of paramount importance given its implications for human health and environmental sustainability. Where high level of NO2 can exacerbate respiratory diseases and decrease lung function. Also Elevated levels of SO2 can lead to the formation of acid rain. Thus, monitoring their concentration is essential to address potential environmental and public health concerns. For a more targeted geographical analysis, a specific location (coordinate of Delhi) was selected with a latitude of 28.7041°N and a longitude of 77.1025°E. Satellite data, with its expansive coverage and detailed granularity, offers a sophisticated tool for such assessments.

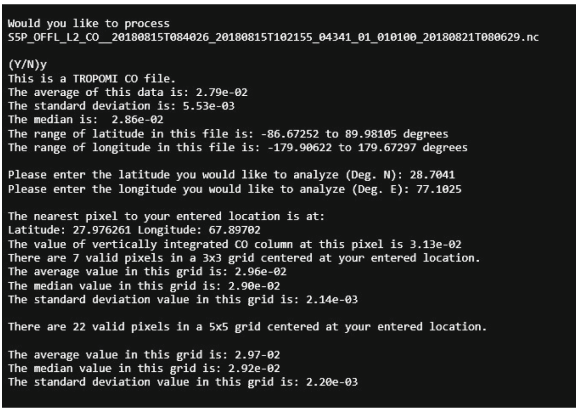


Fig. 3. CO concentration data from TROPOMI file

Figure 3 shows the data from the "Sentinel 5P" satellite, specifically the TROPOMI (Tropospheric Monitoring Instrument) instrument, has been leveraged to analyze CO concentrations in a specific geographical region. Some key statistical insights from the dataset include average CO concentration 2.79e–02, standard deviation is 5.53e–03, the median CO concentration is 2.86e–02. Furthermore, the geographic scope of this dataset encompasses a latitude range from –86.67252 to 89.98105° and a longitude range from –179.90622 to 179.67297°, providing a comprehensive view of CO levels across vast expanses.

The system then identified the nearest data pixel to this location, which was found at a latitude of 27.976261° and a longitude of 67.89702°. The CO concentration at this precise point was determined to be 3.13e–02. The 3 × 3 grid contains 7 valid data pixels. The statistical breakdown of average CO concentration is 2.96e–02, median is 2.90e–02, and standard deviation is 2.14e–03. A more expansive 5 × 5 grid yielded 22 valid data pixels. The statistical insights include average CO concentration of 2.97–02, median of 2.92e–02, and standard deviation of 2.20e–03.

Figure 4 shows the analyses of NO2 concentrations over a specific geographical region. Preliminary statistical analysis of the entire dataset revealed the average NO2 concentration of 6.02e–06, standard deviation of 1.70e–05, and median NO2 concentration of 4.79e–06. The latitude ranges from –86.353195 to 89.97063° and a longitude range from –179.97658 to 179.98184°. The data pixel nearest to the given location was determined to be at a latitude of 28.72764° and a longitude of 77.08238°. The NO2 concentration at this exact coordinate was 4.57e–05. The 3 × 3 grid, centered on the chosen location, had 9 valid data pixels with average NO2 concentration of 5.26e–05, median of 5.39e–05, standard deviation of 4.92e–06. The 5 × 5 grid yielded 25 valid data pixels which encompass an average NO2 concentration of 4.97e–05, median of 4.74e–05, standard deviation of 9.10e–06.

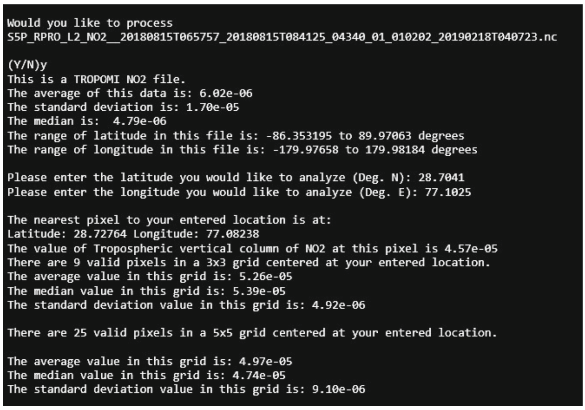


Fig. 4. NO2 concentration data from TROPOMI file

Figure 5 illustrates the SO2 concentrations over specific geographical coordinates. The preliminary statistical assessment of the dataset reveals an average SO2 concentration of 6.26e–07, standard deviation of 3.20e–03, and median SO2 concentration of 1.03e–04. The latitude coverage ranges from –86.40759 to 89.96702° and a longitude range from –179.99757 to 179.9955°. The nearest data pixel to the given location was determined to be at latitude 28.049658° and longitude 68.173775°. The SO2 concentration at this exact coordinate was 2.65e–04, derived from the total slant column under polluted scenarios. The statistical breakdown for 3 × 3 grid of 9 valid data pixels has an average SO2 concentration of 1.77e–04, median of 2.08e–04, standard deviation of 6.08e–04.

For the 5 × 5 grid, there are 25 valid data pixels with an average SO2 concentration of 2.68e–04, median: 2.65e–04, standard deviation of 7.14e–04. The use of TROPOMI satellite data provides a nuanced understanding of CO, NO2, SO2 concentrations, both at a macro and micro level. The detailed insights, ranging from broad overview to localized grid analyses, underscore the potential of satellite data in monitoring atmospheric gases, public health recommendations, and research directions.

```
Would you like to process
S5P_RPR_L2_SO2__20180815T083927_20180815T102224_04341_01_010105_20190218T112347.nc

(Y/N)y
This is a TROPOMI SO2 file.
The average of this data is: 6.26e-07
The standard deviation is: 3.20e-03
The median is:  1.03e-04
The range of latitude in this file is: -86.40759 to 89.96702 degrees
The range of longitude in this file is: -179.99757 to 179.9955 degrees

Please enter the latitude you would like to analyze (Deg. N): 28.7041
Please enter the longitude you would like to analyze (Deg. E): 77.1025

The nearest pixel to your entered location is at:
Latitude: 28.049658 Longitude: 68.173775
The value of total vertical column of SO2 for the polluted scenario derived from the total slant
column at this pixel is 2.65e-04
There are 9 valid pixels in a 3x3 grid centered at your entered location.
The average value in this grid is: 1.77e-04
The median value in this grid is: 2.08e-04
The standard deviation value in this grid is: 6.08e-04

There are 25 valid pixels in a 5x5 grid centered at your entered location.

The average value in this grid is: 2.68e-04
The median value in this grid is: 2.65e-04
The standard deviation value in this grid is: 7.14e-04
```

Fig. 5. SO2 concentration data from TROPOMI file

3.5 Data Visualization

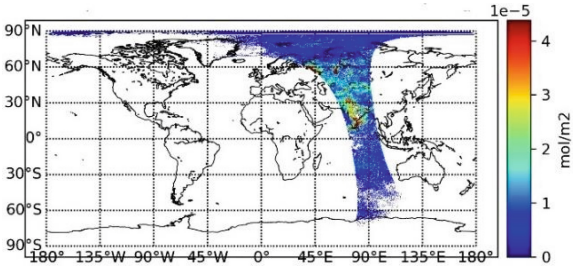


Fig. 6. Visualization of NO2 concentration data

A distinct feature of the visualization is the color bar, which serves as a guide for users to understand the severity of air pollution. As in Figs. 6 and 7, Different shades on this color bar represent varying levels of pollution, and the map is color-coded accordingly. This ensures that users can easily distinguish regions with high pollution levels from those with relatively cleaner air.

Primary visualization tools are used for mapping techniques driven by Python's libraries: Basemap and Folium. These libraries facilitate the creation of detailed and interactive maps. Furthermore, to ensure that the air pollution data is accurately represented through colors on the map, 'pcolormesh' function is used. This function takes in the data and matches it to the appropriate shade on the color bar, ensuring that the visual representation is both accurate and intuitive.

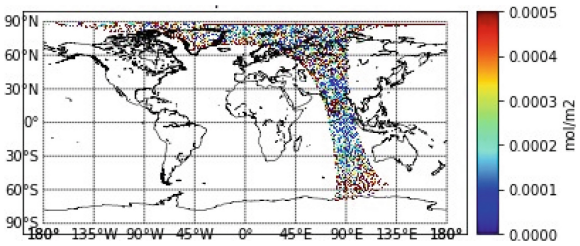


Fig. 7. Visualization of SO2 concentration data

4 Investigation of Satellite Data for Monitoring Air Quality Over Remote Sensing Technology (ISMART) Tool

In this paper, air pollution data is meticulously mapped onto an expansive map of India. When users interact with ISMART tool, the webpage will request to input a specific date for which they wish to view air quality data as in Fig. 8. Upon receiving this date, server processes the request and generates a comprehensive heat map of India. This map vividly illustrates the varying levels of air pollution across different regions. The ISMART tool initiative leverages these satellite-driven insights, combined with sophisticated Machine Learning and Deep Learning models, to offer a comprehensive solution. This approach enables meticulous monitoring, evaluation, and prediction of gas concentrations, notably CO. All results are cohesively presented on a user-friendly, Django-powered platform, fostering public awareness, and aiding governmental intervention strategies as in Fig. 9. Furthermore, the potential of satellite remote sensing has grown substantially over the past decade. Traditional air quality stations, while accurate, lack the spatial distribution that satellite measurements can offer. Given the variable nature of air quality, a synoptic view provided by satellites is invaluable. By bridging this research gap, ISMART tool presents an enhanced, spatially distributed perspective on air pollution, equipping both public and governmental bodies with actionable insights to combat this pressing issue effectively. Hence satellite-based remote sensing technology has undeniably revolutionized the landscape of air quality monitoring. The ISMART tool marks a significant advancement, bridging the gap between technology and practical action. This evolution is poised to drive even greater progress in air quality management, ultimately leading to healthier living environments for all. The journey from the initial use of satellite data for air quality monitoring to the development of ISMART tool has been transformative. They harness the strengths of these studies and offer user- friendly interfaces, making the technology accessible to a wider audience.

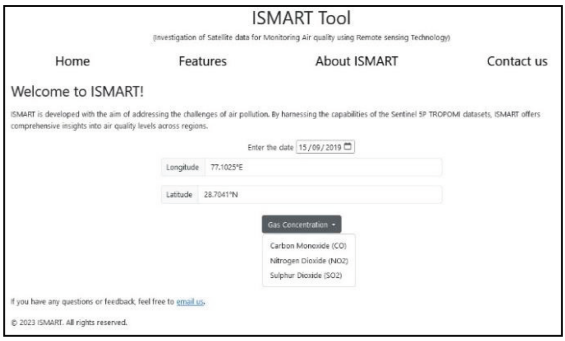


Fig. 8. Homepage of ISMART tool

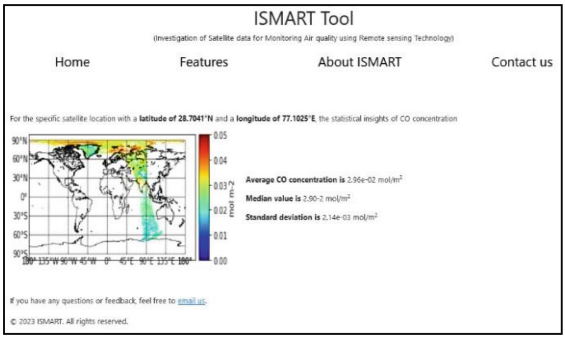


Fig. 9. Visualization of CO concentration data in ISMART tool

5 Results and Discussion

Air pollution concentrations can exhibit temporal fluctuations. For instance, pollutant levels might spike during specific times but tend to decrease during the rainy season, suggesting a seasonal pattern in the data. The data at hand consists of daily recordings, making it imperative to analyse these records over an extended period to ascertain any seasonal trends. To achieve this, time series analysis is employed. Table 1 shows the RMSE values obtained using different machine learning and deep learning models.

Table 1. Comparison of RMSE

Algorithm	Decision tree	Random forest	ARIMA	LSTM
RMSE	1.57	0.027	0.043	**0.006**

For comparison purposes, the decision tree regression, a supervised learning algorithm, has been utilized. to examine object features and train a model structured as a tree to forecast future data. In the provided Fig. 10, illustrates the CO concentration with respect

to date while many actual values align with their predicted counterparts, this model's inherent sensitivity to minor data changes leading to significant structural shifts, thereby introducing instability. Furthermore, its ability to predict continuous values remains questionable.

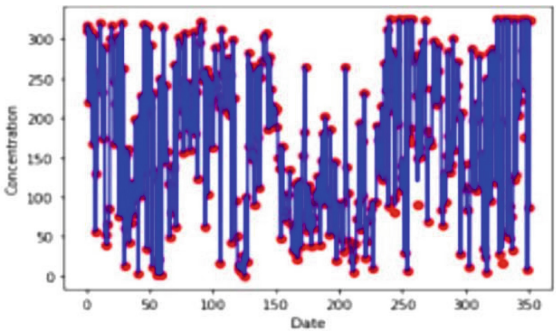


Fig. 10. CO concentration with respect to date using decision tree regression.

Also, random forest regression[13] is used for comparison, which amalgamates multiple decision trees to produce a final output, rather than relying on individual trees. In the Fig. 11, concentration of CO mismatch between many actual and predicted values, which underscores the inadequacy of the method in analysing the dataset's characteristics.

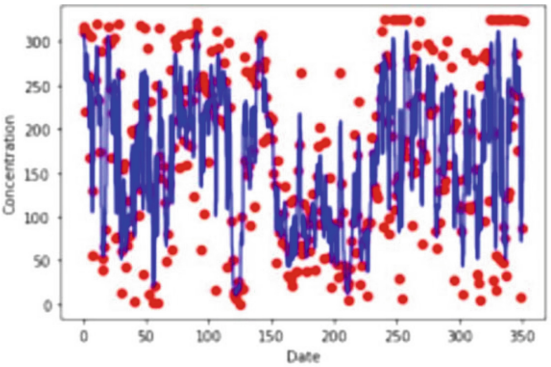


Fig. 11. CO concentration with respect to date using random forest regression.

Auto Regressive Integrated Moving Average (ARIMA) model has also been explored for comparison. To validate the seasonal trend in the data, the ARIMA model is employed. However, before its application, the stationarity of the data is verified using Rolling Statistics and the Augmented Dickey-Fuller (ADCF) Test. A stationary time series exhibits consistent behaviours like constant mean and variance over time, and its auto-covariance shouldn't be time-dependent. ARIMA encompasses Auto Regression (p), Moving Average (q), and Degree of Integration (d). The AR aspect implies that the variable in question is regressed against its previous values. Moving average suggests that the regression

error is a linear combination of error values observed contemporaneously and at various lagged instances. Integration refers to the replacement of data values with the difference between them and their preceding values. Each of these components aims to fit the data as accurately as possible. Given the observed seasonal trend in the data, neural networking is leveraged to predict future pollutant concentrations based on past data (Fig. 12).

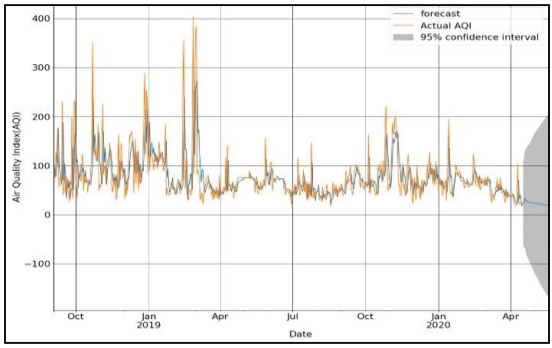


Fig. 12. CO concentration with respect to date using ARIMA model.

Mathematically, a neural network can be described as a differential function mapping one type of variable to another. In classification problems, it transforms vectors into other vectors, while in regression scenarios, it can convert vectors into scalars. The ARIMA Model, as depicted in Fig. 2(a), illustrates the relationship between actual data and its predicted values. This representation suggests that the gas concentration will predominantly fluctuate within the grey region of the graph. However, any abrupt decrease in concentration might result in values that deviate from the anticipated predictions.

Long Short-Term Memory (LSTM) is a widely used architecture within the realm of gated Recurrent Neural Networks (RNN) and has been employed for comparison in this context. Consider a vanilla RNN, here every hidden unit is substituted with LSTM cells. These cells introduce an additional connection, termed the "cell state." LSTM was innovatively crafted to address the challenges of vanishing and exploding gradients that plagued earlier network models.

Distinctively, each LSTM cell not only maintains a hidden vector but also a cell state vector. At each sequential step, the succeeding LSTM has the discretion to read from this state, write to it, or reset the cell, all governed by a specific gating mechanism. Every unit within the LSTM is equipped with three gates of uniform shape:

- Input Gate: Determines if the memory cell should undergo an update.
- Forget Gate: Decides whether the memory cell should be reset to zero.
- Output Gate: Governs the visibility of the current state's information.

The sigmoid activation function characterizes all these gates, primarily because of its smooth curve ranging between 0 to 1, ensuring the model's differentiability. In addition to these gates, there's a vector, denoted as C bar, which influences the cell state. This vector uses the 'tanh' activation function due to its zero-centered range, which facilitates efficient gradient distribution. This characteristic prevents the gradients from vanishing

or exploding too quickly. Each gate processes both the hidden state and the current input x. After concatenating these vectors, a sigmoid function is applied. The C bar vector presents new potential values that could modify the cell states. Subsequently, this vector interacts with the gates; for instance, in the input gate, it merges with C bar, the sole vector with the capability to alter the cell state. Following this, the forget gate interacts with the output gate to produce the hidden vector. In summary, the LSTM architecture's inherent design allows for long- term predictive capabilities, making it adept at forecasting concentrations based on any input data provided. The LSTM [14] curve as shown in figure (), provides insights into the alignment between the training and testing datasets. It achieves a Root Mean Square Error (RMSE) value of 0.006. RMSE represent the standard 13deviation of the residuals. These residuals quantify the difference between observed data points and the regression line, thus offering insights into the data's dispersion around this line. A lower RMSE value indicates that the model's predictions are close to the observed values, signifying the model's effectiveness in capturing the underlying trend of the data (Fig. 13).

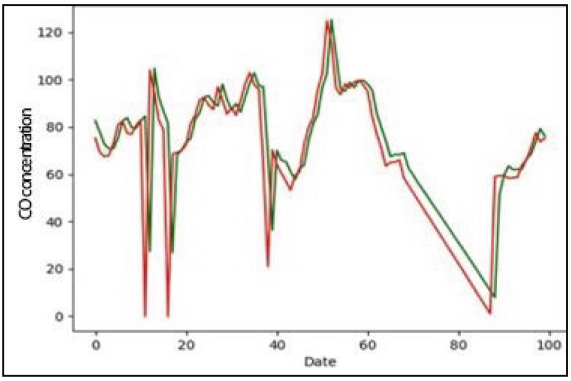


Fig. 13. CO concentration with respect to date using LSTM model.

6 Conclusion

The ever-increasing significance of air quality monitoring finds its roots in the direct implications for human health, environmental sustainability, and economic welfare. By meticulously analyzing the columnar surface of various gases such as CO, NO2, and SO2, this investigation presents an expansive overview of air pollution trends in crucial areas like Delhi, Mumbai, and Punjab. Through the innovative application of methodologies such as the Haversine Formula, a high degree of precision in data extraction and correlation has been achieved. This is further amplified by leveraging machine learning models like ARIMA and LSTM, which demonstrate the potential for nuanced forecasting of pollutant concentrations. The ISMART tool, a tangible outcome of this research, stands as a testament to the power of combining satellite-driven insights with advanced computational models. This tool, made accessible via a Django-powered platform, offers not just

a visualization of current pollution metrics but also predictive insights, thereby arming both public and governmental bodies with the knowledge to make informed decisions. The results derived from the paper underline a crucial observation: while ground-level instruments have their merits, the spatial distribution, vast coverage, and granularity offered by satellite measurements are unparalleled. This research, therefore, doesn't just present data; it offers a transformative lens through which air quality monitoring can be viewed.

For future studies, other pivotal pollutants can be encompassed aiming for a more comprehensive understanding of air quality. By integrating data from IoT devices, finer granularity in datasets can be achieved. This vast coverage offered by satellite data could be further bolstered in terms of accuracy by forging collaborations with ground stations. The realm of deep learning offers opportunities to enhance the predictive prowess of the proposed models, especially by keeping them attuned with real-time data.

References

1. Estimating Near-Surface Concentrations of Major Air Pollutants From Space: A Universal Estimation Framework LAPSO | IEEE Journals & Magazine |
2. Zhu, S., Zhu, H., Xu, J., Zeng, Q., Zhang, D., Liu, X.: Satellite remote sensing of daily surface ozone in a mountainous area. IEEE Geosci. Remote Sens. Lett.Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
3. Mehrabi, M., Scaioni, M., Previtali, M.: Forecasting air quality in Kiev during 2022 military conflict using sentinel 5P and optimized machine learning. IEEE Trans. Geosci. Remote Sens.Geosci. Remote Sens. **61**, 1–10 (2023)
4. Wieczorek, B.: Air pollution patterns mapping of SO2, NO2, and CO derived from TROPOMI over Central-East Europe. Remote Sens. **15**(6), 1565 (2023)
5. Sagar, V.K., et al.: Ground-based remote sensing of total columnar CO2, CH4, and CO using EM27/SUN FTIR spectrometer at a suburban location (Shadnagar) in India and validation of sentinel-5P/TROPOMI. IEEE Geosci. Remote Sens. Lett.Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
6. Zhu, S., et al.: An optimization approach for hourly ozone simulation: a case study in Chongqing, China. IEEE Geosci. Remote Sens. Lett.Geosci. Remote Sens. Lett. **18**(11), 1871–1875 (2021)
7. Kazemi Garajeh, M., Laneve, G., Rezaei, H., Sadeghnejad, M., Mohamadzadeh, N., Salmani, B.: Monitoring trends of CO, NO2, SO2, and O3 pollutants using time-series sentinel-5 images based on google earth engine. Pollutants **3**(2), 255– 279 (2023)
8. Luo, C., Liu, H., Lu, L., Liu, Z., Kong, F., Zhang, X.: Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. J. Integr. Agric.Integr. Agric. **20**(7), 1944–1957 (2021)
9. Kaplan, G., Avdan, Z.Y., Avdan, U.: Spaceborne nitrogen dioxide observations from the sentinel-5P TROPOMI over Turkey. In: 3rd International Electronic Conference on Remote Sensing, MDPI, May 2019, p. 4
10. Scheibenreif, L., Mommert, M., Borth, D.: Toward global estimation of ground- level NO 2 pollution with deep learning and remote sensing. IEEE Trans. Geosci. Remote Sens.Geosci. Remote Sens. **60**, 1–14 (2022)
11. Anitha, M., Kumar, L.S.: Tracking of NO2 and SO2 trace gases emission from Thermal Power Plants in Tamil Nadu using Sentinel 5P Tropomi Satellite with observations from CPCB CAAQM station. In: 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), May 2023, pp. 1–6

12. De Santis, D., et al.: Volcanic SO2 near-real time retrieval using Tropomi data and neural networks: the December 2018 etna test case. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 8480–8483, Jul. 2021
13. Yu, R., Yang, Y., Yang, L., Han, G., Move, O.A.: RAQ–A random forest approach for predicting air quality in urban sensing systems. Sensors **16**(1), Art. no. 1 (2016)
14. Bekkar, A., Hssina, B., Douzi, S., Douzi, K.: Air-pollution prediction in smart city, deep learning approach. J. Big Data **8**(1), 161 (2021)



Analysing Social Media Data for Emotion Detection and Sentiment Analysis

Surbhi Bharti(✉), Divya Verma, Anamika Kumari, and Ashwni Kumar

Indira Gandhi Delhi Technical University for Women, Delhi, India
{surbhi051phd19,divya034bteceai22}@igdtuw.ac.in

Abstract. In the digital era, social media has become a rich source of insights into individuals' emotional state and mental well-being. This paper introduces a cutting-edge framework for sentiment analysis that transcends conventional positive, negative, or neutral annotation of social media text. Sentiment analysis, commonly referred to as opinion mining, is an area of Natural Language Processing (NLP) that has received a lot of attention since it focuses on identifying the emotional tone or sentiment portrayed in textual data. Through rigorous training and fine-tuning of advanced neural networks on a diverse and extensive dataset, our framework achieves an unprecedented level of accuracy in discerning nuanced emotional states and mental health concerns. Our research underscores the potential of leveraging social media as a tool for early intervention and support for individuals navigating mental health challenges. Employing state-of-the-art deep learning models, we were successfully able to predict the sentiment with an accuracy of 78%. In order to comprehend human emotions through textual data, we study several sentiment analysis approaches and applications. We also talk about the difficulties, developments, and potential future directions in this area. By shedding light on the language - emotion relationship in the digital sphere, we contribute to a more empathetic and informed digital landscape.

Keywords: Bidirectional Encoder Representations from Transformers (BERT) · Cutting - edge framework · Deep Learning (DL) · Long Short-term Memory (LSTM) · Sentiment Analysis · Text Classification

1 Introduction

Social media platforms have become a prime conduit for communication, self- expression, and information sharing in today's interconnected digital landscape. Every day, millions of users around the world participate in online dialogues where they regularly share their ideas, opinions, feelings, and experiences [1].

The language and sentiments used in social media content have emerged as an important source of knowledge about the overall emotional health of society within this digital fabric. Twitter, Instagram and Facebook are just a few of the social media sites that have grown to be effective forums for emotional expression [2]. In addition to reinventing

IGDTUW, Delhi.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 367–379, 2025.
https://doi.org/10.1007/978-3-031-75167-7_29

interaction, these platforms have created an immense amount of text-based information that can be used to understand the psychological well-being of their users. This amount of knowledge has an extensive variety of prospective uses, from cutting-edge marketing techniques [3] to the tracking and assistance of mental health. People share their feelings of happiness, grief, achievements, and problems in the form of text-based postings, comments, captions, and tweets.

The possibility to examine and fully comprehend human emotions is distinctive due to this staggering and evolving repository of user-generated content. Text is typically classified as positive, negative, or neutral in context of sentiment analysis, an aspect of NLP [4]. Sentiment analysis classifies and determines emotional undertone of text using NLP techniques. To comprehend the complex nature of human emotions, especially those that have relevance to psychological well-being, this conventional approach frequently falls short. Through the use of text classification, text extraction, and sentiment analysis techniques, researchers and institutions are able to better comprehend public sentiment, enhance services, and offer assistance where it is desired. As social media continues to shape and portray human emotions, sentiment analysis is still an essential approach for identifying the knowledge hidden throughout the vast expanse of online communication [5, 6]. This vast and dynamic repository of user-generated content provides a unique opportunity to analyze and understand human emotions on a large scale.

- Data Availability: Social media sites produce a large amount of text data every day, which makes them an important resource for sentiment research. Users voluntarily provide information that depicts their emotional states [7].
- Real-Time Expression: Social media is an effective instrument for collecting quick reactions to events, the news, and one's own experiences since it allows users to express their feelings in real-time [8].
- Anonymity and Honesty: On social media, users frequently feel more at ease expressing their genuine feelings, which can produce more honest and unfiltered sentiment data.

Text classification [9] is the initial step in the sentiment analysis process. It consists applying one or more predetermined sentiment labels to a specific text document. DL models including Recurrent Neural Networks (RNNs) [10], Naive Bayes (NB), and Support Vector Machines (SVM) [11] utilized for resolving text categorization problems. For developing these models, labelled dataset with textual samples and the requisite sentiment labels were employed. Text extraction [12] is a technique for obtaining relevant text data from social media networks. This involves web scraping or using the Facebook, Twitter, and Instagram Application Programming Interface (APIs) in order to gather posts, comments, and captions. Data pre-processing techniques [13] like maintenance, encoding, and stop word removal are crucial to preparing the text for sentiment analysis. Machine learning (ML) models, as mentioned earlier, are trained on labeled data to predict sentiment. These models can capture complex patterns and nuances in text. Deep learning models have demonstrated excellent performance in sentiment analysis applications, including Recurrent Neural Networks (RNNs) and BERT alike transformer models [14]. They can acquire contextual knowledge and deliver cutting-edge outcomes.

Understanding and deciphering the emotions portrayed in social media posts can be accomplished with the assistance of sentiment analysis. However, it presents its

own peculiar set of challenges particularly when contemplating the complicated and ever-changing nature of online interactions. The inclusion of sarcasm and irony, that individuals frequently employ to cautiously convey their sentiments, makes it harder to discern feelings effectively [15]. Analysts must be competent identify contextual information and user tone in order to successfully negotiate this issue. Moreover, data sparsity is a common hurdle in sentiment analysis, stemming from the frequent use of emojis, misspellings, abbreviations, and informal language in social media text. To address this issue, researchers and practitioners have proposed pre-processing strategies and data augmentation techniques to enhance the robustness of sentiment analysis algorithms. Furthermore, as social media platforms have become more reliant on visual content, the concept of multi-modal analysis, which integrates text and image analysis, has gained prominence. Utilizing multi-modal models that examine both textual and visual elements holds the promise of significantly boosting sentiment analysis accuracy [16, 17]. In this evolving landscape, a comprehensive understanding of sentiments in social media necessitates a holistic approach that considers the interplay between language and visuals (Fig. 1).

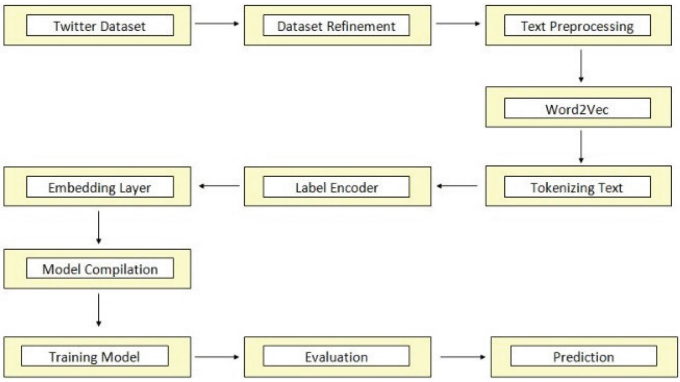


Fig. 1. System Architecture of the Proposed Model

- Our study consists of meticulous data refinement, involving comprehensive text preprocessing techniques. This pivotal step guarantees that the dataset is attentively refined, laying the foundation for a robust sentiment analysis [18, 19].
- We elevate the quality of word representations by harnessing the power of pre-trained Word-2-Vec embeddings. This approach empowers our sentiment analysis with a deeper understanding of semantic nuances within the text [20].
- Our methodology is underpinned by a sophisticated DL model, meticulously crafted using Keras. This model, featuring LSTM layers, serves as the engine behind accurate sentiment classification. We even tried incorporating the BERT model onto the preprocessed data [21, 22].
- We seamlessly progress from model training to rigorous evaluation and, ultimately, the strategic saving of the sentiment analysis model. This well-rounded approach allows for versatile deployment across various real-world applications [14, 23].

This approach comprise of Sect. 1 serving as the Introduction, setting the stage for the research and outlining its significance. In Sect. 2, a thorough Literature Review is conducted, surveying prior works and providing context for the current study. Section 3 talks about the methodology employed in the investigation. The Result section mainly focuses on the insights and findings from related approaches. The analysis providing a deeper understanding of the research outcomes is discussed broadly in Sect. 5. Section 6 summarizes key findings and their implications. Finally, Sect. 7 outlines future Work, suggesting potential directions for further research and development in the field. This structured framework facilitates a coherent presentation of information.

2 Academic Analysis

Recent studies have shown that there has been significant sentiment analysis in social media content related developments. The emphasis on conventional sentiment analysis techniques has traditionally been on classifying text into the obvious classifications of positive, negative, or neutral perspectives. These approaches frequently fail in capturing the subtle complexities of emotional displays, especially those that are affected by mental health. In response, we have gone further into text-based emotion identification in an effort to identify a wider range of emotional states. Our study displays a wide variety of ML algorithms, such as LSTM, effectively classify this expanded range of emotions.

DL techniques have ushered in a new era of sentiment analysis, with models like LSTM and transformative architectures such as BERT enabling more accurate and context-aware sentiment classification. These advancements have not only made sentiment analysis more accurate, but they have also made it possible to comprehend the complex emotional landscape in text data. The interaction of sentiment analysis and mental health detection has been examined in recent articles. Their objective is to determine linguistic cues and patterns that can indicate to issues with mental health, such as anxiety, depression, or even suicidal ideation. Utilising sentiment analysis for early mental health intervention and assistance is an exciting new area of research.

The availability of diverse datasets, including well-known resources like the FER2013 dataset and Twitter sentiment datasets, has played a pivotal role in training and evaluating sentiment analysis models. However, as sentiment analysis in literature pertaining to mental health receives increasing attention, ethical issues have emerged. These factors include things like user permission, privacy, and the proper use of AI technologies. Addressing these ethical issues is crucial as educational institutions strive to optimally utilise sentiment analysis in delicate situations. By expanding on these pillars, our study makes a contribution to this changing environment. We use sophisticated DL methods that enable sentiment analysis and mental health identification at a finer resolution. The project intends to facilitate early intervention and provide assistance to those dealing with mental health difficulties in the digital era by identifying certain indications connected to mental health in social media text. Several studies in the field of sentiment analysis and emotion classification have introduced innovative methods to enhance accuracy and effectiveness. UCRNN model incorporates user qualities alongside textual input, employing CNN for user attribute features and Bi-LSTM for text input features, showing improved classification over TextRNN [4].

	target	ids	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....

Fig. 2. The sample of the used dataset.

A CNN-SVM hybrid approach significantly increases accuracy in sentiment analysis compared to conventional CNN models [3]. Another study suggests word vector and semantic enhancement for text classification, combining word and topic vectors to improve representation[2]. A Dual-Channel CNN utilizes word2vec semantic featureing LDA based model characteristics for enhanced short text sentiment analysis [23]. Logistic Regression is proposed for topic classification in tweets. Additionally, various machine learning techniques are employed for depression prediction and sentiment analysis on Twitter. Semantics are explored as supplementary characteristics for emotion detection and sentiment prediction. SentiRobo employs an enhanced Naive Bayes method for sentiment analysis in large social media datasets [6]. The study uses a variety of ML techniques for evaluating sentiment in tweets. Even when tweets contain a variety of emotions, the goal is to characterise tweet polarity as either positive or negative. The study makes use of a Kaggle dataset that has been pre-labeled as either positive or negative. In the study [8], the twitter analysis is predicted using machine learning approaches. For implementation, ML techniques such the NB algorithm, Logistic Regression (LR), and SVM are utilised. On the tweets acquired from social media, the performance evaluation of the implementation took place in terms of accuracy, precision, recall, and F1-score. These studies collectively contribute to sentiment analysis by introducing novel approaches, techniques, and methodologies for analyzing text sentiment, emotion classification, and topic categorization in social media data.

Table 1. Description of the Dataset.

Attributes	Description
target	It contains the polarity of the tweet whether it is positive, negative or neutral
ids	Unique ID for each tweet
date	The date and time the tweet was posted
flag	It indicates the source of the tweet. If there is no query, then this value is NO_QUERY
user	The username of the Twitter user who posted the tweet
text	The text of the tweet

3 Methodology

The methodology encompasses a multi-stage process designed to enable the comprehensive sentiment analysis of the social media text. To achieve this, we leverage advanced deep learning models and a meticulously curated dataset.

Dataset: We used the dataset [1] named as Sentiment140 sourced from kaggle which contains 1.6 million tweets extracted employing the twitter api. These tweets are labelled to detect the sentiment/emotional tone. Figure 2 represents a sample of the used dataset with all the required attribute. Table 1 gives a brief description of each of the individual attribute from the dataset.

Dataset Exploration - Visualization: This dataset was loaded into a Pandas DataFrame, and exploratory data analysis (EDA) was performed to find out its structure and characteristics. The target variable's distribution, representing the sentiment labels, was visualized using pie charts in the Fig. 3. This step provided insights into the balance or imbalance of sentiment categories in the dataset, which is an essential element for model training.

Data Preprocessing: Text preprocessing is a vital aspect of NLP tasks. This textual data was subjected to various preprocessing steps. In the implementation, we primarily focused on the Keras model and the BERT model and both of them needed a separate set of text data refinements.

– Keras Model: For the Keras model, there are a couple of steps that we accomplished.

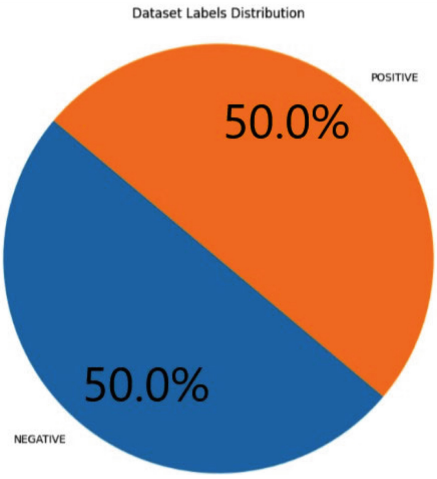


Fig. 3. The Visualization of the target variable using Pie Chart.

- Text Cleaning: This involved removing the Twitter-specific syntax like the removal of special characters, usernames, URLS and other irrelevant information from the text data.

- We have also taken into consideration the tokenization process. Stopword Removal: Eliminating the commonly used words like 'the', 'and' which does not really contribute to the process of sentiment analysis.
- We have used the process of stemming to reduce the word to its root form.
- Padding or Truncation: Ensuring that all text sequences have the same length by either adding padding or truncating. We chose padding for our model.
- Handling emotions and emoji: We mapped the used emojis to their corresponding sentiment labels.
- Word Embeddings: Converting words into numerical vectors using techniques like Word2Vec or pre-trained embeddings. We used the Word2vec techinque for the study.
- Label Encoding: Converted the categorical labels (e.g., "positive," "negative") into numerical values like 0 and 1.

– BERT Model: The following procedures are used by the BERT model to prepare data. We fine-tune our BERT model to the existing model to improve upon the emotion analysis.

  - Tokenization: BERT uses its own tokenizer, which tokenizes text into subwords, and it also adds special tokens like [CLS] and [SEP].
  - Padding or Truncation: Ensuring that input sequences have the same length, similar to Keras.
  - Attention Masks: Creating masks to tell the model which tokens are actual words and which are padding.
  - Label Encoding: Similar to Keras, we converted categorical labels into numerical format.

Sentiment Analysis: We employed the sentiments analised by cutting-edge techniques to benchmark the emotional tone of each text snippet. Traditional positive, negative, and neutral sentiment categories serving the foundation, enabling a baseline understanding of the sentiment landscape.

Deep Learning Models: At the core of our approach lies the utilization of DL models, including RNN and transformers, such as BERT. These fine-tuned models on the sentiment-labeled set to recognize/ predict nuanced emotions and mental health-related signs.

Emotion Recognition: Our model distinguishes between a spectrum of emotions, ranging from happiness and sadness to anxiety and distress. Emotion recognition is performed with a high degree of granularity, enabling the identification of subtle emotional states. This has been done with the use of hummingbird library in the updated model for the analysis of emotions. We have restricted ourselves to major nine types of emotions namely excited, happy, sad, angry, energetic, surprised, relaxed, fearful and calm.

Evaluation Metrics: The accuracy and effectiveness of our model are rigorously evaluated through cross-validation and benchmarking against existing sentiment analysis techniques. Metrics in expression form such as accuracy (1), F1-score (2), precision (3) and recall (4) are used to assess the model's performance in detecting the emotional tone.

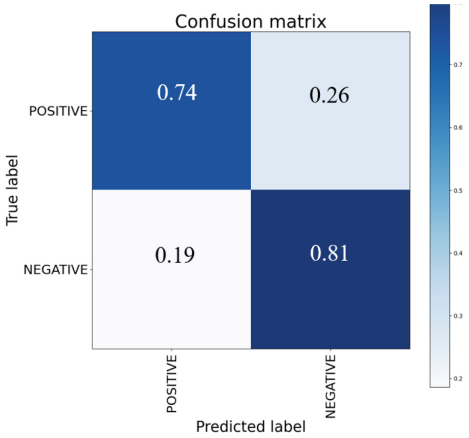


Fig. 4. Heatmap of the model

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of predictions}} \times 100 \tag{1}$$

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4}$$

4 Result and Discussion

This study successfully does the sentiment analysis of a particular text or in this case tweet into the particular emotions in terms of the percentage (Fig. 5; Table 3).

4.1 Findings

This section comprises of the outcomes of the performance metrices on which this analysis has been remarked. The trained sentiment analysis model was evaluated on a separate test dataset to assess its performance. The following metrics were computed:

- Accuracy: The proportion of correctly predicted sentiment labels.
- Confusion Matrix: A visual display of true positive, true negative, false positive, and false negative predictions. Figure 4 shows the Confusion matrix.
- Classification Report: This report consists of a detailed metrics having a blank space in middle, including precision, recall, F1-score, and support for each class. Table 2 displays the classification report of the used model.

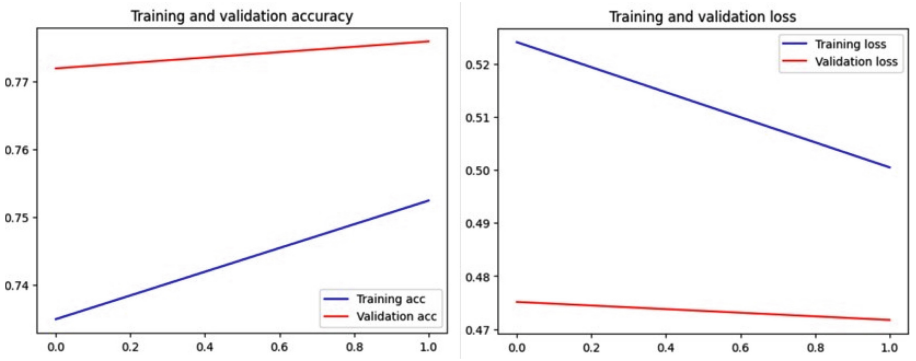


Fig. 5. The representation of the training and validation accuracy and loss.

Table 2. Classification Report of the Model

	Precision	*Recall*	*F1 score*	*Support*
NEGATIVE	0.80	0.74	0.77	159494
POSITIVE	0.76	0.81	0.79	160506
Accuracy			0.78	320000
Macro Avg	0.78	0.78	0.78	320000
Weighted Avg	0.78	0.78	0.78	320000

Table 3. The Input and output of the traditional method of sentiment analysis.

Input Text	Predicted Label	Score	Elapsed Time
"I love the music"	POSITIVE	0.9719099998474121	0.4562 s
"Yayyyy…I'm just super excited"	POSITIVE	0.9816904067993164	0.1961 s
"It's all just ruined"	NEGATIVE	0.06444757431745529	0.1349 s

– Model Evaluation: Our work focuses on the evaluation results demonstrating the effectiveness of the model:

  - The accuracy score indicated the overall correctness of sentiment predictions.
  - The confusion matrix and classification report provided insights into model performance on specific sentiment categories.

Table 4. Prediction's input and respective output of the new cutting edge technique.

Tweet : I'm jumping with excitement
Emotions detected in the tweet :
excited : 42.42%
energetic : 42.42%
surprised : 8.08%
happy : 7.07%
angry : 0.00%
relaxed : 0.00%
fearful : 0.00%
calm : 0.00%
sad : 0.00%

4.2 Discussion

In this section, we do a thorough examination of the findings from this study, looking at their consequences, importance, and compatibility with prior research. With these major considerations we reach towards our conclusion.

- Model Performance: The model achieved competitive accuracy and demonstrated its ability to classify text sentiments accurately. The confusion matrix revealed patterns in classification errors, allowing for potential model improvements.
- Word Embeddings: The use of pre-trained Word2Vec embeddings contributed significantly to the model's performance. These embeddings captured semantic relationships between words, enhancing the model's ability to understand context.
- Table 4 shows that a particular text cannot be specified within a single emotion. So it is represented as a mixture of emotions and the value of each emotion is signified in terms of percentage.

5 Conclusion

This research presented a comprehensive methodology for sentiment analysis using deep learning techniques and pre-trained word embeddings. We explored various approaches to sentiment analysis, ranging from traditional machine learning models to advanced transfer learning techniques with BERT. Our findings defines that the traditional machine learning models can provide reasonably good sentiment analysis results with appropriate pre-processing and feature engineering. Secondly, fine-tuning the BERT model for sentiment analysis yields remarkable improvements in accuracy, showcasing the power of transfer learning in NLP tasks. Thirdly, the choice of sentiment analysis approach should depend on factors like computational resources, dataset size, and the desired level of performance. Traditional models may be suitable for smaller datasets, while BERT-based models are recommended for larger datasets and tasks demanding high accuracy. Lastly, this research lays the foundation for future work in sentiment analysis, including the exploration of other advanced pre-trained models and techniques for further enhancing accuracy and interpretability. Hence, the results demonstrated the

model's effectiveness in accurately classifying sentiment in text data. Moreover, the use of Word2Vec embeddings proved to be a valuable asset in understanding textual context.Further enhancements and real-world applications are promising directions for future research in sentiment analysis.

6 Future Work

We intend to link the results of our sentiment analysis to mental health consequences in an effort to give consumers knowledge about their emotional health and concerns. A more comprehensive approach to user support will emerge from the integration of sentiment analysis with mental health aspects. Enhancing the sentiment analysis methodology to connect users with pertinent mental health assist resources when they are most likely to be experiencing depressive or depressing feelings is one significant advance. However, it's important to acknowledge that data integration and maintaining privacy and data security across multiple platforms can pose challenges in this context. Addressing these issues will be a crucial aspect of our future work to ensure the effectiveness and ethical use of our framework.

References

1. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1, p. 12 (2009)
2. Gao, M., Li, T., Huang, P.: Text Classification Research Based on Improved Word2vec and CNN. In: Liu, X., et al. Service-Oriented Computing – ICSOC 2018 Workshops. ICSOC 2018. Lecture Notes in Computer Science, vol. 11434. Springer (2019)
3. Chen, Y., Zhang, Z.: Research on text sentiment analysis based on CNNs and SVM. In: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 2018, pp. 2731–2734, https://doi.org/10.1109/ICIEA.2018.8398173
4. Xuanyuan, M., Xiao, L., Duan, M.: Sentiment classification algorithm based on multi-modal social media text information. IEEE Access **9**, 33410–33418 (2021). https://doi.org/10.1109/ACCESS.2021.3061450
5. Geetha, S., Kaliappan, V.K.: Tweet analysis based on distinct opinion of social media users. In: 2018 International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, India, pp. 1–6 (2018). https://doi.org/10.1109/ICSNS.2018.8573619
6. Rohani, V.A., Shayaa, S.: Utilizing machine learning in Sentiment Analysis: SentiRobo approach. In: 2015 International Symposium on Technology Management and Emerging Technologies (ISTMET), Langkawi, Malaysia, pp. 263–267 (2015). https://doi.org/10.1109/ISTMET.2015.7359041
7. Susmitha, C., Nikhil, L., Akhil, L., Kavitha, M., Reddy, V.S.N., Shailaja, K.: Sentimental Analysis on Twitter data using supervised algorithms. In: 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 921–925 (2023). https://doi.org/10.1109/ICCMC56507.2023.10084278
8. Mohana, R.S., Kalaiselvi, S., Kousalya, K., M.H.P,L.D., K.A.K.K.: Twitter based sentiment analysis to predict public emotions using machine learning algorithms. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1759–1763 (2021). https://doi.org/10.1109/ICIRCA51532.2021.9544817

9. Kalaivani, P., Dinesh, D.: Machine learning approach to analyze classification result for Twitter sentiment. In: 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, pp. 107–112 (2020). https://doi.org/10.1109/ICOSEC49089.2020.9215278
10. Susmitha, C., Nikhil, L., Akhil, L., Kavitha, M., Reddy, V.S.N., Shailaja, K.: Sentimental analysis on twitter data using supervised algorithms. In: 2023 7th International Conference on Computing Methodologies and Communication (IC- CMC), Erode, India, pp. 921–925 (2023). https://doi.org/10.1109/ICCMC56507.2023.10084278
11. Lee, D.M., Moradi, H.: Knowledge-infused dynamic embedding for predicting the severity of suicidal ideation in social media. In: 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp. 760–766 (2022). https://doi.org/10.1109/CSCI58124.2022.00139
12. Ibrahiem, S.S., Ismail, S.S., Bahnasy, K.A., Aref, M.M.: Multi-emotion Classification Evaluation via Twitter. In: 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, pp. 60–67 (2019). https://doi.org/10.1109/ICICIS46948.2019.9014847
13. Shobana, J., Amudha, S., S.K.A.: Emoji anticipation and prediction using deep neural network model. In: 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, pp. 1–8 (2022). https://doi.org/10.1109/ICPECTS56089.2022.10047692
14. Choudhary, S., Godara, J.: Semantic analysis on social media. In: 2021 International Conference on Computing Sciences (ICCS), Phagwara, India, pp. 239–243 (2021). https://doi.org/10.1109/ICCS54944.2021.00054
15. Kumar, R., Sharma, C.M., Chariar, V.M., Hooda, S., Beri, R.: Emotion analysis of news and social media text for stock price prediction using SVM-LSTM-GRU composite model. In: 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, pp. 329–333 (2022). https://doi.org/10.1109/CISES54857.2022.9844375
16. Bharti, S., Kumar, A.: Low power-area based composite 6T-8T SRAM for soft computing applications. In: 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, pp. 1–6 (2022). https://doi.org/10.1109/ICPC2T53885.2022.9777036
17. Nhlabano, V.V., Lutu, P.E.N.: Impact of text pre-processing on the performance of sentiment analysis models for social media data. In: 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, pp. 1–6 (2018). https://doi.org/10.1109/ICABCD.2018.8465135
18. Likhitha, K., Rekha, K.S., Ramesh, S.: Improved accuracy for exploring text - based emotion recognition in social media conversation generalized linear model compared with random forest. In: 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1–6, https://doi.org/10.1109/ICONSTEM56934.2023.10142393
19. Bharti, S., Kumar, A.: Rooting Of PPN technique on 8T-SRAM for low power multimedia applications. In: 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–5 (2022). https://doi.org/10.1109/IC-CCI54379.2022.9740955
20. Darokar, M.S., Raut, A.D., Thakre, V.: Human temperament and emotions: analytical research for social media. In: 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), NaviMumbai, India, pp. 1–4 (2021). https://doi.org/10.1109/ICNTE51185.2021.9487773

21. Olusegun, R., Oladunni, T., Audu, H., Houkpati, Y., Bengesi, S.: Text mining and emotion classification on monkeypox twitter dataset: a deep learning-natural language processing (NLP) approach. IEEE Access **11**, 49882–49894 (2023). https://doi.org/10.1109/ACCESS.2023.3277868
22. Rai, K., Kumar, P.: Comparative analysis of machine learning and deep learning techniques in text based emotion detection. In: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, pp. 260–263 (2023). https://doi.org/10.1109/CISES58720.2023.10183612
23. Lin, J., Gu, Y., Zhou, Y., Yang, A., Chen, J., Li, X.: Combining convolutional neural networks and word topic features for chinese short text sentiment analysis. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, pp. 422–425 (2018). https://doi.org/10.1109/IC-SESS.2018.8663728



A Comprehensive Approach to AI-Enabled Diagnosis of Lung Diseases: Utilizing Pretrained Models for Fibrosis, Tuberculosis, and Covid-19 Detection

Sapna Yadav[1(✉)], Syed Bilal Abbas Rizvi[2], and Syed Afzal Murtaza Rizvi[1]

[1] Department of Computer Science, Jamia Millia Islamia, New Delhi, India
sapyadav08@gmail.com, sarizvi@jmi.ac.in
[2] Ministry of Defense, Indian Army, AMC, New Delhi, India

Abstract. Chest radiographs are extensively employed for the detection and diagnosis of several pulmonary conditions, including fibrosis, tuberculosis, and COVID-19. The manual processing of chest X-ray images is a time-consuming and error-prone task, yet its analysis is crucial for timely disease identification and diagnosis. This study utilizes Convolutional Neural Networks and five pretrained models viz. ResNet50, MobileNetV2, InceptionNetV3, Xception, and VGG16 to detect lung infectious diseases using chest X-ray. The main purpose of this work is to assess the performance of deep learning models in diagnosing the lung diseases namely Covid-19, Fibrosis, and Tuberculosis. These models are trained on publicly available datasets and assessed for the performance parameters such as accuracy, recall, precision, F1-score, AUC-score, and false positive, false negative counts. The performance of CNN, and five pretrained models are evaluated and compared with respect to these performance metrics. It has been observed from the results that ResNet50, MobileNetV2, and VGG16 show incredibly high performance for all performance parameters used in this work. To provide early diagnosis and patients outcomes, these automated diagnosis methods can be used by the medical professionals.

Keywords: Pretrained Model · Convolutional Neural Network · Fibrosis · Tuberculosis · Covid-19

1 Introduction

Pulmonary infectious illnesses encompass a range of respiratory problems, including as pneumonia, TB, COVID-19, and fungal infections. These conditions manifest through symptoms such as coughing, fever, and respiratory distress. These disorders provide a substantial risk to public health and need precise diagnosis in order to facilitate appropriate treatment. Conventional approaches involve invasive, costly, and error-prone procedures including bronchoscopy, sputum culture, and laboratory testing.

Hence, there exists a necessity for the development of non-invasive diagnostic techniques that include artificial intelligence with medical imaging [1].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 380–392, 2025.
https://doi.org/10.1007/978-3-031-75167-7_30

Medical imaging areas including segmentation, classification, detection, and generation have seen notable success with deep learning approach utilizing artificial neural networks [2]. However, in order to prevent overfitting and efficiently train models, this strategy requires a substantial amount of labelled data. In the context of pulmonary infectious diseases, obtaining a sufficient number of CXR images with reliable annotations is not easy due to ethical, legal, and privacy issues. Therefore, there is a need for methods that can overcome the data scarcity problem and utilize the existing knowledge from other domains or tasks [3].

This study uses CXR pictures to identify lung viral diseases using pretrained models and CNNs. On a chest X-ray dataset of fibrosis, tuberculosis, and Covid-19 illnesses, we evaluate and compare six cutting-edge CNN and pretrained architectures (CNN, VGG16, ResNet50, InceptionV3, Xception, and MobileNet). We then fine-tune them. The performance of models is evaluated using various metrics such as accuracy, recall, precision, F1-score, false positive and false negative counts, and AUC. This study answers the following research questions:

- How do the performance evaluation parameters of the different deep learning models (CNN, VGG16, ResNet50, MobileNetV2, InceptionNetV3, Xception) compare in the detection of fibrosis, tuberculosis, and COVID-19 through chest X-ray images?
- To what extent do the pretrained models, trained on the dataset, generalize to diverse chest X-ray images, and how do they perform on images from different sources or populations?
- How does the selection of datasets influence the deep learning model's performance in detecting lung diseases, and to what extent can these models be used to datasets that possess dissimilar characteristics or distributions?
- Based on the results obtained, is there an optimal deep learning model or a combination of models that can effectively diagnose lung infectious diseases from chest X-rays?

Below are the study's primary contributions:

- Presents a comprehensive study on the application of transfer learning with CNNs for detecting lung diseases (Fibrosis, Tuberculosis, Covid-19) from CXR images.
- Thorough evaluation and performance comparison of five pretrained CNN architectures, conventional CNN. For experimental assessment, recall, precision, accuracy, AUC score and F1-Score performance parameters are em ployed.
- Fine-tune the models on three separate datasets of CXR images with Fibrosis, Tuberculosis, COVID-19.

The ensuing sections are arranged in the following manner: Sect. 2 presents a comprehensive review of the current research pertaining to deep learning and transfer learning methodologies employed in the identification of pulmonary infectious diseases through the analysis of chest X-ray (CXR) images. The methodology and materials employed in this investigation are detailed in Sect. 3. The results of experiments are presented and analyzed in Sect. 4. The conclusion and prospective areas for further research are presented in Sect. 5 of the report.

2 Background Study

In order to diagnose pneumonia in CXR images, the study [4] used Xception, ResNet50, DenseNet121, Incep-tionV3, and other CNNs. The deep learning models shown significant potential in this domain, as seen by their impressive performance metrics of 96.5% accuracy and a corresponding F1-score of 96.4%. [5] created CheXpert, a massive dataset consisting of more than 200,000 X-ray images of the chest labelled for 14 observations employing natural language processing methods based on radiology reports. They also proposed a CNN model that leverages uncertainty labels to handle noisy annotations in the dataset. They evaluated their model on two tasks: multi-label classification and weakly supervised localization. In comparison to earlier techniques, they completed both tasks with state-of-the-art outcomes.

[6] developed LUNGx Challenge, a public dataset of 888 chest CT scans labeled for lung nodule malignancy by four radiologists. They also proposed a CNN model that uses 3D patches to capture both spatial and contextual information from the CT scans. They evaluated their model on the LUNGx Challenge dataset and compared it with the performance of six radiologists. They discovered that the AUC score for their model was 0.99, which was greater than the radiologists' average AUC of 0.79. A reliable Bayesian CNN-based TB diagnosis system (B-CNN) [7] is created and trained using the Montgomery and Shenzhen datasets. Other pre-trained models for feature extraction include AlexNet, VGG16, VGG19, and ResNet with SVM classification. The suggested Acr-2 B-CNN model outperforms other cutting-edge models, according on evaluation findings. Once more, the B-CNN uses Bayesian confidence and model uncertainty to improve model accuracy.

Three techniques were developed by the researchers in reference [8] to improve the visual accuracy of images: Contrast Limited Adaptive Histogram Equalisation, High-Frequency Emphasis Filtering, and Unsharp Masking. The improved dataset was used to train ResNet and EfficientNet models for classification and feature extraction, achieving an AUC score of 89.92% and an accuracy of 94.8%. The authors of this study [9] suggest a unique method for categorizing pneumonia photos using quaternion-valued tensors that uses deep learning on quaternions. The QCNN, a deep learning method, successfully categorizes chest X-ray pictures of pneumonia, achieving high accuracy, sensitivity, and specificity, showcasing the benefits of quaternion deep learning. In the study [10], the use of deep convolutional neural network (CNN) architectures was employed to establish a method for the identification of pneumonia through the analysis of X-ray pictures of the thoracic region. It evaluates the performance of VGG16, ResNet50, InceptionV3, and DenseNet121 as feature extractors and compares them with traditional machine learning classifiers. DenseNet121 achieves the best performance with 97.08% accuracy and 97.22% recall. The paper discusses the advantages and limitations of CNNs in pneumonia detection. A classifier based on machine learning is shown in work [11] that effectively distinguishes between different forms of pneumonia and the chest X-ray images of COVID-19 patients. The COVID-Classifier employs dimensionality reduction techniques to extract characteristics from a dataset consisting of 423 images of COVID-19, 1485 images of viral pneumonia, and 1579 images of conventional chest X-ray images. The effectiveness of the classifier is evaluated using two different methodologies, allowing for efficient prioritization of non-COVID-19 patients.

This paper [12] retrained an open-source pneumothorax detecting algorithm using a diverse dataset of 112,120 X-rays of chest from the RSNA and NIH ChestXray14 Pneumothorax Detection Challenge. The retrained algorithm achieved a higher AUC of 0.94, compared to 0.86 and 0.79 for the original and baseline models. A huge dataset was used to retrain an open-source pneumothorax detection algorithm, showing its usefulness in identifying mistakes and limits. [13] presents an innovative approach for autoencoded deep feature-based pneumothorax detection in x-ray images. The method uses an autoencoder, similarity measure, and classifier to filter out negative images and rank positive ones. It accomplishes classification of pneumothorax with a mean average accuracy of 0.92 and with a sensitivity of 0.95. The method can handle noisy and incomplete labels and generalize to unseen images. The Xception pre-trained model, used in [14], significantly reduces training time in pneumothorax detection. 2027 images in all were utilized, and recall, precision, accuracy, and F1-score were computed.

3 Material and Methods

The methods and datasets utilized to conduct this analysis are presented in the below subsections.

3.1 Dataset Description

Three separate X-ray images of chest public datasets accessible on Kaggle were used in this investigation. There are 11800 normal X-ray images and 1686 fibrosis x-rays of chest in the fibrosis dataset [15] that were used to classify fibrosis. Of the 1686 fibrosis X-rays, 1400 were chosen for this investigation. This compilation consists of a diverse range of openly accessible datasets, including the NIH Chest X-ray collection which showcases 14 Common Thorax Diseases, the Tuberculosis (TB) Chest X-ray Database, the COVID-19 Chest X-ray Database, the Labelled Optical Coherence Tomography (OCT) dataset, and the Chest X-Ray Images for Classification dataset. Two publicly online accessible datasets [16] and [17] were used to compile the COVID-19 radiography dataset. 21 patients with verified COVID-19 were also used in the chest radiography studies [18]. The utilization of "Tuberculosis (TB) Chest X-ray images developed by a collaborative team consisting of researchers from Qatar University, University of Dhaka, Hamad Medical Corporation, and Malaysia" was employed for the purpose of tuberculosis analysis. This dataset's most recent version includes 3500 Normal and 3500 images of Tuberculosis. Both on Kaggle [19] and online [20], the dataset is accessible. 700 TB images and 1200 normal images are taken from this dataset as a subset. Shenzhen and Montgomery's dataset of TB images [21], from which 336 and 58 TB

images were chosen, are further sources of TB images. The dataset's image distribution across classes is seen in Table 1.

Table 1. Distribution of Images of each class

Disease	Sources	Class	Training	Testing	Total
Fibrosis	[15]	Normal	992	408	1400
		Fibrosis	1000	400	1400
Tuberculosis	[19–21]	Normal	780	250	1030
		TB	712	227	939
Covid-19	[16–18]	Normal	864	317	1181
		COVID-19	856	308	1164

3.2 Methodology

The present study utilizes a CNN architecture for the purpose of detecting fibrosis, Covid-19, and tuberculosis in chest X-rays. As seen in Fig. 1, this work encompasses three primary phases.

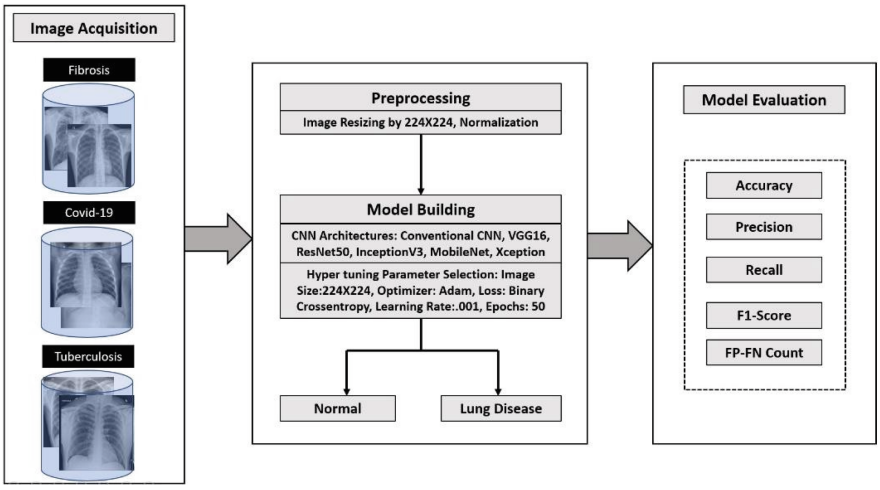


Fig. 1. Workflow of Study

The suggested workflow is broken down into a number of processes, including data collection, preprocessing, training-testing, classification, and performance assessment of model (Fig. 1). This approach is used to normalize the data in a range of [0–255] and is part of the data preparation procedure. The provided input data is subsequently utilized to train pretrained models in order to construct a model that can effectively classify lung disorders. This is followed by model's performance evaluation. The performance of the models is evaluated and compared based on many performance indicators, including accuracy, recall, precision, f1-score, AUC score, and false positive/false negative count.

3.3 Convolutional Neural Network Architecture

The procedure in convolutional neural networks (CNNs) entails the application of a convolutional filter, the accumulation of the resultant sums in a feature map, and the interpretation of the input data as a matrix or array. CNN's performance relies on the number of layers it uses [22]. Numerous tasks, such as labelling, action recognition, posture estimation, object tracking, object identification, visual saliency detection, and text detection, are undertaken by CNN [23]. X-ray images with a size of 224 by 224 are used as input in this investigation at the input layer. With padding, three-by-three filter configurations, and the relu activation function, we used four convolutional layers. The maximum size of a patch was calculated for each feature map using max pooling. And (2,2) is the maximum pooling size. In this CNN architecture, the sigmoid activation function of the outer layer is utilized by the final FC layer. This layer offers the anticipated result, showing whether or not the person has lung opacity.

3.4 Transfer Learning Through Pretrained Models

Transfer learning transfers information between models, allowing them to be precisely adjusted to learn from new datasets by utilizing the knowledge gained from a large dataset [24]. This study uses transfer learning through various pretrained models to examine the performance of VGGNet16, ResNet50, InceptionNetV3, Xception, and MobileNet. To enhance the efficiency of the training process, a strategy is implemented where many pre-trained models with different hyperparameters are utilized to choose the most suitable combination of hyperparameters and models for the specific task of lung disease detection.

3.5 Performance Evaluation Metrics

Various performance metrics are employed to evaluate the effectiveness of Convolutional Neural Networks (CNNs) and pre-trained models. These measurements encompass accuracy, recall, precision, F1-score, as well as the quantification of false positives and false negatives. Four key words are: false positive, true positive, false negative, and true negative. Below are the equations of the metrics utilized in these models' prediction:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Prec = \frac{TP}{TP + FP} \tag{2}$$

$$Rec = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \tag{4}$$

Accuracy is a quantifiable measure of how many correct predictions the algorithm produced. The performance of the model in certain conditions may be assessed by the metrics of "recall" and "precision" [25]. The term "precision" pertains to the level of accuracy in correctly identifying positive labels. And "recall" relates to the percentage of correctly identified positive labels that is accurately predicted by the model.

4 Results and Discussion

The analysis and description of results of CNN and five pretrained models viz. ResNet50, VGG16, Xception, InceptionV3, and MobileNetV2 in detection of three lung diseases (Covid-19, Fibrosis, Tuberculosis) are discussed in this section. The model's performance highly depends on the size, quality of the images and the hyperparameters used. The study assesses model performance in identifying positive and negative categories and projecting true positives using four performance metrics: accuracy, recall, precision, and F1-score. The effectiveness of CNN and pretrained models has been evaluated and contrasted with regard to the detection of lung disorders in chest X-rays. VGG16 and ResNet50 have higher AUC score, F1-scores, precision, accuracy, and recall than any other CNN architectures.

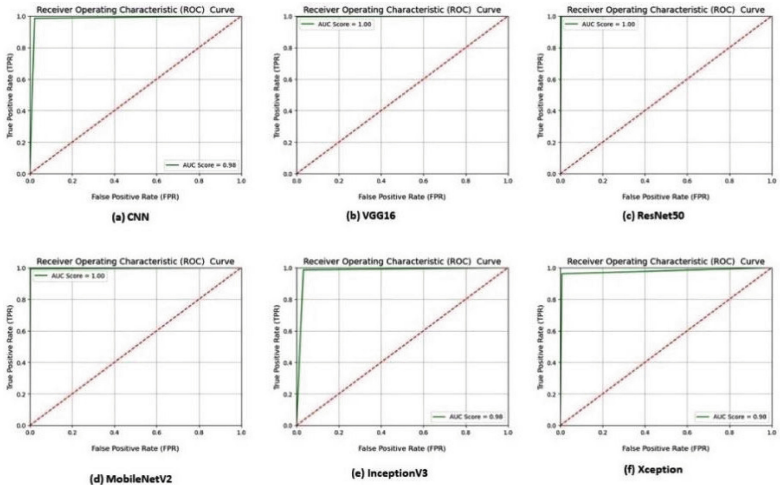


Fig. 2. AUC-ROC Curve for Fibrosis Detection using Deep Learning

Figures 2, 3, and 4 display the receiver operating curve-area under curve (ROC- AUC) of six distinct deep learning models for the detection of Covid-19, TB, and fibrosis. As AUC value of VGG16, ResNet50, and MobileNetV2 for the three-disease detection are high as compared to other models, indicating that these models can identify between primarily positive and negative class items completely. This demonstrates that, despite the non-uniform sample distribution in the dataset, these three models are more reliable and consistent.

From Tables 2, 3, and 4 when compared to CNN and other pretrained classification models, VGG16, ResNet50, and MobileNetV2 outperforms them in terms of total performance. The performance of the 6 CNN models —CNN, VGG16, ResNet50, MobileNetV2, InceptionV3, Xception—is shown in Figs. 5, 6, 7. It can be seen from the graphical comparison of results that the accuracy, precision, and recall of VGG16, ResNet50 is higher for Fibrosis and Covid-19 disease datasets than those of all the classifiers used. And for Tuberculosis dataset, MobileNetV2 model is showing better results for performance metrics than other pretrained models. It has been observed

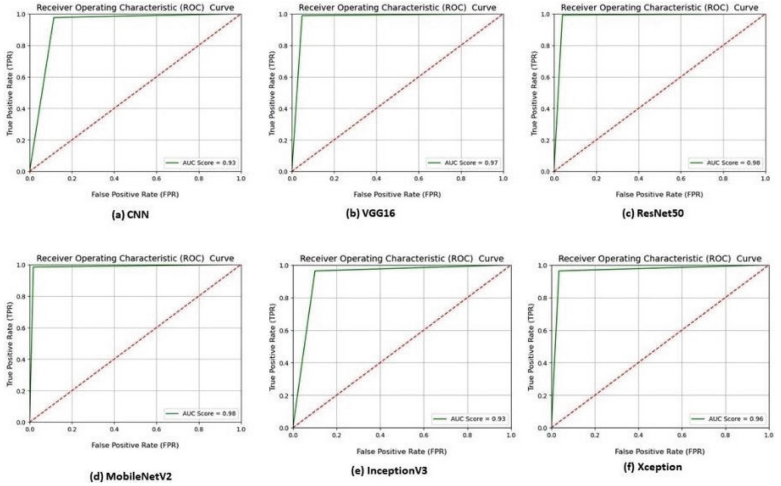


Fig. 3. AUC-ROC Curve for Tuberculosis Detection using Deep Learning

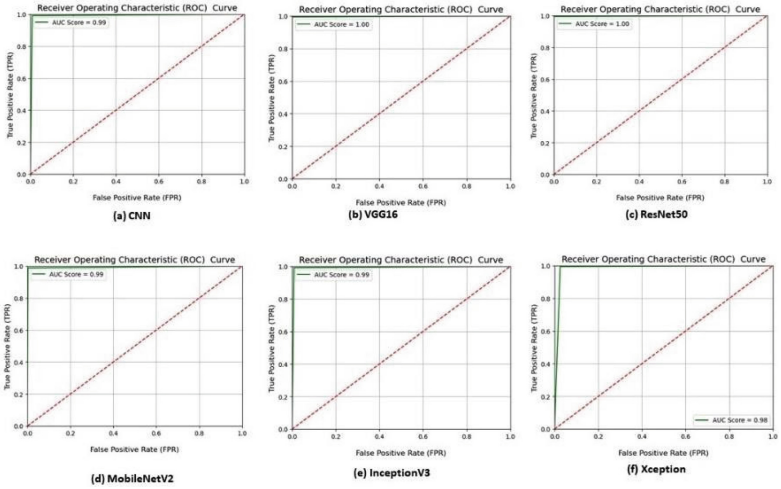


Fig. 4. AUC-ROC Curve for Covid-19 Detection using Deep Learning

that VGG16, ResNet50, and MobileNetV2 gives comparatively higher AUC score than other pretrained models and CNN. Our results show that MobileNetV2 accomplished the highest accuracy of 98.32% on the Tuberculosis disease dataset, followed by ResNet50 with 97.69%, VGG16 with 97.06%, Xception with 96.44%, InceptionV3, and CNN with 93.29%. On the Fibrosis disease dataset, VGG16 achieved the highest accuracy of 99.75%, followed by ResNet50 with 99.63%, MobileNetV2 with 99.50%, CNN with 98.14%, InceptionV3 with 97.65, and Xception with 97.64%. On the Covid-19 disease dataset, VGG16 and ResNet50 also achieved the highest accuracy of 99.68%, followed

Table 2. Performance Evaluation Results for Fibrosis Detection

Model/ Performance Metrics	Accuracy	Precision	Recall	F1-Score	AUC-Score
CNN	98.14	97.75	98.49	98.12	.98
VGG16	99.75	100.00	99.5	99.75	1.0
ResNet50	99.63	99.5	99.75	99.62	1.0
MobileNetV2	99.50	99.75	99.25	99.5	1.0
InceptionV3	97.65	96.75	98.47	97.6	.98
Xception	97.64	99.25	96.13	97.67	.98

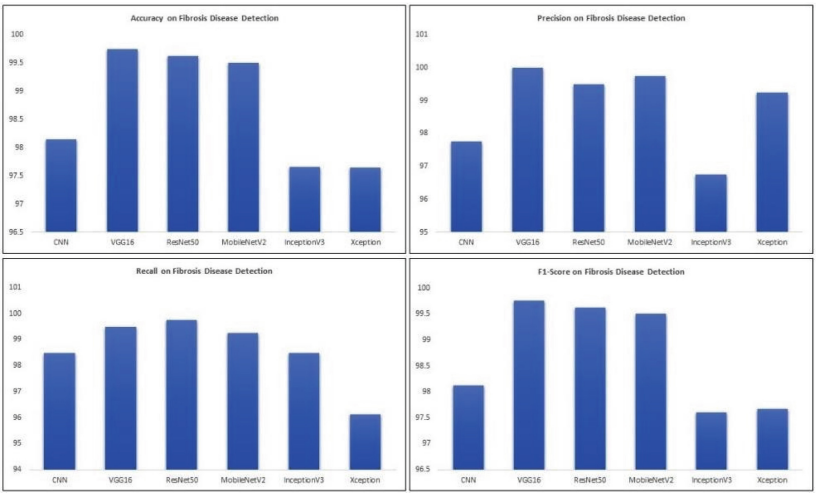


Fig. 5. Performance Evaluation of Pretrained Models for Fibrosis Disease Detection

Table 3. Performance Evaluation Results for Tuberculosis Detection

Model/ Performance Metrics	Accuracy	Precision	Recall	F1-Score	AUC-Score
CNN	93.29	88.55	97.1	92.63	.93
VGG16	97.06	95.15	98.63	98.86	.97
ResNet50	97.69	96.04	99.09	97.54	.98
MobileNetV2	98.32	98.24	98.24	98.24	.98
InceptionV3	93.29	89.87	95.77	92.73	.93
Xception	96.44	96.48	96.05	96.26	.96

by CNN with 99.36%, MobileNetV2 with 99.20%, InceptionV3 with 99.04%, and Xception with 98.40%. These results indicate that VGG16, ResNet50, and MobileNetV2 are

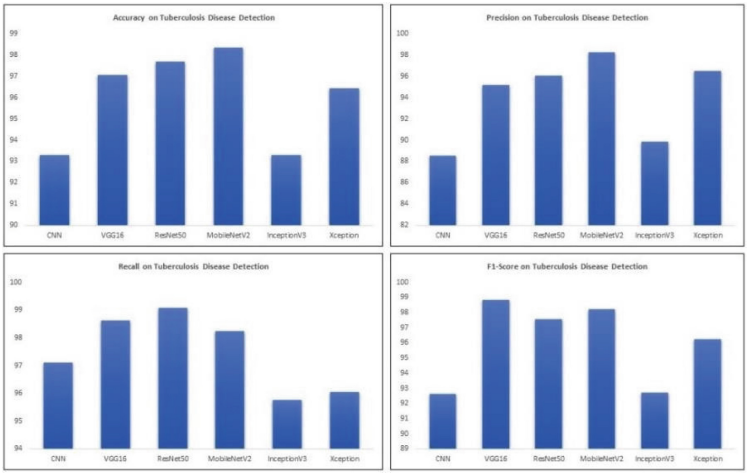


Fig. 6. Performance Evaluation of Pretrained Models for Tuberculosis Disease Detection

Table 4. Performance Evaluation Results for Covid-19 Detection

Model/ Performance Metrics	Accuracy	Precision	Recall	F1-Score	AUC-Score
CNN	99.36	99.35	99.35	99.35	.99
VGG16	99.68	100.00	99.35	99.67	1.0
ResNet50	99.68	100.00	99.35	99.67	1.0
MobileNetV2	99.20	99.68	98.71	99.19	.99
InceptionV3	99.04	99.03	99.03	99.03	.99
Xception	98.40	97.40	99.34	98.36	.98

the effective models for detecting fibrosis, tuberculosis and COVID using X-ray images of chest.

We have also analyzed the confusion matrices, AUC score and the classification report of the models to understand their strengths and weaknesses. We have discovered that COVID-19, fibrosis, and TB can all be detected with excellent recall and precision using VGG16, ResNet50, and MobileNetV2. Along with these performance indicators, false positive and false negative statistics are also examined, with a focus primarily on the false negative count (FN). Since there should be as few False Negatives as possible while handling a healthcare issue. Table 5 provides statistics for FP and FN outcomes for the detection of fibrosis, TB, and COVID-19.

The objective for any healthcare related problem should be to minimise the amount of false negatives, even if it is crucial to reduce both FP and FP count outcomes. When it comes to healthcare, false negatives are more harming than false positives since they imply that a person does not have a condition when they actually do. The assessment findings from this study (Table 5) demonstrate that ResNet50 and VGG16 provide a

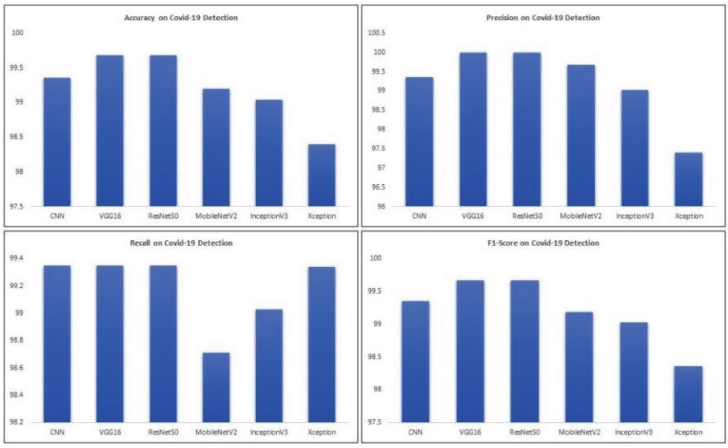


Fig. 7. Performance Evaluation of Pretrained Models for Covid-19 Detection

Table 5. Analysis of False Negative, False Positive for Fibrosis, Tuberculosis, and Covid-19 Detection

Fibrosis			Tuberculosis		Covid-19	
Model						
False Positive (FP)		False Negative (FN)	False Positive (FP)	False Negative (FN)	False Positive (FP)	False Negative (FN)
CNN	9	6	26	6	2	2
VGG16	0	2	11	3	0	2
ResNet50	2	1	9	2	0	2
MobileNetV2	1	3	4	4	1	4
InceptionV3	13	6	23	9	3	3
Xception	3	16	8	9	8	2

low FP and FN count for the identification of fibrosis and Covid-19. On the other hand, MobileNetV2 gives a small number of FP and FN count for Tuberculosis identification as compared to other pretrained models. As a consequence, it has been shown from the findings that in addition to the pretrained architecture, the dataset over which the deep learning model is trained also affects the model's overall performance.

5 Conclusion

The identification of fibrosis, TB, and COVID-19 using chest X-ray images has been tested using five pretrained architectures: VGG16, ResNet50, MobileNet, InceptionNetV3, Xception, and CNN. A total of 7114 X-ray pictures are gathered for this study

from the Montgomery, Shenzhen, and NIH chest X-ray databases. Determining the appropriate model for X-rays classification is among the most crucial problems. Along with suitable model selection, model parameters are also very important to be able to produce improved results. In conclusion, our study has successfully showcased the efficacy of adapting pretrained models for the purpose of identifying lung disorders through the utilization of chest X-rays. Among the six deep learning models we have compared, VGG16, ResNet50, and MobileNetV2 are the suitable ones for this task due to their high accuracy, precision, recall, and interpretability. Our work may be useful in the development of automated systems for the diagnosis of lung disorders utilising chest X-ray pictures. This work adds to the continuing attempts to use deep learning and artificial intelligence for lung disease identification through medical image analysis.

The current study can be expanded in the future by using ensemble learning to combine the predictions of various CNN models and increase the precision and robustness of the classification, as well as multi-task learning can also be included to perform simultaneous detection and localization of multiple diseases on X-ray images of chest. To verify the efficacy of deep learning models in actual healthcare settings, the study may additionally incorporate rigorous clinical trials.

References

1. Bluemke, D.A., et al.: Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board. Radiology **294**(3), 487–489 (2020)
2. Esteva, A., et al.: A guide to deep learning in healthcare. Nat. Med. **25**(1), 24–29 (2019)
3. Rahman, T., et al.: Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Appl. Sci. **10**(9), 3233 (2020)
4. Kundu, R., Das, R., Geem, Z.W., Han, G.T., Sarkar, R.: Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLoS ONE **16**(9), e0256630 (2021)
5. Irvin, J., et al.: Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 590–597 (2019)
6. Ardila, D., et al.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. **25**(6), 954–961 (2019)
7. Abideen, Z.U., et al.: Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks. IEEE Access **8**, 22812–22825 (2020)
8. Munadi, K., Muchtar, K., Maulina, N., Pradhan, B.: Image enhancement for tuberculosis detection using deep learning. IEEE Access **8**, 217897–217907 (2020)
9. Singh, S., Tripathi, B.K.: Pneumonia classification using quaternion deep learning. Multimedia Tools Appl. **81**(2), 1743–1764 (2022)
10. Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., Mittal, A.: Pneumonia detection using CNN based feature extraction. In: Proceedings of the IEEE International Conference On Electrical, Computer and Communication Technologies (ICECCT), pp. 1–7 (2019)
11. Zargari Khuzani, A., Heidari, M., Shariati, S.A.: COVID-classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. Sci. Rep. **11**(1), 9887 (2021)
12. Kitamura, G., Deible, C.: Retraining an open-source pneumothorax detecting machine learning algorithm for improved performance to medical images. Clin. Imaging **61**, 15–19 (2020)

13. Sze-To, A., Riasatian, A., Tizhoosh, H.R.: Searching for pneumothorax in x-ray images using autoencoded deep features. Sci. Rep. **11**(1), 9817 (2021)
14. Singh, R., Sharma, A., Sharma, N., Gupta, R.: Xception model for pneumothorax classification using chest X-ray images. In: Proceedings on 2nd International Conference for Innovation in Technology (INOCON), pp. 1–5. IEEE (2023)
15. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106, IEEE (2017)
16. Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M.: Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988 (2020)
17. https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia, last accessed 2023/03/01
18. Ng, M.Y., et al.: Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiol. Cardiothorac. Imaging **2**(1), e200034 (2020)
19. https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset, last accessed 2023/02/26
20. https://ieee-dataport.org/documents/tuberculosis-tb-chest-x-ray-database, last accessed 2023/02/26
21. Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X., Thoma, G.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg. **4**(6), 475 (2014)
22. Albawi, S., Mohammed, T. A., Al-Zawi, S.: Understanding of a convolutional neural network. In: Proceedings on the International Conference on Engineering and Technology (ICET), pp. 1–6, IEEE (2017)
23. Gu, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
24. Niu, S., Liu, Y., Wang, J., Song, H.: A decade survey of transfer learning (2010–2020). IEEE Trans. Artif. Intell. **1**(2), 151–166 (2020)
25. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240 (2006)

Campus Exploration: Figma-Driven UI Design for a Campus Tour App

Surbhi Bharti(✉), Medha Arora, and Ashwni Kumar

Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi, India
{surbhi051phd19,medha075bteceai22}@igdtuw.ac.in

Abstract. Virtual campus tours have become increasingly popular as a useful resource for prospective students, parents, and other stakeholders to remotely tour educational institutions. In order to develop a user-friendly and engaging interface for virtual campus tours, this research study investigates the convergence of Human-Computer Interaction (HCI) concepts and User Interface (UI) design. The goal is to enhance user interaction and engagement throughout the virtual tour, which will ultimately help students make better decisions when choosing their academic communities. In order to improve the campus experience for both prospective students and visitors, this research study examines the design, development, and implementation of a campus tour app using the Figma design platform. The detailed process of UI design, the tenets of user-centered design, and the seamless integration of technology to produce an interesting and educational app are all covered in this article. It also emphasizes the possibility for future improvement by incorporating AR technology. The article emphasizes the transformative effect of UI design on the educational landscape and explores the triumphs, problems, and lessons learned throughout the app's development. The introduction of augmented reality (AR) is expected to elevate the campus tour experience as the educational technology landscape continues to develop, offering an immersive and interactive way to explore educational facilities. This study lays the groundwork for new advancements in educational technology, with an emphasis on creative UI and AR integration to improve the campus tour experience .

Keywords: Artificial intelligence · Augmented Reality (AR) Integration · Figma · User Interface (UI) Design · Virtual Campus Tours.

1 Introduction

Virtual campus tours have become increasingly popular as a useful resource for prospective students, parents, and other stakeholders to remotely tour educational institutions. In order to develop a user-friendly and engaging interface for virtual campus tours, this research study investigates the convergence of Human-Computer Interaction (HCI) concepts and User Interface (UI) design. The goal is to enhance user interaction and engagement throughout the virtual tour, which will ultimately help students make better decisions when choosing their academic communities.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 393–406, 2025.
https://doi.org/10.1007/978-3-031-75167-7_31

This study starts off with a thorough analysis of the HCI literature that already exists and the best methods for UI design in the context of virtual tours. It draws attention to the particular difficulties and opportunities posed by virtual campus tours and emphasises the significance of an interesting, instructive, and user-centered interface. The design and evaluation parts of the study technique are iterative, and each stage incorporates user feedback. To acquire information and improve the UI design, we use methods including usability testing, heuristic evaluations, and user surveys. User preferences, navigational patterns, and aesthetic components that improve the overall virtual tour experience can be incorporated thanks to the iterative design process. To meet the needs of various user groups, key design considerations include the development of userfriendly navigation controls, interactive campus maps, multimedia integration, and accessibility features. By balancing functionality, aesthetics, and information delivery, the resulting UI design enables users to easily explore campus buildings, academic divisions, and extracurricular opportunities. Additionally, this research analyses user behaviour during virtual tours and conducts user surveys to better understand how the designed user interface affects user engagement and decision-making. The efficiency of the UI design is measured using metrics like user satisfaction through surveys, task completion rates, and time spent on the platform. The results of this study make a contribution to the larger field of HCI by providing understanding of the particular difficulties and design approaches for virtual campus visits. Additionally, the study offers useful suggestions for academic institutions aiming to improve their virtual tour experiences, thereby enhancing their capacity to draw in and inform prospective students. This study report concludes by highlighting the importance of HCI and UI design concepts in producing an engaging and educational interface for virtual campus visits. Institutions may encourage a more inclusive and enjoyable educational experience by putting a priority on user-centered design, which will enable prospective students and other stakeholders to make educated judgements about their educational options.

2 History and Related Work

2.1 Background and Context

Educational institutions are constantly looking for novel approaches to interact with potential students and provide them a thorough sense of campus life and academic offerings in today's quickly changing digital landscape. Virtual campus tours have become an innovative way for people to remotely tour campuses, providing a dynamic and immersive experience. The effectiveness of the user experience is crucial in luring and retaining prospective students as these tours gain popularity. This background section will give an overview of how virtual campus tours have changed over time, as well as how HCI and UI design have become increasingly significant in influencing these experiences. In the digital age, the idea of virtual campus tours is a logical development that fits with the general trend of using technology to make knowledge and experiences that were previously confined by geographic boundaries available. Prospective students depended primarily on static websites and printed materials in the early days of the internet to learn about academic institutions. Although these tools offered crucial knowledge, they

lacked the hands-on component that an actual visit could bring. The first attempts at virtual campus tours consisted of static photos and straightforward verbal descriptions that provided a basic understanding of the school layout. However, these early initiatives failed to adequately engage consumers or deliver an immersive experience. As technology developed, the incorporation of multimedia components including 360-degree photographs, films, and interactive maps started to influence the virtual tour industry.

2.2 Significance of Research

The development of virtual campus tours was hastened by the introduction of smartphones and mobile applications. The potential of specialised applications to give tours catered to the preferences and interests of certain users was recognised by educational institutions. These apps provided prospective students with immediate, on-demand access to campus exploration, allowing them to find academic buildings, housing halls, recreational areas, and other important places of interest. As virtual campus tours developed, it became clear that the effectiveness of the user experience was a key determinant. A prospective student's decision to apply to or enrol in a school might be affected by an interesting and educational tour that makes a lasting impression. The contribution points that can be infered are:

- UI Design principles: A successful user interface depends on being clear and effectively communicating the intended function. Users are guaranteed to understand the purpose of each component with ease because of the combination of clear and simple wording and recognizable graphic components. An interface that is cohesive and pleasant is supported by consistency. Predictability and user confidence are promoted by using a unified design style with constant colors, typefaces, and layouts. Users are guided by a clearly defined visual hierarchy, highlighting importance without overwhelming them by leveraging visual cues like size, color, and location.
- Feedback: A key element of UI design that bridges the gap between user activity and system reaction is feedback. It is a dynamic loop that keeps people informed about the results of their choices, encourages interaction, and makes for a simpler and more user-friendly experience. The UI should provide immediate feedback. The system should react quickly when users engage with an element, whether they are clicking a button or filling out a form. Users are immediately informed that their activity has been acknowledged, eliminating any uncertainty or confusion.
- User-Centered Design Approaches: User-Centered Design (UCD) is a human first methodology that prioritizes user experiences and needs throughout the design process. It is an approach that puts empathy, teamwork, and continual improvement first. Designers that adopt UCD methodologies produce products and user interfaces that deeply connect with customers, improving usability and satisfaction. A key element of UCD is usability testing, which offers clear insights into how people interact with a product. Designers can pinpoint problems, streamline processes, and improve the interface's general usability by analyzing user behavior and gathering feedback.
- Technological Integration in UI Design: Modern User Interface (UI) design relies heavily on technological integration, and Figma is a potent ally in this regard. Figma, a collaborative design platform, revolutionizes how designers conceptualize, iterate,

and create interfaces by seamlessly integrating technology into the UI design process. The adaptability of Figma includes creating for several platforms. Within a single platform, designers may produce designs for desktop, mobile, and online applications. A consistent design language is ensured across numerous user touchpoints thanks to this cross platform feature.

– Educational Impact: Beyond the digital canvas, user interface (UI) design has a significant impact on how people interact with educational institutions in the dynamic world of education. The UI design acts as a link between the online and offline experiences on campus. Prospective students and guests can virtually explore the campus to learn more about its architecture, facilities, and environment using user-friendly interfaces. These interfaces' layout acts as a portal, bridging distances and providing a window into the learning environment.

UI design and HCI come into play in this situation. HCI concepts are crucial for comprehending user behavior and preferences in the context of virtual campus tours, while UI design principles direct the development of interfaces that enable simple navigation and information retrieval. A branch of human computer interaction (HCI), UI design focuses on the visual and interactive elements of user interfaces with the goal of producing interfaces that are user-friendly, beautiful, and effective. Together, these fields make sure that virtual campus tours give visitors a memorable and satisfying experience in addition to information.

2.3 HCI and UI Design's Function in Virtual Campus Tours

The success of virtual campus tours depends on HCI and UI design in a number of ways:

(1) User-Centric Design: HCI principles stress the value of creating technological solutions that are in line with user requirements and expectations. Understanding the objectives, driving forces, and preferences of potential students is critical when designing a user-centric campus tour.
(2) Usability and Accessibility: The design of user interfaces that are intuitive to use and accessible to a wide range of users, including those with disabilities, is influenced by HCI research. The key to effective virtual campus tours is making sure that they are open to everyone and simple to use.
(3) Engagement and Immersion: UI design concepts assist in creating interactive interfaces that captivate users and immerse them in the tour. User engagement is aided by features like simple navigational controls, multimedia integration, and compelling images.
(4) Iterative Design Process: HCI promotes an iterative design approach where usability testing and user input drive ongoing improvements. The tour interface must be improved iteratively in order to match user expectations and keep up with changing technology.

2.4 Augmented Reality (AR)

– **Working**

Modern technology known as AR has great potential for improving HCI. It signifies a paradigm shift in how people engage with both the physical and digital worlds. With AR, computer-generated elements like images, audio, and haptic feedback are seamlessly integrated into the user's actual environment in realtime. By combining the physical and digital worlds, a composite environment is produced that gives users a more engaging and dynamic experience. Overlaying virtual items or data onto the user's view of reality is the fundamental idea behind augmented reality. Depending on the sensory modalities used, this overlay may be visual, aural, or tactile. AR interfaces facilitate more natural and intuitive interactions between people and computers in the context of HCI. Consider the following important factors when investigating AR in HCI research: (a) Visualisation (b) Interaction (c) Context Awareness (d) Usability and Accessibility (e) Cognitive Load.

– **Types of Tracking**

(1) Marker Based Tracking: A key method in AR that enables the exact overlay of digital content on the physical world is marker-based tracking. Fiducial marker tracking, sometimes referred to as marker-based tracking, is the process of identifying and following predefined markers that have been put in the user's field of view. The camera captures an image and scans it to identify a recognized marker (such as a 2D image or a 3D object), which, upon successful detection, renders an augmented scene (such as an object, text, or animation) at the marker's pose [1]. These indicators may appear as 2D barcodes, QR codes, icons, or specially created patterns. These markers are identified and tracked in real-time by the AR application when users point their AR-enabled device, such as a smartphone or AR glasses, towards them.
(2) Marker-less Tracking:-Markerless tracking, sometimes referred to as SLAM or markerless AR tracking, is a core technology in AR that enables digital content to be superimposed onto the physical world without the use of predetermined markers or visual cues.Markerless tracking uses the device's sensors, such as the cameras, gyroscopes, accelerometers, and sometimes depth sensors, to analyse and map the user's environment in real-time rather than depending on markers. With the use of this technique, AR applications may precisely place virtual objects within the user's actual surroundings. The goal of markerless tracking is to offer a more adaptable and dynamic AR experience. Markerless tracking allows users to freely explore their surroundings without having to point their devices at specified markers, and the AR application reacts to the environment in real time. Finally, the development of virtual campus tours emphasizes the expanding significance of HCI and UI design in the field of higher education. These fields have a crucial role in determining the user experience, which eventually affects prospective students' decisions and aids educational institutions in developing more inclusive and effective methods of connecting with their audiences. Many colleges have represented their campus in digital format so many student's can have an idea how college campus looks [2].

The goal of this study article is to delve deeper into these issues by offering perceptions on the creation and assessment of user interfaces for applications including virtual

campus tours. In order to help meet the unmet needs of prospective students who seek a campus tour, the goal of this project is to construct an interactive multimedia, tablet-based digital tour. Tim et al. [3], came up with the first concept of using technology as a tool for campus tours. He discussed the advantages of using technology to explore the campus in 2004. Since the early adopters, technology has advanced significantly, and schools and institutions are now starting to embrace new technologies like augmented reality. AR has been successfully growing nowadays in many industries, mainly in gaming and entertainment, education and navigation [4]. Technologies that overlay digital visuals or information over actual items are referred to as augmented reality. Many universities are utilizing this technology as a mobile campus touring application to make the visitors become familiar with the campus through the self-guided devices [5].

3 Methodology

3.1 The User-Centered Approach to UI Design

Qualitative research conducted at Midwestern State University confirmed that the greatest impact on the decision-making process of a prospective college student is a campus visit (Brown, 2010) [3]. The study conducted by Jill Kruger Brown of the University of Nebraska-Lincoln found that if a student participated in a college tour, they were nearly two times more likely to attend the university than those who did not attend a college tour. The study suggests that additional effort must be placed on the campus tour experience, as the campus visit has the greatest impact of the prospective student's ultimate decision to enroll [6]. The UI, which connects people and computers, is a crucial element in the field of HCI. It includes the look, feel, and interaction components that people interact with when using software programs, websites, or digital gadgets [6]. The user interface, or UI, is essentially how humans convey their intentions to computers, get feedback, and carry out tasks in a digital environment. Key Components of UI in HCI are:

- Visual Elements: The visual elements of an interface, such as buttons, icons, text, images, and layout, are included in UI design. These components ought to be aesthetically pleasing and arranged to help users navigate their interactions [10].
- Navigation: UI design takes user flow across the system into account. Creating a simple and intuitive user experience, entails the placement of menus, links, tabs, and other navigational components.
- Interactivity: Interactivity is a key component of UI. This entails creating the interface so that people can interact with it by clicking, dragging, typing, or utilizing touch gestures. Additionally, it includes the response users get from interacting with objects, such as buttons that change color when clicked.
- Information Presentation: How information is presented to users is the responsibility of UI designers. To effectively communicate information, this includes selecting typefaces, text sizes, color palettes, and data visualization approaches.
- Accessibility: HCI researchers and designers need to make sure that users with different needs, such as those with disabilities, can utilize the UI. This requires taking into account things like screen readers for users who are blind and keyboard navigation for those who are mouse-impaired.

- Uniformity: A well-designed user interface keeps uniformity across the board. Users can comprehend and operate the system more readily when elements are placed consistently and when language and design patterns are used consistently.
- User-Centered Design: User-centered design should be used at all times. Through techniques like usability testing and user interviews, HCI research places a strong emphasis on understanding user needs, behaviors, and preferences. This knowledge directs UI choices to produce user interfaces that meet user expectations.
- Usability: In HCI, usability is of utmost importance. The goal of UI designers is to produce user interfaces that are simple to understand and effective to use. Testing and evaluating usability aid in locating and fixing usability problems.
- The proliferation of different devices and screen sizes makes responsive UI design crucial. This method makes sure that the interface adapts and works well on a variety of gadgets, including PCs, smartphones, and tablets [15].
- Feedback and Error Handling: The user interface (UI) must give users clear feedback regarding the results of their activities. Additionally, it should make error recovery easier by instructing users on how to fix problems.
- Emotional design acknowledges the significance of user emotions and experiences in HCI. UI design can have an impact on users' overall pleasure and engagement with a system by affecting how they feel while engaging with it.

3.2 Propoesd Methodology: FIGMA

The use of Figma as a useful tool for creating the front end of mobile and online applications is examined in this research study. A collaborative environment for designers and developers to develop and iterate on user interfaces (UIs) [7] and user experiences (UX) is provided by the cloud-based design and prototype platform Figma. This essay explores how Figma can improve front-end development, facilitate communication between designers and developers, and ultimately help app projects succeed [14].

3.3 Utilization of Proposed Methodology

For both mobile and web applications, the creation of a front end is essential for user engagement and happiness. Every program must have an intuitive and visually appealing UI/UX to succeed [8]. The popularity of Figma, a cloud-based design and prototype tool, has grown recently due to its capacity to promote cooperation between design and development teams. This novel approach examines the advantages of using Figma for front-end development and how it affects the success of the app as a whole.

1. Overview

   Figma is a flexible platform that combines collaboration, design, and prototype tools in a single cloud-based environment. It provides the following crucial features:

   - Real-time collaboration is made simple for designers and developers to work together and iterate on designs in real time by allowing multiple team members to work on the same project at once. Introduction For both mobile and web applications, the creation of a front end is essential for user engagement and happiness. Every program must have an intuitive and visually appealing UI/UX to succeed.

The popularity of Figma, a cloud-based design and prototype tool, has grown recently due to its capacity to promote cooperation between design and development teams. This essay examines the advantages of using Figma for front-end development and how it affects the success of the app as a whole.

- Design and Prototyping Tools: Figma is a complete front-end development solution since it offers a wide range of design tools, including vector editing, responsive design features, and interactive prototyping capabilities.
- Version control: Figma keeps track of design changes and enables teams to go back to earlier versions if necessary.
- Developer-Friendly Features: To make the transition from design to development easier, Figma provides developer handoff capabilities including CSS and code generation.

2. Streamlined Collaboration: The collaborative capabilities of Figma are quite useful for front-end development. [11] According to research, effective communication between designers and developers is essential for the creation of successful apps. Figma makes collaboration easier by:
   (a) Providing the platform with real-time feedback and communication.
   (b) Allowing designers to establish design systems that developers may access, assuring uniformity in UI elements and styles.
   (c) Supporting version control, which makes it easier to keep track of changes and fix design flaws.
3. Enhanced Design Iteration
   (a) Frontend developers can quickly create and iterate on UI/UX designs thanks to Figma's design and prototype tools. Faster app development cycles and shorter time to market as a result [12].
   (b) Better user experiences as a result of iterative design adjustments.
   (c) Lessened development errors by enabling developers to see the design goals.
4. Developer Handoff

   The ability of Figma to produce CSS and code snippets directly from design files is one of its unique capabilities [13]. This streamlines the transfer of responsibility from designers to developers.
   (a) Within Figma, developers have access to design assets and coding requirements.
   (b) Code generation makes sure that the front end closely matches the design, cutting down on implementation mistakes.
   (c) more effective frontend development that produces a professional app more quickly.
5. Case Studies This section provides actual case studies of frontend development projects that were successful and used Figma as a key tool. In the case studies, we emphasize:
   (a) How Figma facilitated effective teamwork
   (b) How Figma affects the quantity and quality of design iterations
   (c) Experiences with developer handoffs and how they affect front-end development

4 Results and Discussion

4.1 Observations

The user-friendly interface of the campus tour app created on the Figma platform is intended to improve the campus experience. The app will start with our college's emblem and a sizable "Start" button. On the following screen, you can log in if you already have an account. It has areas for your first and last names, phone number, password, and email address. While entering, you have the option of hiding or showing your password. m1 There is a "Login" button for new users that leads to a page where they can register an account. Along with the request for your full name, contact information, and password, it also gives the option to hide the password. The keyboard appears whenever you need to type something. A loading screen featuring our college's logo will appear when you log in, and you'll then be sent to a scrollable menu. There are numerous choices on this menu, including "Start Tour," "Campus Map," "Quizzes," and others. Each choice has a selector button and a "like" button in the top right corner. There is a bar at the bottom of the screen that remains stationary as you scroll. It provides choices for the menu, your profile, and a search feature. Our campus's layout is shown on a map if you select "Campus Map," which will guide you to various locations. To return to the main menu, click the button in the upper left corner. Whether you're a prospective student or just visiting, these elements work together to produce an app that is intended to give you a thorough and educational campus tour experience. We'll go into greater depth about how we created the user interface, put the app through user testing, and solicited feedback in the parts that follow choices on this menu, including "Start Tour," "Campus Map," "Quizzes," and others. Each choice has a selector button and a "like" button in the top right corner. There is a bar at the bottom of the screen that remains stationary as you scroll. It provides choices for the menu, your profile, and a search feature. Our campus's layout is shown on a map if you select "Campus Map," which will guide you to various locations. To return to the main menu, click the button in the upper left corner. Whether you're a prospective student or just visiting, these elements work together to produce an app that is intended to give you a thorough and educational campus tour experience. We'll go into greater depth about how we created the user interface, put the app through user testing, and solicited feedback in the parts that follow (Fig. 1).

4.2 Discussion

Our painstakingly created campus tour app's origins are a narrative of innovation and putting users first in the field of higher education. It was created using the Figma design platform. It becomes clear as we examine the numerous app windows and interfaces that fusing design, technology, and the user experience can fundamentally alter how individuals interact with and view their university. We'll delve into the challenges, revelations, and effects of our campus tour app project during this conversation. We'll examine what made our project successful, how the user experience was skillfully weaved into every aspect of the app, the technological underpinnings that kept everything running smoothly, and the crucial role that Figma played in realizing our vision. Beyond the confines of our project, we'll also take into account how it might influence the overall



Fig. 1. Home Screen and User Authentication Page in the Campus Tour Application.

campus environment and the state of digital innovation in education. Our campus tour app serves as an illustration of how form and function may combine to change how people interact with their educational environment.

Figure 2 displays on selection of "Information about Departments," you will be taken to a page with a list of department names and a picture at the top. Any department can be selected to open a page with a picture and an overview specific to that department. By pressing the "Back" button in the upper left, you may easily return to the previous page. Whether you're a prospective student or just visiting, these elements work together to produce an app that is intended to give you a thorough and educational campus tour experience. We'll go into greater depth about how we created the user interface, put the app through user testing, and solicited feedback in the parts that follow.displays "Quiz" menu item in the app. It's a great opportunity for students to learn more about the university, not simply about meeting places. These points are more than simply abstract values; they may be spent in facilities like the cafeteria, making the experience of being on campus somewhat competitive. Making our software accessible to all pupils is one of its main design tenets. No matter how tech-savvy they are, everyone can use it because of its usability.

We also included a feedback component in the app, providing students and the university with a direct line of communication. Students can comment on the usability of the app and suggest additional features. This strengthens the bond between students and the school, enhancing the learning environment as a whole.

In a nutshell, the way educational institutions engage with students has changed as a result of our campus tour app. It's evidence of the ability of user interface design to transform the educational process and improve the campus experience.

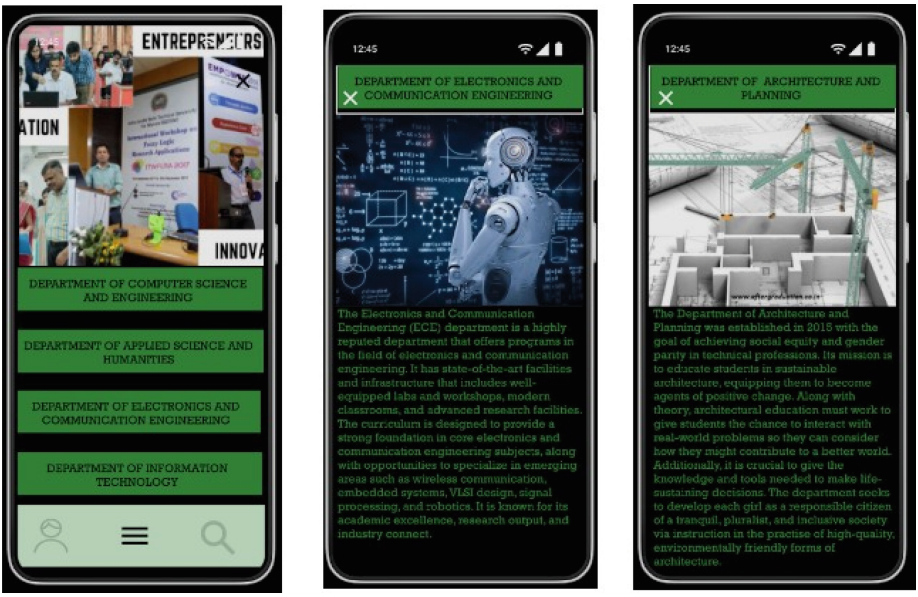


Fig. 2. Exploring Department Information – Users obtain in-depth details on academic departments, promoting a thorough grasp of what the university offers.

AR is a fascinating new area of study and development [9] By bridging the gap between the digital and physical worlds, it has the potential to produce more organic and engaging user experiences. By tackling the issues with interface design, interaction modalities, context awareness, usability, and accessibility to realize the full potential of this game-changing technology, HCI researchers play a crucial role in promoting AR technology. Including AR in various aspects of our app for example virtual tours and student testimonies will enrich the experience of students and help them have a more fun-filled and exciting experience.

In conclusion, our experience utilizing Figma to create a campus tour app has demonstrated the transformative power of user interface design in higher education while also producing a concrete result. We have seen how innovation and user-centricity may come together to re-imagine how students and visitors interact with campus settings. The app's success in fusing design, technology, and user experience with one another successfully serves as a model for educational institutions looking to improve their services. Looking ahead, we see more chances for the app to develop, change, and keep improving the campus experience. Our campus tour app is a tribute to the limitless opportunities that exist at the nexus of technology and education in this era of digital innovation. It is more than just an app; it is a representation of our dedication to developing useful, user-centric solutions that improve the educational process. As we advance, we'll keep working to push the limits of design and technology to produce engaging experiences that inspire, inform, and connect. The journey goes on, and we excitedly look forward to the chapters that lie ahead in the constantly changing higher education scene.

4.3 Results

In this tour, we experience the creation of a campus tour app using the Figma design platform has been a journey characterized by creativity, user-centered design, and a dedication to enhancing the campus experience. We have examined the conception, development, and implications of our campus tour app in this research paper, shedding light on the revolutionary potential of UI design in the field of higher education. Our app, created to satisfy the many requirements of both prospective students and visitors, has successfully tapped into the power of technology to provide a fun, educational, and engaging way to explore our campus. Upon reflection, the following significant lessons become clear Firstly, User Centric Design shows a user-friendly, intuitive, and accessible app is the outcome of our emphasis on user-centric design concepts, usability testing, and feedback loops. The advantages of placing the user at the center of the design process have been discussed. Secondly, the technological Innovation app's technological execution, supported by Figma's design capabilities, illustrates how technology is changing the face of education. It has made it possible for us to develop a dynamic and engaging tool that improves the school environment. Thirdly, thanks to the addition of features Engagement and Gamification based gained knowledge like quizzes and gamified aspects, which have also made the campus tour experience entertaining and engaging. It has created fresh opportunities for communication and education. Lastly, the future Potential through which we are looking ahead, there is a lot of room for growth and improvement for our campus tour app. It is a long-lasting asset for the organization since it can adapt to shifting user preferences and technology improvements. Further, enhanced User Engagement in AR has the ability to give users of the app a more engaging and interactive experience. As the app is based on campus tours incorporating AR into it will enrich the user experience and help them connect more with the campus tour. We can design appealing features that catch user attention and promote sustained app engagement by superimposing digital information onto the actual environment. The main focus of our contributions are-

- Contextual Information: AR can be used to offer information that is pertinent to the user's current location or the objects they are interacting with. This can be used in student testimonies. As compared to reading a testimony the user can be conveyed that information by using marker-based AR. In each department, we can have alumni of the department share their experiences and give advice to freshmen by recording a video that can be incorporated into marker-based AR. This might improve the app's usability and usefulness. This feature would incorporate interaction within the app.
- User-Centered Design: The significance of user-centered design principles has been highlighted by our study on UI design. This idea is in line with integrating AR into the UI since it allows users to customize the app's design to their specific needs and tastes. Every student can sign in to the app and have a customised tour of their department helping them in getting familiar with the department building as well as with the faculty. They can have seniors from their own branch share their experiences and help the students have a smoother experience in college.

Our main aims for including AR in the UI of the app are (1) Design Research: Work with UI/UX designers to investigate augmented reality design ideas that harmoniously combine digital and physical components while upholding a logical and aesthetically

beautiful interface. (2) Development and testing: Work to include AR capabilities into the app while thoroughly evaluating its usability and functionality. (3) User Feedback and Iteration: To improve AR features depending on user preferences and needs, gather user feedback through usability testing and surveys.

We'll take a methodical approach in order to accomplish our goals: User surveys and interviews: (1) We learn more about potential users' needs, wants, and preferences for augmented reality features by surveying and interviewing them. Creating interactive AR prototypes in collaboration with UI/UX designers will help us to visualize the desired user experience. (2) Development: We will use AR development platforms and frameworks to create AR functionalities, assuring interoperability with a range of AR-capable devices. (3) Usability Testing: In order to assess the efficacy and user-friendliness of AR features, usability testing will be carried out with the participation of sample users. (4) Feedback integration: User feedback will be considered as we iteratively improve and optimize the AR components. Future development should head in the interesting and creative route of integrating augmented reality into the user interface of the app. It is consistent with our dedication to providing innovative solutions that raise user engagement and happiness. We think that the incorporation of AR will be a valuable expansion of our work because campus tours will be much more fun with AR and give the freshmen a much-needed boost and a sense of comfort. Future research and development phases will examine the full potential of this cutting-edge technology in relation to our app.

5 Conclusion

Our campus tour app is a technological advancement in the constantly changing higher education scene, but it also demonstrates our dedication to improving the educational experience. It provides as evidence of the influence of cuttingedge UI design and technology in influencing how people interact with and view their educational surroundings. This study paper comes to an end with a sense of success and a vision for the future in which user-centric design and digital innovation will continue to alter and improve the campus experience for all.

6 Future Work

This section describes our justification, goals, and suggested methods for incorporating augmented reality into the UI of the app. Additionally, there is a promising way to improve the usability and usefulness of our software as the integration of AR technologies develops further. We envision an interesting possibility to add AR components into the app's interface in future versions, even if the current version of our programme largely focuses on a standard UI design.

Acknowledgements. We thank Prof. Ashwni Kumar, Indira Gandhi Delhi Technical University for Women, Delhi, India, for their useful discussions and constant support.

References

1. Jindal, P., Park, A.J., Hwang, E.: Augmented reality campus exploration applicationincorporating equity, diversity, and inclusion. In: 15th International Conference on Computer and Automation Engineering (2023)
2. Shah, H., Tupe, V., Rathod, A., Shaikh, S., Uke, N.: A progressive web app for virtual campus tour. In: International Conference on Computing, Communication and Green Engineering (2021)
3. Mendolia-Moore, F.T.A.: The College Path: A Virtual Tour (2018)
4. Nordin, N., Markom, M.A., Suhaimi, F.A.: A web-based campus navigation system with mobile augmented reality intervention. In: Asian Conference on Intelligent Computing and Data Sciences (ACIDS) (2021)
5. Andri, C., Alkawaz, M.H., Sallow, A.B.: Adoption of mobile augmented reality as a campus tour application. Int. J. Eng. Technol. (2018)
6. Pacheco, J., Garbatov, S., Goulão, M.: Improving collaboration efficiency between UX/UI designers and developers in a low-code platform. In: ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C) (2021)
7. Patil, S., Gaikwad, G., Hiran, S., Ikhar, A., Jadhav, H.: metaAR– AR/XR shopping app using unity. In: International Conference for Advancement in Technology (ICONAT) (2023)
8. Alomari, H.W., Ramasamy, V., Kiper, J.D., Potvin, G.: A User Interface (UI) and User eXperience (UX) evaluation framework for cyber learning environments in computer science and software engineering education (2020)
9. O'Shea, P.M., Elliott, J.B.: Augmented Reality in Education: An Exploration and Analysis of Currently Available Educational Apps. Springer International Publishing Switzerld (2016)
10. Bharti, S., Kumar, A.: Review of approximate computing in image processing applications. In: 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST), Delhi, India, pp. 1–6 (2022). https://doi.org/10.1109/AIST55798.2022.10065214
11. Bharti, S., Kumar, A.: Low power-area based composite 6T-8T SRAM for soft computing applications. In: 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, pp. 1–6 (2022). https://doi.org/10.1109/ICPC2T53885.2022.9777036
12. Bharti, S., Kumar, A.: Rooting Of PPN Technique on 8T-SRAM for low power multimedia applications. In: 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–5 (2022). https://doi.org/10.1109/ICCCI54379.2022.9740955
13. Nemati, S.H.H., Eslami, N., Moaiyeri, M.H.: Computing in memory using doubled STT-MRAM with the application of binarized neural networks. IEEE Magn. Lett. **14**, 1–5 (2023)
14. Chen, Y., Mu, J., Kim, H., Lu, L., Kim, T.T.-H.: BP-SCIM: a reconfigurable 8T SRAM macro for bit-parallel searching and computing in-memory. IEEE Trans. Circuits Syst. I Regul. Pap. **70**(5), 2016–2027 (2023). https://doi.org/10.1109/TCSI.2023.3240303
15. Marinelli, T., Pérez, J.I.G., Tenllado, C., Catthoor, F.: COMPAD: a heterogeneous cache-scratchpad CPU architecture with data layout compaction for embedded loop-dominated applications. J. Syst. Archit. **145**, 103022 (2023)

Predicting Startup Exit Strategies with Ensemble Machine Learning

M. Deekshitha Reddy[1], Geetika Vadali[1], Garima Jaiswal[2(✉)], Ritu Rani[3(✉)], and Amita Dev[4]

[1] Department of AI and DS, Indira Gandhi Delhi Technical University for Women, Delhi, India

[2] Department of Computer Science, Bennett University, Greater Noida, India
garima121@gmail.com

[3] Department of Electronics and Communication Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India
ritujangra00@gmail.com

[4] Center of Excellence-Artificial Intelligence, Indira Gandhi Delhi Technical University for Women, Delhi, India
vc@igdtuw.ac.in

Abstract. In the current landscape, Venture capitalists are opting for more data-driven decision-making to achieve high returns while mitigating risks. All VCs seek a home run in their portfolio, for which it is essential to select the optimal exit. The process can be smoothed with the prediction of the type of exit- IPO, M&A, or the next funding round. Further,it is also to gauge which type of exit yields maximum profit for the particular climate. In this paper, through the utilisation of six machine learning models - Elastic Net, SVR, Kernel Ridge, Gradient Boost, XGBoost, and Random Forest, trained on factual data sourced from Crunchbase. The machine learning models are evaluated on the basis of 2 metrics- RMSLE and R-squared Score. We propose an ensemble learning model with Root Mean Square Logarithmic Error (RMSLE) score of 0.29 and R-squared score of 0.76. This weight-averaged ensemble has been validated to predict the optimal exit for a startup under six broad labels - IPO, Acquisition, Leveraged Buyout, Merger, Acquihire, and Management Buyout.

Keywords: Startup Exit Prediction · Ensemble Learning · Machine Learning

1 Introduction

The startup ecosystem experienced an exponential rise in funding prospects after 2019, with the COVID-19 outbreak, primarily because of the reduced interest rates. The majority of central banks throughout the world cut interest rates, and investors now had access to easy money. And the startup's funding saw a hockey stick growth. However, this trend persisted for less than three years; central banks are already raising interest rates to levels seen before the outbreak. As in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 407–416, 2025.
https://doi.org/10.1007/978-3-031-75167-7_32

every scenario, there are outliers; a few countries, like China, are still lowering their interest rates. Indeed, this has impacted the easy money for VCs and, hence, the capital influx in the startup ecosystem. This ongoing funding winter can be attributed to a culmination of factors apart from the low interest rates: the collapse of major US banks like Silicon Valley Bank, the geopolitical issues stemming from the Russia-Ukraine war and other macroeconomic uncertainties. In the current situation, venture capitalists must devise the most effective exit plan to achieve a substantial return multiple.

Venture Capitalists can opt for different types of exits that are available for that particular startup. Exit methods that are common and frequently employed include an initial public offering (IPO), acquisition by or merger with another company, secondary buyout, management buyout, and pursuing the next round of investment. Among other options, liquidation or recapitalisation are less likely to occur, given they only arise when the startup cannot meet its financial obligations or want to reposition the financial structure. Royalty financing is a less popular exit option too, as startup founders prefer to reinvest the revenue generated in the operational runway.

Venture capitalists cannot depend on their intuition alone to assess a startup's exit before making an investment. Over the last two decades, researchers have made substantial contributions by creating various machine learning and deep learning models for evaluating the likelihood of a startup's success. It could also be noticed that very few models have been tested to predict the type of exit.

With access to startup data from platforms such as CrunchBase, Tracxn, etc., refined forecasting models can be designed and trained to predict the appropriate exit for any company and investor. Further, with feature engineering, factors that signal or indicate any type of exit can be categorised and can be used in data-driven decision-making. Xiang et al. 2012 categorised the data features into two types: factual and non-factual features [20]. Factual features include basic features like age, number of milestones, etc., financial features like number of funding rounds, investment size, etc. and managerial features like founders, their experience, etc. On the other hand, non-factual features include topic features like announcements, social media following, etc. Even with such sophisticated data providers, there always exists a gap in the data as most startups keep their information private. Ensemble learning poses an apt option for such a dataset, considering ensemble learning methods can handle the missing data. In this paper, we test different machine learning models trained on factual data and compare them with an ensemble learning model.

Given that there are different kinds of exit options that a startup can opt for with respect to a lot of internal as well as external factors and conditions. This research attempts to explore this relationship between factors like the kind of industry the startup belongs in, the ownership, acquisition or IPO status of the startup and the overall demographic and success of the company.

For our use-case of predicting how a particular startup will exit with respect to its functioning, success and external climate, we have used concepts of regres-

sion [2] and classification adapted from traditional machine learning models like usage of regularisation in terms of L1 and L2, or decision trees in Random Forest Classifier, or more modern models like XGBoost extensively used in classification problems (Fig. 1).

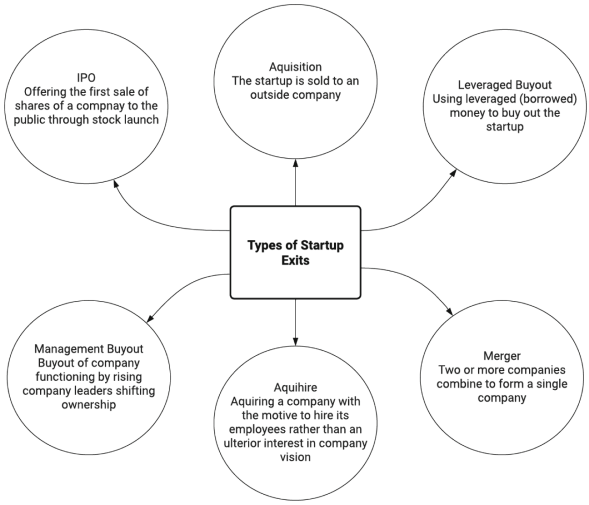


Fig. 1. The various types of Startup Exit Options considered

2 Related Work

Machine Learning models have grown to become a popular choice for predicting startup's success. Initially, primitive models like logistic regression [15] and linear regression models trained on factual and non factual have been employed for the same. Krishna et al. 2016 [14], showed the use of Random Forest, ADTrees and Bayesian Networks to increase accuracy. Notably,the Random Forest regressor proved to be highly accurate with 96.91% accuracy in a study conducted by chhibber et al. 2022 [8]. In order to build more reliable models, factoring in feature importance becomes crucial. Kim et al. 2023 highlight the significance of weighted feature modeling [13].

In recent studies, Deep Learning and Ensemble learning models are showing promising results. Ragothaman et al. 2003, proposed an expert system that utilises uncertainty representation along with inductive logic programming (IXL) induced rules to predict a company's acquisition [17]. In a more recent study, Bangdiwala at al. 2021 implemented MLP Neural networks along with other machine learning models and found MLP Neural network to give similar accuracy as the other models [4]. Furthermore, in a study conducted by Cao et al. 2022, various deep learning models were employed for the identification of Unicorn companies [6]. Unicorns, in this context, refer to privately held startups valued at

over one billion dollars. Deep learning models can be refined by the introduction of time series analysis. Ferrati et al. 2021 implemented a binary deep learning classifier for time series analysis with a recall of 93% [11].

Earlier Ensemble learning models have been used for the predictions in the venture capital ecosystem. Wei et al. 2009 built an ensemble classifier based on patents of companies to predict merger and acquisition [19]. Furthermore, Abhinand et al. 2022, deployed a stacked ensemble model to predict the success of a startup in the Indian Ecosystem [1].

The current research indirectly helps in developing an optimal exit strategy, but direct exit type prediction is crucial for any VC. Hence, this paper explores various machine-learning models and compares them with an ensemble learning model for the same premise.

3 Proposed Methodology

In this paper, we use Regression models instead of Classification models to predict the exit strategy on account of the Fine-Grained Labels in the dataset. Further, these predictions are categorised into six options- IPO, Acquisition, Leveraged Buyout, Merger, Acquihire and Management Buyout using approximation. Moreover, the models are tested using RMSLE (Root Mean Squared Logarithmic Error) as the evaluation metric. As we do not require the precise numeric value but the precise categorisation into six broad labels.

The following section explains the process employed to transform the data to the required form. The next section illustrates the implementation of the six machine learning models- Elastic Net, SVR, Kernel Ridge, Gradient Boost, XGBoost, and Random Forest and the assembling of ensemble model.

3.1 Data Extraction and Preprocessing

The data used for this project is accessed through the startup data provider Crunchbase. The US dataset from 2000–2020 is used for training the models. A total of 5201 data points were utilised, excluding the rows with null values. A total of 32 independent features were extracted from Crunchbase and created with data preprocessing. The features- Startup Name, Total Funding, Description, Founding Date, Exit Date, Headquarters city, IPO status, Acquisition type- Acquisition, Leveraged Buyout, Merger, Acquihire, Management Buyout were directly acquired from Crunchbase.

Further, the Headquarters city was used to find latitude and longitude using the geopy library of Python. The same feature is also used to check if the startup is in New York, California, Texas, Pennsylvania, Illinois, Michigan or Ohio. Furthermore, Based on the description of the startup, a new feature, i.e., Tech_or_NonTech has been created to indicate if the startup is backed by heavy use of technology. With the same description, the startups are differentiated into 17 industries, namely- Technology, Healthcare, Finance, Hospitality and Entertainment, Retail, Travel, Education, Automotive, Food, Sports, Aerospace, Oil

and Energy, Manufacturing and Real Estate, Services, Biotechnology, Transportation, Marketing.

Table 1. Data Description

Feature	Data Description
Total Funding	Total capital received by the startup till exit date in numeric form
IPO Status	Startup is public or private
Acquisition	Type of acquisition as categorized above
Tech_or_NonTech	Utilization of Technology
Age	Age of the startup during exit
Lat	Latitude extracted from geopy
Long	Longitude extracted from geopy
is_[State]	State where the headquarters is located, 1 indicates if it is in [State] otherwise, 0
is_[Industry]	1 indicates if it belongs to [Industry] otherwise, 0
Exit (Target Variable)	Six types of exit: IPO, 'Acquisition', 'Leveraged Buyout', 'Merger', 'Acquihire', 'Management Buyout'

3.2 Models Implemented

1. Elastic Net Regression. Elastic net introduces penalisation in linear regression by combining both L1 and L2 penalties in the training phase [21]. L1 penalty is based on the sum of the absolute coefficient values, and it minimizes the magnitude of these coefficients. L2 penalty is based on the sum of squared coefficient values. A difference between L1 and L2 penalty is L2 doesn't allow the value of any coefficient to become zero. Hence, L1 gives the option of removing some predictors from a model while L2 doesn't. Elastic net often performs better than Lasso [18].

In Elastic Net, optimise

$$\min_{\beta} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - X_i^T \beta)^2 + \alpha\rho \|\beta\|_1 + \frac{\alpha(1-\rho)}{2} \|\beta\|_2^2 \right)$$

where β denotes the vector of coefficients whose optimal value is to be found, y_i is the target value and X_i denotes the feature vector for the i-th sample. N is the number of samples. The hyperparameters used are α which is the regularization strength parameter and ρ controls the trade-off between L1 and L2 regularizations.

2. Support Vector Regression. SVR is considered a non-parametric technique as it makes use of kernel functions [3]. A margin of tolerance is added denoted by epsilon that allows the error to stay in a specific range of values. The main goal of SVR is to minimize error by individualising the plane that lower bounds the margin [9] i.e., find a function that deviates from y_n by a value no greater than ϵ for training point x and subsequently is flat. SVR lets us choose the error tolerance we can let exist in our training. The regression equation for finding the hyperplane approximating the target values typically looks like:

$$f(x) = \langle w, x \rangle + b$$

where w is the weight vector, x the input data and b is the bias term.

3. Kernel Ridge Regression. Ridge regression is used in data with multicollinearity by employing a model tuning method. L2 regularization is used here. Ridge regression seeks to solve the unbiased-ness of the least Kernel Regression regression models are also non-parametric and capable of quantifying the relationship between the predictor variables and the outputs. These relationships can be linear or non-linear in nature and the outcome of the regression might be sensitive to the hyperparameters chosen [10]. There are eight kernel functions and the set of hyperparameters which will be active at a time is determined by the selected function. The hyperparameters in Kernel Ridge regression model include alpha for ridge regularization and all other hyperparameters may be specific to selected kernel function. If more than one kernel functions is selected, a grid search is performed for the cross-validation and this outputs the best fitting model for given data. The kernel functions included in this model's specification are Additive-CHI2, CHI2, Cosine, Laplacian, Linear, Polynomial, RBF and Sigmoid. Kernel ridge regression is generally expressed as a dual problem for a kernel function K. The regression function looks like:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i)$$

where α_i are the dual coefficients for x_i training samples.

4. Gradient Boosting. Hypothesis boosting represents the method of filtering inputs on the basis on which can be handled by weak learners and other more difficult observations. The idea is to minimise model loss by introducing weak learners with a gradient descent-like algorithm [12]. Gradient boosting starts with a loss function that needs to be optimised, a weak learner that will make the predictions from given data and another model that will add weak learners such that the loss function is minimised. Gradient boosting is like an ensemble of numerous weak learner models and combining them to get better predictions. Predictions are improved by focusing on residuals, which are the prediction errors. The goal is to minimise the value of residual so we can get

closer to the true value as much as possible [16]. The regression function comprises numerous weak models which are typically decision trees. A generalistic form can be expressed as:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x)$$

where $F(x)$ is the final regression function that can be specified for a particular use-case. γ_m are the weights for the individual weak learners $h_m(x)$.

XGBoost stands for Extreme Gradient Boosting is just an implementation of the gradient boosting algorithm. It is open-source, it is fast and extremely effective [7]. As in gradient boosting, the models in the ensemble that make the algorithm are fit using random differentiable loss function with the gradient descent optimization.

5. Random Forest Regression. Decision trees are again non-parametric and supervised learning systems that can be used for both multi-level classification and regression problems. As the name suggests, they have tree structure that contains hierarchical decision values. Bootstrapping an algorithm is done by random subset sampling from a given data. When these samples are averaged over the iterations they are sampled in, the result is expected to be close to the true value. Random forest algorithm is bootstrapped and it amalgamates ensemble learning as in gradient boosting with the decision tree model. This way in random forest regression, random decision trees are sampled from the data and averaged over to predict better. The regression function can be written as:

$$f(x) = \frac{1}{N_{\text{trees}}} \sum_{i=1}^{N_{\text{trees}}} T_i(x)$$

where N_{trees} is the number of decision trees in the forest and $T_i(x)$ is the prediction of the i-th decision tree.

6. Ensemble. We propose an ensemble as represented in Fig. 2. of the aforementioned regression models by weight-averaging them based on their individual performance. The ensemble has been generated by assigning equal weights [5] of 0.1 to Lasso regression, Elastic Net regression and Kernel Ridge Regression. Double this weight of 0.2 is assigned to XG Boost and the maximum weight is assigned to Gradient Boosting. This weight averaging is applied on the predictions of respective models and the averaged score is saved as the RMSLE score of the ensemble. In addition to RMSLE, we have explored how the R-squared score ranges from model to model. R-squared score interprets the magnitude of variation in the regressor response (Table 2).

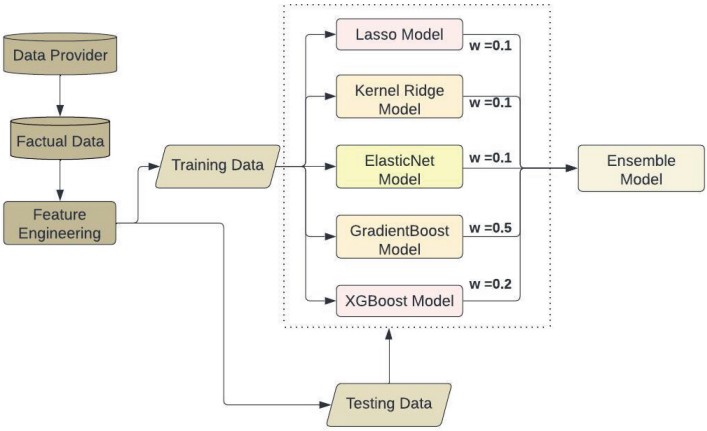


Fig. 2. Generation of Ensemble Model to find the optimal exit option

Table 2. Comparing RMSLE scores of proposed classifiers

Model Used	RMSLE	R2 Score
Elastic Net	0.478495	0.130062
Kernel Ridge	0.476900	0.151631
Support Vector Regression	0.586136	−0.409050
Gradient Boosting	0.432351	0.761315
XGBoost	0.289035	0.700386
Random Forest Classifier	0.313858	0.625402
Ensemble	0.290305	0.760187

4 Results

Based on the Root Mean Square Log Error values of individual regression models from Table 1, the Random Forest algorithm minimises error most efficiently. Support Vector Regression is seen to perform the worst, and we suspect this is because of noise in the data whereby the target exit options may be overlapping. The Ensemble outperforms all individual models by scoring an RMSLE of 0.290305. Elastic Net regression underperforms significantly when there is minimal correlation between the features, as in our dataset. Kernel Ridge regression also depends much on the choice of kernel. The kernel trick often does not predict optimally given the dataset size and the predictor features' relative importance. With respect to the R-squared score, Gradient Boosting proved to be unbeatable even by the proposed ensemble although by a small margin. It was found to be as good as the gradient boosting model while minimizing error by over 67% which is major for a regression model. R-squared even so, rarely completely interprets

the goodness of fit of a model and works on very different parameters than Error which is the essential aim of any regressor model.

5 Conclusion and Future Work

Considering the evaluation metrics used and overall computational applicability, the proposed ensemble of weight-averaged regression models is concluded to predict the exit strategies most efficiently and with the slightest error. XGBoost works next best as an individual model, and Gradient Boosting and Random Forest Regressor follow this in minimizing error.

Future prospects for this work include the usage of dynamic data extracted from Twitter or other platforms where startups interact. As the startup ecosystem is ever changing and highly dependent on shorter changes in market, it will be better for prediction than using historical data.

More exit strategies can also be explored as we have explored only 6 in current research. More such exit strategies include, Employee stock ownership plans, friendly buyouts which are often latent or third-part-sales. It would be interesting to gauge the exit strategy employed based on the startup health at a given time.

Since the dataset used is geographically limited, it can be extended to include startup ecosystems from Asia and India. This would help in understanding the ever-expanding innovation sector in India and hence help in the growth of more such startups and this tool can be utilized as a measure of startup success and health and interpret the future trajectory of the technology, product or service they propose.

As the research here has been mainly via regression and machine learning paradigm, it will be interesting to regard at deep learning and neural networks to improve this further and effectively build a highly optimized predictor that can be utilised by startup founders and venture capitalists.

Acknowledgements. This work has been undertaken as a part of the Centre of Excellence - Artificial Intelligence sponsored by the Department of Science & Technology (DST) Curie-AI Project, Government of India.

References

1. Abhinand, G., Poonam, B.N.: An efficient stacking ensemble technique for success prediction of Indian ventures. In: 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), vol. 1, pp. 1–6. IEEE (2022)
2. Agarwal, A., Jain, K., Dev, A.: Modeling and analysis of data prediction technique based on linear regression model (DP-LRM) for cluster-based sensor networks. Int. J. Ambient Comput. Intell. **12**, 98 (2021)
3. Awad, M., Khanna, R.: Support Vector Regression, pp. 67–80. Apress, Berkeley (2015)

4. Bangdiwala, M., Mehta, Y., Agrawal, S., Ghane, S.: Predicting success rate of startups using machine learning algorithms. In: 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), pp. 1–6. IEEE (2022)
5. Banner, K.M., Higgs, M.D.: Considerations for assessing model averaging of regression coefficients. Ecol. Appl. **27**(1), 78–93 (2017)
6. Cao, L., von Ehrenheim, V., Krakowski, S., Li, X., Lutz, A.: Using Deep Learning to Find the Next Unicorn: a Practical Synthesis (2022). arXiv preprint arXiv:2210.14195
7. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
8. Manasi Chhibber. Startup profit predictor using machine learning techniques. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 2*, pages 49–58. Springer, 2022
9. Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support vector regression machines. In: Mozer, M.C., Jordan, M., Petsche, T. (eds.) Advances in Neural Information Processing Systems, vol. 9. MIT Press (1996)
10. Exterkate, P.: Model selection in kernel ridge regression. Comput. Stat. Data Anal. **68**, 1–16 (2013)
11. Ferrati, F., Chen, H., Muffatto, M.: A deep learning model for startups evaluation using time series analysis. In: European Conference on Innovation and Entrepreneurship, pp. 311. Academic Conferences International Limited (2021)
12. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
13. Kim, J., Kim, H., Geum, Y.: How to succeed in the market? predicting startup success using a machine learning approach. Technol. Forecast. Soc. Chang. **193**, 122614 (2023)
14. Krishna, A., Agrawal, A., Choudhary, A.: Predicting the outcome of startups: less failure, more success. In: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 798–805. IEEE (2016)
15. Lussier, R.N., Pfeifer, S.: A crossnational prediction model for business success. J. Small Bus. Manag. **39**(3), 228–239 (2001)
16. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front. Neurorobot. **7**, 21 (2013)
17. Ragothaman, S., Naik, B., Ramakrishnan, K.: Predicting corporate acquisitions: an application of uncertain reasoning using rule induction. Inf. Syst. Front. **5**, 401–412 (2003)
18. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodolo. **58**(1), 267–288 (1996)
19. Wei, C.-P., Jiang, Y.-S., Yang, C.-S.: Patent analysis for supporting merger and acquisition (m&a) prediction: a data mining approach. In: Designing E-Business Systems. Markets, Services, and Networks: 7th Workshop on E-Business, WEB 2008, Paris, France, December 13, 2008, Revised Selected Papers 7, pp. 187–200. Springer (2009)
20. Xiang, G., Zheng, Z., Wen, M., Hong, J., Rose, C., Liu, C.: A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on techcrunch. Proc. Int. AAAI Conf. Web Soc. Med. **6**, 607–610 (2012)
21. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. **67**, 301 (2005)



An Inclusive Analysis on Deep Learning Hinged Malware Detection Techniques

Vinisha Sumra[1(✉)], Naveen Malik[1], and Santosh Kumar[2]

[1] Computer Science Engineering, MMDU Mullana, Ambala, Haryana, India
vinishamalik317@gmail.com, naveenmalik317@gmail.com
[2] Computer Science Engineering, Chandigarh University, Mohali, Punjab, India
santosh.e11120@cumail.in

Abstract. Malwares are touching the sky in connection with their numbers and this accretion is contemporaneous with the amelioration in usage of Android smart phones. Nearly everything is a blessing and curse at the same time. This applies to innumerable applications available in play stores too. Malware developers earmark this platform to consummate their thrust. These malicious programs can torment the devices in disparate ways such as absconding with the privileged information, superintending the device, wreaking havoc on the battery and maneuvering the data stockpiled in the device. Malware has become proficient in coping with various customary detection techniques. Static detection methods are powerless to seize the malware that utilizes dynamic features such as network traffic features whereas dynamic detection procedures discover it strenuous to detect malware that lodges static features like permissions, images and intents etc. Therefore, the exigency surfaces to maneuver some advanced methodologies in this direction. Deep learning coupled with numerous techniques fulfill this objective. This paper analyzes numerous deep learning techniques which are worthwhile for the detection of these stubborn malwares. The analysis throws light on various deep learning models also.

Keywords: Malwares · deep laerning · machine learning · static analysis · dynamic analysis · hybrid analysis · Abdroid · neural network

1 Introduction

Deep learning refers to the contrivance that is a wedge of the machine schooling genealogy. This possesses a ground in neural network. Deep learning stands as a methodology that upskills the device to perform the actions precisely anticipated by humans. It appertains to AI family tree which goes around with hierarchical schooling method that is chiefly consigned to unsupervised and supervised learning. This cavernous learning is hinted as mysterious box in the arena of Artificial Intelligence [1]. Deep learning embodies affiliated neurons as components. Edifice of neural network may be cultivated wielding associated layers. A latticework of neurons is floated by multitudinous veneers such as hidden layer, input and output coating. The hidden coat amounts to the layers

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 417–425, 2025.
https://doi.org/10.1007/978-3-031-75167-7_33

intermediate to input layer and output layer. A malicious program is the software piece that possesses the aptness of spawning defacement to, or oust the normal service of the device. This is a prevalent appellation which narrates adware, virus, trojan, spyware and some additional forms of vindictive codes [2]. A malignant detection setup endeavors to ascertain whether a specific code is malicious or not [2, 3]. Deep schooling techniques are launched to possess loftier accomplishment over machine investigation techniques [4].

The endowment in this paper comes after in the following manner. We have moulded exploration of foremost and up to the minute literature to possess cognizance on malware detection utilizing deep invesigation. Research gaps have also grabbed the place which may be fruitful for the coming research.

The rump of the paper is assembled as depicted in numerous sections. Section 2 highlights innumerable analytical techniques for malware detection. Section 3 reviews numerous deep learning models that are employed to detect malwares. Section 4 shelters the conclusion of the work.

2 Analytical Techniques

2.1 Static Analysis

W. Li et al. [5] Carried out research on malware recognition system embedding deep learning method that utilizes both the perilous permission and hazardous API calls amalgamation as attributes to fabricate a DBN model. This model can impetuously acknowledge malignant applications from the normal programs. The authors performed the testing on Drebin dataset and the model attained 90% accuracy.

The researchers in [6] proposed a deep investigation apparatus targeted at malware detection in android device. The research utilizes static inspection for obtaining attributes named permissions embedded in the applications. Thereafter, the permissions have been transformed into a figure file which is easy to handle through a deep investigation system. Next, the system is passed through training of numerous image files. The system was tested on huge dataset including 500 benign applications and 2000 malware applications. The depicted model achieved an accuracy around 93%. The researchers recommended deep investigation and learning for malware detection as it ascertains malware reckoning on pattern sequence instead of utilizing signatures.

The researchers in [7] advanced DeepClassifyDroid for singling out Android rooted malware hinged on CNN. The model utilizes static investigation to procure five sundry attributes. The model travelled through testing on 10770 applications comprising of 5546 malicious and 5224 normal ones. The results depicted an accuracy of 97.4%.

The authors in [8] proposed DeepFlow, which stands for a malignant program detection model that employs data freshet of malevolent applications. These data streams are in disparity with those of benign ones. DeepFlow makes use of this disparity to eventually segregate malware applications from the benign ones utilizing deep learning. Thereafter, a segregation model was fabricated grounded on DBN. The developed system was tested on 8000 malicious applications and 3000 normal ones. The model achieved FI-score of 95.05%.

The investigators in [9] enacted a program portrayal MalInsight to gauge malware. Portrayal has been carried out by making allowance for shallow level and inflated level observable attributes of malevolent programs. Authors are also considerate about the structure for the profiling purpose. Structural data is perceived from basic profiling. Actions on network, registry and files accelerate to high level profiling. The classifiers like RF, DT and KNN are trained for recognition of the malwares. The authors have used dataset of nearly 4200 samples including numerous platforms such as Windows 7 Pro and VirusShare. The results depict 97.21% accuracy.

The investigators in [10] advanced a multifaceted deep investigation procedure for stumbling across the Android stemmed mobile malevolent files. The research utilized numerous attributes such as function opcode, method API, and string etc. Copious ingredients are retrieved and a vector is engendered grounded on these features. Initial networks are trained with this vector. Final networks are trained by making use of initial networks. Shabtai et al. [11] dispensed an assortment to detect malwares by employing machine investigation algorithms. The authors employed ingredient types and ingredient selection for the detection. The researchers embedded feature selection procedures involving report prevalence, Fisher Tally, Gain Quotient and Hierarchical ingredient selection. Assortment tacks such as Naïve Bayes, Bayesian Meshworks, K-Nearest Neighbor and Artificial Neural Networks also found their place in the work.

2.2 Dynamic Analysis

Vinaya Kumar et al. [12] erected a framework for malware detection consisting of two junctures. The authors provided a hugely ductile framework by engaging an amalgamation of customary machine investigation techniques and extensive learning frameworks hinged on image processing, static and dynamic methods. In the first instance, malware categorization was carried out utilizing static and dynamic analysis methods. Thereafter, these malwares were clustered into numerous heads using image processing techniques.

Pascanu et al. [13] trained RNN for retrieval behavior. Thereafter CNN was trained to categorize images. Projection level employed the amalgamation of RNN and Echo State Network to retrieve the features. The training was performed on 150k samples. The researchers in [14] brought forward a multi-tasking intensive learning architecture for segregation of benign and malware application. This detection was grounded on spontaneous language tailoring which focused on the dialect of malwares for feature retrieval purpose.

Kolosnjaji et al. [15] depicted that the fusion of neural networks and convolution networks' recurrent layers ameliorates modelling and categorization of system call sequences and surmounts SVM and Markov Models. Rhode et al. [16] devised an RNN-grounded malware projection model. The research envisioned the malicious files before the onset of payload execution and this was successful to prevent the strike.

2.3 Hybrid Analysis

Kabakus et al. [17] probed into network traffic and permissions of the samples in their research. The authors perceived that malicious samples spawn fewer API whoops clubbed with permissions in contrast with the ordinary samples. Vinisha et al. [18]

devised a hybrid model for android malware assessment. The authors in the research clubbed static and dynamic features for analyzing the samples. The research utilized the amalgamation of permissions and intents with network traffic features. The detection accuracy was higher when hybrid method was used instead of the static or dynamic methods alone. The results depicted an accuracy of 94.2% with the use of hybrid analysis technique.

The investigators in their work in [19] brought forward a hybrid malevolent program detector which utilizes fusion of system permissions and network traffic to intensify the detection rate. The authors labelled the model NTPDroid that unsheathes permissions and network attributes from the applications. Training and testing phases of the work employed FP- Growth algorithm to engender recurrent patterns encompassing permissions and traffic attributes.

Shijo et al. [20] lodged an integrated model indulging a weave of steady and vigorous techniques to detect the malignant programs. The research analyses and classifies an undivulged executable file. The feature vector is settled on the analysis of dynamic deportment and binary code.

2.4 Image Derived Analysis

Yongkang et al. [21] devised a YongImage method for the deep learning grounded malware categorization. The research implants malware where injunction-level counsel is coupled with disassembly metadata. This metadata has been fabricated through the IDA dissimulator targeted at the conversion of image vectors. The authors floated a deep neural network termed as malVecNet. The proposed malVecNet is further optimized by employing the proposition of sentence-based codification in Natural Language Processing [22].

Kancherla et al. [23] dispensed a visualization rooted slant to ascertain malwares. Firstly, the executable is reworked into gray scale image termed as byteplot. Thereafter, low level attributes are wrenched out. Intensity and texture hinged features come under this category. These features are clubbed with computational intelligent methods for malware detection. The research utilized Support Vector Machines and procured the accuracy around 95%.

The researchers in [24] methodically accumulated malware detection procedures hinged on image rooted analysis from numerous perspectives. The research focused on evolution of computer vision techniques targeted at malware detection along with depiction of attribute images.

The researchers in [25] probed the efficacy of Convolutional Neural Networks inimical to expendable API injection as prosaic deep learning methods chiefly reckon on Recurrent Neural Networks which might be assailable to redundant API injection. The authors delineated a malevolent program detection model which transfigures malicious programs into figure portrayals and categorizes images utilizing CNN. The research employed spatial pyramid pooling layers to handle differing input size with CNN.

Tan et al. [26] devised a semi-supervised learning grounded malware detection system. The model revolves around three heads: malware conception, feature extrication and stratification. Primarily, malware conception transfigures malevolent files into grayscale figures. Thereafter, ingredients of figures are wrenched out to mirror the garbling motifs

of the targeted malwares. Eventually, a learning model is embedded to detect malwares utilizing labeled and unlabeled samples. The results depicted enhancement in detection accuracy coupled with reduction in cost of sampling.

3 Deep Learning Techniques

Machine Learning coupled with artificial intelligence with its roots in samples for learning paves the path for Deep Learning. Numerous layers in hierarchy are utilized for wrenching out the features. Intensive investigation models are constructed either on superintended or unconquered learning [27–29].

3.1 Deep Neural Network

A deep neural meshwork is embodied with numerous coats amidst input and output coatings of the given meshwork. With the usage of random prognosis, input magnitude of an exhaustive web is lessened with a facet of 45 employing nearly 26 lacs samples. This depicts fallacy rate of 0.42% and 0.49% for the quartet of neural networks and solitary neural networks correspondingly [30, 31]. The detection accuracy does not rely on escalating the tally of shrouded layers. Performance of neural meshworks degrades with two or more veneers in the network in contrary to solitary layer networks [31].

Yuan et al. [32] devised a model named Droid-Sec which availed the deep learning techniques for android malwares. The research throws light on the comparison of the model with the erstwhile machine learning systems such as Naïve Bayes, C4.5, MLP, LR and SVM.

The attributes of two dimensional bipartite programs are utilized to detect malignant programs. This work makes use of intensive neural meshwork [33]. The entire scaffolding is split into three phases. In the beginner phase, four different attributes are fetched from malicious and benign twofolds. Second phase is aim at the construction of deep neural meshwork with one input coating, one output coating and two concealed layers. Score calibrator discovers its place in the third phase which is employed to measure the likelihood of malevolent programs. This is achieved by transliterating the output of the targeted neural web. Results depict 0.1% False Positive Rate and 95% Detection Rate with nearly 4 million binary files.

3.2 Recurrent Neural Network

Recurring Neural Web is implemented in conjunction with remembrance of speeches utilizing Spontaneous Language Distilling. Consecutive outcomes are discerned to foretell the succeeding event. RNN has been utilized to schlep the wholesome counsel from the inception to the termination time and this facility has been termed as Long-Short Term Memory Module. This process finds its position to fight shy of the gradient issue [31, 34].

Athiwaratkun et al. [35] utilized API call trace to detect malicious programs. RNN in the research is amalgamated with GRU and LSTM. The system has two phases to accomplish its target. Firstly, GRU and LSTM scoop up the features. In the next phase,

these features are catalogued. Consecutive dependencies are patterned in the API calls and the outputs are hinged on the precursory inputs [13].

Yin et al. [36] put forward a proposal to fight shy of the Gradient issue. This problem can be avoided by amalgamating four layers of LSTM with Recurrent Neural Networks (RNN). Prior to the onset of training phase of Convolutional Neural Network (CNN), RNN is imposed to fabricate a behavioral system with data. Thereafter, outcome of RNN has been transfigured into attribute effigies for CNN. RNN has been employed to detect unrevealed malwares in IoT environment and opcodes are utilized to categorize the malwares. Thereafter Support Vector Machine (SVM) is embedded to gauge the opcodes and further discovery of the attribute set of the malware. Results present an accuracy of 96% with the employment of n-gram tack.

3.3 Convolutional Neural Network

A convolutional neural network is procured from Deep Neural Meshwork that is utilized for augmenting visualization. CNN has found its place in training of nearly 12 lacs HD effigies for categorization [37]. Gilbert et al. [28] put forward hierarchical CNN to treat the hierarchy in configuration of executable files that utilizes CNN mnemonics to keep track of attributes at the functional level.

Chia-Mei et al. [38] adopted CNN to shoot up malware detection accuracy and efficiency under huge volume and numerous families of malwares. The demonstration depicts the accuracy of nearly 90% for identification of benign and malicious codes. It can further detect malware that is implanted into benign programs.

Arindam et al. [39] devised distinctly efficient and accurate malware detection model grounded on one-dimensional CNN. The system clutch binary file as input and relegate it as benign or malicious. Only nominal pre-processing of the binaries is the prerequisite of the process. Moreover, attribute discovery is handled in the training phase. As a means to compare the system with other models, TF-IDF benchmark rooted detector has also been implemented.

V. Malik et al. [40] nowadays android security has become more tangy due to the sensitivity of the system design. This paper includes various malware attacks in android phones and also their detection techniques.

Venkata et al. [41] proposed a detection system that accepts binary files as input with minutest preprocessing. Thereafter it predicts whether the file is benign or harmful. Attributes are tracked down during training phase. The model is highly efficient and accurate.

4 Conclusion

Security upskills us to be enterprising instead of just being reactive to malwares. Younger generation malwares have gyrated resistant to the malware detection system which makes this development challenging. State-of-the-art evasion methods have sanctioned the adaptation of younger generations of malwares. Deep learning grounded malware detection methods shrink the imperfection of prosaic detection techniques. This paper puts forward an organized exploration of malware detection utilizing Deep-learning

methods. Contemporary techniques to detect malwares on iOS, Android, APTs and Windows are recced.

References

1. Karresand, M.: Separating Trojan horses, viruses, and worms – A proposed taxonomy of software weapons. In: IEEE Systems, Man and Cybernetics Information Assurance Workshop, pp. 127–134 (2003)
2. Idika, N., Mathur, A.P.: A survey of malware detection techniques. In: Purdue University, vol. 48, no. 2 (2007)
3. Christodorescu, M., Jha, S., Seshia S.A., Song D., Bryant, R.E.: Semantics-aware malware detection. In: Proceedings of the 2005 IEEE Symposium on Security and Privacy, pp. 32–46, Oakland, CA, USA, (2005)
4. Majid, A.A.M., Alshaibi, A.J., Kostyuchenko, E., Shelupanov, A.: A review of artificial intelligence based malware detection using deep learning. In: Materials Today Proceedings, vol. 80, Part 3, pp 2678–2683 (2023)
5. Li, W., Wang, Z., Cai, J., Cheng, S.: An android malware detection approach using weight-adjusted deep learning. In: 2018 International Conference on Computing, Networking and Communications, pp 437–441 (2018)
6. Ganesh, M., Pednekar, P., Prabhuswami, P., Nair, D.S., Park, Y., Jeon, H.: CNN-based android malware detection. In: 2017 International Conference on Software Security and Assurance, pp. 60–65 (2017)
7. Zhang, Y., Yang, Y., Wang, X.: A novel android malware detection approach based on convolutional neural network. In: Proceedings of the 2nd International Conference on Cryptography, Security, and privacy, ACM (2018)
8. Zhu, D., Jin, M. H., Wu, D., Yang, M. Y., Chen, W.: DeepFlow: deep learning-based malware detection by mining android application for abnormal usage of sensitive data. pp. 0–5 (2017)
9. Han, W., Xue, J., Wang, Y., Kong, Z., Mao, L.: Maldae: detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. pp. 208- 233, Comput. Secur. , 208–233(2019)
10. Kim, T., Kang, B., Rho, M., Sezer, S., Im, E.: A multimodal deep learning method for android malware detection using various features. IEEE Trans. Inf. Forensics Secur. (2018)
11. Shabtai, A., Moskovitch, R., Elovici, Y., Glezer, C.: Detection of malicious code by applying machine learning classifiers on static features: a state-of-the-art survey. In: Information Security Technical Report, vol. 14, no. 1, pp. 16–29 (2009)
12. Vinayakumar, R., Mamoun, A., Soman, K.P., Prabaharan, P., Sitalakshmi, V.: Robust intelligent Malware detection using deep learning. IEEE Trans. **7**(2019), 46717–46738 (2019)
13. Pascanu, R., Stokes, J.W., Sanossian, S., Marinescu, M., Thomas, A.: Malware classification with recurrent networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP (2015), pp. 1916–1920 (2015)
14. Huang, W., Stokes, J.W.: Mtnet: a multi-task neural network for dynamic malware classification. In: Proceedings International Conference Detection Intrusions Mal-ware, Vulnerability Assessment, Springer, Cham, pp. 399–418 Switzerland (2016)
15. Kolosnjaji, B., Zarras, A., Webster, G., Eckert, C.: Deep learning for classification of malware system call sequences. In: Advances in Artificial Intelligence: 29th Australasian Joint Conference, Springer, Cham, pp. 137–149, Switzerland (2019)
16. Rhode, M., Burnap, P., Jones, K.: Early - stage malware prediction using recurrent neural networks. Comput. Secur. **77**(2018), 578–594 (2018)

17. Kabakus, A., Talha, A., Dogru, I.: An in-depth analysis of android malware using hybrid techniques. Digital Investigation vol. 24 (2018)
18. Malik, V., Goyal, S.K., Malik, N.: A hybrid model for Android Malware Detection. IJITEE **8**(12) (2019)
19. Arora, A., Peddoju, S.K.: NTPDroid: a hybrid android malware detector using network traffic and system permissions. In: 2018 17th IEEE International Conference On Trust. Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), pp. 808–813. NY, USA, New York (2018)
20. Shijo, P.V., Salim, A.: Integrated static and dynamic analysis for malware detection. In: International Conference on Information and Communication Technologies (2014)
21. Jiang, Y., Li, S., Wu(B), Y., Zou, F.: A novel image-based malware classification model. In: 26th International Conference, ICONIP 2019 Sydney, NSW, Australia, December 12–15 (2019)
22. Mikolov, T., Sutskever, I., Chen, K., Corado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process. Syst. **(2013)**, 3111–3119 (2013)
23. Kancherla K., Mukkamala, S.: Image visualization based malware detection. In: 2013 IEEE Symposium on Computational Intelligence in Cyber Security (CICS), Singapore, 2013, pp. 40–44 (2013)
24. Lan, Q., Du, Y., Gao, C., Zheng, H.: Status and outlook of image-based malware detection technology. In: 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 2023, pp. 598–603 (2023)
25. He, K., Kim, D.S.: Malware detection with malware images using deep learning techniques. In: 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 2019, pp. 95–102 (2019)
26. Gao, T., Li, X., Chen, W.: Co-training for image-based malware classification. In: 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2021, pp. 568–572 (2021)
27. Li, Y., Xiong, K., Chin, T., Hu, C.: A machine learning framework for domain generation algorithm-based malware detection. IEEE Access **7**(2019), 32765–32782 (2019)
28. Gilbert, D., Mateu, C., Planes, J.: A hierarchical convolutional neural network for malware classification. In: The International Joint Conference on Neural Networks 2019, IEEE, pp. 1–8 (2019)
29. Dib, M., Torabi, S., Bou-Harb, E., Assi, C.: A multi-dimensional deep learning framework for IoT malware classification and family attribution. IEEE Trans. Netw. Serv. Manag. **18**(2), 1165–1177 (2021)
30. Dahl, G. E., Stokes, J.W., Deng, L., Yu, D.: Large-scale malware classification using random projections and neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (2013), pp. 3422–3426 (2013)
31. Gopinath, M., Sethuraman, S.C.: A comprehensive survey on deep learning based malware detection techniques. Elsevier **47** (2023)
32. Yuan, Z., Lu, Y., Wang, Z., Xue, Y.: Droid sec: deep learning in android malware detection. ACM SIGCOMM Comput. Commun. Rev. **44**(4) 920140, 371–372
33. Saxe, J., Berlin, K.: Deep neural network based malware detection using two dimensional binary program features. In: 2015 10th International Conference on Malicious and Unwanted Software, MALWARE 2015, pp. 11–20 (2015)
34. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8)(1997), 1735–1780 (1997)

35. Athiwaratkun, B., Stokes, J.W.: Malware classification with LSTM and GRU language models and a character-level CNN. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP (2017), pp. 2482–2486 (2017)
36. Yin, C., Zhu, Y., Fei, J., He, X.: A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access **12**(5)(2017), 21954–21961 (2017)
37. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **2012**, 1097–1105 (2012)
38. Chen, C.M., Wang, S.H., Wen, D.W., Lai, G.H., Sun, M.K.: Applying convolutional neural network for malware detection. In: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, 2019, pp. 1–5 (2019)
39. Sharma, A., Malacaria, P., Khouzani, M.: Malware detection using 1-dimensional convolutional neural networks. In: 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Stockholm, Sweden, vol. 2019, pp. 247–256 (2019)
40. Malik, V., Malik, N., Goyal, S. K.: Analysis of android malwares and their detection techniques. In: 2016 Fourth IEEE International Conference on Parallel Distributed and Grid Computing (PDGC), Juit, Solan, India, vol. 2017, pp. 597–602 (2017)
41. Kalyan, E. V. P., Adarsh, A. P., Reddy, S.S.L., Renjith, P.: Detection of malware using CNN. In: 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, vol. 2022, pp. 1–6 (2022)



Advances in Text Summarization Techniques: A Comprehensive Review and Future Prospects

Harsh Vishwakarma, Yashika Kukkar, Aditi Chauhan(✉), Abhinav Maheshwari, Dharmender Saini, and Preeti Nagrath

Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi, India
aditichauhan50@gmail.com, {dharmender.saini, preeti.nagrath}@bharatividyapeeth.edu

Abstract. This research work provides a concise comparative analysis of prominent content summarization models, namely Textrank, T5, Pegasus, and Bart. The models are evaluated for their specific strengths: Textrank excels in sentence extraction, T5 showcases versatility, Pegasus demonstrates superior abstractive summarization, and Bart proves robust across diverse content types. Evaluation metrics, including ROUGE scores and human assessment, offer insights into each model's performance. Beyond performance, the paper considers computational efficiency. The study contributes to content summarization by incorporating machine learning and LSTM techniques, advancing automated text processing. This concise reference aids professionals in selecting optimal summarization models for specific use cases and resource constraints. Evaluation metrics, including ROUGE scores, highlight Pegasus as the top performer out of the implemented models. Pegasus demonstrates exceptional capability in generating summaries aligned with references.

Keywords: Long Short-Term Memory (LSTM) · Encoder · Decoder

1 Introduction

The field of natural language processing is one where the goal of effective content summarization models has assumed a pivotal role, particularly in the face of an unrelenting deluge of textual information. This research paper embarks on a comprehensive exploration, with a keen focus on four leading contenders: Textrank, T5, Pegasus, and Bart. In the era of machine learning and advanced techniques like Long Short-Term Memory (LSTM), the demand for precise content summarization algorithms is not merely theoretical but an imperative for industries across journalism, artificial intelligence, and information retrieval.

Within the tapestry of content summarization, each model—Textrank, grounded in graph-based algorithms, T5, a transformer-based workhorse; Pegasus, the vanguard of abstractive summarization; and Bart, fortified with its denoising autoencoder architecture—stands as a testament to the progress in this domain. With accuracy as the bedrock

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 426–437, 2025.
https://doi.org/10.1007/978-3-031-75167-7_34

metric, these models employ intricate neural architectures, harnessing machine learning and LSTM methodologies to discern the most pertinent information.

This research paper serves as a crucible for rigorous comparative assessments. Beyond examining their summarization prowess, we scrutinize their computational efficiency and real-world applicability. In this age of information inundation, where the demand for concise and contextually-rich summaries is insatiable, our analysis guides researchers, developers, and industry practitioners in the selection of a content summarization model that aligns with their precise requisites and resource constraints.

Throughout the ensuing sections, we delve deep into the methodologies, evaluation metrics, and empirical results, providing a vantage point into the dynamic realm of content summarization. The goal is to not just unveil the relative merits of these models but also to push the boundaries of content summarization, fostering innovation at the intersection of machine learning, LSTM techniques, and the ever- expanding expanse of textual data.

2 Methodologies Used

PEGASUS, or "Pre-training with Extracted Gap-sentences for Abstractive Summarization," is a transformative deep learning model developed by Google Research. It excels in abstractive summarization by pre-training on gap-sentences extracted from documents, enabling coherent and human-like summarization. PEGASUS is exceptional for simultaneous extractive and abstractive summarization, delivering state-of-the-art results across various benchmarks. Its versatility allows fine-tuning for specific tasks, attracting researchers and developers alike. In our research, we harnessed PEGASUS to enhance long dialogue understanding and summarization, achieving exceptional control over summaries based on user preferences and keywords. PEGASUS stands as a pivotal model, revolutionizing the summarization landscape.

T5-model, an abbreviation for "Text-to-Text Transfer Transformer," stands out as a significant breakthrough in the field of natural language processing. Developed by Google AI, it introduces a novel approach by treating a broad spectrum of language tasks as text-to-text tasks, where both input and output are represented as text strings. This adaptability positions T5 as a standout performer in various NLP applications, encompassing tasks such as translation, summarization, question- answering, and text generation. In our research endeavors, we capitalized on the capabilities of T5 to facilitate controllable summarization in the context of lengthy dialogues. Through the strategic use of keywords and prompts as conditioning factors, we harnessed T5's power to deliver flexible and customizable summaries. This approach significantly improved the comprehension and extraction of crucial information from extensive conversational exchanges, providing users with a more tailored and nuanced summary experience.T5 has firmly established itself as a foundational model in NLP research, presenting a unified and highly effective methodology for addressing a diverse array of language-related tasks. Its ability to handle text in a versatile manner has contributed to its widespread adoption and success in various natural language processing applications.

Text-rank, The Textrank model is a robust technique that is rooted in graph theory and linear algebra, providing a potent solution for automatic text summarization and

information retrieval. This algorithm employs a transformative approach by representing a document as a mathematical graph, where sentences serve as nodes, and their relationships are depicted through weighted edges. At its core, TextRank draws inspiration from the PageRank algorithm, originally designed for ranking web pages within search engines. By adapting a similar principle to sentences, Textrank assigns importance scores based on the interconnectedness between sentences. This process closely resembles the computation of eigenvector centrality, converging to a stable state that reflects the significance of each sentence in relation to others. The Textrank model's effectiveness lies in its ability to identify key sentences that contribute significantly to the overall meaning of a document. Following the score propagation process, the sentences with the highest importance scores emerge as the top-ranked, forming a concise summary that encapsulates the essential content of the original document.

BART (Bidirectional and AutoRegressive Transformers) model, a cutting-edge neural architecture, is employed for various NLP tasks. Known for its bidirectional design and auto-regressive training, BART demonstrates exceptional capabilities in text generation tasks, making it a preferred choice for our research. Utilizing BART's pre-trained knowledge, we fine-tuned the model on our specific dataset, tailoring it to excel in tasks like summarization, answering the questions, and dialogue understanding. The model's proficiency in producing content that is logical and appropriate for the setting played a crucial role in achieving superior results across multiple domains and control aspects in our research, highlighting its significance in contemporary NLP research.

3 Literature Review

In [1], the model uses the transformer-based model for abstractive summarization. It uses a special self-supervised pre-trained objective called Gap-Sentence- Generation (GSG) which involves generating summary-like text from given input & Masked-Language-Model where 15% of tokens are selected in the input text and 80% of the selected tokens are replaced by masked token and left 10% by random token and another 10% remains unchanged. Two significant text corpora, C4 and HugeNews, are used to train the model.

In [2], This paper conducted a study on text summarization in Indian languages by scraping news stories and associated descriptions from publicly accessible news websites to create the dataset. The paper provides a detailed overview of the sequence-to-sequence paradigm employed for Indian languages. Utilizing multilingual translation models such as MT5, MBart, and IndicBart, the study performed k-fold cross-validation on the training dataset to prevent overfitting. The dataset is divided into 10 folds, with 9 folds utilized for training and 1 fold for validation.

In [3], This Paper analyzes the text's accuracy level is the goal of using models like BART. Word- and sentence-level granularity is used to break a document into components, and each element gets allocated to the head among the SANs that compose the BART encoder. The outcomes produced by the SANs layers are then combined to capture information that has several granularities, with a hyperparameter controlling the proportion of combining information at the word and sentence levels.

In [4] In the experiment of the current study, extractive text summarization was employed, specifically leveraging the TextRank algorithm to automatically condense

various types of articles. Although it uses textual information as its input rather than web pages, the traditional TextRank model is based on the PageRank algorithm. It works by building an undirected, weighted graph in which the nodes reflect the tested sentences in the text. Each node is given a relevance value based on the total number of associations it has, and the edge weight linking any two phrases acts as the link between those nodes.

In [5] The research performs the Tweet summarization of news articles which is an extractive text summarization with the LEXRANK model and the dataset used is the News article dataset and tweet dataset with the measurement of performance using the Rouge-1, F1-score: 0.664, Rouge-2, F1-score: 0.548 scores. The downside is that the categorization of perspective from tweets is not applied, which makes it harder to pinpoint the appropriate tweet set by better capturing the varied perspectives.

In [6], The research emphasizes the T5 model, conceived by Google Research, which is a versatile transformer model designed to tackle various NLP tasks, reframing summarization as a text-to-text problem, simplifying the summarisation process.

In [7], The research introduces DIALOGLM, a previously trained model designed for comprehension and summary of lengthy conversations. Current dialogue- oriented models fall short in handling lengthy conversations, which are prevalent in real-world scenarios. The authors suggest a window-based denoising technique prior to training methodology and an integrated mechanism of attention to close this gap. The study performs an additional 20,000 warm-up steps using UniLm and the window-based denoising framework. The batch counts 64 along with learning rate 2e-5 are hyperparameters. With the maximum window size set to 512 tokens and the pre-trained window size set to 10% of the input length.

In [8], CTRLsum introduces a pioneering approach to controllable summarization, addressing the limitation of generic, user-disconnected summaries. Users can now influence summaries via textual input, such as keywords or prompts, allowing precise control over multiple summary aspects. The model learns from keywords without requiring human annotations or predefined controls during training. CTRLsum showcases its flexibility and effectiveness across various domains, including news articles, scientific papers, and patent documents, encompassing aspects like entity-centric summaries, length control, contribution summaries, invention purpose, and question-guided summarization. Furthermore, it achieves cutting-edge outcomes based on the CNN/Dailymail dataset by outperforming strong BART models in conventional summarization.

In [9] The paper employs a diverse array of methodologies and confronts numerous challenges in the domain of text summarization. It rigorously evaluates its proposed methods through a multifaceted lens, incorporating Co-Selection and Content- Based metrics. Throughout its exploration of extractive summarization, the paper sheds light on inherent issues, including but not limited to redundancy, irrelevance, and the challenge of achieving comprehensive coverage. Within this comprehensive framework, the paper delves into innovative approaches to address these challenges. Noteworthy methodologies include Graph-Based methods, which leverage intricate relationships within the text; Maximal Marginal Relevance-based techniques, which prioritize the most informative content; and Metaheuristic Relevance-based strategies, introducing a nuanced layer of optimization.

Ln [10] This research critically examines essential factors in document summarization, specifically focusing on saliency, fluency, coherence, and novelty requirements. The study investigates how neural summarization models uncover salient features within documents, drawing inspiration from graph-based models and introducing a novel graph-based approach. Addressing challenges associated with seq2seq models, the research proposes a new hierarchical model for generating references. Experimental validation is conducted on a large-scale corpus with human-generated summaries, showcasing the effectiveness of the introduced methods in enhancing document summarization.

In [11] This paper introduces an innovative query-focused summarization system named ATSUM, which concurrently addresses saliency ranking and relevance ranking. The system automatically generates distributed representations for both sentences and document clusters. It incorporates an attention mechanism designed to emulate human attentive reading behavior in response to a given query. The paper presents extensive experiments conducted on DUC query-focused summarization datasets. Notably, AttSum achieves competitive performance without relying on hand-crafted features. Furthermore, the study observes that the sentences identified as focusing on the query effectively fulfill the query requirements.

ln [12] The paper proposes a data-driven summarization framework with an encoder-extractor architecture, presenting two model categories for sentence and word extraction. Key ideas include designing hierarchical neural structures aligned with summarization task characteristics and adopting a generation-by-extraction approach. This strategy successfully navigates challenges in generating content within a vast vocabulary, covering the entire dataset, encompassing many low- frequency words, and named entities.

In [13] This research explores neural network approaches to document summarization, covering a spectrum from purely extractive to abstractive systems.

The proposed model combines the strengths of neural extractive systems with the flexibility of compression, allowing interpretability through discrete compression options. The extractive and compressive components are jointly trained, overcoming challenges in constructing oracle summaries for supervision. Experimental results on standard datasets demonstrate the model's superiority, achieving notable improvements in ROUGE scores. Additionally, human evaluations affirm the model's grammaticality and fluency, marking a significant advancement in extractive and compressive summarization.

In [14] Extractive methods select salient content, traditionally addressed via greedy, graph-based, and constraint optimization approaches. Neural networks, including recursive autoencoders and CNNs, enhance extractive summarization. SummaRuNNer, a leading document summarization model, combines extractive and abstractive techniques. It excels in interpretability, with a distinct classification layer allowing a nuanced understanding of sentence selection factors. Visualization of normalized scores enhances user comprehension. Examining Daily Mail and DUC corpora, SummaRuNNer effectively identifies key document points, demonstrating qualitative success in summary generation.

4 Experimental Setup and Methods

In this research study, we conducted a comprehensive evaluation of the contemporary model utilizing the Daily Mail/CNN dataset, assessing the performance through the calculation the rouge scores as a means the calculate the summarization capabilities. The various steps involved are data collection, data exploration, evaluation parameters, and parameter setting.

4.1 Data Collection

This study evaluates the summarization capabilities of different models using the CNN/Daily Mail dataset. This dataset, known for its diverse collection of news articles and summaries, serves as our testing ground for assessing the summarization quality of different models. This research aims to quantify and evaluate these models' respective performances, shedding light on their summarization efficacy. By leveraging the dataset's rich textual content, this analysis seeks to provide valuable insights into the comparative strengths and weaknesses of these models in summarizing news articles, contributing to the advancement of automatic text summarization techniques (Table 1).

Table 1. This table represents the details about the DataSet used and split in the DataSet

Dataset Split	Number of Instances in Split
Train	287,113
Validation	13,368
Test	11,490

- Article with mean token count: 781
- Highlights with the summary: 56

4.2 Evaluation Parameters

We used reliable assessment frameworks like ROUGE-1, ROUGE-2, and ROUGE- L in our research. The precision, recall, and F1-score are quantified and the summarizing performance is evaluated by comparing the n-grams and LCS among the summary that was generated and the reference summary (Table 2).

- **ROUGE-1(Unigram Overlap)**: ROUGE-1 calculates the amount of unigram (single word) overlap between the reference summary, the generated summary, and the ground truth.
  - Precision: (number of overlapping unigrams in created summary) / (total unigrams in generated summary)
  - Recall: (number of overlapping unigrams in the produced synopsis) / (total number of unigrams in the reference summary).

- F1-score: (Precision*Recall)/(Precision + Recall)

- **ROUGE-2(Bigram Overlap)**: ROUGE-2 measures the overlap of bigram (pair of adjacent words) between the reference summary and the (ground truth) and the generated summary.

  - **Precision:**(Total number of overlapping bigrams in ref summary)/(Total number of bigrams in ref summary).
  - **Recall:** (overlapping bigram in generated summary)/(Total number of bigram in reference summary)
  - **F1-score:**(Precision*Recall)/(Precision + Recall)

- **ROUGE-L (Longest Common Subsequence)**: calculates the longest common subsequence between the reference and the generated summaries. It checks the longest common subsequence of words that are common in both summaries

  - Precision:(Length of LCS) / (Total number of bigram in generated summary)
  - Recall:(Length of LCS) / (Total number of bigram in reference summary)
  - F1-score:(Precision*Recall)/(Precision + Recall)

Parameter Setting:

Table 2. The Table represents the maximum target length, minimum target length, and Length penalty parameters that were set during the testing of Pegasus, T5, BART

Parameter	Pegasus	T5	BART
Max target length	120	120	120
Min target length	70	70	70
Length penalty	2.0	2.0	2.0

5 Figures and Tables

We conducted extensive experiments to evaluate the performance of the T5, BART, PEGASUS, and TextRank models on our dataset, wherein we adjusted the maximum length parameter to an average value corresponding to the sentence length distribution within the dataset, yielding the following results. The following result were tested on 100 random CNN Daily mail dataset (Table 3).

Google/pegasus-Large

The three main ROUGE scores ROUGE-1, ROUGE-2, and ROUGE-L, each assesses different aspects of text similarity between a reference text (typically human-generated) and a machine-generated output.

Table 3. Scores when the adjusted length parameter to an average value corresponding to the sentence length distribution within the dataset

Scores	Reported Scores	Scores
ROUGE-1	45.20	36.99
ROUGE-2	22.06	12.404
ROUGE-L	36.99	20.592

1. **ROUGE-1**: In this instance, the disclosed ROUGE-1 score stands at 45.20, suggesting that approximately 31.495% of the unigrams in the reference text are also present in the generated text. This indicates that the machine-generated summary successfully captures a noteworthy portion of individual words.
2. **ROUGE-2**: In this scenario, the documented ROUGE-2 score is 22.06, indicating that around 12.404% of the bigrams in the reference text coincide with those in the generated text. The lower score implies a challenge for the machine in producing coherent sequences of words.
3. **ROUGE-L**: The provided ROUGE-L score is 36.99%, signaling that approximately 20.592% of the content in the reference text is present in the generated text. The diminished score suggests a difficulty for the machine in creating overlapping sequences (Table 4).

T5-Model

Table 4. Scores when we adjusted the length parameter to an average value corresponding to the sentence length distribution within the dataset

Scores	Reported Scores	Predicted Scores
ROUGE-1	43.52	30.82
ROUGE-2	21.55	11.11
ROUGE-L	40.69	20.39

The three main ROUGE scores ROUGE-1, ROUGE-2, and ROUGE-L, each assesses different aspects of text similarity between a reference text (typically human-generated) and a machine-generated output.

1. **ROUGE-1**: In this case, the reported ROUGE-1 score is 43.52, indicating that around 43.52% of the unigrams in the reference text also appear in the generated text. The lower score indicates that the model struggles to generate
2. **ROUGE-2**: Here, the reported ROUGE-2 score is 21.55, suggesting that approximately 17.03% of the bigrams in the citation text match those within the text that was generated.
3. **ROUGE-L**: The reported ROUGE-L score is 40.69, indicating that about 34.99% of the reference text's content can be found in the generated text (Table 5).

Text-Rank Model

Table 5. Scores when we didn't adjust the length parameter to an average value corresponding to the sentence length distribution within the dataset

Scores	Reported Scores	Predicted Scores
ROUGE-1	42.07	38.70
ROUGE-2	29.91	23.57
ROUGE-L	40.66	31.34

The three main ROUGE scores ROUGE-1, ROUGE-2, and ROUGE-L, each assesses different aspects of text similarity between a reference text (typically human-generated) and a machine-generated output.

1. **ROUGE-1**: In this case, the reported ROUGE-1 score is 42.07, indicating that around 42.07% of the unigrams in the reference text also appear in the generated text. The higher Rouge1 scores indicate that the machine is able to generate the matching single words.
2. **ROUGE-2**: Here, the reported ROUGE-2 score is 29.91, suggesting that approximately 29.91% of the bigrams in the reference text match those in the generated text. This lower score indicates that the machine struggles to generate consecutive words.
3. **ROUGE-L**: The reported ROUGE-L score is 40.66, indicating that about 40.66% of the reference text's content can be found in the generated text. The higher score indicates that the machine is able to capture a significant amount of sequence (Table 6).

BART Model

Table 6. Scores when we adjusted the length parameter to an average value corresponding to the sentence length distribution within the dataset

Scores	Reported Scores	Predicted Scores
ROUGE-1	29.52	21.43
ROUGE-2	17.03	12.35
ROUGE-L	34.99	25.529

The three main ROUGE scores ROUGE-1, ROUGE-2, and ROUGE-L, each assesses different aspects of text similarity between a reference text (typically human-generated) and a machine-generated output.

1. **ROUGE-1**: In this case, the reported ROUGE-1 score is 29.52, indicating that around 29.52% of the unigrams in the reference text also appear in the generated text. The lower score indicates that the model struggles to generate overlapping words.

2. **ROUGE-2**: Here, the reported ROUGE-2 score is 17.03, suggesting that approximately 17.03% of the bigrams in the reference text match those in the generated text. The lower score indicates that the model struggles to generate consequent words.
3. **ROUGE-L**: The reported ROUGE-L score is 34.99, indicating that about 34.99% of the reference text's content can be found in the generated text (Figs. 1, 2, 3).

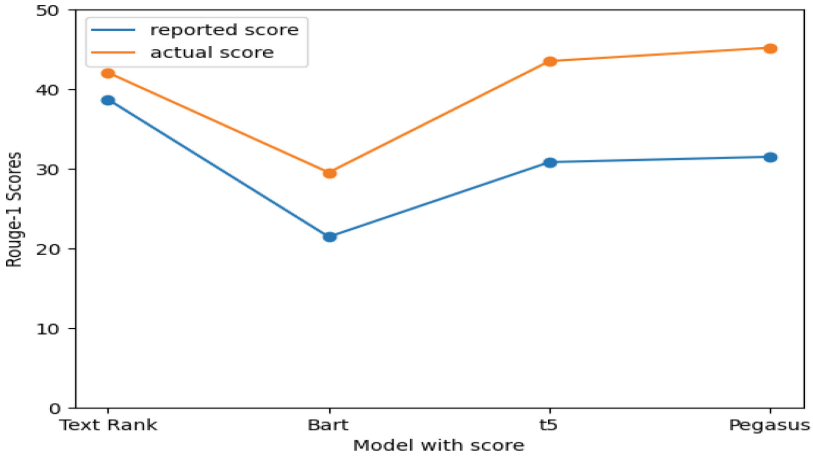


Fig. 1. The Graph represents rouge scores of Text Rank, Bart, T5, Pegasus, where Pegasus shows the highest score among all other models

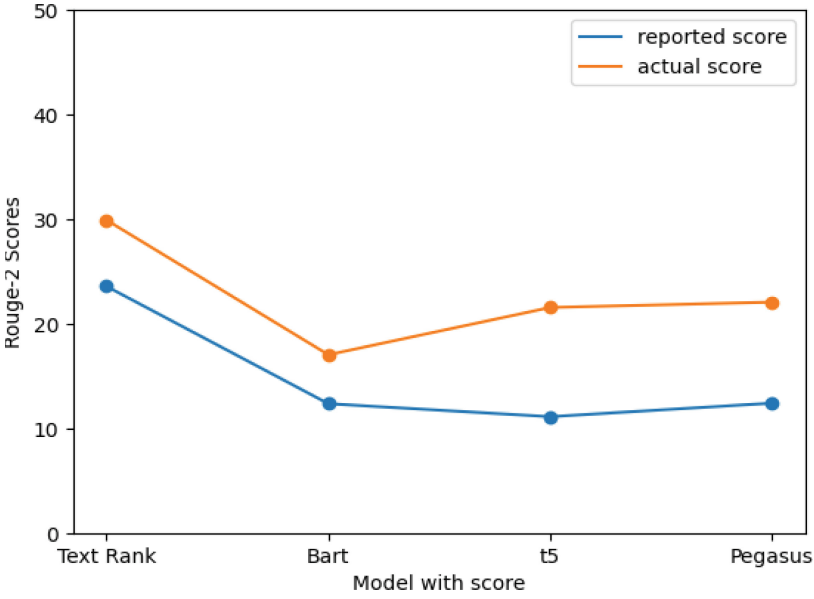


Fig. 2. The graph represents Rouge 2 Scores of the different models such as Text Rank Bart T5 and pegasus were T5 and pegasus Shows almost the highest result

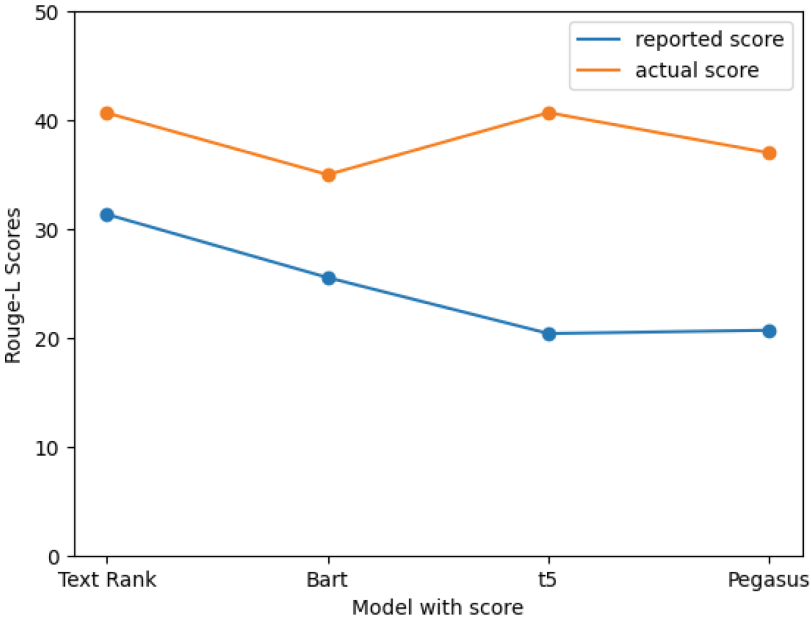


Fig. 3. The Graph represents rougeL scores of Text Rank, Bart, T5, and Pegasus, where T5 Shows the highest score among all other models.

6 Conclusion

In conclusion, this research paper comprehensively analyzes four content summarization models: Textrank, T5, Pegasus, and Bart. Each model demonstrates unique strengths in natural language processing and content summarization. Textrank excels in sentence extraction, while T5 showcases versatility. Pegasus stands out in abstractive summarization, and Bart exhibits resilience across varied content.The evaluation metrics, including ROUGE scores and human assessment, shed light on their performance. Analyzing the ROUGE scores, the best-performing model is Pegasus, with the highest ROUGE-1 score of 45.20, ROUGE-2 score of 22.06, and ROUGE-L score of 36.99. These results show Pegasus' exceptional capability in generating summaries that closely align with reference summaries. Beyond performance, computational efficiency is also considered, enhancing real-world applicability. By incorporating machine learning and LSTM techniques, this study advances automated text processing, providing valuable insights for researchers, developers, and practitioners. In summary, Pegasus emerges as the top-performing model in content summarization based on the ROUGE score analysis.

References

1. Liu, P.J., Zhao, Y.: PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization.arXiv:1912.08777 (2020)
2. Urlana, A., Bhatt, S.M., Surange, N., Shrivastava, M.: Indian language summarization using pretrained sequence-to-sequence models (2023). *arXiv preprint* arXiv:2303.14461

3. Akiyama, K., Tamura, A., Ninomiya, T.: Hie-BART: document summarization with hierarchical BART. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, pp. 159–165 (2021)
4. Vlainić, M.: A Comparative Study of Automatic Text Summarization System Performance (2013)
5. Chakraborty, R., Bhavsar, M., Dandapat, S., Chandra, J., A network based stratification approach for summarizing relevant comment tweets of news articles. In: Bouguettaya, A., et al. Web Information Systems Engineering – WISE 2017. WISE 2017. Lecture Notes in Computer Science(), vol. 10569. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68783-4_3
6. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization (2016). arXiv preprint arXiv:1607.06450
7. Zhong, M., Liu, Y., Xu, Y., Zhu, C., Zeng, M.: Dialoglm: pre-trained model for long dialogue understanding and summarization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 10, pp. 11765–11773 (2022)
8. He, J., Kryściński, W., McCann, B., Rajani, N., Xiong, C.: Ctrlsum: towards generic controllable text summarization (2020). *arXiv preprint* arXiv:2012.04281
9. Verma, P., Verma, A.: A review on text summarization techniques. J. Sci. Res. **64**(1), 251–257 (2020)
10. Tan, J., Wan, X., Xiao, J.: Abstractive document summarization with a graph-based attentional neural model. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1171–1181 (2017)
11. Cao, Z., Li, W., Li, S., Wei, F.: Attsum: Joint learning of focusing and summarization with neural attention. arXiv preprint arXiv:1604.00125 (2016)
12. Cheng, J., Lapata, M.: Neural summarization by extracting sentences and words (2016). *arXiv preprint* arXiv:1603.07252
13. Chen, Y., Wu, L., Zaki, M.J.: Reinforcement learning based graph-to- sequence model for natural question generation (2019). *arXiv preprint* arXiv:1908.04942
14. Nallapati, R., Zhai, F., Zhou, B.: Summarunner: a recurrent neural network based sequence model for extractive summarization of documents. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1 (2017)
15. See, A., Liu, P.J., Christopher, D.: Manning.: a novel architecture that augments the standard sequence-to-sequence attentional model in two orthogonal ways (2017). arXiv:1704.04368v2

Multiple Object Detection and Tracking Using DeepSORT

Rashmi Gandhi[1], Yash Pasrija[1], Amisha Chauhan[1], and Ritu Rani[2(✉)]

[1] CSE, ASET, Amity University, Noida, India
[2] ECE, Bhagwan Parshuram Institute of Technology, Delhi, India
ritujangra00@gmail.com

Abstract. The traffic surveillance and control have always been a major concern. Detecting and tracking vehicles or automobiles through satellites and surveillance cameras has made the task of tracking the defaulters and managing the vehicles on the roads a lot easier. Only detecting a moving object is not enough rather keeping track of it could possibly help in better management of traffic. So, the main objective is to detect and track the automobiles, person on the roads and provide an effective model for the same. The main challenge is to increase the detection count of the objects so that the tracking could be done efficiently and for that purpose we need to use proper detection algorithm. The dataset with around 4680 drone images has been used for the model training. After training the detection models; YOLOv5, YOLOv7, YOLOv8 the results are fed to the tracking algorithm i.e. DeepSORT. The results of the experiment reflects that the proposed object detection method has higher accuracy for faster real-time detection than the other models and the number of objects detected are also higher which helped in efficient tracking. Among all the detection algorithms YOLOv8 produced the best results with higher accuracy and maximum detections.

Keywords: Object Detection · Object Tracking · YOLOv8 · DeepSORT · RCNN

1 Introduction

Any object tracking programme starts with object detection. It alludes to a computer vision method for detecting and recognising specific item instances in pictures or movies. Prior to processing objects further, apps must be able to detect them. Finding any instances of objects in an image frame and, if so, identifying the position of the object is the primary goal of detection.

An entirely new field of object detection has been opened up by deep learning. It has greatly accelerated, simplified, and improved the accuracy of the detecting process. One well-known method of object recognition is the "you only look once (YOLO)" technique, which employs a convolutional neural network (CNN) to identify objects. The quick processing speed of the YOLO algorithm is its benefit. High accuracy and efficiency are provided by YOLO, which occasionally even exceeds the R-CNN method. The tracking process begins when object detection is complete. It is employed to determine

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 438–448, 2025.
https://doi.org/10.1007/978-3-031-75167-7_35

each frame's detected item placements. One may refer to the tracking's inputs as the detection's outputs. Bounding boxes contain the items that have been detected. The dimensions and characteristics of the identified item are then ascertained for each box. After then, the tracking phase keeps track of the things it has found till the end of the video stream. Four distinct bounding boxes are generated, for instance, if three automobiles and a truck are identified. These bounding boxes are then monitored in all ensuing video frames.

Object tracking is the process of automatically detecting and classifying objects in a video and interpreting them as a sequence of trajectories based on frame analysis. Finding a moving item's location inside a single video frame is the aim of object tracking. In essence, it's a computer vision task that involves determining an object's trajectory by extracting its movement from a set of images. These days, object tracking is used in a wide range of industries, including traffic, security, automobiles, cameras, and retail.

In this research paper we have discussed about object detection and tracking in traffic scenarios, where the moving vehicles are tracked through the camera and then analyzed and sorted with the algorithm. An algorithm analyses sequential frames in a video and the shows us the target motion of the objects between frames to execute the tracking. Various algorithms exist, each with strengths and weaknesses. A visual tracking mainly holds two important parts first is target representation and location and then filtering the information for better interpretation.

2 Literature Review

There is significant amount of work that has been done in the field of object detection and tracking. A lot of research studies and experiments have produced significant results (Table 1).

Table 1. Literature Review

Author	Purpose	Summary
Akansha Bathija, Prof. Grishma Sharma (2019)[1]	Analysis of tracking-by-detection approach	Object detection was done using YOLOv3 and tracking using DeepSORT
Anand John, Dr. Divyekant Meva (2020) [2]	A Comparative Analysis and Performance Evaluationof DifferentObject Detection Algorithms	Various detecton algorithms like RCNN, Fast-RCNN, Faster RCNN, SSD, YOLOv3 were studied and YOLOv3 was implemented YOLOv3 gave the best results
Tuan Lin Dang, Gia Tuyen Nguyen and Thang Cao (2020) [3]	Tracking objects with an enhanced Deep Sort-Yolov3 architecture	Insertion of dlib in tracking architecture helped in reducing identity switches

(*continued*)

Table 1. (*continued*)

Author	Purpose	Summary
Jun Deng, Xiaojing Xuan, Weifeng Wang, Zhao Li, Hanwen Yao, Zhiqiang Wang (2020) [4]	Review of various deep learning object detection algorithms	All the algorithms were studied and compared and analysis was done about which detection algorithm gives better results
Lou, Xuehu & Guo, Junme & Liu, Jason & Bi, Lingyun & Chen, Haiying & Gu, Haonan, Haitong & Duan (2023). [5]	Introduces an algorithm that improves small size target detection in camera sensors by addressing limitations in human observation, increasing precision, and ensuring accuracy across target sizes	This paper presents a small size target detection algorithm that surpasses existing algorithms in precision and accuracy for different target sizes. It incorporates context feature preservation, improved feature fusion, and a novel network structure, outperforming YOLOX, YOLOXR, YOLOv3, scaled YOLOv5, and YOLOv8 on authoritative datasets
Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi (2023) [6]	The purpose of this paper is to present a generalized and refined model for real-time detection of flying objects, utilizing the YOLOv8 architecture, to address the challenges posed by object variance and achieve improved performance	This study presents a YOLOv8 architecture based real-time flying object detection methodology. The authors use a dataset including 40 types of flying objects to train a generalised model, and then they enhance the model using transfer learning on a dataset that is more realistic. The study tackles issues including occlusion, grouped backdrops, and spatial size variation. On 1080p films, the final adjusted model retains a high inference speed while achieving enhanced mean Average Precision (mAP). The YOLOv8 architecture and its features are also explained in the paper
Li Y, Fan Q, Huang H, Han Z, Gu Q. (2023) [7]	In order to improve UAV multitarget recognition efficiency and resilience in the face of obstacles such target changes and occlusion, this research proposes an aerial image detection model	In order to overcome obstacles in UAV multitarget recognition, such as variations in target size, form, occlusion, and lighting conditions, this research suggests an aerial image detection model. The algorithm is evaluated on the VisDrone2019 dataset and undergoes various experiments to demonstrate its effectiveness and feasibility, providing a promising approach for deep learning deployment in UAV multitarget detection

To retrieve relevant hidden features, the majority of researchers employ deep learning algorithms. They have demonstrated exceptional performance in a range of challenging tasks that have traditionally required bespoke features, such as object tracking, traffic management, identification, monitoring, and human crowd recognition. With the advent of driverless cars, advanced video surveillance, face recognition, and other people-counting applications, there is an increasing demand for fast and accurate object detection systems. For these systems to work, every item in an image must be identified, classed, and found by tracing around the relevant bounding box. This makes object identification far more difficult than its traditional computer vision; image recognition [8].

YOLO is very fast in terms of detecting objects. It is a very good option if you require real-time detection without sacrificing too much precision. YOLO predicts several bounding boxes and category probabilities for these boxes simultaneously using a single neural network.

Object detection is an essential field of study in computer vision, deep learning, artificial intelligence, and related fields. It looks to precisely identify the category, find the target of interest inside the picture, and give each target's bounding box. It has been extensively employed in a variety of applications, including intelligent video surveillance, medical image analysis, automatic driving of vehicles, video and image retrieval, and industrial inspection [9]. Countless techniques and algorithms can be used to detect objects. R-CNN, YOLO, SSD, YOLO v8, and other well-known deep learning-based techniques/methods that employ convolutional neural networks (CNNs) are taught to automatically detect objects inside the images [10, 11]. We can either construct a new object detector or utilize as pretrained one. Analysis through tracking-by-detection approach have shown that the results of tracking through DeepSORT are more accurate when detection is performed using YOLOv3 rather than other detection algorithms [12]

Comparative studies followed by the experiments of various object detection algorithms have shown that YOLO algorithms are faster than the other algorithms and detection results are better. With the insertion of Dlib in the process of tracking, identity switches can be significantly reduced. In order to overcome the constraints of human perception and cognition, the research introduces a unique method for tiny size target recognition in camera sensors. The algorithm achieves higher precision and accuracy for small targets while maintaining performance across different target sizes. It introduces innovative techniques such as context feature preservation and improved feature fusion. The experimental results demonstrate its superiority over existing algorithms, making it a valuable contribution to target recognition technology [5]. A generalized model and a refined model for real-time detection of flying objects is proposed. The generalized model achieves a mean Average Precision (mAP), 0.685, while the refined model improves upon this with an mAP, 0.835. Both models maintain a high inference speed on 1080p videos, making them suitable for real-world implementation. The paper contributes to the field of object detection, specifically for flying objects, by utilizing the YOLOv8 architecture and addressing challenges such as object variance, occlusion, and clustered backgrounds [6].

In order to overcome obstacles in UAV multitarget recognition, such as variations in target size, form, occlusion, and lighting conditions, this research suggests an aerial

image detection model. The model incorporates the Bi-PAN-FPN concept to enhance feature fusion, utilizes the GhostblockV2 structure to reduce information loss and model parameters, and applies the WiseIoU loss with a dynamic nonmonotonic focusing mechanism for bounding box regression. The algorithm is evaluated on the VisDrone2019 dataset and undergoes various experiments to demonstrate its effectiveness and feasibility, providing a promising approach for deep learning deployment in UAV multitarget detection [7].

The most widely used and well-known advanced object tracking system is Deep SORT, an improvement on Simple Real-time Tracker. In deep SORT, the Kalman filter is an important component. It operates in a recursive manner, where we first anticipate the current state using the current readings, then we utilized the measurements to revise our predictions.

3 Methodology

Objects in a visual can be instantly detected and identified by humans. Human vision is quick and precise, and it can perform challenging tasks like detecting barriers and differentiating numerous objects with little conscious effort. So, for image detection purposes several models have to be fed for with the images as dataset format so that they could observe and predict for the same. Our project was divided into several stages of production which includes training and testing models. We used YOLOv8 for the detection of the object followed by tracking using DeepSORT.

3.1 Architecture

See Fig. 1.

3.2 Data Collection and Training

We downloaded a vehicle dataset from Kaggle. Once we have the dataset, we tagged the photos using a labelling tool; LabelImg. We generated bounding boxes around various items and assigned labels to each box using the popular open-source labelling application LabelImg. The dataset had 4680 drone images of the road containing multiple objects. It had 10 different classes including people, cars, trucks, plane, boat, motorcycle, pickup, tractor, van. The dataset was split into three settest (11 images), valid (619 images) and train (4050 images). A collection of labelled training data in the form of picture files and associated annotation files is also needed for YOLO (v5 and above). The annotation files, which often have the same name as the associated picture files, provide details on the positioning and kind of objects in the image. The COCO format is the most typical format for annotation files, although YOLO (v5 and higher) also supports additional formats including Pascal VOC and YOLO. We've adopted the YOLO format. We trained our dataset on the YOLOv8 model for 640 images and 50 epochs.

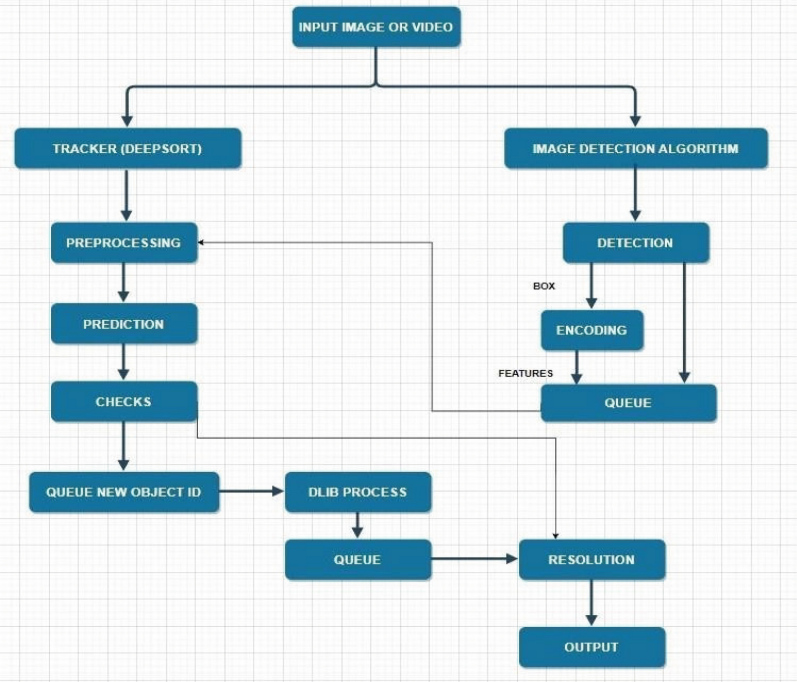


Fig. 1. Details of proposed architecture

3.3 Object Detection

Ultralytics' newest generation of YOLO-based Object Detection models, dubbed YOLOv8, provides state-of-the-art performance. The YOLOv8 model offers a uniform framework for training models while being quicker and more accurate. It can perform image classification, instance segmentation, and object detection. YOLOv8 is quite versatile and scalable offering flexibility to switch through different versions of YOLO for object detection. YOLOv8 also make use of Non-Maximum Suppression (NMS) in order to maximize the number of detections in each frame.

- First, import the necessary libraries.
- Build a function to filter the boxes according to their probability and threshold in step two.
- Build a function to determine the IoU between the two boxes in step 3.

We also trained our dataset for various other detection algorithms such as YOLOv5, RCNN. We then compared the results of each detection algorithm and found out that YOLOv8 is producing best results. Then we used output of YOLOv8 to do tracking using DeepSORT.

3.4 Object Tracking

After each object is detected and a bounding box is formed around it. The detection results are offered to the tracking algorithm. The tracking Algorithm with we are using here is DeepSORT. The assumption that every object is tracked per-frame forms the basis of the IOU tracker, leaving either no or very little space between detections. An ID is given to each object as it is tracked using the computer vision tracking technique DeepSORT. An improvement to the SORT (Simple Online Realtime Tracking) algorithm is DeepSORT.

Through the addition of an appearance descriptor to the SORT algorithm, DeepSORT integrates deep learning to improve tracking by reducing identity switches. The queue is used so as to track the new object entering the frame and pop the object which moves out of the frame. The trails are drawn behind each object that enters the frame and the trail disappears as it leaves the frame. The model was tested on a video input of the road with vehicles.

4 Experimental Results

The custom dataset we used was trained on YOLOV8, YOLOv5, YOLOv7, RCNN for object detection and following results were obtained. YOLOv8 is not only faster but also efficient in detecting smaller objects (Fig. 2).

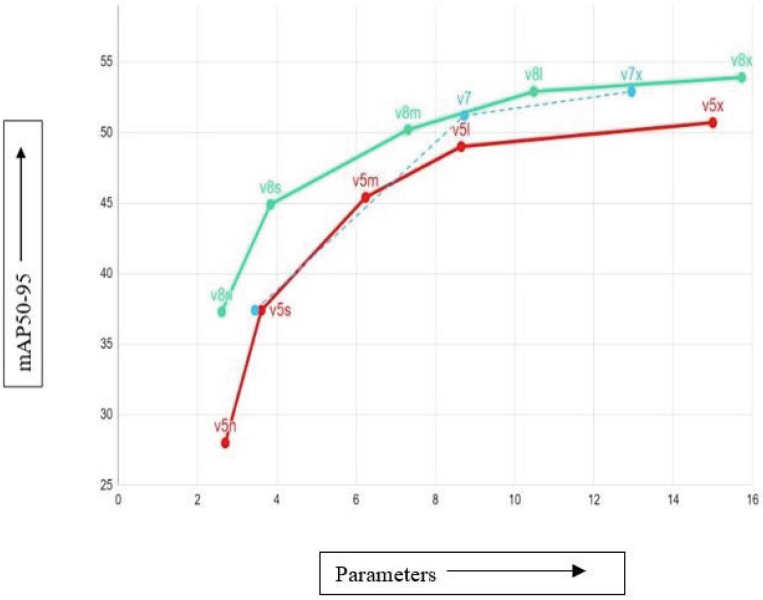


Fig. 2. Comparison of YOLOv8 vs YOLOv7 vs YOLOv5

The graph above shows the mAP value of YOLOv5, YOLOv7, YOLOv8 and shows that YOLOv8 is best amongst all of them. The plot has various pretrained models of the

stated algorithms and displays the corresponding mean average precision (Table 2 and Fig. 3).

Table 2. Comparison table for YOLOv8 and other detection algorithms

Algorithm	Precision	Recall	mAP50	mAP50–95
RCNN	0.651	0.568	0.521	0.328
YOLOV5	0.678	0.529	0.565	0.372
YOLOV7	0.705	0.611	0.678	0.398
YOLOV8	0.757	0.687	0.764	0.429

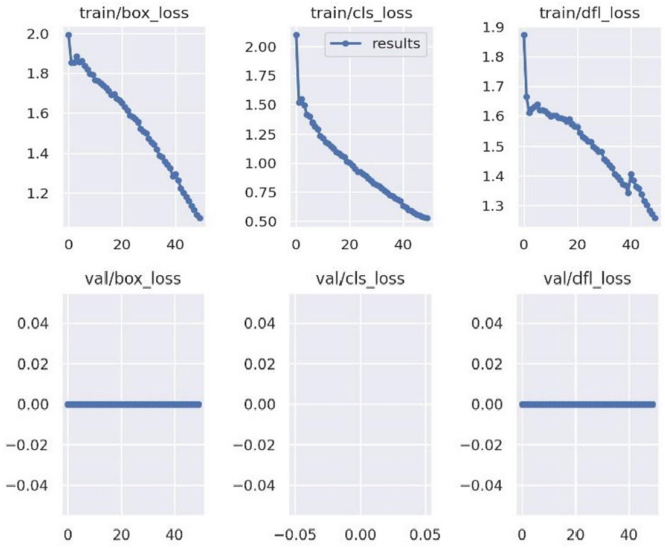


Fig. 3. Training graphs for YOLOv8 model

The above training graphs represents the box loss, classification loss during training and validation. The X-axis represents the number of epochs, here 50. The Y-axis represents the percentage value. The box_loss represents how close is the bounding boxes are to the labelled images of the dataset. The lower box_loss values represents that the model is improving and more accurate bounding boxes are being generating for the labelling of the objects.

Classification loss (cls_loss) helps in measuring the accuracy of the bounding box classification. One bounding box may contain object from different class or same classes or may contain a null object.

The following training metrics Precision, Recall, mAP-50, mAP50-95 are plotted for YOLOv8 model. mAP uses the model predictions, get the average accuracy for each

class in your data. mAP value here is about 0.757 for YOLOv8. It clearly points towards the higher acccuracy of the model (YOLOv8) (Figs. 4 and 5).

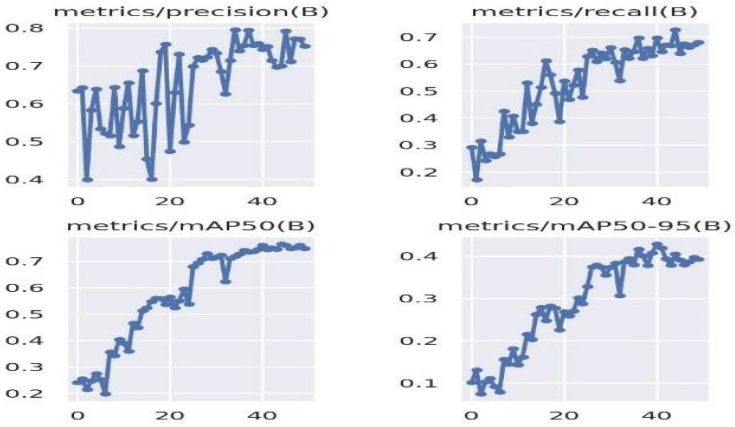


Fig. 4. Training Metrics (Precision, Recall, mAP-50, mAP50–95) for YOLOv8 model

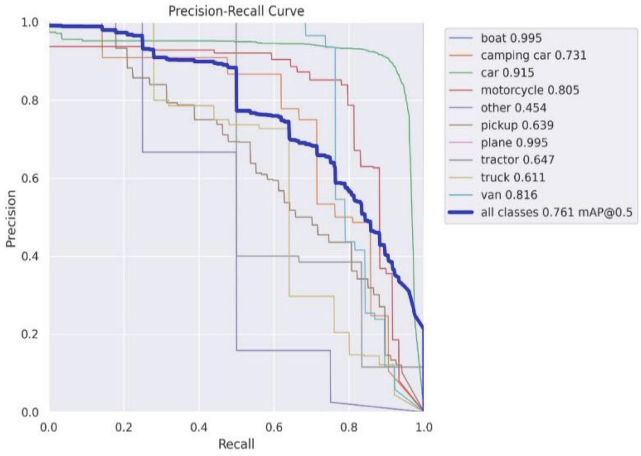


Fig. 5. Precision-Recall Curve for YOLOv8 model

The precision-recall curve shows the accuracy vs. recall tradeoff for different levels. Low false positive and false negative rates are associated with good accuracy and high recall, respectively. High accuracy and recall are shown by a large area under the curve. This curve clearly represents that class car have an completely ideal P-R curve thus indicates maximum number of correct detections. The DeepSORT algorithm used for tracking proved to be quite efficient as studies shows that the tracking accuracy of DeepSORT is about 78.7%. When the model was implemented on the video input following out was obtained (Fig. 6).

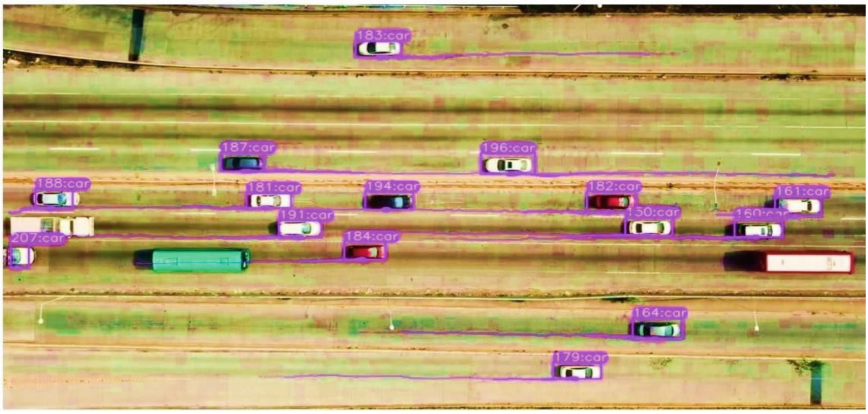


Fig. 6. Output Image showing tracked vehicles

5 Conclusion

Through our experiments, it is very clear that YOLOv8 even after being the latest algorithm which is in the its latest phase provides better detection results than the existing detection algorithm for the smaller objects. It is even one of the fastest detection algorithm. This MOT system may detect specific items in surroundings where there are many other things, track a certain type of object, identify a group of object classes, or count a specific object, among other real-time applications. The results of the tracking algorithm; DeepSORT solely depends on the results of the detection algorithm. If the detections are high then it will increase the efficiency of the tracking algorithm and vice-versa. Due to improved detections though YOLOv8 and use of NMS, Dlib have helped in better tracking and lesser ID switching. In future work, using YOLOv8 for Multiple Object Tracking (MOT) could involve exploring techniques to integrate detection and tracking seamlessly, incorporating temporal information and motion cues, enhancing association algorithms for handling occlusions and complex scenarios, and optimizing the model for real-time performance in MOT applications.

References

1. Bathija, A., Sharma, G.: Visual object detection and tracking using YOLO and SORT. Int. J. Eng. Res. Technol. (IJERT) ISSN: 2278-0181 IJERTV8IS110343 **8**(11) (2019)
2. John, A., Meva, D.: A comparative study of various object detection algorithms and performance analysis. Int. J. Comput. Sci. Eng. **8**, 158–163 (2020). https://doi.org/10.26438/ijcse/v8i10.158163
3. Dang, T.L., Nguyen, G.T., Cao, T.: Object tracking using improved Deep_Sort_YOLOv3 architecture. ICIC Express Lett. **14**, 961–969 (2020). https://doi.org/10.24507/icicel.14.10.961
4. Jun, D., et al.: A review of research on object detection based on deep learning. J. Phys.: Conf. Ser. **1684**, 012028 (2020)

5. Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., Chen, H.: DC-YOLOv8: small size object detection algorithm based on camera sensor (2023). https://doi.org/10.20944/preprints202304.0124.v1
6. Reis, D., Kupec, J., Hong, J., Daoudi, A.: Real-Time Flying Object Detection with YOLOv8 (2023). [cs.CV] arXiv:2305.09972
7. Li, Y., Fan, Q., Huang, H., Han, Z., Gu, Q.: A modified YOLOv8 detection network for UAV aerial image recognition. Drones. **7**(5), 304 (2023). https://doi.org/10.3390/drones7050304
8. Moon, S., Lee, J., Nam, D., Kim, H., Kim, W.: AComparitivestudy on multi-object tracking methods for sports events. In 2017 19th International Conference On Advanced Communication Technology (ICAT) (2017)
9. Liu, Y., Yao, L., Shi, Q., Ding, J.: Optical flow based urban road vehicle tracking. In: 2013 Ninth International Conference on Computational Intelligence and Security. IEEE (2014)
10. Pramanik, A., Pal, S.K., Maiti, J., Mitra, P.: Granulated RCNN and multi-class deep SORT for multi-object detection and tracking. IEEE Trans. Emerg. Top. Comput. Intell. **6**(1), 171–181 (2022). https://doi.org/10.1109/TETCI.2020.3041019
11. Bin Zuraimi, M.A., Kamaru Zaman, F.H.: Vehicle detection and tracking using YOLO and DeepSORT. In: 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, pp. 23–29 (2021).https://doi.org/10.1109/ISCAIE51753.2021.9431784
12. Azhar, M.I.H., Zaman, F.H.K., Tahir, N.M., Hashim, H.: People tracking system usingDeep-SORT. In: 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, pp. 137–141 (2020). https://doi.org/10.1109/ICCSCE50387.2020.9204956

An Ensembled Travel Recommender Model Using Autoencoder and Machine Learning Algorithms

V. K. Muneer(✉)

Research Department of Computer Science, Sullamussalam Science College, Areekode
Affiliated to University of Calicut, Kerala, India
vkmuneer@gmail.com

Abstract. This study investigates the use of autoencoders to analyse travelogues posted in Malayalam on Facebook. The main goal is to use autoencoders to provide a compact representation of the input data, which will be used to train different machine learning models to improve accuracy and efficiency. To address the lack of a standardized dataset in Malayalam for the tourist sector, NLP techniques need to be used to unstructured, long, and uneven travel logs. This is handled by using extra filtering methods and using a specialized POS Tagger in conjunction with lookup dictionaries. The study used a two-step technique. An autoencoder neural network architecture is used to encode the travelogues into a lower-dimensional latent space representation. The encoder's compressed representation is transmitted to the decoder to reconstruct the original travelogues. The encoded model is utilized for training several machine learning models, including MLP, logistic regression, RFC, SVM, KNN, and SGD. The model attains a remarkable validation accuracy of 96.84%.

Keywords: NLP · Deep level Autoencoders · Recommender model · Low resourced Languages

1 Introduction

Travel and tourism are essential for promoting cultural interchange, economic development, and individual improvement. People today have unparalleled access to a vast amount of travel-related knowledge and experiences shared by other travellers over social networking platforms. Facebook travel groups have become lively centers for documenting trips, sharing engaging travel stories, and offering helpful advice and suggestions.

This research aims to create a personalized trip recommender module using Facebook travel groups. The system uses machine learning algorithms and artificial intelligence (AE) to suggest travels based on user interests. A corpus of 12,500 Malayalam language transcripts was collected from travel-related groups on Facebook. Malayalam's complex morphology, agglutinative character, and rich inflected structure make it challenging for text and speech processing. The lack of benchmark datasets and consistency in sentence

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 449–459, 2025.
https://doi.org/10.1007/978-3-031-75167-7_36

structure and spelling between state sources further hinders the development of reliable models.

The dataset was processed and organized to guarantee high quality and dependability. Information like Travel Type, Travel Mode, Location Type, Location Climate, Users, and destinations mentioned in the travelogue are extracted from the unstructured dataset. The personalized trip recommender model was constructed via an autoencoder neural network architecture to capture crucial patterns and characteristics. This encoded format is utilized to train different machine learning models, allowing them to learn from the concise and significant representation of trip data.

The autoencoder model that has been constructed achieved an impressive validation accuracy of 96.84%. The study aims to enhance travel experiences by providing personalized recommendations based on individual preferences. Key contributions include the introduction of an autoencoder model, analysis of data compression, evaluation of compressed data performance, enhanced travel recommendation accuracy, and comparative analysis of various machine learning models trained using encoded travel representations.

2 Background and Related Work

Text understanding has been greatly impacted by recent developments in NLP, which have enabled robots to comprehend and analyse spoken language with extraordinary accuracy [1]. Mary Priya S [2] carried out SMT techniques meant for foreign languages to a Dravidian language like Malayalam and examined in experimental research on Statistical Machine Translation (SMT) for the Malayalam language. R.K. Thandil addressed the difficulties in decoding Malayalam language in a similar work, emphasizing the language's complicated agglutinative structure and morphological complexity [3]. A thorough research on the creation of recommender systems by D. Roy made a significant contribution to the area by demonstrating the kinds, features, and related issues of different knowledge filtering algorithms [4]. Travel systems that offer recommendations with a personalized touch have gained prominence due to the increasing desire for customized experiences while travelling [5]. To improve personalized travel suggestions, F. Lu and W. Zhang introduced a novel approach in their study [6].

In research comparing AE based RS to conventional models, Guijuan ZHANG and Yang LIU found that such systems perform better than conventional approaches [7]. Using AE based CF, Ferreira D. et al. presented an item recommendation system that outperformed Singular Value Decomposition [8]. Ding introduced a novel method for maintaining confidentiality while guaranteeing data privacy with better predictive accuracy [9]. He did this by modifying autoencoder for federated shared filtering.

Guo et al. presented AutoLFA, a cutting-edge fusion system for recommendations that blends autoencoder with LFA techniques. To take use of the advantages of both approaches, AutoLFA combines two independent RS that each function in a different metric representational space via a customized self-adaptive weighting mechanism [10]. Y. Bougteb et al. looked at a combination of recommender systems that uses SVD++ segmentation to record connections among various feature elements and a deep autoencoder to learn user preferences and rebuild missing ratings [11]. In a different work,

Kamble used data mining methods to investigate recommender system performance. An SVM-based recommender system was used, and trials were conducted on a range of product datasets [12].

Hybrid recommender systems are crucial since they combine many machine learning and deep learning models such as SVM, KNN, logistic regression, decision trees, Random Forest, and Multilayer Perceptron [13]. Understanding the evolving landscape of recommendation systems is achieved via the integration of many approaches.

3 Methodology

This study developed a travel recommendation system using an autoencoder. The process starts with data collection from community sites and online travel groups, followed by preprocessing to convert unstructured information into a consistent format. Features are extracted to prepare the dataset. An AE model is created and trained to learn condensed representations of travelogue data. Self-supervised learning is utilized to improve effectiveness. The machine learning recommender system is created to work with both raw data and encoded representations, using conventional and deep learning methods [14]. Experimental results and comparisons provide insights into the system's effectiveness and potential enhancements. Figure 1 describes the entire process and phases. Different phases of the proposed work include Data Collection, Data Preprocessing, Feature engineering, Training of AE, development of self-supervised frameworks, Machine learning based RS model, ML and Ensembled model recommendations, Experimental result, and its analytical evaluation.

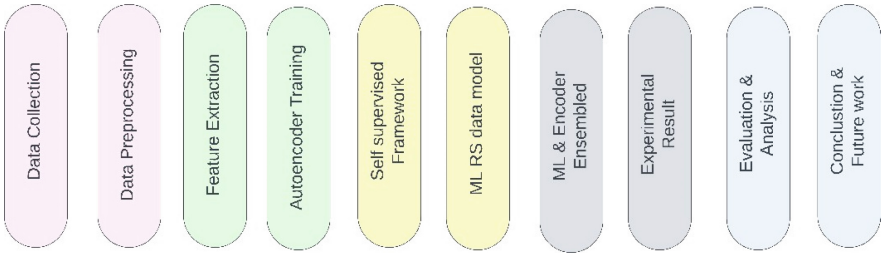


Fig. 1. Diagrammatic representation of aimed work.

4 Data Collection

The data collection involves collecting Malayalam travelogues from online data specifically from the Malayalam travel group of Facebook platform. The entire travelogue in the purest form is extracted. The size, language and presentation varied from person to person. Different travelogues from same person were also collected to learn the travel behavior of the person with more precision. As a result, a dataset of different size and differently expressed travel experiences were gathered in the most unstructured form.

From among this the tougher part was to extract the most relevant information to learn the travel behavior of each individual user. The information like the mode of travel (road, air, water, rail), type of travel (solo, family, friends, groups or communities), location climate (rainy, winter, summer, spring, mist, fog) and location type (mountains, nature, beach, woods, snow) are captured for each person. A travel DNA of each individual is extracted from all the travelogues and stored in a csv file with username as label. At this point a structured representation of the earlier data is obtained.

The research gathered information from a range of online communities, with a particular emphasis on the most sizable Malayalam Travel group on Facebook, known as 'Sanchari', as well as travel websites. Web scraping techniques were employed to retrieve online articles; however, the process encountered difficulties such as client-side rendering, memory leakage, and malware detection. The data exhibited characteristics of being unstructured, chaotic, and inconsistent. The retrieved posts, which are displayed in Figure 2, were systematically arranged in an Excel sheet, which served as the basis for gathering data and preparing the dataset.

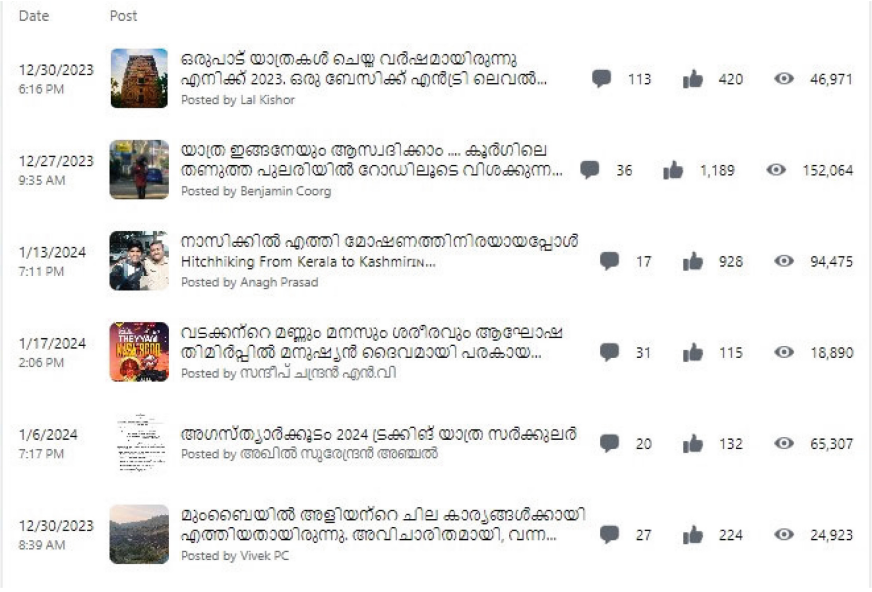


Fig. 2. Extracted form of travelogues from Facebook group.

The organization used a file-based approach to store travel write-ups, ensuring easy retrieval and management of data. The system also accommodated new user input by capturing travel-related text and storing it separately. Following the preprocessing stage, the 12,500 unstructured travelogues underwent a conversion into a structured tabular format. This conversion incorporated travelogue-specific information, comprehensive post details, and user responses such as shares and comments. This enhanced the organization and accessibility of the collected data.

5 Travelogue Processing

The text preprocessing process is crucial for developing a personalized travel recommender model. The data is initially cleaned for any punctuations, emojis, and text other than Malayalam. The data is then cleaned to remove extraneous characters, symbols, or special characters, streamlining the analysis. Stopwords removal is used to reduce noise in the dataset. The stopwords used in the Malayalam language is outlined in Figure 3. Then the travelogues underwent tokenization that segmented the text into individual tokens or words. These words are again lemmatized to extract the lemma/root word of each token.

To preserve linguistic consistency, stemming or lemmatization is performed by reducing words to their root forms. By mapping words from a predefined dictionary and resolving synonyms and related terms into a standardized representation, domain-specific knowledge is integrated. Part of Travelogue Tag [POTT] is a tool utilized for the purpose of labelling and distinguishing travel-related entities, such as modes of conveyance, destinations, and activities. The data that has been tagged is methodically arranged and saved in CSV or Excel files, which enables comprehensive analysis. The rigorous preprocessing methodology described herein forms the bedrock for subsequent research undertakings.

Malayalam	English	Tokens	Lemmas	English
Stopwords		Tokens and root words in Malayalam and English		
അത്	that	പോയിരിക്കും	പോകുക	Go
അങ്ങനെ	so	കുടുംബമായി	കുടുംബം	Family
മതി	enough	കാറിലേക്ക്	കാർ	Car
എങ്കിൽ	if	തണുപ്പിൻറെ	തണുപ്പ്	Cool
മറ്റു	other	മരത്തിൻറെ	മരം	tree

Fig. 3. Few of the stop words in Malayalam, Root words.

6 Feature Extraction

The feature extraction process is crucial in organizing and meaningfully analyzing travelogues. The Part of Travelogue Tagger (POTT) is used to annotate each token within travelogues, categorizing them into distinct feature categories. By implementing this methodology, a well-organized dataset is generated, encompassing critical data associated with diverse facets of travel experiences. By ensuring that extracted features correspond to the context of travelogues written in the Malayalam language, the Part of Travelogue Tagger improves the model's comprehension of content pertaining to travel. The architectural design of the structured dataset is depicted in Table 1.

Word embedding and one-hot encoding are widely employed methods to convert textual data into a format that machines can comprehend. These techniques are frequently implemented in CSV files containing columns containing textual information.

By employing one-hot encoding, every word or term in the text is converted into a binary vector, with each dimension representing a distinct word or term present in the complete dataset. The corresponding dimension of a word is initialised to 1 if it is present in a specific row; otherwise, it is initialised to 0. This technique generates a sparse matrix representation, which is frequently applied to categorical features. In contrast, word embedding operates as a dense representation method wherein continuous vectors in a high-dimensional space are utilised to encode the semantic significance of words. In contrast to one-hot encoding, word embedding maintains the interconnections among words, thereby facilitating algorithms in comprehending the textual context and semantics. This method is frequently employed in natural language processing endeavours including text generation, language translation, and sentiment analysis, owing to its capacity to discern subtle meanings and semantic resemblances among words. In processing and analysing textual information contained in CSV files, one-hot encoding and word embedding each provide unique benefits, contingent upon the particular specifications and characteristics of the data.

One-hot encoding transforms the names of users, into a binary representation, where each category is represented by a binary vector. This technique is suitable for the 'user' column as it effectively encodes discrete user identifiers without imposing any ordinal relationship between them. On the other hand, word embedding, such as Word2Vec is applied to the other four columns (TT, TM, LC, LT) containing textual information. Word embedding techniques map words or phrases from a vocabulary to high-dimensional vectors, capturing semantic relationships between words based on their contextual usage. This approach is ideal for the text data in these columns, enabling the algorithm to understand and analyze the semantic meaning and context of the text, which is crucial for various natural language processing tasks like sentiment analysis or text classification. By utilizing one-hot encoding and word embedding techniques tailored to the specific characteristics of each column, the CSV data can be effectively encoded for further analysis or machine learning tasks.

Table 1. Sample of feature set from Malayalam Travelogues

No	LC	TT	LT	TM	Locations
1	ചൂട്	ഫ്രണ്ട്സ്	ഫ്രകൃതി	വിമാനം	കശ്മീർ
2	സമ്മർ	ചങ്ങാതി	സാഹസികം	ഫ്ലൈറ്റസ	ലഡാക്കസ
3	വിന്റർ	സഹഫ്രവർതകർ	ഫ്രകൃതി	കപ്പല്	കകാവളം
4	ഫ്ലൈതയം	ഓഫീ്	സാഹസികം	കോഡ്	ക ാവ
5	വസന്തം	ഭാര്യ	ഫ്ലഹകഞ്ചസ	കടല്	വയനാട്

7 AE and Self-supervised Learning Model Preparation

A specialized AE architecture developed for this study, employing pre-processed Malayalam travelogue data. This architecture encompasses both a decoder and an encoder. The encoder takes on the responsibility of transforming input values to the condensed latent space model with decreased dimensions. The decoder is tasked with reconstructing the original data from this compressed representation.

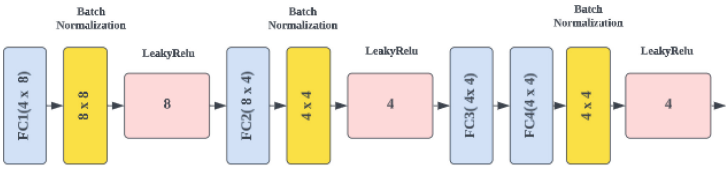


Fig. 4. AE Architecture designed Without any Compression.

In Figure 4, the autoencoder architecture is illustrated, comprising a decoder network and an encoder network. The encoder network operates in two layers to reduce the dimensionality of the input data and extract significant features. An additional layer is applied to the output of the initial layer, which generates a compressed version of the input travelogue and captures intricate patterns. By mirroring the structure of the encoder in the opposite direction, the decoder network increases dimensionality and captures more details. The initial input data is reconstructed in the ultimate output layer.

Rather than utilizing a synthetic dataset, the autoencoder model compresses textual data in the Malaya-lam language using a pre-processed one. Encoder and decoder components are each composed of two dense layers, with batch normalization and leaky ReLU activation techniques incorporated. By configuring the bottleneck layer to have a dimension that is half of the input features, it becomes possible to produce a compressed representation of the textual data.

To optimize the performance of the model, Adam is compiled as the optimizer, the MSE loss function, and accuracy. The primary aim is to acquire knowledge of a condensed rendition of the textual data that retains important attributes such as context, tone, semantics, and Malayalam-specific linguistic complexities.

8 Ensembled Model – AE with Various Machine Learning Approaches

Moving on to the next phase, to assess the efficacy of various machine learning algorithms in predicting destinations based on the provided dataset. The output of the Autoencoder model will serve as the input feature for these algorithms. Figure 5 illustrates a diagrammatic representation of this integration.

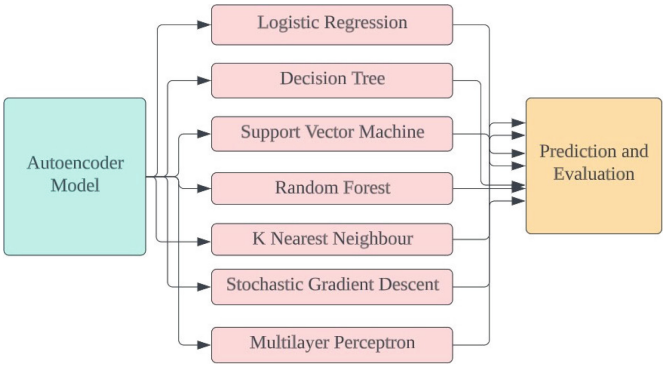


Fig. 5. Block diagram of ensembled model of RS.

9 Evaluation and Performance Metrics

During our investigation, we assessed many ML Models, each selected based on its own advantages and features of efficiency. The SVM looks for the best hyperplane to separate classes. The dataset was divided into segments using decision trees based on attribute values; RF, an ensemble learning technique that used multiple decision trees; KNN, which relied on the majority of nearby data points; SGD, which updated coefficients iteratively to minimize the cost function; and MLP, a kind of feedforward artificial neural network with multiple layers of nodes.

All ML models had rather high accuracy rates, ranging from 83.4% to 91%, where accuracy was the metric of choice. SVM had the lowest accuracy at 84%, while MLP and logistic regression had the best accuracy at 91%.

Table 2. Ensembled model performance evaluation

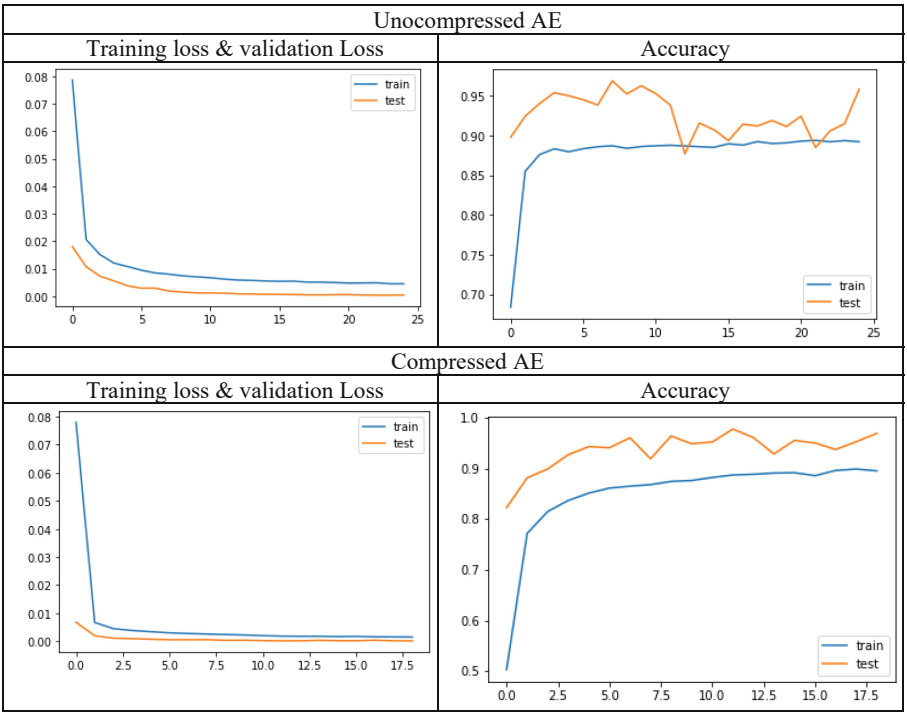
No	Methodology		Performance Accuracy		
1	Uncompressed Encoder		96.84		
2	Compressed Encoder		95.96		
Encoder as input to ML models					
No	Model Name	Accuracy	Precision	F1-score	Recall
1	L.R	87.96	91	87	87
2	D.T	88	86	86	86
3	S.V.M	84	85	84	84
4	R.F	87.66	88	87	88
5	K.N.N	87	89	87	89
6	S.G.D	86	88	86	86
7	M.L.P	87.66	91	88	88

At 88%, the MLP model exhibits the highest F1 Score, suggesting that it maintains an optimal balance between precision and recall. The F1 Scores of alternative models vary between 84% and 87%. KNN attains the maximum recall score of 89%, signifying its capacity to accurately classify a greater percentage of instances as positive. 84% is the lowest recall result for the SVM model. An exhaustive breakdown is provided in Table 2.

10 Result and Discussion

The autoencoder is implemented using two distinct methodologies: one that requires compression and the other that does not. The Encoder Without Compression technique attains a remarkable accuracy rate of 96.84%, effectively preserving most of the original information while capturing crucial features and patterns. To generate a more condensed representation in the bottleneck layer and reduce the dimensionality of input data, the Encoder with Compression method attains a marginally reduced accuracy of 95.96%. A graphical illustration of the accuracy and loss curves for both methodologies is presented in Table 3.

Table 3. Compressed and uncompressed Autoencoder graph.



11 Conclusion

This research paper presents a novel personalised travel recommender model that integrates machine learning algorithms with an autoencoder for the benefit of the Malayalam-speaking community. The model extracts significant attributes from unstructured travelogues, including travel type, travel mode, location climate, and specific designations. The model demonstrates substantial improvements in accuracy when compared to conventional approaches, with the most effective models being Logistic Regression and MLP. Subsequent endeavours will be directed towards enhancing the model's adaptability and precision through the incorporation of user feedback and realtime data.

References

1. Basha, M.J., Vijayakumar, S.: Advancements in natural language processing for text understanding. In: E3S Web Conference, vol. 399, p. 04031 (2023). https://doi.org/10.1051/e3sconf/202339904031
2. Sebastian, M.P., Santhosh Kumar, G.: Malayalam natural language processing: challenges in building a phrase-based statistical machine translation system. ACM Trans. Asian Low-Resour. Lang. Inf. Process. **22**(4), Article 117 (2023). https://doi.org/10.1145/3579163
3. Thandil, R.K., Basheer, M., VK, M.: Natural language processing of Malayalam text for predicting its authenticity. Soc. Sci. Res. Network (2021). https://doi.org/10.2139/ssrn.3972525
4. Roy, D., Dutta, M.: A systematic review and research perspective on recommender systems. J. Big Data **9**(1) (2022). https://doi.org/10.1186/s40537-022-00592-5
5. Sawicki, J., et.al.: The state of the art of Natural Language Processing a systematic automated review of NLP literature using NLP techniques. Data Intelli. (2023).https://doi.org/10.1162/dint_a_00213
6. Lu, F., Zhang, W.: Personalized travel recommendation based on the fusion of TGI and POI algorithms. Wirel. Commun. Mob. Comput. **2022**, 1–9 (2022). https://doi.org/10.1155/2022/4058729
7. Zhang, G., Liu, Y., Jin, X.: A survey of autoencoder-based recommender systems. Front. Comp. Sci. **14**(2), 430–450 (2019). https://doi.org/10.1007/s11704-018-8052-6
8. Ferreira, D., et al.: Recommendation system using autoencoders. Appl. Sci. **10**(16), 5510 (2020). https://doi.org/10.3390/app10165510
9. Ding, X., et. al.: Combining autoencoder with adaptive differential privacy for federated collaborative filtering. In: Wang, X., et al. Database Systems for Advanced Applications. DASFAA 2023. Lecture Notes in Computer Science, vol 13943. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-30637-2_44
10. Guo, S., Liao, X., Li, G., Xian, K., Li, Y., Liang, C.: A hybrid recommender system based on autoencoder and latent feature analysis. Entropy **25**(7), 1062 (2023). https://doi.org/10.3390/e25071062
11. Bougteb, Y., Ouhbi, B., Frikh, B., Zemmouri, E.M.: A deep AutoEncoder-Based hybrid recommender system. Int. J. Mob. Comput. Multimed. Commun. **13**(1), 1–19 (2022). https://doi.org/10.4018/ijmcmc.297963
12. Nitin, K.: Product recommendation system using machine learning. Int. J. Innov. Res. Comput. Commun. Eng. **9**(11) (2021). https://doi.org/10.2139/ssrn.4245401

13. Parvattikar, S., Parasar, D.D.: Recommendation system using machine learning. In: Proceedings of the International Conerence on Recent Advances in Computational Techniques (2020).https://doi.org/10.2139/ssrn.3702439
14. Muneer V.K, Mohamed Basheer, K.P.: A hybrid travel recommender model based on deep level autoencoder and machine learning algorithms. J. Adv. Zool. **44**(5), 1220–1229 (2023). https://doi.org/10.53555/jaz.v44i5.3571

Future-Proofing Careers: The AI Era and the Imperative of Reskilling

Shristi Agarwal, Garima Jaiswal, Mala Saraswat(✉), and Drishti Rai

Bennett University, Greater Noida, Uttar Pradesh, India
{E20CSE480,2E20CSE495}@bennett.edu.in, malasaraswat@gmail.com

Abstract. Foundational disciplines of the wide, interdisciplinary study of Artificial Intelligence (AI) include statistical inference and data, software application, rational analysis, information theory neurology, linguistics, and information theory. The way AI is implemented is a crucial factor in how it will change the workforce. It follows that AI systems can grow. Unlike earlier systems that typically involved human labor, these systems are intelligent and capable of learning beyond pre-defined structures and logic. This creates the opportunity for AI systems to take on formerly reserved for humans, experience, and understanding tasks. Organizations' prospective use of cutting-edge AI technologies raises serious questions concerning both security and the welfare of their workers. In the future, technology will probably replace a significant portion of the workforce, which, if not effectively managed, might result in shockingly high unemployment rates throughout society. The skills required today might not be significant in the future as much. Our research focuses on the future of work and how artificial intelligence (AI) will change it, particularly in terms of job displacement, reskilling, and upskilling. Everyone agrees that AI is profoundly changing the workplace and is poised to redefine what it means to be a worker. Artificial intelligence (AI) has the potential to boost production and efficiency, but there is also a concern that it may lead to the replacement of human labor as computers take over tasks that they were previously performing.

Keywords: Artificial Intelligence · Automation · Job Displacement · Reskilling · Upskilling · Workforce Development

1 Introduction

The goal of the rapidly progressing field of AI, which is part of computational science, develops artificially intelligent machines that are designed for carrying out responsibilities that usually call for human intelligence. Use cases of AI are decision-making, natural language processing, visual perception, speech recognition, and technology for robotics. As shown in Fig. 1, A wide range of sectors, including healthcare, travel, financial services, and entertainment is possessed by AI [1]. A substantial impact has already been made on the development of technology and science. The process of creating artificially intelligent algorithms and models capable of imitating human cognitive processes like

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 460–470, 2025.
https://doi.org/10.1007/978-3-031-75167-7_37

perception, learning, and reasoning is AI. Machine learning is a branch of AI that uses a lot of knowledge to teach algorithms to make predictions or judgments [2]. By using neural networks Deep learning duplicates the workings of the human brain this has been helpful in activities like image recognition as well as speech recognition. A variety of benefits could result from AI, including enhanced healthcare consequences, more effective transportation systems, and increased technological findings. Since machine learning models and algorithms have the potential to reinforce or simply magnify pre-existing social biases and inequalities, bias, and discrimination are also important problems in AI. This emphasizes the significance of creating fair and impartial AI systems and making sure that they are used in a way that supports socially equitable policies. It is essential to make sure that these systems are created and applied responsible and ethical way as AI becomes more pervasive in society.

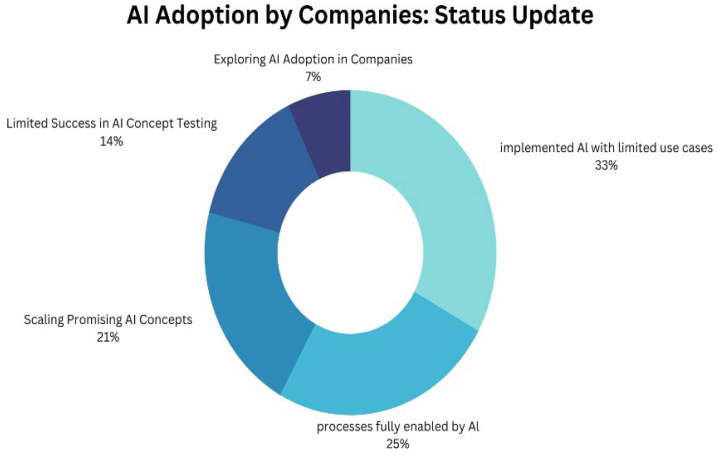


Fig. 1. AI Adoption by Companies: Status Update

2 Related Work

Acemoglul et al. [1] examined how AI affects economic growth, modeling it as the most recent type of automation that can automate previously inconceivable tasks. The study seeks to provide answers to several concerns, including how AI will impact economic growth, how income is distributed between labor and capital, and what effect firm-level factors have on the relationships between AI and growth. The paper tries to create an agenda for the area by focusing on several important topics involving AI and economic development. One theme that emerges is based on Baumol's "cost disease" idea, which contends that rather than what we are strong at, what is necessary but difficult to improve upon may be what limits growth. Overall, the study depicts how AI has the power to regenerate the economy, and it stresses the fact that its effect on employment and wealth distribution needs to be taken into consideration. Through the lens of an industry perspective, Hammer et al. [2] examined the effects of AI on employment via a swift

study of reports from government and business sources that were released till August 2017. According to the study, automation of some job and desk duties and the substitution of regular and non-regular tasks would likely have an impact on employment soon. It is anticipated that new collaborations between people and machines would necessitate adjustments to employable skills. The report emphasizes the necessity for enterprises to rethink their human resource function to correspond with the realities of AI and advocates for concerted solutions that go beyond organizations' and industries' boundaries to include other pertinent stakeholders.

Manyika et al. [3] evaluated the likelihood of computerization for 702 specific occupations using a novel methodology based on a Gaussian process classifier to investigate the computerization susceptibility of jobs. With a prime focus on the various jobs at risk and the association among the earnings, educational attainment, and probability of computerization, the authors deduced this methodology to foresee the expected impact of future computerization on the outcomes of the US labor market. According to the findings, 47% of the US jobs could be mechanized in the coming years. The authors point out that employment with fewer educational requirements and lower pay carries a higher risk. Focusing on the future work implications, Illanes et al. [4] focused to examine the extensive field of AI and determine its effect on job opportunities in the future. The study depicts how machine learning (ML) has enabled applications of AI that have already assimilated into human day-to-day life and have a profound impact on socioeconomic structures.

Lane et al. [5] examined how robots and AI may change the nature of labor. With consideration for the socio-technical and financial limitations of its use, it analyses the idea of technological singularity and its relevance to robots and AI. The article compiles case studies and data from the International Federation of Robotics to demonstrate the "take-up" of robots. It quickly examines the two "tests" of artificial "intelligence" set out by the father of computer science Alan Turing and discusses their usefulness in modern applications. This study looks at several issues of public policy concerning AI and robotics. Based on that it concludes that the technological singularity is not at all imminent. In this research paper, Pradhan et al. [6] has discussed the workforce challenges posed by automation and digitization, as well as the critical need for retraining and reskilling of mid-career, and mid-aged workers. According to a recent McKinsey poll cited in the article, executives increasingly view retraining and "upskilling" current employees as a top corporate priority (Table 1).

Table 1. Related work can be concluded in following table.

Reference	Main Objective	How This Research Paper is Different
[1]	To analyze the potential impact of AI on economic growth	While our research article investigates the effects of AI on the labor market and considers ways to reduce job displacement through reskilling and upskilling programs, this paper focuses explicitly on the impact of AI on economic growth
[2]	To examine the potential impact of AI on job losses	While our research article examines methods to lessen job displacement through reskilling and upskilling programs, this paper focuses explicitly on the risk for employment losses caused by AI
[3]	To estimate the susceptibility of jobs to computerization	While our research article addresses the effects of AI on the labor market and considers ways to reduce job displacement through reskilling and upskilling programs, this paper focuses explicitly on the computerization susceptibility of jobs
[5]	To explore the impact of AI on the workforce and the potential for new job roles	While our research article investigates methods to lessen job displacement through reskilling and upskilling programs, this paper addresses the possibilities for new work roles owing to AI
[4]	To examine the effects of AI on the workforce and economy	While the McKinsey report covers a wider variety of subjects relating to the impact of AI on employment, this article focuses particularly on job displacement, reskilling, and upskilling measures in response to AI disruption
[6]	To analyze the need for reskilling in response to automation and identify effective strategies for reskilling workers	While the emphasis of this study is on reskilling in reaction to automation, our research paper focuses on the effects of AI on the workplace and analyses upskilling

3 AI-Capable Jobs

The technology known as AI has fundamentally altered the way we work, communicate and do business. As the usage of AI spreads across various sectors, worries about the displacement of human jobs are becoming more prevalent [3]. AI can, however,

do time-consuming, monotonous activities, freeing up human attention for professions that require creativity, critical thought, and problem-solving. We'll examine the types ofactivities that AI is capable of throughout this piece along with the ways that might change the way individuals work in the future [4]. AI can perform a variety of tasks, including entering information, analysis of information, photo acknowledgment, natural language processing, and making choices. One of the primary objectives of AI has been to automate everyday and recurrent tasks, especially entering information as well as interpretation.

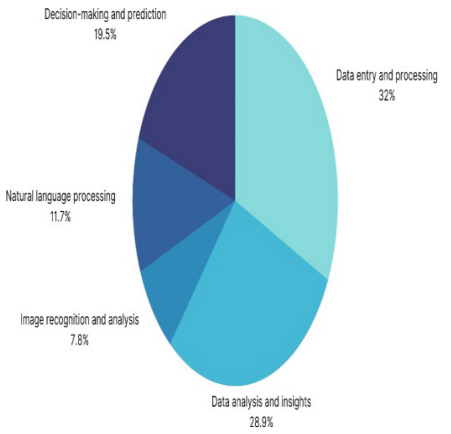


Fig. 2. AI-Capable Jobs

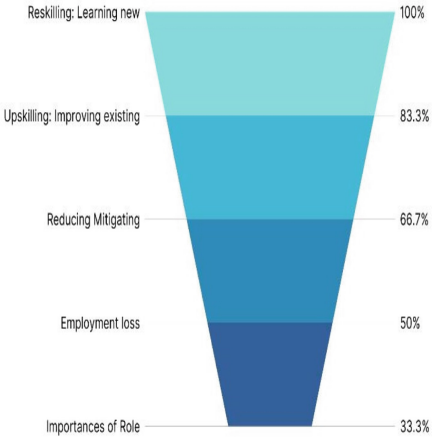


Fig. 3. Reskilling and Upskilling needed

Another application for AI is the automatic identification of images. AI could have been trained to recognize individuals, things, as well as sentiments in videos and images. In industries like security and retail, where cameras may be used to monitor consumer habits and identify possible risks, such information is extremely useful [4]. AI could be applied in this area, much like natural language processing, which permits computers to understand and interpret words spoken by humans. As shown in Fig. 2, AI is used in a variety of applications. It is capable of being utilized, for instance, to improve the oversight of supply chains, anticipate stock values, and diagnose medical ailments. It is crucial to keep in the forefront that AI was only as effective as the information that it has been educated on. The AI will make decisions that are incomplete or biased if the data is biased or insufficient. Consequently, it's crucial to guarantee that the information that is used to train AI is reliable and impartial [5]. Even though AI can completely transform how humans are employed, there are nonetheless pertaining regarding the way technology will affect the job market in the future. Certain researchers predict that AI will eventually substitute for human work, particularly that of people who perform monotonous as well as regular tasks.

Employees in industries like banking and healthcare are capable of being trained to work alongside powered by AI systems in a manner that's comparable to this, increasing consequences and enabling workers to formulate more informed choices. So, in the long run, AI possesses an opportunity to transform the way humans work, however, it is critical to focus on reskilling and upskilling strategies for the purpose to limit the

possible adverse effects on the economy. Employees possess a greater opportunity to focus on activities that require creativity, critical thought, and problem-solving while routine and monotonous tasks are taken care of through AI. However, it's imperative to guarantee that the information utilized to train AI is trustworthy and unbiased. In the modern era of AI, how work is done ultimately depends on our capacity to change and adapt.

4 The Role of Reskilling and Upskilling in Mitigating Job Displacement

A lot of professions are becoming obsolete because of technological improvements, automation, and increased global competition, which is having a significant impact on the workplace. This trend has been intensified by the COVID-19 pandemic, as many industries are being negatively affected by the economic crisis. In this setting, reskilling and upskilling have developed into crucial techniques for preventing job displacement and putting workers in the best possible position for the future of work.

As shown in Fig. 3, there is a huge impact of upskilling and reskilling. The procedures of obtaining new skills and information to perform a different job or enhance one's present job performance are referred to as reskilling and upskilling, respectively [7]. These tactics are essential since technology is altering the way we work and making some abilities that were once in high demand obsolete. Furthermore, as automation, AI, and machine learning become more common, there will certainly be a continuing increase in the need for new talents. Recognizing the skills required for a given industry or position and offering training and educational programs to assist employees in acquiring these abilities are both parts of the reskilling and upskilling process. There are various ways of giving courses, including online learning, classroom instruction, and on-the-job training.

In the process of upskilling and reskilling employers are essential. To help their employees gain the skills required by their business or sector, they must identify those needs and invest in training and education programs. Also, employers should give staff members chances to use their newly acquired talents in the workplace and foster a culture that encourages lifelong learning. The creation of regulations and incentives that encourage companies to invest in workforce development as well as funding for training and education programs are other ways that governments can assist in reskilling and upskilling.

There are several tactics that can be used to achieve the goal of reskilling and upskilling the workforce. Governments sponsor training and education initiatives and work with educational institutions to produce easily accessible, scalable training programs. Identifying the skills needed in the job market, creating training programs to teach these skills, and giving workers access to these programs are all parts of the reskilling and upskilling process [8]. To achieve the goal of upskilling and reskilling is necessary to collaborate with a variety of stakeholders, including governments, employees themselves, businesses, and educational institutions. The advantages of reskilling and upskilling are numerous and can result in increased social mobility, greater pay, better career opportunities, and improved job security.

5 Current Jobs are Less Affected by AI and New Job Opportunities Coming into Demand

By eliminating the existence many jobs are becoming automated in this age of AI. Some professions are less susceptible to automation as more jobs are becoming available. Following are a few examples of recent employment that has not been significantly disrupted by AI, as well as some high-demand new job openings:

As clearly visible in Fig. 4, AI has less of an influence on jobs:

- Workers in the medical field: Even though AI is making inroads into the healthcare industry, doctors and other healthcare professionals are still in high demand. These experts range from physicians to nurses to caregivers [9]. AI cannot replace the human touch and critical thinking skills required in the healthcare sector.
- Teachers and educators: Teaching is less likely to be an automated job because it takes a lot of interpersonal connection and mental capacity. Teachers are in high demand online as e-learning has grown in popularity.
- Creative occupations: It's less likely that AI will replace creative occupations like writing, painting, and graphic design [10]. These positions call for a high degree of originality and inventiveness that robots cannot match.
- High-skilled trades: It is less likely that manual labor-intensive jobs like woodworking, and plumbing, including the welding procedure will be mechanized. These jobs require a high level of dexterity and analytical skills that are indispensable to machines.

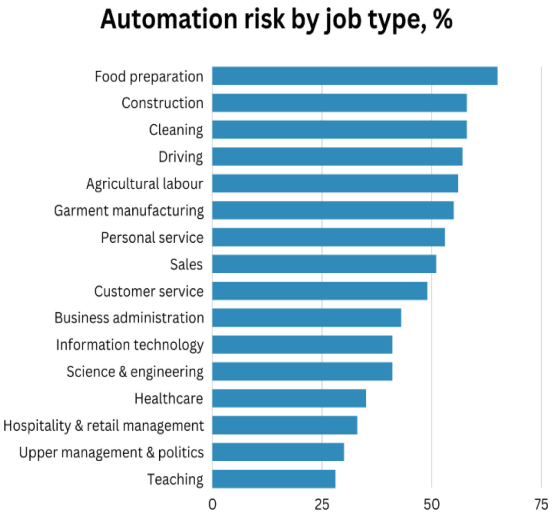


Fig. 4. Automation risk by job type as per percentage

Newly available positions in great demand:

- Data scientists and analysts: With the amount of data that AI generates, there is an increasing demand for professionals who can sort through it and interpret it. When making decisions based on insights generated by AI, organizations can profit from the experience of data scientists and analysts.
- experts in cyber security: Cybersecurity flaws are becoming more prevalent as AI and the Internet of Things (IoT) are utilized more frequently. There is a rising need for cybersecurity skills to protect firms from internet threats.
- AI educators: They are in greater demand as the field of AI develops due to the complexity of these systems. The ethical and responsible design and implementation of AI systems can be aided by the understanding of these individuals.
- Customer experience specialists: They are in high demand because they can offer personalized and human-like interactions with clients in the age of AI-powered chatbots and virtual assistants.
- Environmental specialists: With the concentration on ecological responsibility expanding, this is a demand for professionals who can help firms implement sustainable concepts. There is a considerable demand for experts in energy conservation as well as administrators of sustainability programs.

While some occupations risk being automated out of existence, there are a lot of others that are less at risk. Several new positions are in high demand including cybersecurity professionals, data scientists, and AI trainers. For individuals to remain competitive in the labor market, they must be able to adapt and learn new skills as the workforce changes.

6 Policy Implications for a Fair and Inclusive Transition to an AI-Powered Economy

A fair and inclusive transition to an AI-driven economy must be planned for by policymakers. Programs for retraining and upgrading skills are crucial, as are initiatives to open new employment prospects. To ensure that the advantages of automation are dispersed fairly, policymakers should also consider measures like a basic income. It is important to consider ethical issues, such as data security and accessibility requirements. Invest in AI education and training: The need for people with Alrelated abilities will increase as Al becomes more commonplace in the workplace. The absence of workers with these abilities, however, could cause the skills gap to increase and result in job displacement for those who lack them. Governments can solve this issue by instructing staff members on how to create, build, and maintain Al systems by providing funds for education and training [11]. This can include both formal training initiatives like college degrees in computer science or AI and informal initiatives like online courses or apprenticeships. To provide uniform training courses suitable for the Al-powered economy, governments may also work with enterprises and trade associations.

- Support research and development: Since AI is a rapidly developing technology, continual study, and development are necessary to enhance its capabilities and guarantee its moral and ethical application. Contribution can be done to projects dedicated to AI research like public-private partnerships or university research centers if government

encourages it. Tax benefits or other forms of financing can also be offered so that companies can be incentivized and can participate in AI research and development. It can be made sure by the government that technology can be made for the benefit of society by giving support to AI research and development.

- Promote ethical AI: If they are not developed and put into use ethically, AI systems have the potential to promote prejudice and discrimination. Governments should adopt policies and rules that support moral AI practices to guarantee that AI is employed for the benefit of society. This may entail setting criteria for AI system accountability and transparency, guaranteeing that AI systems are free from prejudice and discrimination, and encouraging the creation of AI systems that are in line with social and environmental objectives. Governments can ensure that the advantages of AI are equitably distributed and that the technology is utilized for the greater good by encouraging ethical AI practices.
- Address the digital divide: A gap between those who can access technology and the ones who cannot access technology is known as the digital divide. Governments can work in closing the gap by providing access to technology and resources including computer software and hardware, and high-speed internet so that it can provide an opportunity for everyone to get engaged in an AI-powered economy.

This may involve sponsoring initiatives that give low-income homes, schools, and other public institutions free or inexpensive access to technology [12]. Governments can help ensure that no one is left behind as the economy grows more digital by tackling the digital divide.

- Create a social safety net: Some workers will probably lose their jobs as the economy shifts to one powered by AI. To transition new careers of displaced workers government should create some social safety nets which offer healthcare benefits, and financial aid [13]. To ensure that workers have the skills necessary to engage in the AI-powered economy, governments can also collaborate with businesses to develop job retraining programs. To contribute to an AI-powered economy, the government can establish a social safety net for an equitable and fair transition.
- Foster collaboration: Some workers will probably lose their jobs as the economy shifts to one powered by AI. Governments should create a social safety net for these people that offers financial aid, healthcare benefits, and job training programs that aid displaced workers in transitioning to new careers [14].

7 Case Studies of Successful AI Adoption and Workforce Development Initiatives

Numerous companies are seeing a requirement for additional training and education for their workers to remain relevant when Intelligence grows increasingly prevalent throughout work environments. The methodologies and strategies which have proved productive in assisting employees as well as fostering expansion in the economy are going to be highlighted when we look at many case studies of successful adoption of AI and workplace development projects in this section.

- Amazon's employment of robots with AI within its shipment fulfillment centers constitutes one of the finest and most popular examples of effectiveness in the adoption of AI. Although some jobs have been lost due to advancements in technology, Amazon additionally has put together several other workplace growth initiatives to promote its staff [15].
- The usage of AI-powered chatbots in customer support is yet another instance of the effective adoption of AI. Businesses like Capital One and American Express have started using chatbots to respond to frequently asked queries from customers, freeing up human employees to handle more challenging tasks.
- AI is being used in the healthcare industry to improve patient experiences and cut costs. The Cleveland Clinic, for example, has implemented a machine learning-powered AI system to predict treatment outcomes while identifying patients with heightened risks who may require further precautions.
- The financial sector is implementing AI to improve fraud detection and reduce operating costs. For instance, JPMorgan Chase has developed a machine learningpowered AI system that uses patterns of suspicious behavior to identify fraudulent transactions [16].
- AI is also being used in the manufacturing sector to boost productivity and cut expenses. For instance, General Electric has put in place a system that is AI-powered and uses predictive analytics to improve maintenance schedules and reduce downtime.

Each of these circumstances necessitated a comprehensive strategy that considered the needs and worries of the workforce to properly integrate AI. Organizations can cooperate to provide a more hospitable and inclusive environment that fosters economic growth and guarantees that people have the resources and tools they need to prosper in an AI-powered economy [17]. Sharma et. al in their work proposed LSTM for sentiment analysis of tweets [18]. Some more general trends and techniques have been successful in promoting AI adoption and workforce development in addition to these case studies. Gamification and other innovative training techniques are one trend being used to make learning more interesting and interactive.

8 Conclusions

In summary, it is projected that the development of AI will significantly affect how people operate in the future. Understanding the ramifications, coming up with solutions to lessen job displacement, and enabling a fair and equitable transition to an AI-powered economy are critical as AI and automation continue to eliminate jobs across industries. According to findings, programs to reskill and upskill people are essential if they are to be able to adapt to new and upcoming AI-driven employment. Such programs must be reasonable and accessible to be inclusive and reach all workers, particularly those from underserved communities. Initiatives to reskill and upskill people must also consider contemporary developments like the gig economy, remote work, and hybrid employment. The shift to an AI-driven economy must be shaped by policymakers. They must create regulations that support inclusive and fair economic growth and support employees who have been adversely affected by job displacement. This entails funding programs for learning and training, enhancing access to social safety nets, and pressing businesses to give reskilling

and upskilling initiatives a top priority. Case studies concerning effective employment growth as well as the implementation of AI initiatives provide valuable information about how firms can successfully handle the challenges of working environments powered by AI. By prioritizing working together, spending on staff development, and embracing an innovation culture, businesses may not only survive but also thrive in an AI-powered mark.

References

1. Acemoglu, D., Restrepo, P.: Automation and new tasks: the implications of the task content of production for labor demand. J. Econ. Perspect. **33**(2), 3–30 (2018)
2. Hammer, A., Karmakar, S.: Automation, AI and the future of work in India. Empl. Relat.: Int. J. **43**(6), 1327–1341 (2021)
3. Manyika, J., et al.: Jobs lost, jobs gained: what the future of work will mean for jobs, skills, and wages (2017)
4. Illanes, P., Lund, S., Mourshed, M., Rutherford, S., Tyreman, M.: Retraining and reskilling workers in the age of automation. McKinsey Global Institute (2018)
5. Lane, E.Y.: A new age of reskilling: why learning matters more than ever.
6. Pradhan, I.P., Saxena, P.: Reskilling workforce for the AI age: challenges and the way forward. In: The Adoption and Effect of AI on Human Resources Management, Part B, pp. 181–197. Emerald Publishing Limited (2023)
7. Manyika, J., Sneader, K.: AI, automation, and the future of work: ten things to solve for (2018)
8. Madakam, S., Holmukhe, R.M., Jaiswal, D.K.: The future digital work force: robotic process automation (RPA). JISTEM-J. Inf. Syst. Technol. Manag. **16** (2019)
9. Skidelsky, R., Nan, C. (eds.) Work in the Future: The Automation Revolution. Springer Nature (2020)
10. Donepudi, P.K.: Application of AI in automation industry. Asian J. Appl. Sci. Eng. **7**(1), 7–20 (2018)
11. Jagannathan, S., Ra, S., Maclean, R.: Dominant recent trends impacting on jobs and labor markets-An Overview. Int. J. Train. Res. **17**(sup1), 1–11 (2019)
12. Ra, S., Shrestha, U., Khatiwada, S., Yoon, S.W., Kwon, K.: The rise of technology and impact on skills. Int. J. Train. Res. **17**(sup1), 26–40 (2019)
13. Germain, M., Grenier, R., Moats, J.: Preparing for the future of work and the development of expertise. Expert. Workcurrent Emerg. Trends, 197–224 (2020)
14. Szinovacz, M.E.: Introduction: the aging workforce: challenges for societies, employers, and older workers. J. Aging Soc. Policy **23**(2), 95–100 (2011)
15. Miao, M., Holmes, W., Huang, R.: And Hui Zhang A guidance for policymakers. UNESCO Publishing AI and education (2021)
16. Thakkar, D., Kumar, N., Sambasivan, N.: Towardsan AI-powered future that works for vocational workers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2020)
17. Pedro, F., Subosa, M., Rivas, A., Valverde, P.: AI in education: challenges and opportunities for sustainable development (2019)
18. Sharma, A., Khajuria, D., Rani, R., Jaiswal, G., Saraswat, M.: LSTM based model for classification of tweets. In: 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), pp. 1–7. IEEE (2023)

Author Index

A
Abeysinghe, Binu I-311
Abhilash, Goli I-379
Agarwal, Himani II-179
Agarwal, Pooja I-238
Agarwal, Shristi II-460
Agraval, Rahul I-20
Agrawal, Sneha I-171
Ahlawat, Anil II-332
Ahuja, Laxmi II-18
Akshayalakshmi, V. K. I-346
Alam, Arham I-426
Alam, Golam Rabiul II-30
Alam, Sheba I-335
Alvi, Syed Ahad Murtaza II-303
Anand, Khushi I-394
Arora, Medha II-393
Ashraf, Faisal Bin II-30
Ayub, Rashid II-303

B
Balamurali, B. T. I-311
Bansal, Poonam I-90, I-143, I-323, I-405, I-418, I-440, II-191
Barathi Ganesh, H. B. I-66
Barde, Dewanshu I-20
Barik, R. C. I-365
Basu, Joyanta I-195
Basu, Tapan Kumar I-195
Bawa, Rajesh K. II-42
Bedi, Punam I-301
Behura, Subhranshu I-426
Bhagat, Dhananjay I-20
Bharti, Surbhi I-335, II-367, II-393
Bhatt, Nidhi I-227
Bhatt, Shobha I-227, I-253

C
Channawar, Rashi I-20
Chantilyan, M. II-104
Chaudhary, Himanshi II-332
Chauhan, Aditi II-426
Chauhan, Amisha II-438
Chhabra, Niharika I-394
Chopra, Ankita II-118

D
Deekshitha Reddy, M. II-407
Deepa, R. I-3
Desai, Shravani II-292
Dev, Amita I-143, I-323, I-405, I-440, II-407
Dhak, Bharat S. II-131
Dinesh, P. II-104
Dwivedi, Sakshi II-223
Dwivedi, Vimal I-55

E
Easwarakumar, K. S. I-346

G
Gabrani, Goldie I-426
Gambhir, Pooja I-143
Gandhi, Rashmi II-438
Garg, Geetanjali I-227
Garg, Nikita I-55
Geethakarthi, A. II-56
Goel, Nimish I-210
Gomathi, C. II-104
Gopal, Piyush II-351
Gupta, Anushka I-394
Gupta, Jahanvi II-262
Gupta, Manisha I-440
Gupta, Shivam II-272
Gupta, Suraj Kumar II-272

H
Habib, Adria Binte II-30
Harsha Prada, M. D. I-78
Hasan, Zafarul II-303
Hossain, Muhammad Iqbal II-30

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

A. Dev et al. (Eds.): AIST 2023, CCIS 2268, pp. 471–473, 2025.
https://doi.org/10.1007/978-3-031-75167-7

J
Jagtap, Srushti I-210
Jain, Bhawna I-394, II-281
Jain, Shradha I-159
Jaisinghani, Omprakash A. II-131
Jaiswal, Garima II-407, II-460
Jalala, P. V. I-451
Jalan, Deepanshi I-159
James, Jesin I-311
Jana, Sandip I-195
Jandhyala, Ananya I-238
Jha, Prerna I-335
Jindal, Meemansa II-319
Jindal, Soumya II-319
Joshi, Anuradha II-292

K
Kadian, Karuna I-55
Kakodkar, Ishan G. S. I-210
kale, Prathamesh I-20
Kambli, Gitesh I-116
Karmakar, Amiya I-195
Kathiroli, Panimalar II-351
Kaur, Sukhandeep I-55
Khan, Shakir II-303
Khandale, Kalpana B. I-284
Khatter, Kiran I-379
Khedikar, Kanchan A. II-143
Kikan, Vaishali I-159
Komal, I-184
Koul, Apeksha II-42
Krishna, Kata Venkata Sai I-379
Krishnan, Gokul I-238
Kukkar, Yashika II-426
Kumar, Ankit I-301
Kumar, Ashwani I-159
Kumar, Ashwni I-335, II-157, II-367, II-393
Kumar, Avadhesh II-212
Kumar, Rituraj I-131
Kumar, Santosh II-417
Kumar, Tanuj I-365
Kumar, Vishwajeet I-365
Kumar, Yogesh II-42
Kumari, Anamika II-367
Kumari, Ritika II-191
Kumari, Tanya II-191

L
Laxmaiah, Bagam II-237
Liu, Junchen I-311

M
Madhu Priya, G. I-78
Mahender, C. Namrata I-284
Maheshwari, Abhinav II-426
Maity, Amit I-116
Makhnytkina, Olesia I-102
Malik, Kiran I-90
Malik, Naveen II-417
Manikandan, V. II-104
Manisha, II-212
Marwah, Khyati I-272
Matveev, Yuri I-102
Mehra, Saumya II-281
Mehrotra, Monica I-40
Mishra, Atul I-426
Mittal, Puneet I-272
Moghe, Snehal II-3
Moid, Abdul II-246
Muneer, V. K. II-449
Murpani, Ronit II-292

N
Nagrath, Preeti I-184, II-426
Narayanaswamy, Akash I-66
Nayak, Lakshmish I-131
Nema, Malay Kumar I-131
Niruban, R. I-3
Nisha, B. A. I-346

O
Oza, Jay I-116

P
Panda, Sivananda II-351
Pandey, Shashank II-223
Pareek, Piyush Kumar II-143
Pasrija, Yash II-438
Paudel, Shushil I-210
Phutela, Nishtha I-426
Posokhov, Pavel I-102
Preethi, M. I-451

R
Raghuveera, T. I-346
Rai, Drishti II-460
Raju, K. Srujan II-237
Ramteke, P. L. II-131
Ramteke, Prabhakar II-201
Ramudu, Bharathi I-131

Rani, Ritu II-407, II-438
Ranvijay, II-223
Rao, Dhruthi I-238
Rao, Mohit M. II-167
Rastogi, Sugandh I-171
Rathod, Nachiket II-201
Relan, Devanjali I-379
Revathi, R. II-85
Rindhya, C. I-66
Rizvi, Syed Afzal Murtaza II-380
Rizvi, Syed Bilal Abbas II-380

S
Sahu, Sakshi I-335
Saini, Dharmender II-426
Saini, Ishika I-184
Saini, Madan Lal II-118
Sakarkar, Gopal II-65
Sandur, Anshu I-238
Saraswat, Mala II-460
Sasankar, Prachi II-65
Sawarkar, Kalyani I-20
Serajuddin, Mohammad II-303
Sethi, Nandini I-418
Shah, Akash I-40
Shah, Sheel I-66
Sharma, Anunya I-90
Sharma, Arun II-303
Sharma, Astha II-157
Sharma, Narendra II-246
sharma, Ravish I-171
Sharma, Richa I-301
Sharma, Sukhwinder I-272
Shete, Sakshi II-292
Shrivastava, Virendra Kumar II-167
Singh, Harsh Pratap II-3
Singh, Himanshu II-262
Singh, Vijay Bhan II-223
Singhal, Mayank I-184
Singhal, Shweta II-179
Sinha, Anunay I-365
Sivasubramanian, Arrun I-66
Skrylnikov, Stepan I-102
Sowpari, Niranjana I-418
Sree Sharmila, T. I-3
Sri Saila, A. S. I-78
Suhani, I-405
Sumathi, V. P. II-56
Sumra, Vinisha II-417

T
Talmale, Mayur I-20
Taneja, Divya I-253
Thakur, Ayush II-18
Thandil, Rizwana Kallooravi I-451
Tiwari, Apoorva I-55
Trivedi, Shivani I-171, II-262, II-272
Tuli, Anandita I-159
Tyagi, Kajal II-332
Tyagi, Vartika I-323

U
Udgire, Suvarna II-292

V
Vadali, Geetika II-407
Vanisri, Kesana II-237
Varshney, Sapna I-40
Venkata Subramani, A. T. I-78
Verma, Ananya I-394
Verma, Divya II-367
Verma, Neetu II-223
Vijayalakshmi, V. II-351
Vishwakarma, Harsh II-426

Y
Yadav, Ankit II-272
Yadav, Dileep Kumar II-212
Yadav, Sapna II-380

Z
Zahid, Rahbar I-451
Zeeshan, Mohammad I-184