Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 34 (S1) 2026

A Special Issue devoted to

Technologies and Applications for Sustainable Development

Guest Editors

I. Siva, K. Reddy Madhavi, and Nagaraj Ramrao

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

PERTANIKA JOURNAL OF SCIENCE & TECHNOLOGY

About the Journal

Overview

Pertanika Journal of Science & Technology is an official journal of Universiti Putra Malaysia. It is an open-access online scientific journal. It publishes original scientific outputs. It neither accepts nor commissions third party content.

Recognised internationally as the leading peer-reviewed interdisciplinary journal devoted to the publication of original papers, it serves as a forum for practical approaches in improve quality on issues pertaining to science and engineering and its related fields.

Pertanika Journal of Science & Technology currently publishes 6 issues a year (*February, April*, *June, August, October, and December*). It is considered for publication of original articles as per its scope. The journal publishes in **English** and it is open for submission by authors from all over the world.

The journal is available world-wide.

Aims and Scope

Pertanika Journal of Science & Technology aims to provide a forum for high quality research related to science and engineering research. Areas relevant to the scope of the journal include: bioinformatics, bioscience, biotechnology and bio-molecular sciences, chemistry, computer science, ecology, engineering, engineering design, environmental control and management, mathematics and statistics, medicine and health sciences, nanotechnology, physics, safety and emergency management, and related fields of study.

History

Pertanika Journal of Science & Technology was founded in 1993 and focusses on research in science and engineering and its related fields.

Vision

To publish a journal of international repute.

Mission

Our goal is to bring the highest quality research to the widest possible audience.

Quality

We aim for excellence, sustained by a responsible and professional approach to journal publishing. Submissions can expect to receive a decision within 90 days. The elapsed time from submission to publication for the articles averages 180 days. We are working towards decreasing the processing time with the help of our editors and the reviewers.

Abstracting and Indexing of Pertanika

Pertanika Journal of Science & Technology is now over 33 years old; this accumulated knowledge and experience has resulted the journal being abstracted and indexed in SCOPUS (Elsevier), Journal Citation Reports (JCR-Clarivate), EBSCO, ASEAN CITATION INDEX, Microsoft Academic, Google Scholar, and MyCite.

Citing Journal Articles

The abbreviation for Pertanika Journal of Science & Technology is *Pertanika J. Sci. & Technol.*

Publication Policy

Pertanika policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. It prohibits as well publication of any manuscript that has already been published either in whole or substantial part elsewhere. It also does not permit publication of manuscript that has been published in full in proceedings.

Code of Ethics

The *Pertanika* journals and Universiti Putra Malaysia take seriously the responsibility of all of its journal publications to reflect the highest in publication ethics. Thus, all journals and journal editors are expected to abide by the journal's codes of ethics. Refer to *Pertanika*'s **Code of Ethics** for full details, or visit the journal's web link at http://www.pertanika.upm.edu.my/code_of_ethics.php

Originality

The author must ensure that when a manuscript is submitted to *Pertanika*, the manuscript must be an original work. The author should check the manuscript for any possible plagiarism using any programme such as Turn-It-In or any other software before submitting the manuscripts to the Editorial Office, *Pertanika* Journal Publication Section.

All submitted manuscripts must be in the journal's acceptable similarity index range:

≤ 20% – PASS; > 20% – REJECT.

International Standard Serial Number (ISSN)

An ISSN is an 8-digit code used to identify periodicals such as journals of all kinds and on all media–print and electronic.

Pertanika Journal of Science & Technology: e-ISSN 2231-8526 (Online).

Lag Time

A decision on acceptance or rejection of a manuscript is reached in 90 days (average). The elapsed time from submission to publication for the articles averages 180 days.

Authorship

Authors are not permitted to add or remove any names from the authorship provided at the time of initial submission without the consent of the journal's Chief Executive Editor.

Manuscript Preparation

For manuscript preparation, authors may refer to Pertanika's **INSTRUCTION TO AUTHORS**, available on the official website of *Pertanika*.

Editorial Process

Authors who complete any submission are notified with an acknowledgement containing a manuscript ID on receipt of a manuscript, and upon the editorial decision regarding publication.

Pertanika follows a **double-blind peer-review** process. Manuscripts deemed suitable for publication are sent to reviewers. Authors are encouraged to suggest names of at least 3 potential reviewers at the time of submission of their manuscripts to *Pertanika*, but the editors will make the final selection and are not, however, bound by these suggestions.

Notification of the editorial decision is usually provided within 90 days from the receipt of manuscript. Publication of solicited manuscripts is not guaranteed. In most cases, manuscripts are accepted conditionally, pending an author's revision of the material.

The Journal's Peer Review

In the peer-review process, 2 to 3 referees independently evaluate the scientific quality of the submitted manuscripts. At least 2 referee reports are required to help make a decision.

Peer reviewers are experts chosen by journal editors to provide written assessment of the **strengths** and **weaknesses** of the written research, with the aim of improving the reporting of research and identifying the most appropriate and highest quality material for the journal.

Operating and Review Process

What happens to a manuscript once it is submitted to *Pertanika*? Typically, there are 7 steps to the editorial review process:

1. The journal's Chief Executive Editor and Editor-in-Chief examine the paper to determine whether it is relevance to the journal's needs in terms of novelty, impact, design, procedure, language as well as presentation and allow it to proceed to the reviewing process. If it is not appropriate, the manuscript is rejected outright and the author is informed.

2. The Chief Executive Editor sends the article-identifying information having been removed, to 2 to 3 reviewers. They are specialists in the subject matter of the article. The Chief Executive Editor requests that they complete the review within 3 weeks.

   Comments to authors are about the appropriateness and adequacy of the theoretical or conceptual framework, literature review, method, results and discussion, and conclusions. Reviewers often include suggestions to strengthening of the manuscript. Comments to the editor are in the nature of the significance of the work and its potential contribution to the research field.

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

3. The Editor-in-Chief examines the review reports and decides whether to accept or reject the manuscript, invite the authors to revise and resubmit the manuscript, or seek additional review reports. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the authors) are forwarded to the authors. If a revision is indicated, the editor provides guidelines to attend to the reviewers' suggestions and perhaps additional advice on how to revise the manuscript.

4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the Chief Executive Editor along with specific information describing how they have addressed' the concerns of the reviewers and editor, usually in a tabular form. The authors may also submit a rebuttal if there is a need especially when the authors disagree with certain comments provided by the reviewers.

5. The Chief Executive Editor sends the revised manuscript out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the Editor-in-Chief examines their comments and decides whether the manuscript is ready to be published, needs another round of revisions, or should be rejected. If the decision is to accept, the Chief Executive Editor is notified.

7. The Chief Executive Editor reserves the final right to accept or reject any material for publication, if the processing of a particular manuscript is deemed not to be in compliance with the S.O.P. of *Pertanika*. An acceptance letter is sent to all the authors.

   The editorial office ensures that the manuscript adheres to the correct style (in-text citations, the reference list, and tables are typical areas of concern, clarity, and grammar). The authors are asked to respond to any minor queries by the editorial office. Following these corrections, page proofs are mailed to the corresponding authors for their final approval. At this point, **only essential changes are accepted**. Finally, the manuscript appears in the pages of the journal and is posted online.

Pertanika Journal of

SCIENCE & TECHNOLOGY

Vol. 34 (S1) 2026

A Special Issue devoted to
Technologies and Applications for Sustainable Development

Guest Editors
I. Siva, K. Reddy Madhavi, and Nagaraj Ramrao

A scientific journal published by Universiti Putra Malaysia Press

PJST
Pertanika Journal of Science & Technology
AN INTERNATIONAL PEER-REVIEWED JOURNAL

Journal Section Office
Putra Science Park
1st Floor, IDEA Tower II
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia

General Enquiry
Tel. No: +603 9769 1622
E-mail:
pertanika.proceeding@upm.edu.my
URL:
http://www.pertanika.upm.edu.my/

PUBLISHER
UPM Press Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia
Tel: +603 9769 3606
E-mail: dir.penerbit@upm.edu.my
URL: http://penerbit.upm.edu.my

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

Pertanika Journal of Science & Technology is indexed in Journal Citation Reports (JCR-Clarivate), SCOPUS (Elsevier), EBSCO, ASEAN Citation Index, Microsoft Academic, Google Scholar and MyCite.

Pertanika Journal of Science & Technology
Vol. 34 (S1) 2026

Contents

Preface

The 2nd International Conference on Advances in Technologies and Applications for Sustainability (ICATAS 2024), organised by Mohan Babu University, Tirupati, India, on 21–22 December 2024, served as a vibrant platform where scholars, researchers, practitioners, and industry experts came together to deliberate on the role of science, engineering, and technology in shaping a sustainable future.

This conference was envisioned to provide a premier interdisciplinary forum to discuss recent innovations, address pressing challenges, and share practical solutions aligned with the United Nations' Sustainable Development Goals (SDGs). The deliberations reflected the diverse yet interconnected dimensions of sustainability spanning sustainable engineering, energy, environment, material sciences, digital transformation, and societal well-being.

The proceedings of ICATAS 2024, now published by Pertanika, encapsulate the scholarly contributions presented during the conference. The selected works not only highlight theoretical advancements but also emphasise real-world applications, providing pathways to achieve balance between technological growth and environmental responsibility. This special issue thus stands as a testimony to the collective intellectual effort directed towards sustainability, demonstrating how innovations can be harnessed for long-term societal and ecological benefits.

We are deeply grateful to team members of Universiti Putra Malaysia's Pertanika Journal of Science and Technology for their unwavering support in realising this publication. Our sincere appreciation is also extended to the management of Mohan Babu University, chief patrons, patrons, advisors, reviewers, session chairs, and all committee members whose dedicated efforts ensured the success of ICATAS 2024. Above all, we thank our contributors and participants from across the globe for bringing diverse perspectives and for making this conference a truly enriching experience.

We hope that this special issue will inspire researchers, academics, and practitioners to explore novel directions and foster collaborations that contribute to a more sustainable and equitable world.

Guest Editors
I. Siva (Prof. Dr.)
K. Reddy Madhavi (Prof. Dr.)
Nagaraj Ramrao (Prof. Dr.)

Pertanika J. Sci. & Technol. 34 (S1): 1 - 26 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Deep Residual Networks with Synthetic Minority Over-Sampling Technique Edited Nearest Neighbours (SMOTE-ENN) for ECG Monitoring for Arrhythmia Classification

K. Ghamya[1]* and K. Reddy Madhavi[2]

[1]Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Chittoor District, Andhra Pradesh, India
[2]School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Chittoor District, Andhra Pradesh, India

ABSTRACT

The electrocardiogram (ECG) is one of the most important physiological markers utilised in arrhythmia diagnosis. The ECG arrhythmia classification devices, which include multi-module sensors, clustering algorithms and neural networks, have established themselves as vital monitoring and detection systems for cardiovascular disorders during the last few years. The current ECG arrhythmia classification algorithms encounter two main difficulties because they have complicated models and they need extended periods to complete their operations. The identification of arrhythmias in ECG data needs precise evaluation to perform successful cardiac monitoring and receive correct medical care. The main problem with ECG dataset analysis through deep learning methods stems from the unbalanced distribution of arrhythmia types, which makes some arrhythmia types less common than others. This research presents a new method to enhance arrhythmia detection from ECG data by integrating deep residual networks (ResNets) with the synthetic minority over-sampling technique - edited nearest neighbours (SMOTE-ENN). The ResNet architecture serves to extract deep hierarchical features from ECG data, which enables the model to learn complex arrhythmia patterns with high accuracy. The system uses SMOTE-ENN to create artificial minority class examples while it eliminates unimportant and borderline data points, which produces a dataset that is both clean and balanced. The proposed framework demonstrates superior performance to conventional methods according to experimental results, which were conducted on ECG benchmark datasets. The combination of ResNet with SMOTE-ENN produces better model generalisation, which results in improved accuracy, F1-score, and sensitivity performance, especially for the minority arrhythmia classes. The research shows that deep learning systems, which use improved data balancing techniques,

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 02 May 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.01

E-mail address:
ghamyakotapati@gmail.com (K. Ghamya)
kreddymadhavi@gmail.com (K. Reddy Madhavi)
* Corresponding author

e-ISSN: 2231-8526

can successfully detect arrhythmias in ECG monitoring systems which function in actual clinical settings.

Keywords: Arrhythmia classification, deep learning, deep residual networks (ResNets), electrocardiogram (ECG), synthetic minority over-sampling technique-edited nearest neighbours (SMOTE-ENN)

INTRODUCTION

The public health situation has become more critical because arrhythmias and other cardiovascular diseases now represent major health risks (Hong et al., 2022). While ventricular tachycardia and ventricular fibrillation (Sangha et al.,2022) necessitate prompt medical attention, atrial premature beats are examples of arrhythmias that require regular monitoring or surveillance. The unpredictable pattern of arrhythmia events, which last only briefly, makes it challenging to detect these events through 24-hour ambulatory ECG monitoring or standard ECG recordings. The system provides vulnerable people with current cardiac information and correct medical guidance through real-time ECG monitoring, which runs on wearable or portable devices to reduce medical facility workload and decrease their chances of getting sick (Liu et al.,2022).

The 12-lead ECG serves as a clinical image processing tool which cardiologists use to detect cardiovascular diseases while performing precise cardiac health evaluations of their patients. The research by Al-Zaiti et al. (2025) investigates medical diagnosis through their study and prediction using pre-hospital 12-lead ECGs, which have been the focus of prior studies. Several challenges persist in achieving successful non-clinical self-monitoring ECG detection. Nonmalignant arrhythmias, which include atrial premature beats and nonpersistent sinus tachycardia, ventricular premature beats and sinus bradycardia, show small variations from their typical ECG readings because of various clinical and physiological factors. The evaluation of these minimal differences requires either specialised equipment or medical professionals with advanced training for accurate assessment. The new technology enables customers to monitor their heart rate precisely through wearable devices, which alert them about abnormal heart rhythms that include atrial fibrillation and premature beats. Modern monitoring systems use robust programs, which Liu et al. (2024) describe in their research. The system lets users check their heart health status while they monitor their medications and complete various health tasks. To improve the precision and effectiveness of organisational outcomes, further study is needed into the integration of lightweight neural networks with ECG monitoring methods (Liu et al.,2022; Qaisar et al.,2022).

The system for ECG-based arrhythmia classification becomes effective through the combination of deep residual networks (ResNets) with the synthetic minority over-sampling technique and edited nearest neighbours (SMOTE-ENN). The ResNet network solves

vanishing gradient problems while extracting hierarchical features, but SMOTE-ENN addresses class imbalance, which occurs frequently in ECG datasets. The SMOTE algorithm produces artificial examples which improve minority class representation, while Enn eliminates unnecessary data points to develop an improved dataset that produces superior model performance. The hybrid method enables doctors to identify arrhythmias with high precision, which leads to better real-time cardiac monitoring and diagnosis results.

Research of Our Work

The research investigates a deep residual network architecture, which incorporates SMOTE-ENN for ECG monitoring and arrhythmia organisation to determine how combining state-of-the-art deep learning techniques with data preprocessing methods will enhance arrhythmia detection accuracy. The research uses deep residual networks to extract and analyse complex patterns in ECG signals, combined with SMOTE-ENN to address class imbalance by synthesising minority class samples and eliminating noisy data. The method enhances both precision and system stability for cardiovascular health tracking, which enables doctors to detect arrhythmias at early stages before they develop into severe conditions, thus resulting in improved patient outcomes.

Motivation of this Research

Scientists need to develop reliable arrhythmia detection systems which identify dangerous heart rhythm problems for this research purpose. The process of monitoring electrocardiograms (ECGs) enables doctors to make early diagnoses, but ML models struggle because their training data contains unbalanced information, which makes aberrant instances occur rarely. The research combines deep residual networks (ResNets) with SMOTE-ENN to solve data imbalances, which results in better classification results. The main objective of this research is to improve diagnostic accuracy through the combination of SMOTE-ENN preprocessing strength with ResNet's feature extraction capabilities. The system will lead to better medical results for patients while it develops new automated healthcare systems.

The Main Contributions of this Work

- To develop a unique strategy for enhancing arrhythmia detection from ECG data by merging deep residual networks (ResNets) with SMOTE-ENN
- To address class imbalance, SMOTE-ENN is employed to generate synthetic minority class samples while eliminating noisy and borderline data points, resulting in a cleaner and more balanced dataset.

- Finally, experimental outcomes on benchmark ECG datasets demonstrate that the suggested framework outperforms traditional approaches.

Structure of Our Article

The rest of the article is structured as follows: Section 2 provides a thorough literature assessment, while Section 3 outlines the proposed technique. The research results and analysis appear in Section 4, followed by a conclusion that presents the study's findings and suggests additional research directions.

Literature Survey

The research investigates new ECG monitoring systems which identify arrhythmias through an analysis of current diagnostic approaches and classification methods, and their application in healthcare settings. The identification of cardiovascular health issues at their earliest stages requires ECG-based monitoring systems that can detect arrhythmias with high precision and speed. The study investigates present-day arrhythmia detection methods which use machine learning (ML) and deep learning (DL), and wearable device technology, while it determines their present-day obstacles and their prospective uses.

Zheng et al. (2020) created an arrhythmia recognition system which used CNNs and LSTM to identify eight ECG data categories, including normal sinus rhythm. The article used ECG data from the MIT-BIH arrhythmia database. The model received its input data through the conversion of one-dimensional signals, which became two-dimensional greyscale images. To detect and categorise the incoming data, the combination model was used. The benefit of this technique is that the ECG signal does not need to be feature-extracted or noise-filtered.

A new feature extraction technique was proposed by Zou et al. (2022) for better heartbeat categorisation. Segment label features were learned by a convolutional neural network and then fed into a random forest classifier with other standard features. Using the MIT-BIH arrhythmia database, the model was trained to classify three different heartbeat types: NEB, VEB, and SVEB. The method obtained a 96% accuracy rate and an F1 score of 88.3% for each of these classes. The research established that segment labels play a vital role in accurate heart rhythm classification because they provide essential contextual information.

Ganguly et al. (2020) created a DL system which employed LSTM methods to perform automatic arrhythmia detection and ECG recording extraction. A feature-based bidirectional LSTM (bi-LSTM) system took the role of the conventional LSTM network architecture. The analysis used bidirectional multifractal detrended fluctuation analysis to extract relevant features. Multifractal parameters were utilised to analyse the complex, stochastic, and nonlinear fluctuations of the online-accessible ECG data. A single-layer bi-LSTM method was fed ten characteristics extracted from the segmented ECG beats.

The research by Atal and Singh (2020) developed a deep convolutional neural network (deep CNN) which uses optimisation to perform automatic arrhythmia organisation. They presented the Bat-Rider Optimisation Algorithm (BaROA), a new optimisation method that combines the Rider Optimisation Algorithm (ROA) and the Multi-Objective Bat Algorithm (MOBA). Wavelet and Gabor features were recovered in the first step since the signals represented different ECG components. These features were then processed by a BaROA-based DCNN classifier to differentiate between arrhythmia and non-arrhythmia cases. The techniques were evaluated utilising the MIT-BIH arrhythmia database. The corresponding outcomes were 93.19%, 95%, and 93.98% for accuracy, specificity, and sensitivity, respectively.

Fatimah et al. (2024) proposed an inter-patient paradigm to address the challenges of generalisation, where the test dataset was collected from a participant group rather than the training dataset. Their suggested method achieved an F1 score of 89.35% in detecting supraventricular ectopic beats (SVEB) compared to existing methods. They analysed ECG data on multiple scales using the Fourier decomposition method (FDM). The narrow-band signal elements produced by FDM were used to extract time-domain and statistical features. To eliminate redundancy and rank the most significant attributes, two feature selection methods, the Kruskal-Wallis test and minimum redundancy maximum relevance (MRMR), were employed.

Katal et al. (2023) investigated the accuracy, specificity, precision, and F1 score of their DL-based techniques for identifying arrhythmias from ECG signals. To evaluate its performance against pretrained models such as GoogLeNet, they propose building a compact CNN. The comparative analysis highlights the promising potential of diagnosing arrhythmias using ECG signals and deep learning.

Kumar et al. (2022) proposed an optimized dl classifier for the automatic identification and categorisation of arrhythmias. First, ECG data collected by IoT nodes is used to extract the QRS complex and the RR interval. These are then utilised to create the feature vector for arrhythmia organisation. The ECG signal anomalies are found using a deep convolutional neural network based on coy-grey wolf optimisation (coy-gwo-based deep CNN) classifier. To efficiently update the classifier's parameters, the suggested coy-gwo technique makes use of hybrid data, such as social hunting hierarchies and canid hunting experiences.

Varshney et al. (2022) and their co-authors developed a method to analyse ECG data through software which organises heart rhythm information by arrhythmia type. The researchers applied an undecimated discrete wavelet transform (UDWT) to break down each signal. The system performed ECG waveform reconstruction to detect R-peaks through its application of specific frequency bands. To categorise various arrhythmia forms, the heart rate was determined. The capacity of the UDWT-based method to dissect signals at numerous levels makes it beneficial for a variety of processing applications. In the frequency and time domains, it also makes simultaneous localisation easier.

Research Gap

The research field of ECG monitoring for arrhythmia classification has achieved significant advancements, but it still encounters various unresolved difficulties (Mekruksavanich & Jitpattanakul, 2024). The current systems depend on human-made features and particular datasets, which restrict their ability to work with different patient groups in actual clinical settings. The current methods fail to identify both infrequent and concurrent arrhythmia types, which leads to lower detection precision. The lack of effective real-time monitoring systems which work with wearable devices prevents most medical facilities from adopting these technologies. Researchers need to develop complex ML models which will automatically identify important features to solve these existing problems.

Problem Identification in the Existing System

Many systems face challenges in accurately classifying arrhythmias in real time due to noise, artefacts, and variability in ECG signals across patients.

- Most models are trained on datasets that fail to fully represent diverse patient populations, resulting in biases and reduced performance in real-world applications.
- Current systems often depend on complex hardware and expensive computational resources, limiting accessibility, particularly in low-resource settings.
- Wearable or moveable devices encounter issues with battery life, restricting long-term monitoring and continuous data collection.

Our Proposed Model

The research presents an original method to enhance ECG-based arrhythmia detection through the integration of deep residual networks (ResNets) with SMOTE-ENN. ResNets are employed to extract deep hierarchical features from ECG data, enabling accurate learning of complex patterns connected with various arrhythmias. The system uses SMOTE-ENN to create artificial minority class examples while removing unimportant and uncertain data points, which produces a dataset that is both clean and balanced. This work introduces a hybrid framework for ECG-based arrhythmia classification that combines residual deep learning architectures with a data-level imbalance correction strategy using SMOTE-ENN.

The block diagram for the suggested model is shown in Figure 1. This architectural design displays a pipeline for categorising arrhythmias utilising the MIT-BIH arrhythmia database. The system starts with data preprocessing, which uses a six-layer wavelet transform to clean ECG signals and produce ECG images. Ngo et al. (2022) the process of heartbeat segmentation enables researchers to extract individual cardiac cycles from the recorded data. The system uses deep residual networks (ResNets) to extract features through their convolutional layers, which include residual connections for achieving both high accuracy and efficient feature extraction. Finally, the SMOTE-ENN method is employed

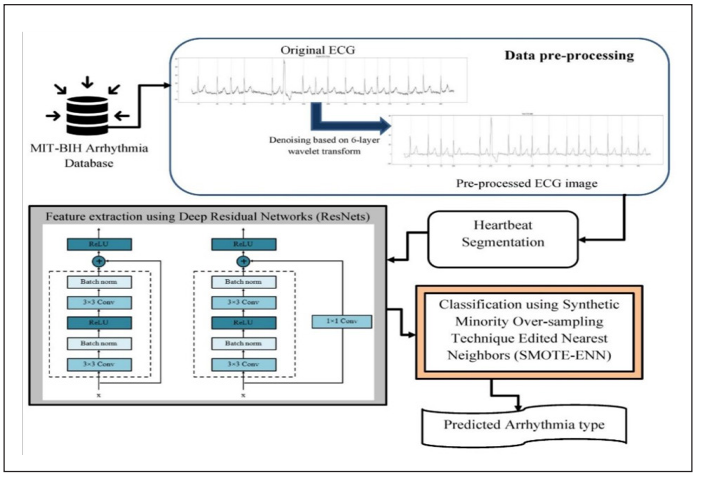


Figure 1. Block diagram for the proposed model

for classification to address class imbalance in the dataset and predict arrhythmia types. The method generates powerful feature extraction results while it effectively manages unbalanced data, which produces superior classification results.

Dataset

The experiments were conducted using the MIT-BIH arrhythmia database, a widely accepted benchmark for evaluating arrhythmia detection methods. The database consists of annotated ambulatory ECG recordings obtained from multiple subjects under controlled conditions shows in Figures 2 and 3. For consistency across experiments, only the modified lead II signal was utilised in this study. Heartbeat annotations were grouped into five standard classes in accordance with AAMI recommendations: normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unclassifiable beat (Q) show in Table 1. Due to the dominance of normal beats, the dataset exhibits significant class imbalance, motivating the use of advanced resampling strategies.

Table 1
Testing and training are divided for every class

Class	Meaning	Training samples	Samples	Proportion (%)	Test samples
S	Supraventricular premature or Ectopic beat	2223	2779	2.54	556
N	Normal beat	72,471	90,589	82.77	18,118
V	Premature ventricular contraction	5788	7236	6.61	1448
F	Fusion of ventricular and normal beat	641	803	0.73	162
Q	Unclassifiable beat	6431	8039	7.35	1608
Total		**87,554**	**109,446**	**100.00**	21,892

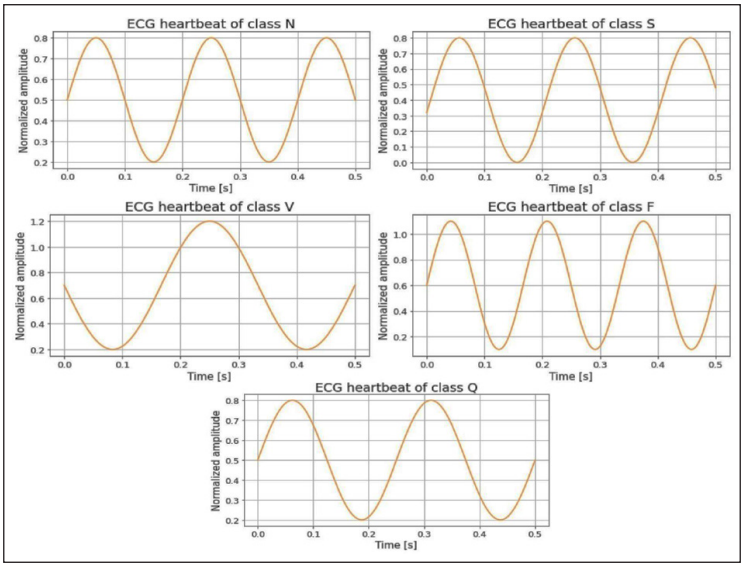


Figure 2. Samples of random ECG heartbeats

Data Preprocessing

Before feature extraction, ECG signals were processed using a multi-level wavelet-based denoising approach (Qin et al.,2024). This step reduces baseline drift and high-frequency interference while preserving clinically relevant waveform characteristics. R-peak detection was then performed to segment the ECG signals into individual cardiac cycles.

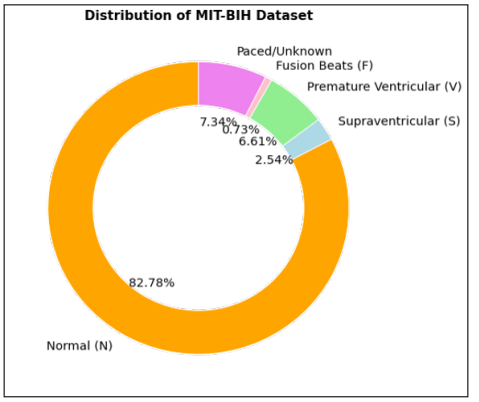


Figure 3. Distribution of the dataset

Each extracted heartbeat was normalised and transformed into a standardised input format suitable for deep learning. These steps ensure uniform representation across samples and improve the stability of the training process in Equation 1 and Equation 2.

$$cA_{i+1}(y) = \sum_x cA_j\,(x)h(x-2y) \quad [1]$$

$$cD_{j+1}(y) = \sum_x cA_j\,(x)\,g(x-2y) \quad [2]$$

The following is the reconstruction equation defined in Equation 3:

$$A_j\,(y) = \sum_x cA_{j+1}(x)h\,(y-2x) + cD_{j+1}(x)g(y-2x) \quad [3]$$

The analysis of ECG data requires the identification of essential waveforms, which doctors need to diagnose cardiac diseases. The process required the following sequence of operations.

1. The positive and negative peak values of the ECG signal were identified.
2. The maximum values from the calculation allowed us to determine the peak and valley points of the ECG waveform. The signal waveform shows heart voltage changes during a cardiac cycle, but its peak values represent the maximum and minimum points. The most noticeable aspects of the ECG indication, the R-waves of the QRS complex, which typically correspond to peaks, are indicative of ventricular polarisation.
3. The T-waves, which are an element of the repolarisation process and come after the P-waves, can be regarded as troughs.
4. The research team removed excess r-waves which appeared by accident while they added any absent R-waves to preserve the complete ECG signal.
5. The QRS complex, T-wave, and P-wave were detected through the current threshold settings.

Healthcare providers use systematic methods to diagnose ECG waveforms, which enable them to detect cardiac conditions. Based on the R waves, the test results were classified into several cycles: 200 R, 200 A, 300 V, 400 N, and 300 L in total. Ten ECG signals from each classification, totalling 50, were selected as the fitting dataset. There were 250 ECG signals, 50 from each group, in the training dataset. 466 ECG signals, in total: 80 A, 110 N, 110 L, 110 R, and 56 V, were selected for study. To prevent falsely high classification accuracy, there was no overlap between the fitting, training, and testing datasets.

Deep Residual Networks (ResNets)

Deep residual networks form the backbone of the proposed feature extraction process. Unlike traditional deep neural networks, residual architectures incorporate shortcut connections that allow information to bypass certain layers (Fan et al., 2020). This design facilitates efficient gradient propagation and enables the training of deeper models without performance degradation. The network is composed of multiple residual blocks, each containing convolutional layers followed by normalisation and activation functions shows in Figure 4. By learning residual mappings, the model effectively captures temporal and morphological patterns associated with different arrhythmia types.

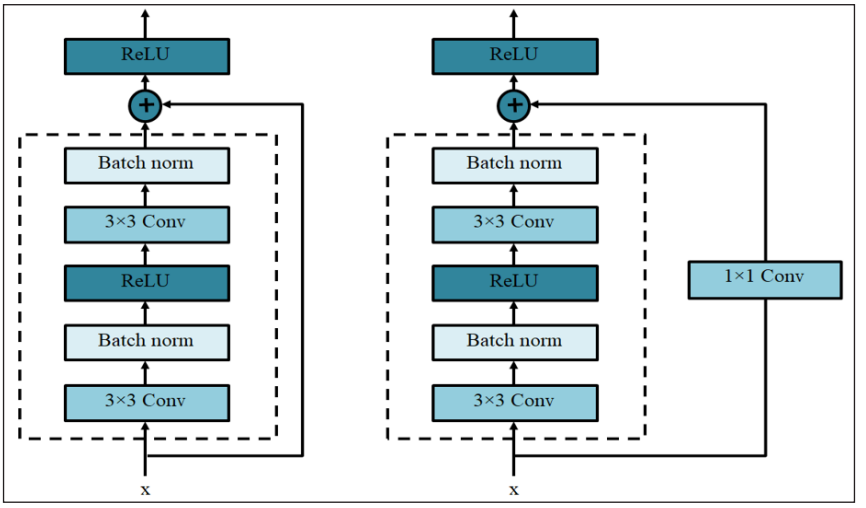


Figure 4. The architecture of the ResNet model

Residual Learning Framework

ResNets explicitly reformulate the layers to learn residual functions $F(x) = H(x) - x$ instead of directly learning $H(x)$, where $H(x)$ is the desired mapping, and x is the input. The residual function is easier to optimise because it assumes the identity mapping $H(x) = x$ as an initial approximation. Thus, the output of the block becomes, according to Equation 4:

$$y = F(x) + x \quad [4]$$

Identity Shortcut Connections

These are direct pathways that bypass one or more layers in the network. For example, in a block with two convolutional layers, the input x skips these layers and is added to the output of the second layer. Mathematically, defined in Equation 5:

$$y = \sigma(F(x, \{W_k\}) + x \quad [5]$$

wherever W_k are the learnable parameters of the convolutional layers and σ is the activation function of ReLU.

Block Structure

Resnets are composed of multiple residual blocks, which typically include:

- a convolutional layer
- batch normalisation

- non-linear activation (ReLU)
- an identity shortcut connection

For example, a simple block can be represented by Equation 6:

$$y = ReLU\left(BN\left(W_2.ReLU\left(BN(W_1.x)\right)\right) + x\right) \quad [6]$$

where W_1 and W_2 are weight matrices for two convolutions, and bn represents batch normalisation.

Depth and Performance

The key advantage of ResNets lies in their ability to train extremely deep networks without experiencing a degradation in performance. For example, a 152-layer ResNet outperformed shallower models on the ImageNet dataset.

Gradient Flow

In standard deep neural networks, gradients tend to diminish as they backpropagate, leading to poor learning in the earlier layers. In ResNets, the gradient of the loss function *L* with respect to parameters *W* in a residual block container can be expressed as Equation 7:

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y}.\left(\frac{\partial F(x)}{\partial W} + \frac{\partial x}{\partial W}\right) \quad [7]$$

The second term ensures that gradients can propagate effectively even if $F(x)$ is close to zero.

Training Objective

The loss *L* is minimised by optimising both the residual function $F(x)$ and the identity mapping. For a network with nnn layers, the output at layer *k* can be recursively expressed as Equation 8:

$$x_k = x_{k-1} + F(x_{k-1}, W_k) \quad [8]$$

Variants and Improvements

Bottleneck Residual Blocks

To reduce computational cost in very deep networks, bottleneck structures are used, where the block consists of 1 × 1, 3 × 3, and 1 × 1 convolutions according to Equation 9:

$$y = W_3.ReLU\big(W_2ReLU(W_1.x)\big) \quad [9]$$

ResNet with Pre-Activation

An improved variant includes batch normalisation and ReLU applied before convolutions, which enhances optimisation.

Synthetic Minority Over-Sampling Technique Edited Nearest Neighbours (SMOTE-ENN)

SMOTE

SMOTE is a resampling method that creates synthetic samples in the feature space of the minority class to rectify class imbalance. This is accomplished using linear interpolation between the minority class's current samples and their k-nearest neighbours (Hai Ly et al., 2022).

Mathematically, given a sample x_i since the minority class and one of its k-nearest neighbours x_k, a synthetic sample x_{new} is generated in Equation 10:

$$x_{new} = x_i + \lambda(x_k - x_i) \quad [10]$$

wherever λ Is a random number drawn from a uniform distribution. $\lambda \sim U(0,1)$. This ensures the synthetic sample lies along the line segment connecting x_i and x_k, introducing diversity in the feature space while maintaining realistic boundaries for the minority class.

Edited Nearest Neighbours (ENN)

ENN is a data cleaning method applied after oversampling to remove potential noise and borderline cases. It operates by examining the k-nearest neighbours of each sample and removing samples whose class labels differ from the majority of their neighbours' labels.

Given a sample x_i, let $N_k(x_i)$ represent its k-nearest neighbours. The decision for retention is based on the agreement of x_i's label *yi* with the majority label $y_{majority}$ in $N_k(x_i)$ according to Equation 11:

$$y_{majority} = mode\big(\{y_j | x_j \in N_k(x_i)\}\big) \quad [11]$$

if $y_i \neq y_{majority}$, the sample x_i is removed from the dataset. This process cleanses noisy data and enhances the decision boundaries, especially for imbalanced datasets.

SMOTE-ENN

SMOTE-ENN links the advantages of ENN and SMOTE. The ENN system removes both incorrect and noisy data points which exist in both synthetic and original data sets. The SMOTE system generates artificial data points to achieve class distribution equilibrium.

Let the combined dataset after SMOTE be D_SMOTE as defined in Equation 12, and the processed dataset after ENN be D_(SMOTE-ENN) as defined in Equation 13. Formally:

1. Apply SMOTE to obtain:

$$D_{SMOTE} = D_{original} \cup D_{synthetic} \quad [12]$$

where $D_{synthetic}$ includes newly generated samples.

2. Use ENN for data cleaning:

$$D_{SMOTE-ENN} = \{x_i \in D_{SMOTE} | y_i = mode(\{y_j | x_j \in N_k(x_i)\})\} \quad [13]$$

This two-step process improves class balance and enhances decision boundary clarity, leading to better classifier performance, particularly on imbalanced datasets.

Advantages of the Proposed Method

- SMOTE-ENN ensures better representation of the minority class, a critical component in the grouping of arrhythmias, by resolving the problem of imbalanced distribution of ECG datasets.
- A combination of deep residual networks (ResNets) with SMOTE-ENN allows deep learning models to discover intricate patterns from balanced data, which produces superior prediction accuracy.
- The ENN component eliminates unnecessary data points, which produces better training data quality and makes the model more stable.

RESULT AND DISCUSSION

Implementation Details

PyTorch and Python were used to implement the framework. The training process for all neural networks occurred on a laptop which contained an Intel i7-7820HK CPU and an NVIDIA GeForce GTX 1080 GPU. The confusion matrix from the dataset appears in Figure 5.

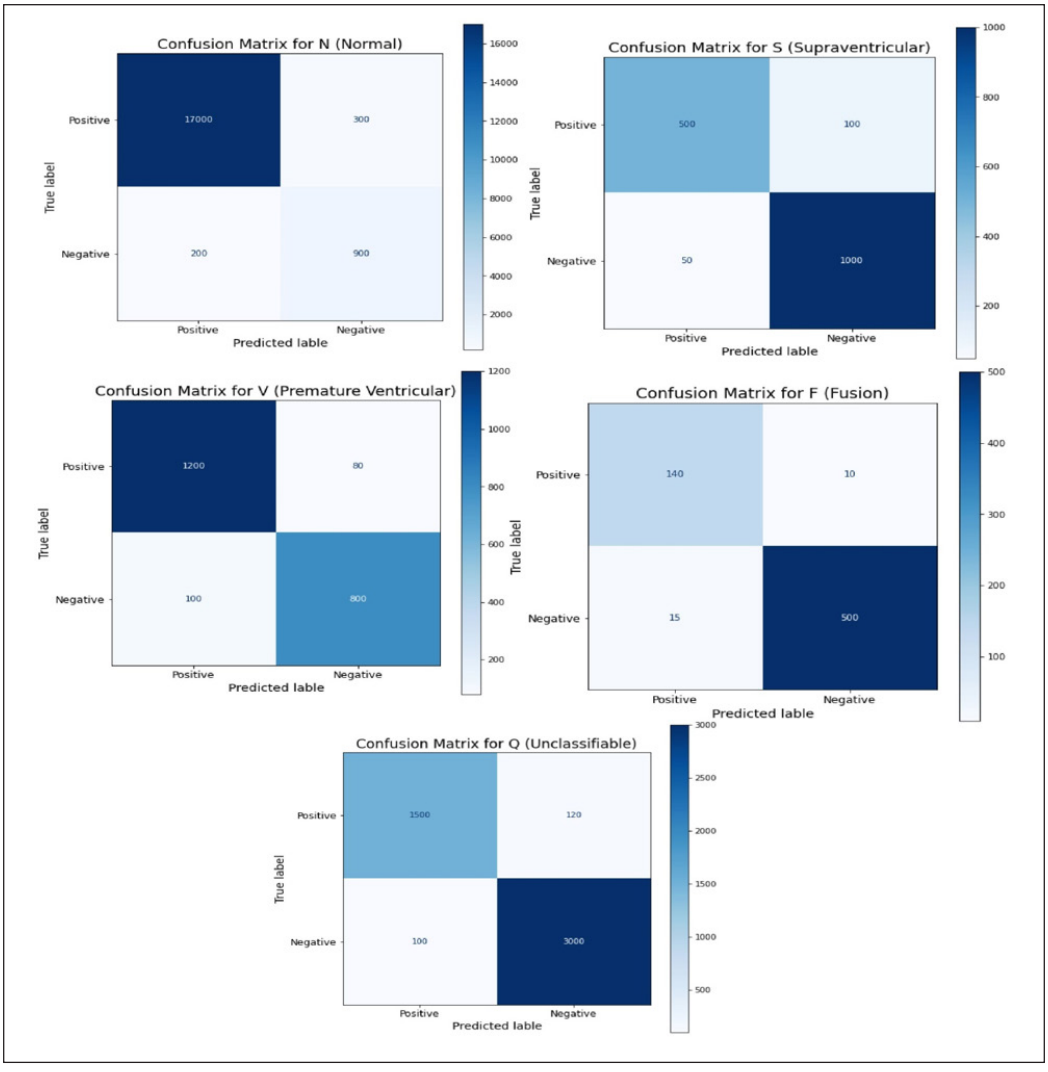


Figure 5. Confusion matrices for each ECG beat class (n, s, v, f and q)

Performance Metrics

Organisations achieve accurate organisational models through their use of precise measurement systems, which enable them to assess their operational performance. The system calculates the percentage of correctly predicted positive events from all the positive events that were projected. The level of precision in a system determines how many incorrect results it produces according to Equation 14.

$$\Pr ecision = \frac{TP}{TP + FP} \qquad [14]$$

The classification process requires recall to identify the correct proportion of true positive examples which the system correctly identifies according to Equation 15. The true positive rate (TPR) and sensitivity are other names for it.

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \quad [15]$$

A classification model achieves its accuracy through the ratio of correct predictions to all its generated forecasts, which serves as its main performance evaluation metric according to Equation 16.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad [16]$$

The performance of binary classification systems is assessed using a statistic called the Matthews correlation coefficient (MCC). The system provides a fair assessment through its evaluation of false negatives, true positives, true negatives and false positives according to Equation 17. The method delivers its best results when working with datasets which contain unbalanced data points.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{\left(TP + FP\right)\left(TP + FN\right)\left(TN + FP\right)\left(TN + FN\right)}} \quad [17]$$

An indicator of the discrepancies between expected and observed values is the root mean square error (RMSE). The method generates a single value which represents the square root of the average squared changes to determine forecast precision according to Equation 18.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad [18]$$

The accuracy of a predictive model can be evaluated using a metric called MAPE (mean absolute percentage error). The mean of the absolute proportion errors between the expected and actual values is calculated according to Equation 19.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100 \quad [19]$$

The duration which a program or algorithm needs to finish its execution cycle determines its running time. It is often expressed as a function of the input size n and varies depending on the algorithm's complexity as expressed by Equation 20.

$$T(n) = O(f(n)) \quad [20]$$

Comparative Methods

FRM-CNN (Fusing Rhythmic Morphological-Convolutional Neural Network)

Fan et al. (2020) FRM-CNN outperforms state-of-the-art recognition approaches in tracking people's health problems using mobile devices.

MRFO-SVM (Manta Ray Foraging Optimisation-Support Vector Machine)

Houssein et al. (2021) note that while MRFO is used to optimise the limits of SVM and select the subset of pertinent characteristics that guarantees the optimal organisation presentation, SVM is utilised for organisation in the MRFO-SVM technique.

AE-biLSTM (Auto- Encoder and Bidirectional Long Short-Term Memory)

Ramkumar et al. (2022) proposed the AE-biLSTM technique, which consists of a decoder that utilises a biLSTM network to reconstruct the electrocardiogram arrhythmia signals, and an encoder that employs a biLSTM network to extract higher-level features from the signals.

Table 2 and Figure 6 present the comparative analysis of the models FRM-CNN, MRFO-SVM, AE-biLSTM, and the proposed model across Precision, Recall, Accuracy,

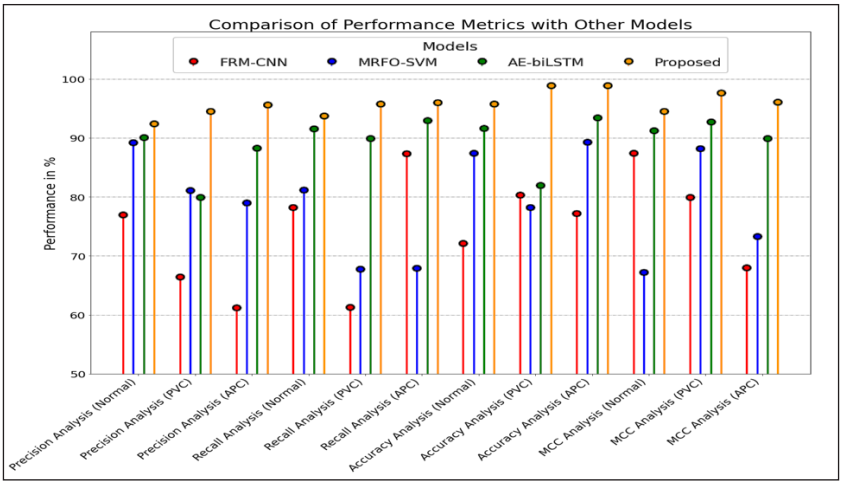


Figure 6. Comparison of performance metrics with other models

Table 2
Comparison of performance metrics with other models

Models	Precision Analysis			Recall Analysis			Accuracy Analysis			MCC Analysis		
	Normal	PVC	APC	Normal	PVC	APC	Normal	PVC	APC	Normal	PVC	APC
FRM-CNN	76.98	66.44	61.22	78.22	61.33	87.33	72.11	80.33	77.21	87.44	79.91	67.98
MRFO-SVM	89.22	81.11	78.98	81.22	67.76	67.91	87.44	78.22	89.33	67.22	88.22	73.33
AE-biLSTM	90.11	79.91	88.31	91.55	89.91	92.99	91.66	81.98	93.45	91.22	92.76	89.91
Proposed	92.43	94.55	95.66	93.76	95.76	95.98	95.76	98.87	98.90	94.55	97.65	96.11

and MCC metrics, revealing notable differences in performance. The FRM-CNN model exhibited lower performance, particularly in Precision and MCC, with values of 66.44 for Premature Ventricular Contraction (PVC) and 61.22 for Atrial Premature Contraction (APC), and an MCC score of 67.98 for APC. The MRFO-SVM model achieved better results than FRM-CNN because it produced 89.22 Normal Precision and 89.33 APC Accuracy scores. However, it still lagged the other models in Recall and MCC. The AE-biLSTM model achieved excellent results in all evaluation criteria by reaching 90.11 Precision for Normal data and 88.31 Precision for APC data while maintaining high Accuracy and Recall performance, but did not match the performance of the recommended model. The suggested model performed better than all other models in every evaluation metric by achieving the highest Precision (95.66 for APC), Recall (95.98 for APC), Accuracy (98.90 for APC) and MCC (96.11 for APC). The proposed model demonstrates better performance than other models because it achieves the best combination of Precision and Recall and classification Accuracy.

Table 3 and Figure 7 demonstrate that the models were evaluated through RMSE and MAPE metrics, which measured their ability to make accurate predictions. The FRM-CNN model produced high RMSE results, which reached 34.56 for PVC and 24.76 for APC, while the MAPE values reached 22.11 for PVC and 25.98 for APC. The MRFO-SVM model achieved average results, but its RMSE values reached 31.11 for Normal and 29.98 for PVC, and its MAPE values reached 39.11 for PVC and 35.87 for APC. The AE-biLSTM model produced results that were evenly distributed because its RMSE values spanned from 29.98 to 33.87 and its MAPE values spanned from 26.76 to 30.22. The proposed model achieved the best results in RMSE and MAPE performance because it produced

the lowest RMSE values of 10.67 for Normal and 12.11 for APC, and MAPE values of 15.77 for PVC and 19.54 for APC.

Table 3
Comparison of RMSE and MAPE analysis

Models	RMSE Analysis			MAPE Analysis		
	Normal	PVC	APC	Normal	PVC	APC
FRM-CNN	23.87	34.56	24.76	32.87	22.11	25.98
MRFO-SVM	31.11	29.98	29.11	29.11	39.11	35.87
AE-biLSTM	29.98	30.33	33.87	30.43	26.76	30.22
Proposed	10.67	14.22	12.11	18.22	15.77	19.54

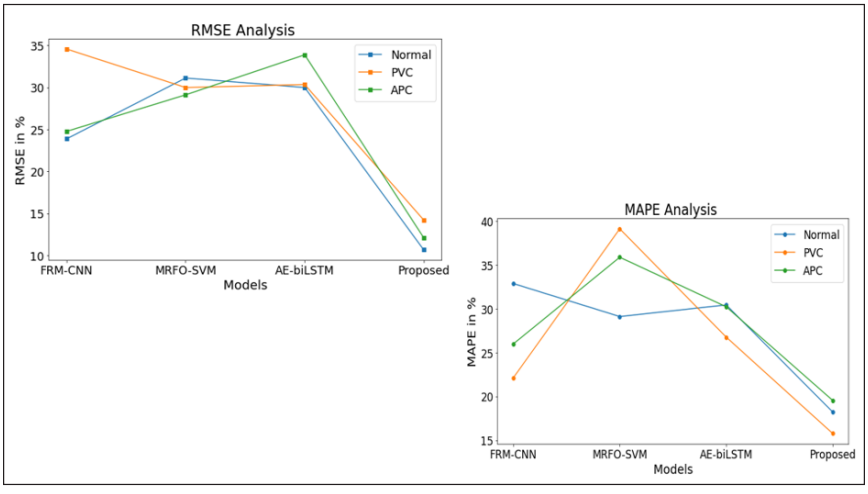


Figure 7. Comparison of RMSE and MAPE analysis

The running time analysis of the models appears in Table 4 and Figure 8, which demonstrates substantial variations between the models. The FRM-CNN model produced the longest running times, which reached 10.654 seconds for Normal, 12.843 seconds for PVC and 14.765 seconds for APC. The MRFO-SVM model showed slightly lower running times for PVC at 6.765 but generally higher values for Normal and APC. The AE-biLSTM model showed typical execution times because Normal produced the longest time of 14.876 seconds, while PVC and APC produced shorter times. The proposed model delivered the fastest results in every

Table 4
Running time evaluation

Models	Running Time Analysis		
	Normal	PVC	APC
FRM-CNN	10.654	12.843	14.765
MRFO-SVM	13.562	6.765	12.765
AE-biLSTM	14.876	8.543	8.113
Proposed	1.654	3.765	5.564

test scenario because it produced the shortest execution times, which reached 1.654 seconds for Normal, 3.765 seconds for PVC and 5.564 seconds for APC.

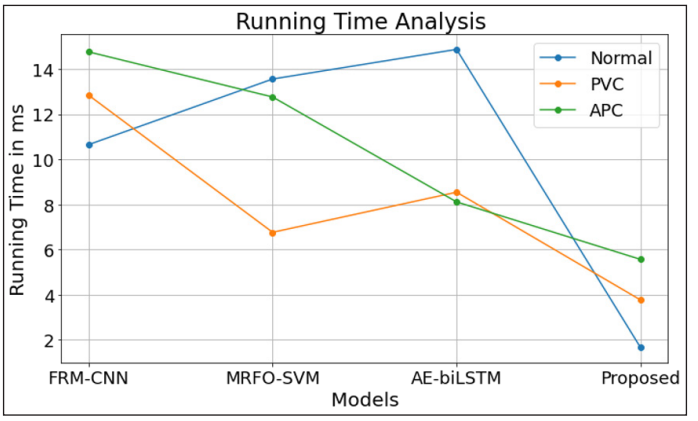


Figure 8. Running time analysis of the proposed model

DISCUSSION

The comparison of various models across multiple presentation metrics highlights the superior presentation of our model. The research by Fan et al. (2020) produced results that were lower than ours for Precision, Recall, Accuracy, MCC, RMSE and MAPE. Our model achieved superior results to all other models in the study. The precision of our model was 95.66 for APC, which surpassed 61.22 for APC. The model achieved significant enhancements in Recall and Accuracy, and MCC performance, which reached 95.98, 98.90 and 96.11 for APC, respectively. The model achieved high prediction accuracy through its RMSE scores of 10.67 for Normal and MAPE scores of 15.77 for PVC. The model achieved the best performance regarding running time because it demonstrated the lowest execution times of 1.654 for Normal and 5.564 for APC among all tested models. The complete evaluation demonstrates that our model achieves superior results to other models because it achieves both high precision and fast computation speed.

Ablation Study Analysis

Table 5 and Figure 9 show the accuracy of three different methods used for performance evaluation. The ResNet model achieved an accuracy of 92.76%, while the SMOTE method, which addresses class imbalance, improved the accuracy to 95.56%. The combination of ResNet with SMOTE-ENN produced better results, achieving an accuracy of 97.84%. The results indicate that SMOTE

Table 5
Ablation study analysis

Method	Accuracy
ResNet	92.76
SMOTE	95.56
ResNet + SMOTE-ENN	97.84

integration with ResNet produces better results when SMOTE-ENN is used for improved data balancing.

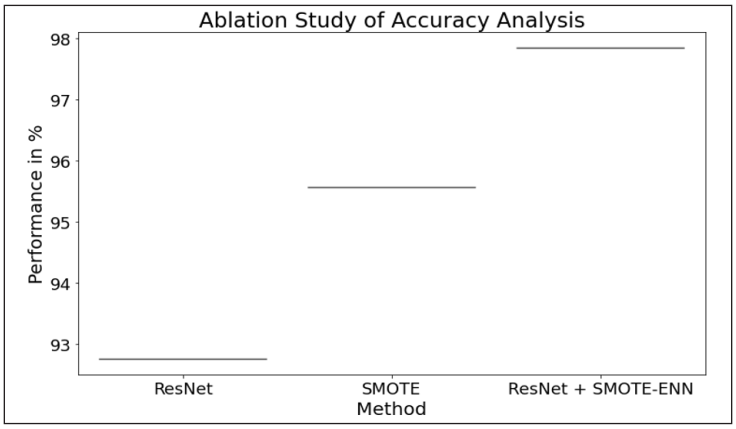


Figure 9. Ablation study and performance analysis

Training Validation Loss and Accuracy Analysis

The system operates by training itself while it tracks both validation loss and accuracy values. Figure 10 illustrates that ECG monitoring for arrhythmia classification involves training ML models to accurately distinguish among different types of arrhythmias. Metrics such as training and validation accuracy, along with training and validation loss, are utilised to estimate the efficacy of these models. The training loss shows how well the model learns, but the validation loss demonstrates its performance on unseen data points. The training accuracy of the model shows its performance on the training data, while validation accuracy demonstrates its capacity to identify unknown ECG signals. The training process

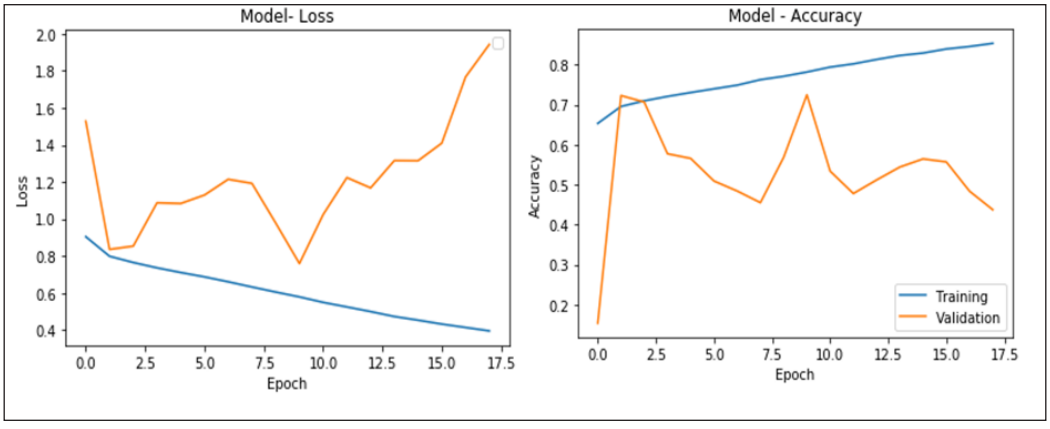


Figure 10. Training vs. validation loss and accuracy

allows users to monitor these performance indicators, which help them adjust the model for better results and improved arrhythmia organisation accuracy.

Table 6
10-fold Cross-Validation of ResNet-SMOTE-ENN Analysis

k-folds	ResNet-SMOTE-ENN Accuracy
1-fold	0.95
2-fold	0.97
3-fold	0.98
4-fold	0.98
fold-5	0.96
fold-6	0.97
fold-7	0.98
fold-8	0.98
fold-9	0.97
fold-10	0.98
10-fold mean	0. 97

Influence of SMOTE-ENN

The SMOTE-ENN algorithm proves its value because it enables improved model visualisation through its capability to manage unbalanced class distributions in data. ENN removes noisy cases by considering the nearest neighbours, while SMOTE generates synthetic examples for the minority class. This combined method improves the quality of the dataset, resultant in more accurate and reliable model forecasts, particularly in scenarios where class imbalance can significantly affect performance. SMOTE-ENN has proven to be effective in refining models by reducing overfitting and enhancing generalisation.

K-Fold Matrix

The application of 10-fold cross-validation significantly enhanced the consistency and resilience of the methods utilised in ECG monitoring for arrhythmia organisation. This statistical technique involves dividing the dataset into 10 segments, or subsets, where nine are used for training and one for validation. The proposed ResNet-SMOTE-ENN shows in Table 6 technique outperformed competing methods, achieving an accuracy of 97.84% on our input data utilising 10-fold cross-validation.

Comparison of the Proposed Model

Table 7 and Figure 11 present a comparison of accuracy across different methods and references for the MIT-BIH database. Swetha and Ramakrishnan (2021) achieved an accuracy of 91.50% using the KC-FLC technique. Essa and Xie (2021) improved performance to 95.81% by utilising Bagging Models, while Chen et al. (2020) reached 94.35% with the MACN+UDA approach, and Pokharel et al. (2024) achieved 96.71% accuracy using a CNN+ML Hybrid pipeline. Our proposed model, which combines ResNet with SMOTE-ENN, demonstrated the highest accuracy of 97.84%. The results demonstrate that ResNet-SMOTE-ENN outperforms traditional methods because it achieves better results.

Table 7
Comparison of the proposed model

Reference	Database	Method	Accuracy
Essa & Xie, 2021	MIT-BIH	Bagging Models	95.81%
Chen et al., 2020	MIT-BIH	MACN+UDA	94.35%
Swetha & Ramakrishnan, 2021	MIT-BIH	KC-FLC	91.50%
Pokharel et al., 2024	MIT-BIH	CNN + ML Hybrid Pipeline	96.71%
Our Model	MIT-BIH	ResNet-SMOTE-ENN	97.84%

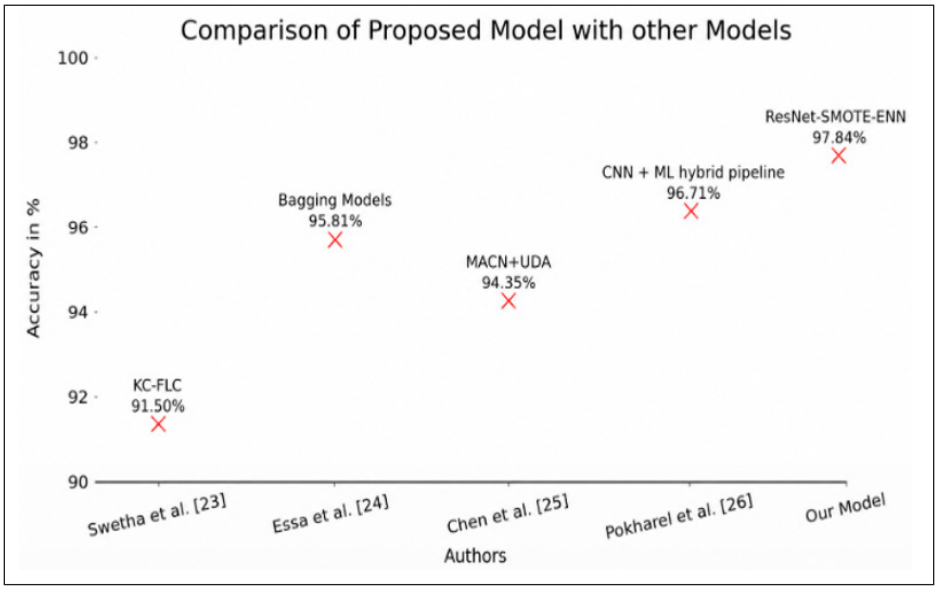


Figure 11. Comparison analysis of the proposed model with other models

Limitations

Despite the promising outcomes of utilising deep residual networks (ResNet) in conjunction with SMOTE-ENN for arrhythmia organisation, several limitations persist. First, SMOTE-ENN is heavily reliant on the amount and calibre of training data because the artificial data it produces frequently falls short of accurately capturing the intricacy of real-world ECG signals. The process of training deep neural networks requires significant computational resources, which leads to longer processing times and requires substantial hardware equipment. The method encounters two main difficulties when working with datasets that contain extreme imbalances between classes and outliers because SMOTE-ENN generates occasional noisy data points. The model maintains its ability to generalise data from the MIT-BIH database, but it needs additional training to work efficiently with different types of datasets.

CONCLUSION

The system, which uses deep residual networks with SMOTE-ENN methods, operates efficiently to analyse ECG data and identify arrhythmias. The hybrid approach solves two major problems which occur when ECG datasets contain unbalanced data and noisy information by creating a model which achieves better generalisation and higher classification accuracy. The proposed solution achieves better results than existing methods through its combination of SMOTE-ENN data balancing capabilities with deep residual networks' advanced feature extraction methods. The research demonstrates how this method functions as a dependable system which detects arrhythmias at their beginning stages for better clinical diagnosis. Future research should focus on two main areas, which include testing the system for real-time ECG monitoring and developing methods to merge the system with wearable health devices for ongoing patient monitoring.

ACKNOWLEDGEMENT

Author contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

ECG	:	Electrocardiogram
ResNet	:	Residual Network
SMOTE	:	Synthetic Minority Over-sampling Technique
ENN	:	Edited Nearest Neighbours
ML	:	Machine Learning
DL	:	Deep Learning
NEB	:	Normal Ectopic Beats
VEB	:	Ventricular Ectopic Beats
SVEB	:	Supraventricular Ectopic Beats
LSTM	:	Long Short-Term Memory
bi-LSTM	:	Bidirectional Long Short-Term Memory
ROA	:	Rider Optimisation Algorithm
BaROA	:	Bat-Rider Optimisation Algorithm
MOBA	:	Multi-Objective Bat Algorithm
MRMR	:	Minimum Redundancy Maximum Relevance
ReLU	:	Rectified Linear Unit
FRM-CNN	:	Fusing Rhythmic Morphological-Convolutional Neural Network

MRFO-SVM	:	Manta Ray Foraging Optimisation-Support Vector Machine
AE-biLSTM	:	Auto- Encoder and Bidirectional Long Short-Term Memory
PVC	:	Premature Ventricular Contraction
APC	:	Atrial Premature Contraction
MCC	:	Matthews Correlation Coefficient
RMSE	:	Root Mean Square Error
MAPE	:	Mean Absolute Percentage Error
MACN	:	Multi-Attention Convolutional Network
UDA	:	Unsupervised Domain Adaptation
KC-FLC	:	Knowledge-Centric Fuzzy Logic Controller
CNN	:	Convolutional Neural Network
MIT-BIH Database	:	Massachusetts Institute of Technology–Beth Israel Hospital Database

REFERENCES

Al-Zaiti, S., Besomi, L., Bouzid, Z., Faramand, Z., Frisch, S., Martin-Gill, C., Gregg, R. E., Saba, S., Callaway, C., & Sejdić, E. (2020). Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nature Communications, 11*, Article 3966. https://doi.org/10.1038/s41467-020-17804-2

Atal, D. K., & Singh, M. (2020). Arrhythmia classification with ECG signals based on the optimisation-enabled deep convolutional neural network. *Computer Methods and Programs in Biomedicine, 196*, Article 105607. https://doi.org/10.1016/j.cmpb.2020.105607

Chen, M., Wang, G., Ding, Z., Li, J., & Yang, H. (2020). Unsupervised domain adaptation for ECG arrhythmia classification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 304-307). IEEE. https://doi.org/10.1109/EMBC44109.2020.9175928

Essa, E., & Xie, X. (2021). An ensemble of deep learning-based multi-models for ECG heartbeats arrhythmia classification. *IEEE Access, 9*, 103452–103464. https://doi.org/10.1109/ACCESS.2021.3098986

Fan, X., Hu, Z., Wang, R., Yin, L., Li, Y., & Cai, Y. (2020). A novel hybrid network of fusing rhythmic and morphological features for atrial fibrillation detection on mobile ECG signals. *Neural Computing and Applications, 32*(12), 8101-8113. https://doi.org/10.1007/s00521-019-04318-2

Fatimah, B., Singhal, A., & Singh, P. (2024). ECG arrhythmia detection in an inter-patient setting using Fourier decomposition and machine learning. *Medical Engineering & Physics, 124*, Article 104102. https://doi.org/10.1016/j.medengphy.2024.104102

Ganguly, B., Ghosal, A., Das, A., Das, D., Chatterjee, D., & Rakshit, D. (2020). Automated detection and classification of arrhythmia from ECG signals using a feature-induced long short-term memory network. *IEEE Sensors Letters, 4*(8), Article 6001604. https://doi.org/10.1109/LSENS.2020.3006756

Hong, J., Li, H.-J., Yang, C.-C., Han, C.-L., & Hsieh, J.-C. (2022). A clinical study on atrial fibrillation, premature ventricular contraction, and premature atrial contraction screening based on an ECG deep learning model. *Applied Soft Computing, 126*, Article 109213. https://doi.org/10.1016/j.asoc.2022.109213

Houssein, E. H., Ibrahim, I. E., Neggaz, N., Hassaballah, M., & Wazery, Y. M. (2021). An efficient ECG arrhythmia classification method based on manta ray foraging optimisation. *Expert Systems with Applications, 181*, Article 115131. https://doi.org/10.1016/j.eswa.2021.115131

Katal, N., Gupta, S., Verma, P., & Sharma, B. (2023). Deep-learning-based arrhythmia detection using ECG signals: A comparative study and performance evaluation. *Diagnostics, 13*(24), Article 3605. https://doi.org/10.3390/diagnostics13243605

Kumar, A., Kumar, S., Dutt, V., Dubey, A. K., & García-Díaz, V. (2022). IoT-based ECG monitoring for arrhythmia classification using Coyote Grey Wolf optimisation-based deep learning CNN classifier. *Biomedical Signal Processing and Control, 76*, Article 103638. https://doi.org/10.1016/j.bspc.2022.103638

Liu, J., Li, Z., Fan, X., Hu, X., Yan, J., Li, B., Xia, Q., Zhu, J., & Wu, Y. (2022). CRT-Net: A generalised and scalable framework for the computer-aided diagnosis of electrocardiogram signals. *Applied Soft Computing, 128*, Article 109481. https://doi.org/10.1016/j.asoc.2022.109481

Liu, Y., Liu, J., Tian, Y., Jin, Y., Li, Z., Zhao, L., & Liu, C. (2024). Pruned lightweight neural networks for arrhythmia classification with clinical 12-lead ECGs. *Applied Soft Computing, 154*, Article 111340. https://doi.org/10.1016/j.asoc.2024.111340

Liu, Y., Qin, C., Liu, C., Liu, J., Jin, Y., Li, Z., & Zhao, L. (2022). Multiple high-regional-incidence cardiac disease diagnoses with deep learning and its potential to elevate cardiologist performance. *iScience, 25*(11), Article 105434. https://doi.org/10.1016/j.isci.2022.105434

Mekruksavanich, S., & Jitpattanakul, A. (2024). Deep residual network with a CBAM mechanism for the recognition of symmetric and asymmetric human activity using wearable sensors. *Symmetry, 16*(5), Article 554. https://doi.org/10.3390/sym16050554

Ngo, H. L., Nguyen, H. D., Loubiere, P., Tran, T. V., Şerban, G., Zelenakova, M., Bachofer, F., & Laffly, D. (2022). The composition of time-series images and using the technique SMOTE-ENN for balancing datasets in land use/cover mapping. *Acta Montanistica Slovaca, 27*(2), 407-425. https://doi.org/10.46544/ams.v27i2.05

Pokharel, A., Dahal, S., Sapkota, P., & Chhetri, B. B. (2024). *Electrocardiogram (ECG) based cardiac arrhythmia detection and classification using machine learning algorithms*. *arXiv*. https://doi.org/10.48550/arXiv.2412.05583

Qaisar, S. M., Khan, S. I., Dallet, D., Tadeusiewicz, R., & Pławiak, P. (2022). Signal-piloted processing metaheuristic optimisation and wavelet decomposition-based elucidation of arrhythmia for mobile healthcare. *Biocybernetics and Biomedical Engineering, 42*(2), 681-694. https://doi.org/10.1016/j.bbe.2022.05.006

Qin, S., Liu, L., Wang, X., Dong, N., Li, N., & Zheng, Q. (2024). ECG arrhythmia classification based on the fast ant colony clustering algorithm with improved spatiotemporal feature perception ability. *Heliyon, 10*(17), Article e37111. https://doi.org/10.1016/j.heliyon.2024.e37111

Ramkumar, M., Kumar, R. S., Manjunathan, A., Mathankumar, M., & Pauliah, J. (2022). Auto-encoder and bidirectional long short-term memory-based automated arrhythmia classification for ECG signal. *Biomedical Signal Processing and Control, 77*, Article 103826. https://doi.org/10.1016/j.bspc.2022.103826

Sangha, V., Mortazavi, B. J., Haimovich, A. D., Ribeiro, A. H., Brandt, C. A., Jacoby, D. L., Schulz, W. L., & Khera, R. (2022). Automated multilabel diagnosis on electrocardiographic images and signals. *Nature Communications, 13*, Article 1583. https://doi.org/10.1038/s41467-022-29153-3

Swetha, R., & Ramakrishnan, S. (2021). K-means clustering optimised fuzzy logic control algorithm for arrhythmia classification. In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICAECT49130.2021.9392494

Varshney, A., Kolhe, R., Gatne, S., & Ingale, V. V. (2022). Arrhythmia classification of ECG signals using undecimated discrete wavelet transform. In *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE. https://doi.org/10.1109/I2CT54291.2022.9824433

Zheng, Z., Chen, Z., Hu, F., Zhu, J., Tang, Q., & Liang, Y. (2020). An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology. *Electronics, 9*(1), Article 121. https://doi.org/10.3390/electronics9010121

Zou, C., Müller, A., Wolfgang, U., Rückert, D., Müller, P., Becker, M., Steger, A., & Martens, E. (2022). Heartbeat classification by random forest with a novel context feature: A segment label. *IEEE Journal of Translational Engineering in Health and Medicine, 10*, Article 1900508. https://doi.org/10.1109/JTEHM.2022.3203807

Pertanika J. Sci. & Technol. 34 (S1): 27 - 43 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Neuroimaging of Brain Activation in Emotional Decision-Making and Mental Health

Sailaja G[1]* and B. Narendra Kumar Rao[2]

[1]*Department of Computer Science and Engineering, School of Computing, Mohan Babu University, 517102 Tirupati, Andhra Pradesh, India*
[2]*Department of AI&ML, Sri Ramachandra Institute of Higher Education and Research, Chennai, India*

ABSTRACT

The fast development of metropolitan life and the quick development of assembled conditions frequently disregard the metropolitan personal satisfaction and the mental prosperity of occupants. Thusly, the conditions in which we invest a large portion of our energy are seldom inspected for their consequences for emotional wellness and prosperity. This paper plans to review and dissect existing research on what the assembled climate means for mental states, especially from the perspective of brain activity measurements that show transient perspectives. The writing audited reliably shows that natural habitats promote more relaxed and thoughtful brain activity, while constructed conditions are associated with increased feelings of anxiety in the human brain. This methodology and procedure spans two domains: the inner, abstract insight of feelings and considerations, and the outer universe of brain electrical activity. Using a clever occasion-related mind enactment imaging method, we show that transient mental cycles, happening within roughly 100 milliseconds, can be clearly connected to notice cerebrum enactment in controlled tests. Our technique utilises an ipsative evaluation approval process, which coordinates self-report information with ongoing EEG accounts, giving a far-reaching perspective on both mental and mind movement during transient responses, feelings, and choices considering controlled upgrades. To characterise an individual (or any framework) in terms of their likelihood of responding in a specific manner, one can utilise the progress probabilities within this range of possible reaction states. This model's potential uses and value in the domains of directing, brain research, and psychiatry are discussed and examined.

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 26 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.02

E-mail address:
gsailaja3@gmail.com (Sailaja G)
narendrakumarraob@gmail.com (Narendra Kumar Rao)
* Corresponding author

Keywords: Brain activation imaging method, brain electrical activity, EEG recordings, EEG signals, ipsative assessment, psychiatry

INTRODUCTION

Human feelings include a mix of mental and physiological reactions connected to

e-ISSN: 2231-8526

emotional sentiments, disposition, character, motivational tendencies, conduct responses, and physiological excitement. These feelings influence cognition, navigation, and relational communications. Positive feelings can improve day-to-day work effectiveness, while gloomy feelings can disrupt ordinary life. As of late, there has been a growing spotlight on cerebrum PC interfaces (BCIs), which use brain activity to empower consistent client PC cooperation (Vecchio, F., 2019). Essentially, the ascent of emotionally-man-made consciousness in human-machine interaction (HMI) has collected expanding interest.

The mix of feelings in human-PC communications has developed into the multidisciplinary field of emotional computing, which incorporates software engineering, neuroscience, brain research, and psychology. One critical issue within full feeling registration is feeling acknowledgement, with expected applications in sickness evaluation, checking exhausted driving, and assessing mental responsibility. However, previous in-depth examinations of the flourishing effects of natural change have typically focused on unambiguous environmental impacts (such as the impact of wildfires on thriving), success influences (such as word-related success results), or countries, effectively limiting our enthusiasm for what is currently known about the various flourishing effects of natural change across the globe (Collura, Thomas F, 2019). To guide future assessment and action to lessen and mitigate the prosperity impacts of ecological change and its usual results, a comprehensive summary of the most recent research on the prosperity impacts of natural change is required. Although there has been a significant increase in the number of special studies examining the economic effects of natural change over the past ten years, these studies do not consider the topic to be a central theme of the ongoing research on the subject. Clearly, deliberate evaluations provide the piece with a more effective arrangement strategy. Although past audits have inspected the well-being effects of environmental change, they normally centre around unambiguous ecological variables (e.g., the effect of rapidly spreading fires on prosperity), explicit well-being results (e.g., word related wellbeing results) (Craik et al., 2019), or individual nations, and are in many cases not extensive. This restricted focus restricts our general comprehension of the different wellbeing impacts of ecological change on a worldwide scale.

Environmental change has the potential to harm human well-being in two ways:

i. In the first place, by reducing the severity or recurrence of medical problems they are already impacted by climatic or meteorological conditions.

ii. Next, by causing exceptional or unforeseen Risks or health hazards in spots or seasons when they have never occurred there.

Environmental change adversely affects well-being; a few populations will be especially hard hit. These gatherings incorporate poor people, individuals of various backgrounds, and individuals with limited English proficiency and workers, native groups of people,

youths, pregnant women, more settled adults, feeble word related social affairs, people with disabilities, and people with diseases.

A healthy brain exhibits transitions prioritising safety and survival, such as fear and withdrawal, while maladaptive ones, like paranoia or antisocial tendencies, can be identified. These maladaptive transitions can hinder social relationships and overall well-being. It is important to address and work through these tendencies with the help of therapy or other interventions to promote a healthier mindset (Koelstra et al., 2012). These adaptive transitions may be indicative of underlying mental health issues or trauma that requires professional intervention and support. It is important to recognise and address these warning signs to promote overall well-being and functioning.

LITERATURE REVIEW

Theoretical Framework

Neuroscience offers a novel perspective on human behaviour and mental health. We can learn more about why people think, feel, and act the way they do by understanding how the brain works and what its priorities are. Although people may feel in charge of their lives, making their own choices and establishing their own priorities, this sense of control is totally reliant on the brain operating healthily (Alarcão & Fonseca, 2019). This dependence raises the possibility that our ideas about autonomy might not entirely match reality.

Deviated EEG enactment over the cerebrum is the primary focus of EEG-based investigation regarding the beginning and mind handling of feelings, according to the valence hypothesis. Alpha asymmetry is the characteristic most frequently utilised to identify variations in activation between the two cortical hemispheres. The F3 and F4 locations, which are above the dorsolateral prefrontal cortex, are the main locations where this alpha activity is observed. On the other hand, research that embraces the theory of discrete emotions frequently fails to establish a link between suggested EEG characteristics and neurophysiological theories (Katsigiannis & Ramzan, 2018). The research proposes that while EEG-based investigations normally depend on layered speculations, neuroimaging concentrates, as a rule, depend on discrete feeling hypotheses.

Neurophysiology of Emotions

To understand the neurophysiology of emotions, observations of brain reactions must be closely monitored. The limbic system, which consists of the hippocampus, hypothalamus, and amygdala, is one of the important structures. These areas are critical for memory formation, emotional processing, and controlling emotion-related physiological reactions. Sensation-based pathways allow emotional cues to reach the thalamus and then the amygdala with information. The amygdala is a key brain region that triggers emotional

reactions, especially when it senses danger or fear. It has a close relationship with the prefrontal cortex, which manages higher-order cognitive processes like controlling emotions and making decisions.

Neurotransmitters that modulate emotional responses and regulate mood include norepinephrine, dopamine, and serotonin. Connecting the brain's emotional responses to the autonomic nervous system and endocrine system, the hypothalamus (central nervous system) plays a critical role in regulating physiological reactions like breathing, heart rate, and hormone release (Bashivan et al., 2016). Basic emotions and brain activation regions are most closely linked in the following ways:

i. Fear and the amygdala;

ii. Disgust and the insula, amygdala, and ventral prefrontal cortex;

iii. Sadness and the medial prefrontal cortex;

iv. Anger and the orbitofrontal cortex;

v. Contentment and the rostral anterior cingulate cortex; and

vi. Individual differences like age and gender can also affect how the brain works.

The literature demonstrates that while EEG-based studies primarily rely on dimensional theories, neuroimaging studies primarily rely on discrete emotion theories. These different theoretical frameworks have led to varying interpretations of the neural correlates of emotion processing. EEG studies often focus on patterns of brain activity associated with valence and arousal, while neuroimaging studies may highlight specific brain regions activated during emotional experiences.

Research Method

The examination system for this study consolidated a recently developed mind initiation imaging procedure with deep-rooted self-assessment and self-report techniques in administration and training spaces. The target of this approach was to work on transient and spatial goals by developing prior examinations in mind unevenness and profound direction (Li et al., 2018). With only one preliminary, the method permitted brain activation to be estimated at a spatial goal of many millimetres and a transient goal of roughly 100 milliseconds.

The research methodology employed in this study combined a recently developed brain activation imaging technique with well-established report and rating techniques. To improve both temporal and spatial resolution, this method was built on earlier research on the asymmetry of the brain and making emotional decisions. The technique measured brain activation at tens of millimetres for spatial resolution and approximately 100

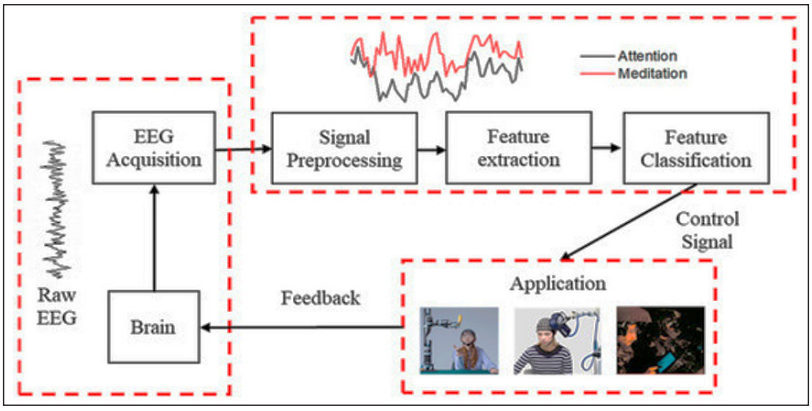


Figure 1. Response of the process validation protocol

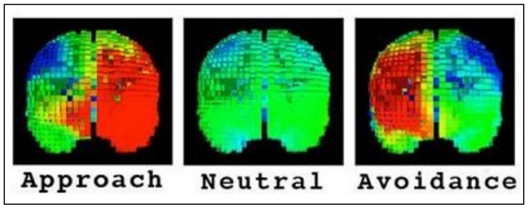


Figure 2. Frontal lobe gamma asymmetry summary

milliseconds for temporal resolution in just one trial. This creative system is considered a more extensive comprehension of the brain processes engaged with independent direction, revealing insight into the many-sided interchange between mental and profound variables (Diwadkar et al., 2017). The combination of techniques provided a unique perspective that could potentially lead to new insights in both educational and managerial settings. Determining whether assessment respondents are correctly processing and understanding the purpose of the assessment items can be aided by using response processing data, such as gamma asymmetry and sLORETA imaging, as shown in Figure 1.

Approach, neutral, and avoidance responses are shown in Figure 2, which illustrates frontal lobe gamma asymmetry. The brain is oriented anteriorly, with the left half of the picture addressing the right hemisphere. Red (white areas) in the neutral image denotes increased gamma activity, blue (darkest areas) denotes decreased activity, and green (greyest tones) denotes little to no activation (Yuan et al., 2012). These images depict the brain's activity at specific moments. The approach response is distinguished by increased gamma activity in the left frontal lobe, whereas the avoidance response has increased activity in the right frontal lobe. These findings indicate that different emotional states are associated with distinct patterns of frontal lobe brain activity.

The research has expanded on these methods by using a sequence of images (8 per second), with a time-frame-based depiction of emotional and decision-making processes in the frontal lobes.

i. **A summary of the architecture of sLORETA**: Researchers and clinicians can analyse sLORETA results to understand the spatial distribution of brain activity associated with cognitive processes, neurological disorders, or clinical conditions.

ii. **Acquisition of EEG Data:** The first step in the procedure is gathering an individual's EEG data. Electrodes are placed on the scalp to monitor the electrical activity of the brain. EEG data is typically made up of voltage measurements from various electrodes taken over a set period.

iii. **Model Process:** From the sLORETA's architecture, it is inferred that the simulation of Electrical signals, which are propagated from the brain to tissues to scalp electrodes, is used for analysing the EEG Data as well as neural processes by taking into consideration the geometry and conductivity of brain tissues. By precisely knowing the electrical signals' movement, researchers can accurately model the propagation of electrical signals (Michel et al., 2009) and come to a better understanding of the neural dynamics associated with problems related to brain disorders. From this information, advanced imaging technologies will be developed for better treatment of neurological conditions.

iv. **Pre-Processing EEG:** Removal of Systematic fluctuations in the EEG signal is a must and should be achieved by making baseline corrections, whereas noise and artefacts can be avoided by different techniques and advanced artefact rejection algorithms. After doing this, the EEG data is ready for further analysis. For future steps, the collected EEG data must be cleaned by repeating artefact removal as well as baseline correction.

v. **Source Localisation:** The sLORETA architecture mainly investigates the sources of brain activity from EEG data. sLORETA uses different models to assume the different activities of the brain coming from different areas of the brain. It finds the current density at each point in a three-dimensional brain space using a linear inverse solution. sLORETA is widely used for precise estimation of neural sources in neuroimaging research to accurately localise brain activity by using both spatial and temporal information.

vi. **Normalisation:** sLORETA's important characteristic is its normalisation. It utilises cathode facilities and a normalised head model for better performance among studies and exploration groups for reproducible outcomes. sLORETA's advanced procedures lead to helping in minimizing error sources and bias, which gives prominent results.

vii. **Resolution:** Compared to other EEG source imaging techniques, sLORETA has made its mark for effectively managing as well as localising the inverse problem and

mitigating volume conduction by estimating brain activity on a coarse grid of voxels using spatial resolution. Besides, sLORETA has been very good at localising deep brain sources.

viii. **Yield Representation:** sLORETA is calculated by assessing the thickness values of all through the cerebrum using three layers of map. From this, different brain activities as well as the conveyance of power of movement are addressed. It helps researchers and clinicians to better understand brain activity patterns and their implications, which leads to improvement in the diagnosis and advanced developments in neurological conditions. sLORETA maps give better insights into the inner brain workings by providing detailed brain activities.

8 maps per second or sample the filtered sLORETA current-source density data at 125 ms intervals would be enough to capture state changes. This is confirmed visually, from the sequential maps clustering into groups with gradual changes (Schirrmeister et al., 2017). Various maps have characteristics in common with nearby instances shown in Figure 3. This infers smooth transition between states. sLORETA current-source density data reveals the effectiveness of the chosen sampling interval for capturing state changes.

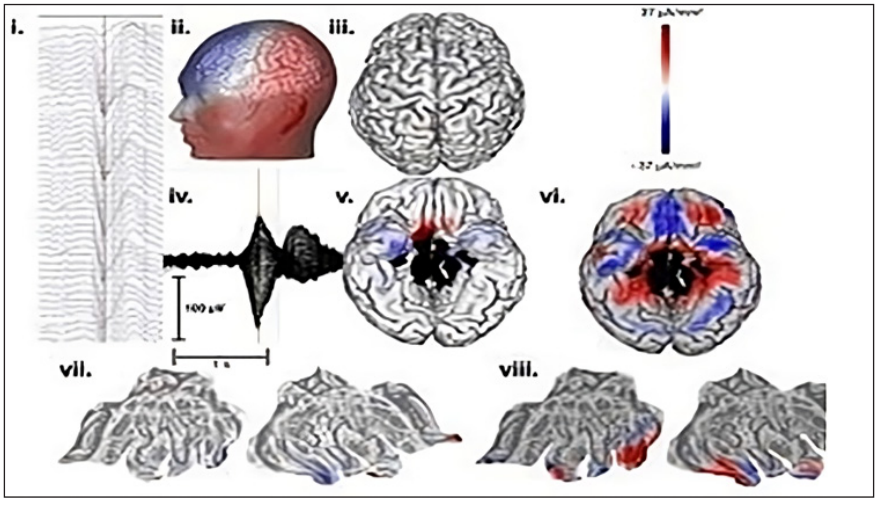


Figure 3. Frontal-negative slow oscillation

Haidt's Intuitive Model

By incorporating brain pathways and providing insight into brain activities associated with decision making, Haidt's model, which is described in detail in "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," improves Resolution Methods. His approach to decision-making questions the conventional "rationalist

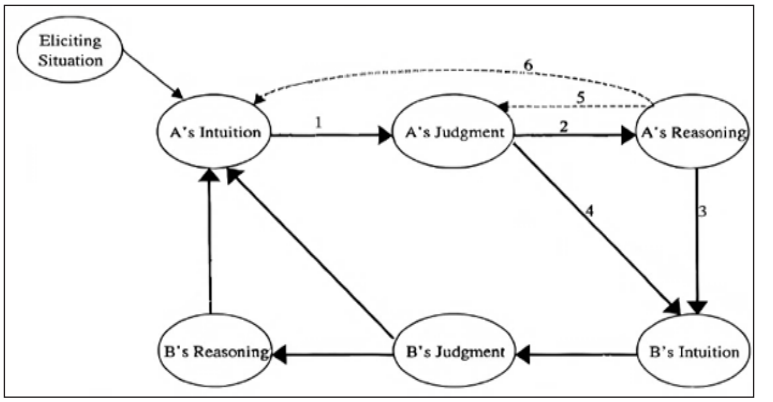


Figure 4. Haidt's intuitive model

approach" and substitutes an intuitionist model. According to this model, Moral intuitions include moral emotions. The first inevitably results in judgments that match them. These intuitions are referred to by Haidt as "primary emotions/sensations," and the ultimate assessment is referred to as "secondary emotions/perceptions." Haidt's research emphasises the importance of emotions in shaping moral judgments, highlighting the Function of intuition in decision-making processes (Li et al., 2022). By recognising the influence of emotions on moral reasoning, his model provides a more comprehensive understanding of human behaviour and decision-making.

Figure 4 demonstrates that emotionally intense situations first evoke an intuitive response or feeling, which then gives rise to a judgment. In this framework, judgments are formed before deliberate reasoning (Trompetto et al., 2019). Consequently, the concluding phase of cognitive processing is less about systematically evaluating evidence and more about explaining and rationalising the judgment that has already been made. This framework covers interactions between individuals, like that of A and B, as well as individual A's response to a triggering circumstance. These people could be a couple having a conversation, a client and therapist, or any two individuals with a clear and distinct relationship.

This model, which can be applied to human cooperation in different settings, serves as the reason for the creators' findings, providing a close-to-home and objective direction. In interpersonal communication, for example, one person's behaviour towards another person evokes an emotional response which leads to judgment, eventually, a kind of reasoning (Yang et al., 2020). The emotional reactions, decision-making, and so on are the different stages of this cyclical process, which repeats itself. Knowing this process can provide insights into how individuals go through complex social situations and make decisions based on both emotional and rational factors.

The dashed lines labelled 5 and 6 show that reasoning should be added to intuition and ultimate judgment, instead of being treated as an afterthought. This addition leads to a holistic decision-making process, which includes both logical analysis and gut feelings. By mixing these two aspects, individuals can make better choices.

Toward a Model Obviously

Fundamental ideas put forth by the authors lead to the model "Toward a Functional Model of Navigation, which is close to home guidelines and psychological wellness effects. The work is based on advanced and superior models.

The left and right hemispheres of Figure 5 show the decision-making process. The right side of the equator generally takes part in a worldwide design acknowledgement check considering previous encounters, geared towards quickly distinguishing expected dangers. The left hemisphere is serial in the process.

The significant obstacles are directly represented by the applications of neuroscience principles to mental health issues (Islam et al., 2021). The integrative model advances to a more profound degree of understanding by consolidating a key useful differentiation between the hemispheres. This model suggests that the left hemisphere oversees positive emotional states and related behaviours like approach, while the right hemisphere oversees negative emotional states and related behaviours like withdrawal. This is further developed by incorporating various processing modes (parallel vs. sequential) and processing levels (sensation, perception) as key dimensions in our model.

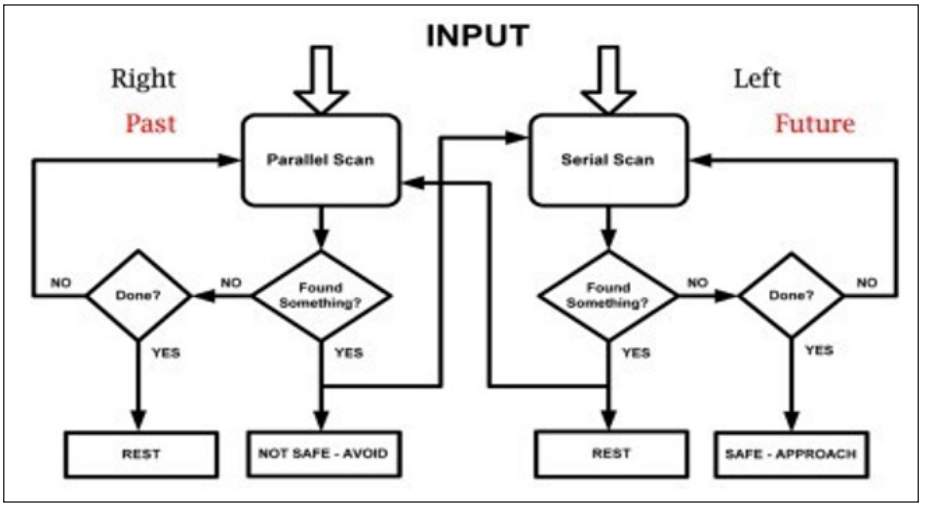


Figure 5. Emotional decision-making model

RESULTS AND DISCUSSIONS

As a result of this investigation, several studies have been published that validate that the proposed model serves as a framework for coordinating and deciphering in-depth experiences close to remotely estimated physiological responses. In particular, the subsequent guides and connections outline not only people's responses and choices but also how these connect with their internal thoughts and discourses. To outline the wide material of this methodology, a few central trials were conducted, each utilising a particular strategy to correlate physiological measures with in-depth reactions. Pilot concentrates on investigating different components of personal dynamic cycles.

Stress Detection Methods

The anxiety level factors, such as psychological, physiological, environmental and social is analysed using key trends and patterns. The correlations between these anxiety factors are shown through heat maps, which mark the features most strongly attached to stress.

Psychological Factors : The heat map for psychological factors gives different comparisons between anxiety levels and variables such as mental health history and depression (Roy et al., 2019). These factors have a great impact on stress, which tends to make individuals with a history of mental health challenges more likely to face higher levels of anxiety.

Environmental Factors : The environmental analysis indicates a medium negative correlation between noise levels and anxiety, which leads to greater stress and anxiety.

Social Factors : The social factors heat map shows a strong correlation between anxiety levels, the lack of social support, as well as higher peer pressure, which tends to result in elevated anxiety.

Psychological Factors

i. Visual Analysis:

- Visual representation of self-esteem levels in the dataset is given by the histogram using code. It includes:
- A distribution curve (KDE) overlaid on the histogram represents the density of self-esteem scores.
- A red dashed line tells the average self-esteem level, which gives a visual assessment of how many individuals fall below this threshold.

ii. Analytical Approach:

The analytical method gives the grouping of the data, which tells whether everyone's self-esteem level is below or above the average shown in Figure 6.

$$\mu SE = 1/N \sum_{i=1}^{n} SEi \quad [1]$$

Where SE_i = self-esteem score of the i^{th} participant, $\mu_{SE=}$ mean self-esteem score. Equation 1 gives the group individuals count (below average vs. above average).

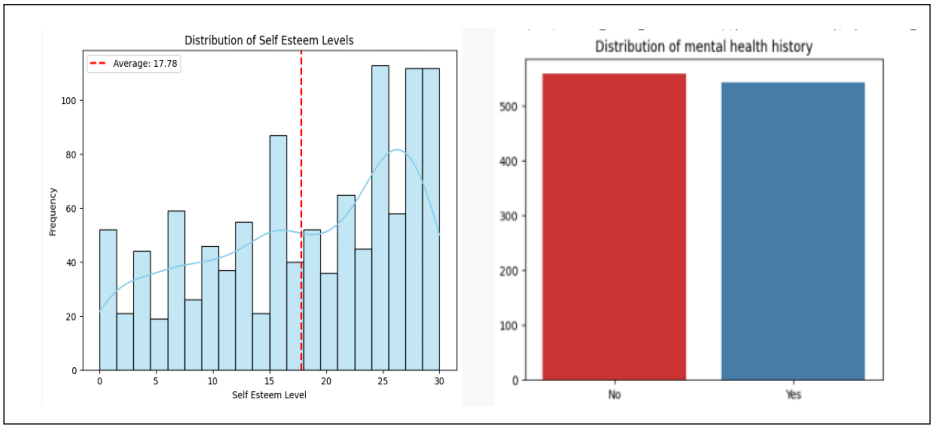


Figure 6. Distribution of self-esteem level vs. mental health history

The distribution of blood pressure categories observed in the dataset is presented in Figure 7. Additionally, the proportion of individuals experiencing stress compared to those without stress is illustrated in Figure 8. The average blood pressure is 2.1818.

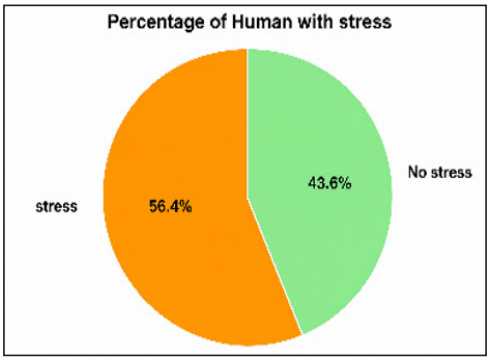


Figure 7. Blood Pressure Distribution

blood_pressure	count
3	500
1	300
2	300

Figure 8. Stress Status Distribution

sleep_quality	
1	328
5	192
2	188
4	188
3	170
0	34

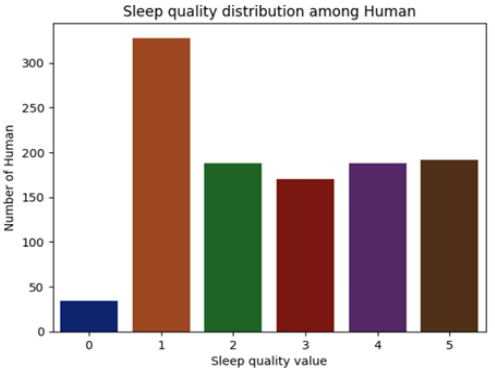


Figure 9. Sleep quality distribution

The distribution of sleep quality levels among the participants is illustrated in Figure 9, showing the frequency of individuals across different sleep quality categories.

Numbers equal to or lower than 2 will be considered 'poor '. Human Depression percentage is 56.4.

Environmental Factors

The findings say that the maximum number of individuals report high noise levels, and the visual representation effectively represents the impact of noise on the population, which leads to areas for improvement in community health and environmental conditions is shown in Figure 10.

	noise_level
count	1100.000000
mean	2.649091
std	1.328127
min	0.000000
25%	2.000000
50%	3.000000
75%	3.000000
max	5.000000

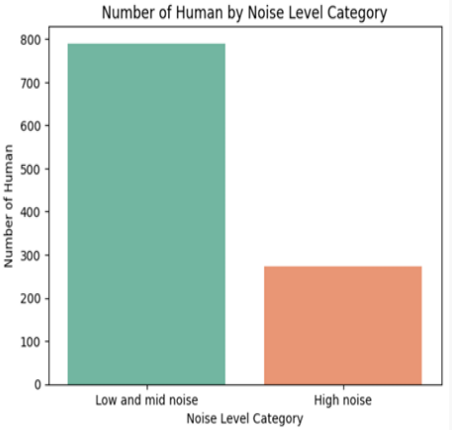


Figure 10. Humans live in conditions with high noise levels

The data is divided into two groups:

- Low and Mid Noise: Noise levels from 0 to 3.
- High Noise: Noise levels of 4 and 5.

This division gives a better understanding of the distribution of noise levels among the respondents.

Visualisation with Bar Plot. It is divided into two types of visual breakdown.

- The x-axis represents the noise level type.
- The y-axis indicates the number of individuals in each type.

Social Factors

A deep study revealed key insights into social and extracurricular factors among individuals. A maximum score of 3 on the scale says that strong social support was reported by 458 participants, which says the value of reliable support networks. 32.73% of respondents faced bullying or similar forms of persecution that require action is shown in Figure 11.

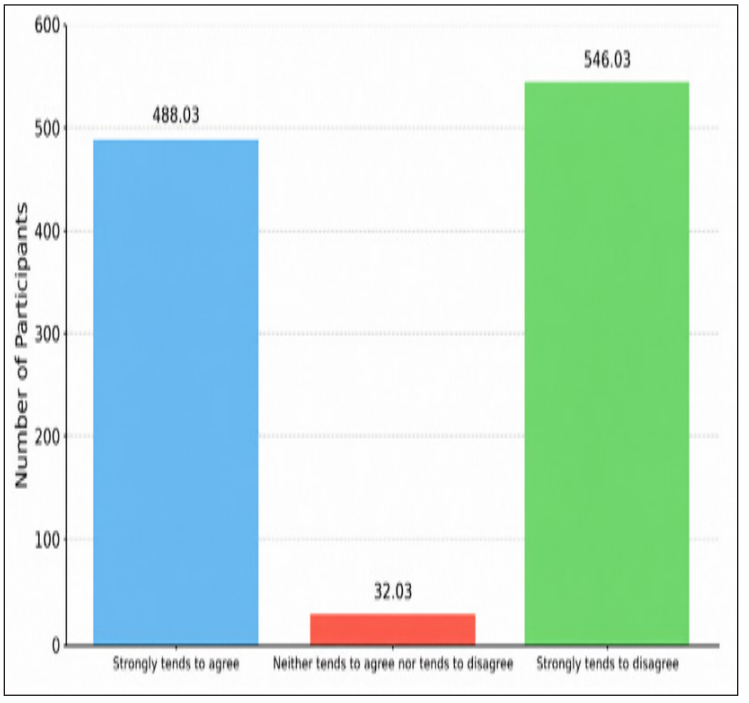


Figure 11. Analysis of social and extracurricular factors among individuals

This visualisation highlights the distribution of factors such as strong social support, experienced persecution and participating in extracurricular activities, emphasising areas requiring attention.

General Exploration

A thorough finding revealed negative experiences across psychological, physiological, environmental, academic, and social factors. The findings presented in Tables 1 to 4 emphasise the critical need to prioritise psychological well-being and strengthen social support systems

Table 1
Psychological

Negative Experiences for Psychological	
Feature Sums	
Anxiety level	512
Self-esteem	670
Mental health history	558
Depression	652
dtype	int64
Factor Total	2392

Table 2
Physiological

Negative Experiences for Physiological	
Feature Sums	
Headache	544
Blood pressure	500
Sleep quality	550
Breathing problem	547
dtype	int64
Factor Total	2141

Table 3
Environmental

Negative Experiences for Environmental	
Feature Sums	
Noise level	537
Living conditions	549
Safety	535
Basic needs	552
Noise level	int64
Factor Total	**2173**

Table 4
Social

Negative Experiences for Social	
Feature Sums	
Social support	600
Peer pressure	573
Extracurricular activities	550
Bullying	541
Social support	int64
Factor Total	**2141**

The findings suggest the need to give importance to psychological well-being and enhance social support systems to greatly reduce these negative experiences. These findings suggest promoting better mental and social health outcomes.

Comparative Analysis

The bar plot of Figure 12 shows the correlation coefficients between various psychological and health-related factors:

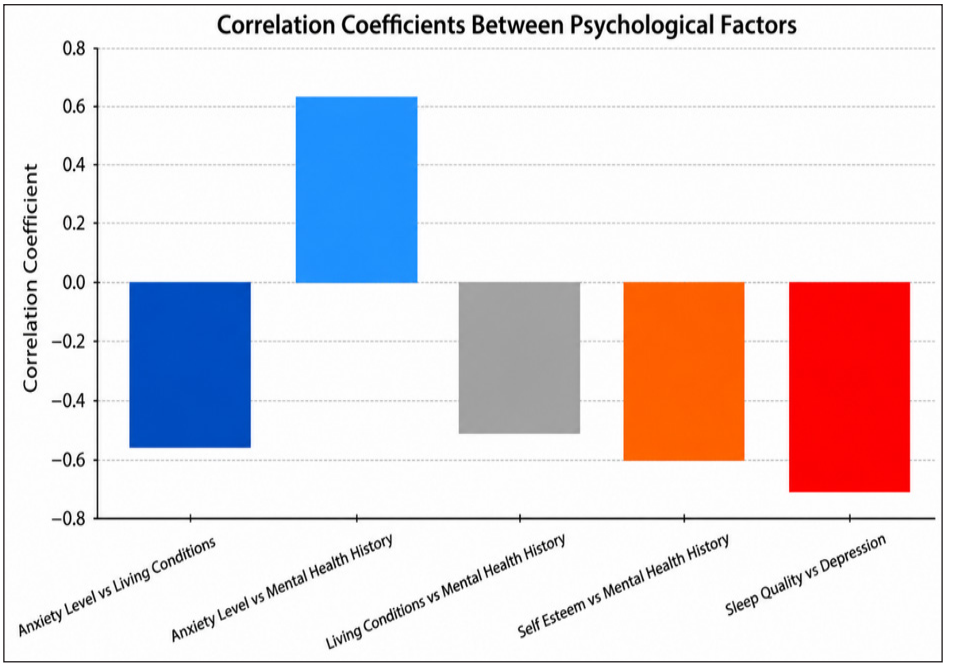


Figure 12. Correlation between anxiety, mental health, and other factors

1. Anxiety Level vs. Living Conditions:

   A moderate negative correlation of -0.568 indicates that individuals with lower anxiety levels are likely to have good living conditions.

2. Anxiety Level vs. Mental Health History:

   A strong positive correlation of 0.634 indicates that those with a history of mental health issues often experience higher anxiety levels.

3. Living Conditions vs. Mental Health History:

   A moderate negative correlation of -0.509 reveals that worse living conditions are associated with more mental health challenges

4. Self-Esteem vs. Mental Health History:

   A negative correlation of -0.604 shows that lower self-esteem tends to correlate with a higher incidence of mental health issues.

5. Sleep Quality vs. Depression:

   A strong negative correlation of -0.693 points to the fact that poor sleep quality is strongly linked to higher depression levels

CONCLUSION

Neuroscience and psychological wellness experts perceive the vital role of feelings in navigation, yet pragmatic application is obstructed by an absence of a unified language and a model that clearly represents possible Connections to neurons. Through obtaining a more profound comprehension of how the cerebrum decides and the impact of feelings on these cycles, we can start to unwind the moment-to-moment elements of human way of behaving, including the job of subliminal considerations. SLORETA's normalised approach guarantees consistency and equivalence across various investigations and examination groups. Although it works at a lower spatial goal, it displays a grand common objective, making it proper for investigating dynamic psyche processes. Its applications range from mental neuroscience research to clinical applications in diagnosing and observing neurological and mental conditions.

Neuroscience and psychological wellness experts recognise the role of feelings in navigation but lack a consistent language and model for neurological pathways. SLORETA's normalised approach ensures consistency across studies and is suitable for studying dynamic mind processes. Its applications include mental neuroscience research and clinical diagnosis of neurological and mental conditions.

This new design allows for the interpretation of confirmation-based and dispersed strategies such as rethinking, reworking, testing, reflecting, and attentive listening. Joint efforts and the aide-client relationship are understood as associations between two frontal cortices, each of which picks up on, considers, and forms a connection with a cycle. These interactions give a better understanding and a shared decision-making process, which enhances strategy implementation. Through a collaborative environment, both parties can provide their unique perspectives and skills to achieve successful outcomes. In summary, these heat maps offer valuable insights into the various factors, such as human stress, highlighting the intricate connections between mental health, physical health, the environment, and social support. The psychological and social factors were found to have a great impact on anxiety, especially in areas such as mental health history, depression, and the availability of social support.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

REFERENCES

Alarcão, S. M., & Fonseca, M. J. (2019). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing, 10*(3), 374-393. https://doi.org/10.1109/TAFFC.2017.2714671

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2016). Learning representations from EEG with deep recurrent-convolutional neural networks. In *International Conference on Learning Representations (ICLR 2016)*. https://doi.org/10.48550/arXiv.1511.06448

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering, 16*(3), Article 031001. https://doi.org/10.1088/1741-2552/ab0ab5

Diwadkar, V. A., Asemi, A., Burgess, A., Chowdury, A., & Bressler, S. L. (2017). Potentiation of motor sub-networks for motor control but not working memory: Interaction of dACC and SMA revealed by resting-state directed functional connectivity. *PLOS ONE, 12*(3), Article e0172531. https://doi.org/10.1371/journal.pone.0172531

Islam, M. R., Moni, M. A., Islam, M. M., Rashed-Al-Mahfuz, M., Islam, M. S., Hasan, M. K., Hossain, M. S., Ahmad, M., Uddin, S., Azad, A., Alyami, S. A., Ahad, M. A. R., & Lio, P. (2021). Emotion recognition from the EEG signal, focusing on deep learning and shallow learning techniques. *IEEE Access, 9*, 94601-94624. https://doi.org/10.1109/ACCESS.2021.3091487

Katsigiannis, S., & Ramzan, N. (2018). DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics, 22*(1), 98-107. https://doi.org/10.1109/JBHI.2017.2688239

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18-31. https://doi.org/10.1109/T-AFFC.2011.15

Li, X., Zhang, P., Song, D., Yu, G., & Hou, Y. (2022). Emotion recognition from EEG using multi-scale deep neural networks. *Biomedical Signal Processing and Control, 72*, Article 103299. https://doi.org/10.1016/j.bspc.2021.103299

Li, Y., Zheng, W., Cui, Z., Zong, Y., & Ge, S. (2018). EEG emotion recognition based on graph regularised sparse linear regression. *Neural Processing Letters, 49*, 555-571. https://doi.org/10.1007/s11063-018-9827-7

Michel, C. M., Koenig, T., Brandeis, D., Gianotti, L. R. R., & Wackermann, J. (Eds.). (2009). *Electrical neuroimaging*. Cambridge University Press. https://doi.org/10.1017/CBO9780511596889

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering, 16*(5), Article 051001. https://doi.org/10.1088/1741-2552/ab260c

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualisation. *Human Brain Mapping, 38*(11), 5391-5420. https://doi.org/10.1002/hbm.23730

Trompetto, C., Marinelli, L., Puce, L., Mori, L., Serrati, C., Fattapposta, F., & Currà, A. (2019). "Spastic dystonia" or "inability to voluntarily silence EMG activity"? Time for clarifying the nomenclature. *Clinical Neurophysiology, 130*(6), 1076-1077. https://doi.org/10.1016/j.clinph.2019.03.009

Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2020). Continuous emotion recognition based on EEG signals using LSTM networks. *Applied Soft Computing, 96*, Article 106697. https://doi.org/10.1016/j.asoc.2020.106697

Yuan, H., Zotev, V., Phillips, R., Drevets, W. C., & Bodurka, J. (2012). Spatiotemporal dynamics of the brain at rest: Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *NeuroImage, 60*(4), 2062-2072. https://doi.org/10.1016/j.neuroimage.2012.02.031

Pertanika J. Sci. & Technol. 34 (S1): 45 - 68 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Evaluation of Classifiers for Non-Invasive Heart Rate Measurement Using Video Magnification

Rupinder Saini* and Pooja Sharma

Department of Computer Science and Engineering, Rayat Bahra University, 140301 Kharar, Punjab, India

ABSTRACT

As the current world approaches a better lifestyle, medical problems have also increased and gained a lot of attention in the last couple of years. Among others, heart rate detection has become a vital part of remote healthcare monitoring and diagnostic systems. The contactless heart rate measurement is a recent concept that monitors the colour changes in the skin, specifically the facial region, to assess the heart rate of the human. In this paper, the authors have presented an analysis of various machine learning classifiers and independent component analysis methods to support this concept, as the naked eye visualisation cannot detect minor changes in the face. Hence, machine learning and magnification of the frame become a vital step. The paper describes the general procedure of detecting the heart rate by monitoring the forehead of a live object. The training and classification procedure is discussed in detail, and a comparative analysis based on quantitative parameters for different machine learning classifiers is also presented in this paper. The analysis among various classifiers shows that the neural network with the highest accuracy proved to be a better machine learning classifier for heart rate detection work.

Keywords: Artificial intelligence, data mining, heart rate detection, machine learning, magnification

INTRODUCTION

The contemporary society is currently striving towards enhancing the field of Human-to-computer (H2C) interaction to attain maximum benefits from machines. However, recent decades have also witnessed major health issues, such as cancer, heart failure, and brain disorders, due to technological advancements. Rochester medical research suggests that heart-related problems have increased due to several

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 27 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.03

E-mail address:
errupindersaini2@gmail.com (Rupinder Saini)
pooja.rani@rayatbahrauniversity.edu.in (Pooja Sharma)
* Corresponding author

e-ISSN: 2231-8526

factors, including economic and mental stress (Krajnak, 2014) and food intake behaviour. Individuals in the workforce and office settings are particularly vulnerable to heart attacks due to excessive workloads, long shifts, and mental pressure that increase blood pressure (Alishahi et al.,2022). This research aims to establish a relationship between changes in facial appearance and heart rate due to stress. Modern life has resulted in individuals competing with millions of others, leading to anxiety, depression and other psychological and physical health issues. Unhealthy lifestyle choices, such as stress and poor diet, can lead to metabolic disorders, diabetes, joint problems, cardiovascular diseases, tumours, hypertension and obesity (Virtanen et al., 2012). Although technology has made life easier, it has drawbacks as well. Globally, heart-related deaths are increasing, accounting for 32 per cent of all deaths in 2019.

The World Health Organisation (2025) says that bad food, not getting enough exercise, smoking, and drinking too much alcohol are the main behaviours that put people at risk for heart problems (WHO,2025). Therefore, it is necessary to research cardiac problems and use technology to transform available data into valuable knowledge to facilitate better assessments of patients' health. This paper aims to monitor patients suffering from heart diseases, as changes in blood pressure can be fatal (Hwang & Lee, 2026). Facial colour changes are the first visible symptom, with the face turning red as blood pressure increases. At early stages, these changes are not easily identifiable, making it crucial to observe changes through magnified video. In such cases, the video is extracted into frames, and the forehead is selected as the Region of Interest (RoI). Male facial hair may interfere with the colour change detection, making the forehead the preferred RoI. Data mining refers to mining desired information. A data mining architecture is divided into three blocks, namely the input block, the processing engine, also referred to as the inference engine and the output block. The inference engine processes the supplied input values based on a stored repository that is supplied to the engine before putting it to any test. The general work diagram of data mining architecture is provided in Figure 1.

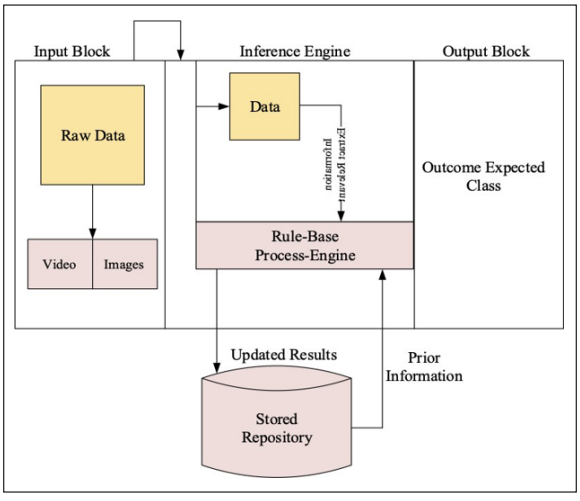


Figure 1. The general data mining architecture

Figure 1 shows the general data mining architecture for heart rate estimation, which highlights the stages involved in data acquisition, preprocessing, feature extraction, and classification. A data mining architecture is divided into three blocks, namely the input block, the

processing engine, also referred to as the inference engine and the output block. The main components of a data mining architecture system are as follows:

Input Set : The input set contains the objects and their descriptions. The input might be a video, frames of a video, audio segments, images, etc. When it comes to medical data analysis, surveillance through videos has gained a lot of popularity (Ben et al.,2020). The visual aspects (Špetlík et al., 2018) provide a significant amount of information that becomes useful when it comes to training the system (Diller et al.,2014).

Inference Engine : The inference engine is the main working block of any data mining architecture, in which the inference engine processes the input values and decides based on the previous information stored in the repository. A stored repository is created using a training algorithm.

Output Set : The output set gives the final decision produced by the inference engine. As per prior knowledge stored in the repository and after applying basic rules and algorithms, the system generates outcome-based expected classes, which are utilised for classification and detection of the provided input data. This output is then utilised for analysis, validation or decision-making.

This study represents a structured comparative analysis of machine learning classifiers for non-invasive heart rate estimation within an ICA and video magnification-based framework.

The rest of the paper is organised in the following manner. Section 2 represents the search methodology employed in conducting the detailed literature review presented in Section 3, focusing on contactless heart measurement. Section 4 discusses the concept of training and classification algorithms and statistical machine learning metrics. The comparative analysis of the classifiers was conducted in Section 5 based on reported results from existing studies. The paper is finally concluded in Section 6.

The rigorous evaluation of lightweight machine learning classifiers in conjunction with an ICA-based signal extraction and video magnification framework for non-invasive heart rate measurement is the primary originality of this work. Unlike previous studies that primarily focus on deep learning-based approaches (Ben Ghezala, H. (2020) requiring large-scale datasets and high computational resources, this work focuses on computationally efficient models ideal for real-time and resource-constrained scenarios. Additionally, by providing a comprehensive comparison analysis within a single experimental flow, the paper exposes the trade-offs between computational complexity and accuracy. This makes the proposed method particularly relevant to edge-based healthcare monitoring systems.

METHODOLOGY

The review analysed the existing literature on the topic "heart rate detection using video samples", and the focus was given to the articles based on machine learning and independent component analysis. The academic databases referred to for the literature search include IEEE Explorer, Scopus, Google Scholar, etc in addition to Google search to ensure comprehensive coverage of the most relevant literature. The search statistics have involved articles published against two keywords, namely, "Heart Rate", "video samples", "machine learning", "independent component analysis", etc.

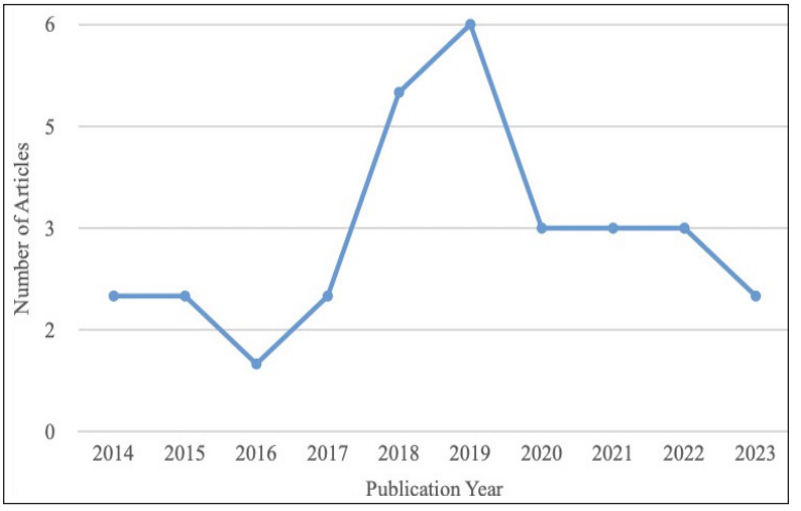


Figure 2. Distribution of articles used in the study

Figure 2 represents the number of articles considered in this study. In addition to several inclusion and exclusion criteria, the whole survey comprises papers published from 2014 to the present, and the number of papers covered from each year is graphically depicted in Figure 2. The screening process used here involves titles, abstracts, and full-text availability in addition to being a recent-year publication.

LITERATURE REVIEW

Contactless techniques to measure heart rate have gained more attention in recent years due to their non-invasive and remote features (McDuff et al.,2015). It has gained high popularity in the healthcare industry to monitor fitness and handle healthcare concerns using human-computer interaction. To measure heart rate, thermal imaging, facial videos, and photoplethysmography (Acharya et al., 2025), signals play a very important role. In a study by Xiaochuan et al. (2016), a digital camera was used to capture conventional wrist movies, and the Eulerian video magnification (EVM) method was used to detect

the wrist pulse signals non-intrusively. The signals were analysed using the 2-Gaussian curve modelling approach, and the frames of the captured films were subjected to spatial decomposition and temporal filtering. The results of the experiments showed that the EVM approach may be used to capture the properties of the wrist pulse signal and can be utilised to anticipate major cardiovascular events. Rizvi et al. 2018 modified the EVM video post-processing control settings using the PCB-IDE and used the differential algorithm using selective mutation (DE/rand/1/bin) to provide an EVM control parameter tuning method. The suggested PCB-IDE aids specialists in choosing the best settings for the EVM process, and the interactive parameter estimation through PCB-IDE is viable, although it suffers from user fatigue limitations. Next research should focus on developing application-specific meta-models to alleviate user fatigue and further the optimisation process's exploration of the population space. There are several researchers who have taken advantage of Independent Component Analysis as a powerful signal processing technique to analyse mixed and complex signals generated due to skin colour, physiological signals and motion artefacts (Casalino et al.,2019).

An innovative technique for remote photoplethysmography signal measurement from facial movies is presented by Yu, Li, et al. 2019. Their method, which makes use of Spatio-Temporal Networks, shows encouraging results in non-invasive physiological monitoring. The work could benefit from addressing potential issues with lighting and subject diversity; however, it lacks thorough validation on a variety of datasets. Overall, this study is an important step towards convenient and accessible health monitoring, but more study is required to increase its robustness and generalizability.

The use of highly compressed facial recordings (Peng et al., 2019) offers an amazing deep learning method for remote heart rate assessment (Debnath et al.,2025). Their approach has great potential for non-invasive health monitoring. The method is impressively able to extract heart rate data in challenging video conditions. However, it's vital to keep in mind that extra testing and validation on a larger dataset are valuable in establishing the practical utility and generalizability of the proposed system. Overall, the field of computer vision for healthcare applications has advanced significantly because of this study.

In the area of remote photoplethysmography, Hu et al. (2019) demonstrated a cutting-edge strategy. They demonstrated the potential of their technique for non-invasive health monitoring applications by showcasing notable developments in the spatial-temporal information capture field. The study, however, was hampered by a lack of readily available data, which might have affected how broadly applicable their findings were. Despite this, the novel convolutional neural network design by the scientists and its encouraging results represent a substantial advancement in the field of remote PPG research.

Zarezadeh and Rezaeian (2016) offered a novel method for recognising face micro-expressions using the EVM approach to extract the fine movements of the face. The eigenface methodology was applied to detect micro-expressions in the SMIC database and classify detected micro-expressions as negative vs positive in the SMIC database and

pleasure versus disgust in the CASME database. The findings demonstrated that using the eigenface approach with the EVM method to retrieve minor facial movements improves micro-expression identification performance. Yu, Peng, et al. (2019) demonstrated how video sequences may be used to get dynamic heart rate readings, which are generally collected from sensors placed near the heart. An independent component analysis (ICA) combined with mutual information was used to make sure that accuracy is not lost while using short video durations. Experimental findings demonstrated that the suggested approach may be utilised to quantify dynamic heart rates, with RMSE and correlation coefficients of 1.88 BPM and 0.99, respectively. A system proposed by Antink et al. (2019) multimodal sensor fusion approach for discrete vital sign monitoring. The technique used skin colour fluctuations and the head motion signal taken from a webcam video. Their findings prove that the proposed technique is effective for unobtrusive and continuous monitoring of both heart rate and respiration rate. A Multi-hierarchical Convolutional Network (MCN) was developed by Zhang et al. (2021) for effective remote photoplethysmograph (rPPG) signal and heart rate prediction from facial video. The main objective of the study was to improve the accuracy and computational efficiency of rPPG-based heart rate estimation. MCN extracted the spatial and temporal features from facial video data by using hierarchical convolutional layers to predict reliable heart rates from video recordings. The study generates promising results. However, variations in illumination and facial expressions make the model susceptible to noise, and it may affect the quality of rPPG signals and, consequently, heart rate measurement.

Hu et al. (2021) proposed ETA-rPPGNet, an Effective Time-Domain Attention Network to enhance accuracy and robustness for remote heart rate measurement from facial images. The network employs time-domain attention procedures that focus on relevant temporal features, which help improve heart rate estimation accuracy. The study proved a considerable improvement in heart rate estimation accuracy when it is compared to existing techniques. The increase in computational complexity of the attention mechanism may affect its feasibility for real-time processing on resource-constrained systems.

Cotes et al. (2022) proposed an observational research study on assessing schizophrenia and depression using multiple modalities, incorporating heart rate technologies. The objective of the work was to explore whether multimodal data, including heart rate, can support diagnosis and evaluate mental conditions. The report presents findings that highlight important information about how to integrate heart rate information into mental health assessments. The complexity of analysing and processing multimodal data is a significant challenge, and it requires data fusion and analysis methodologies.

Hu et al. (2022) introduced a Spatial-Temporal Attention Network to provide a robust heart rate estimate from face images. The research focused on enhancing the accuracy of heart rate measurement under challenging conditions like facial movements and noise. The network utilises spatial-temporal attention procedures to focus on essential regions

and frames in the video frame. The study demonstrated notable improvements in heart rate estimation and accuracy, however, in difficult situations. A major problem could be the requirement for a large amount of training data to generalise well across varied individuals and conditions.

A technique to estimate heart rate and heart rate variability in real-time from photographs is put forth by Su et al. (2023). The authors aimed to make picture data available for real-time vital sign monitoring. Heart-related signals from image sequences are extracted using an approach that combines independent component analysis and particle swarm optimisation. The creation of an unobtrusive, real-time heart rate monitoring device is one possible result of this research. The accuracy of heart rate variability estimation, however, could be problematic because it can be impacted by several things, such as motion (Estepp et al.,2014), artefacts and picture data noise.

ViT-rPPG, a network based on vision transformers for remote heart rate measurement, is presented by Sun et al. (2023) with the sole aim of using vision transformers to boost the precision of remote heart rate estimation. The study shows how well ViT-rPPG captures temporal characteristics from video data, yielding precise heart rate estimations. The computational demands of vision transformer networks, however, provide a major obstacle and may prevent their use on devices with limited resources.

A detailed comparative analysis of recent studies aimed at heart rate measurement based on face videos is summarised. Independent component analysis has played a very important role in improving the capabilities of machine learning-based mechanisms involved in contactless heart rate measurements.

Table 1 shows that a few public datasets have been widely used in the simulation analysis by several researchers in recent years. Further, the observed accuracy by which the HR has been measured and the correlation coefficient are also common evaluation parameters that helped the research community to evaluate their work.

The development of non-invasive techniques for monitoring cardiovascular health depends heavily on video files for heart rate measurement investigations. These datasets are used to create and test algorithms that can predict heart rates without having direct contact with a person's skin or pulse. Typically, these datasets are video recordings of people's faces or bodies. As an outcome of the detailed analysis of the review, several video datasets came to light, including PURE, OBF, MAHNOB-HCL, etc., that have demonstrated several applications in addition to some of the self-recorded sample datasets. In addition to this, there also exist a large number of datasets with a variable number of subjects and videos to support training and evaluation of machine learning models and Popular datasets such as DEAP (Koelstra et al., 2012), MAHNOB-HCI (Soleymani et al., 2012), and UBFC-rPPG (Bobbia et al., 2017) are widely used for rPPG evaluation. A handful of them are summarised in Table 2 to provide a broad option to the reader for the selection of the dataset for future research.

Table 1
Summary of heart rate measurement techniques

Year	Citation	Objective	Technique	Dataset	Parameter
2017	Singh et al., 2017	Contactless real-time heart rate measurement	Kalman-Lucas-Tomasi (KTL) Face De-tection Algorithm, Independent Component Analysis (ICA), Fast Fourier Transform (FFT)	Self-Recorded Dataset	Heart Rate Measurement and Accu-racy
2017	Alghoul et al., 2016	Heart rate traction variability ex-	Independent Component Analysis (ICA) and Empirical Mode Decomposition (EMD)	Not Specified	Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Correlation Coefficient
2018	Špetlík et al., 2018	Visual Heart Rate Estima-tion with CNN	Convolutional Neural Network	204 fitness-themed videos of 20 to 53-year-old subjects	HR accuracy (MAE) Mean Absolute Error
2018	Chen & McDuff, 2018	Video-based physiological measurements (HR, blood volume pulse, breathing rate)	Convolutional Attention Networks	3 datasets of RGB videos and 1 dataset of IR videos	Measurement accuracy, Signal to Noise Ratio, Correlation Coefficient
2018	Casalino et al.,2019	Monitoring cardiovascular risk in real-time without contact	Independent Component Analysis (ICA), Fast Fourier Transform (FFT)	Self-Recorded Dataset of 116 patients	Heart Rate, Breath Rate, Accuracy, Positive Predictive Value, True Positive Rate
2019	Yu, Li, et al., 2019	HR measurement from facial videos	Spatio-Temporal Networks	OBF and MAHNOB-HCI	Accuracy, Pearson Correlation Coefficient
2019	Yu, Peng, et al., 2019	HR measurement from com-pressed facial videos	Deep learning and video enhancement	2 video datasets	Accuracy, Pearson Correlation Coefficient
2019	Hu et al., 2019	Remote HR measurement	Spatio-Temporal Convolutional Neural Networks	PURE and COHFACE datasets	Accuracy, Pearson Correlation Coefficient

Table 1 *(continued)*

Year	Citation	Objective	Technique	Dataset	Parameter
2019	Qiu et al., 2022	Heart rate measurement from face videos	Independent (ICA) Component Analysis	112 videos recorded with 201600 frames	Mean absolute deviation (MAD), root mean square error (RMSE) and the Pearson correlation coefficient
2020	Yu et al., 2020	HR measurement with neural searching	End-to-end neural searching	VIPL-HR datasets and MAHNOB-HCI	Accuracy, Pearson Correlation Coefficient
2021	Hu et al., 2021	Remote HR measurement	Time-domain Attention Network	UBFC rPPG Dataset, COHFACE Dataset, Pulse Rate Detection Dataset, and MMSE-HR dataset	Accuracy, Pearson Correlation Coefficient
2022	Hu et al., 2022	HR measurement from facial videos	Spatio-Temporal Networks	Pulse Rate Detection Dataset and COHFACE dataset	Accuracy, Pearson Correlation Coefficient
2022	Cotes et al., 2022	Multimodal Assessment using EEG, HR technology, etc	Machine Learning Model	Self-Recorded Dataset	Pearson Correlation Coefficient
2023	Su et al., 2023	Measure Heart Rate Variability	Independent Component Analysis (ICA) and Particle Swarm Optimisation (PSO)	Self-Recorded Data from 20 subjects	RMSE and mean absolute error percentage (MAPE), Correlation Coefficient
2023	Sun et al., 2023	Remote HR measurement	Deep learning Model for a vision transformer-based network	Mixed dataset	Computational Complexity and Signal Fitting
2023	Li et al., 2023	HR measurement from video clips	Multi-hierarchical Convolutional Network	UBFC-RPPG and COHFACE datasets	Accuracy, SNR, MAE

Table 2
Datasets for heart rate measurement

Name of the Dataset	Number of Subjects	Number of Videos	Description
PURE (Stricker et al., 2014)	10	60	Recorded for different activities
OBF (Li et al., 2018)	106	2120	Facial videos recorded in reference to physiological signals
MAHNOB-HCI (Soleymani et al., 2012)	27	527	Short videos with human actions
DEAP (Koelstra et al., 2012)	32	120	Frontal face recordings of human affective states
MMSE-HR (Qiu et al.,2022)	40	102	Captured under various facial expressions
VIPL-HR (Niu et al., 2019)	107	3130	Collected using smartphones to support remote heart rate estimation
UBFC-RPPG (Bobbia et al., 2017)	Typically small	10-20	Collected conditions under variable
EmotiW (Dhall et al., 2020)	30-50	Variable	Videos of different facial expressions for heart rate estimation

Further, the included papers showcase the promising possibility for non-invasive and real-time monitoring and demonstrate the ongoing developments in contactless vital signs measurement, particularly in remote heart rate estimation. Recent studies have explored deep learning-based approaches such as CNN and rPPG, while these methods give high accuracy but require a large dataset. For further study in this area, difficulties, including computational complexity and susceptibility to external influences, must still be considered. Datasets should be based on their specific goals and requirements.

Techniques and Methods

The review reflected that ICA is not a mainstream method chosen for heart rate or variability measurement. It offers versatility and shows its greater implications when dealing with some specific contexts, such as multi-sourced data. As such, it is majorly integrated with machine learning techniques for validation to assure utilizing critical aspects of independent component analysis in the context of heart rate measurement (Gao et al., 2024). The most popular ICA methods used in this context are spatial, temporal, frequency domain, multichannel and physiological priors. A comparison among these is given in Table 3 to have a clear idea of possible limitations and suitability before their implication.

The method chosen is determined by the specific application and constraints. The first approach may be adequate for scenarios with low motion; the second or third methods may be more appropriate for more complex scenarios, particularly those with motion. In the end,

the accuracy of the heart rate measurement will be determined by parameters such as data quality, camera resolution, and illumination conditions, in addition to the ICA method used.

Table 3
Popular ICA methods

ICA Method	Scalability	Suitable Sample Size	Accuracy	Advantages	Limitations
Spatial ICA	Yes	Large	Good to Excellent	Proficient in handling spatial variations	A very careful filter selection is required
Temporal ICA	Yes	Medium	Moderate to Good	Simple and computationally efficient	Highly vulnerable to motion artefacts
Frequency Domain ICA	Yes	Medium	Good to Excellent	Effective in frequency domain	Sensitive to artifacts and noise
Multi-channel ICA	Yes	Large	Excellent	Efficient in utilising data from multiple cameras	Calibration and synchronisation is required among cameras
ICA with Physiological Priors	Yes	Large	Excellent	Uses preexisting knowledge of physiological signals	May not be adapt the variations

Training and Testing Mechanisms

Independent component analysis is one of the most popular statistical signal processing techniques that has been associated with most of the popular heart rate measurement methods, such as photoplethysmography and electrocardiography, to analyse physiological sensor data.

This involves temporal dynamics of the extracted features of heart rate-related components; however, it faces many challenges with complex and noisy data. Hence, combined with machine learning techniques. The training and classification are done based on the training engine and classification engine. Artificial Intelligence is a very fascinating concept that is strongly associated with the concepts of data mining and rule mining. Machine Learning, A branch of AI, is divided into two stages: training and testing. The first phase suggests a learning process based on several well-known dataset properties. Phase two uses the knowledge picked up in the first to develop predictions about unidentified features. "Training and testing" are also referred to as "learning and prediction" from this perspective. The goal of ML is to develop a model that can be used to predict future events using a learning algorithm. As a result, this practice is known as forecasting. These terms are widely used in present-day research to improve the standards of work as well as an individual's day-to-day life. In rule mining, the learning concept of various machine learning algorithms mentioned in Table 4 is used to analyse the correlation among the frequently occurring patterns or the categorical data from the databases or repositories.

Table 4
Comparative analysis of popular machine learning techniques

Technique	Category	Type	Areas of Application	Advantages	Limitations
Linear Regression	Supervised	Regression	Predicting numeric values	Simple, efficient for linear data	Limited to linear relationships
Logistic Regression	Supervised	Classification	Binary classification	Simple and provides probability estimates	Limited to classification binary
Decision Trees	Supervised	Classification /Regression	Classification, Regression	Can handle non-linear relationships	Prone to overfitting, unstable for small changes
Random Forest	Supervised	Ensemble	Classification, Regression	Reduces over-fitting, handles complex relationships	Less interpretable, slower training
Support Vector Machines (SVM)	Supervised	Classification /Regression	Classification, Regression	Effective For high- dimensional data, versatile kernel functions	Very sensitive to hyperparameters, slow training
K-Nearest Neighbours (KNN)	Supervised /Un-supervised	Classification, Regression, Clustering	Classification, Regression, Clustering	Simple, no training required, works well with small datasets	Computationally expensive for large datasets
Naive Bayes	Supervised	Classification	Text classification, spam detection	Fast and works well With high-dimensional data	Assumes independent features, not suitable for complex relationships
Neural Networks	Supervised	Deep Learning	Image recognition, NLP, complex tasks	Versatile and can learn complex patterns.	Complex architecture, require large datasets
K-Means Clustering	Unsupervised	Clustering	Data segmentation, customer segmentation	Simple and efficient with large datasets	Very sensitive to initial cluster centroids
Principal Component Analysis (PCA)	Unsupervised	Dimensionality Reduction	Feature selection, noise reduction	Reduces dimensionality, retains most variance	Assumes linear relationships, loss of interpretability

To assess heart rate using video data, machine learning approaches have become increasingly popular. They solve several major problems in this field. The following are some of the principal difficulties that machine learning methods can aid in overcoming:

- **Skin tone and lighting conditions:** A person's skin colour and brightness can be influenced by a variety of factors, including skin tone, ambient lighting, and camera quality. By normalising for skin tone and correcting for lighting conditions, machine learning algorithms may be trained to adapt to these differences and retrieve reliable heart rate information.
- **Face Tracking:** To detect heart rate accurately, continuous detection of the face and tracking are required. Machine learning algorithms can accurately detect and track even when the person is moving or only partially visible.
- **Calibration:** To improve the precision of heart rate measurement, machine learning models can be adjusted and fine-tuned. This process may entail collecting ground truth data to adjust and validate the system.
- **Interference and Noise:** Videos may contain multiple sources of noise and interference, such as background movements, music, illumination changes or any surrounding environmental conditions. Machine learning models can be trained to exclude irrelevant signals and keep meaningful components, which are required for heart rate extraction.
- **Motion Artefacts:** Motion artefacts occur when the camera or the person moves during recording, which can distort facial features used to estimate heart rate. Machine learning models are trained to remove these artefacts, which helps to increase the overall accuracy of heart rate readings.

Hence, by using machine learning algorithms, several problems can be resolved with heart rate measurement from video data. One of the biggest issues is the need for exact and non-invasive measurements. Machine learning algorithms can determine heart rate by detecting minute differences in blood flow, motion, and facial colour without the use of intrusive instruments. Another issue is robustness to multiple circumstances, such as different skin tones, lighting, and facial expressions. Machine learning models can be trained on a variety of datasets to boost generalisation and account for these differences

Estimation Matrices

When it comes to machine learning, the computation of the error metrics is involved in the analysis. The predictions and the estimations performed using machine learning techniques are analysed using correlation metrics such as standard error, root mean square error, and standard deviation. Statistical Machine Learning (S-ML) refers to calculating the parametric values using stats viz. numeral values. Mean Squared Error (MSE) and

Standard Error (SE) are perfect examples of S-ML architecture. Sometimes additional validation is also performed using some similarity indices, for instance, cosine similarity, soft cosine, Jaccard, etc.

Standard Error (SE)

This statistical term is used to quantify the accuracy with the help of a sample distribution to represent a population using the standard deviation. It is used to refer to the standard deviation of different sample statistics, like the mean or median. To compute the value of standard error, Equation 1 is used.

$$SE = \sigma/\sqrt{n} \quad [1]$$

Where SE is defined as the standard error of the sample, n is defined as the number of samples, and σ defined as the sample standard deviation.

Root Mean Square Error (RMSE)

It is the standard deviation of the predicted errors found when estimating effort. The distance from the distribution to the regression line is calculated. Equation 2 is used to compare the experimental results to the project's estimated and desired output.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (E_{predicted} - E_{actual})^2} \quad [2]$$

Mean Squared Error (MSE)

It is the measure of the quality of the estimator used in the prediction analysis. It is mathematically represented as the mean value of the squared difference between the predicted and the observed values of the parameter under study. In this prediction, the lower the value, the better the predictions reflect the better quality of the estimator. To compute the MSE, Equation 3 is used.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Observed_{value} - Predicted_{value})^2 \quad [3]$$

Standard Deviation (SD)

The degree of dispersion seen in a set of values is indicated by this parameter. It indicates that the calculated values are near the predicted values when they are low. It is calculated using the following Equation 4.

$$SD = \sqrt{\frac{\sum E_{value} - Mean_{value}}{Proj_{total}}} \quad [4]$$

Where E_{value} is the computed value and $Mean_{value}$ is the mean of the computed values over the number of project files $Proj_{total}$. To compare the effectiveness of the classifiers, a quantitative parameter analysis is done.

Precision

It is the metric that is used to measure or assess the correctness of the designed methodology to reflect positive predictions made during testing. The formula for precision calculation is given in Equation 5.

$$Precision = True_{positive} / (True_{positive} + False_{positive}) \quad [5]$$

Recall

It's a metric that quantifies the completeness of positive predictions made during the testing phase. It is computed by dividing the number of true positives by the sum of true positives and false negatives. Equation 6 is mentioned below to find the Recall value.

$$Recall = True_{positive} / (True_{positive} + False_{negative}) \quad [6]$$

F-measure

It is the statistic that creates a single score by combining recall and precision. It is computed by taking the harmonic mean of precision and recall. Equation 7 represents the formula to compute F-measure.

$$F - measure = 2 * (Precision * Recall)/(Precision + Recall) \quad [7]$$

Accuracy

It is a number that assesses the test's overall accuracy. It is computed by dividing the total number of cases by the sum of true positives and true negatives, as mentioned in Equation 8.

$$(TP + TN)/(TP + FP + TN + FN) \quad [8]$$

These are the popular metrics used for assessing the effectiveness of the designed classification approaches for heart rate measurement. The outcomes are usually tested in reference to the ground truth and expressed in terms of these metrics.

RESULT AND DISCUSSIONS

Although deep learning methods such as CNNs and Transformers have demonstrated remarkable accuracy in rPPG-based heart rate estimation, they often require large, annotated datasets and significant processing resources. In contrast, this study focuses on evaluating lightweight machine learning models that are more suitable for real-time deployment on edge devices with limited processing capability. This trade-off between processing performance and accuracy is critical for practical healthcare applications.

There are numerous algorithms in the large field of machine learning, each of which is created to tackle a particular class of issues. The effectiveness and performance of a machine learning model can be significantly impacted by the algorithm of choice. Therefore, it is crucial to conduct a comparative analysis before selecting an algorithm.

To evaluate the effectiveness of various machine learning techniques, the analysis is based on precision, recall and F-measure based on the training and classification data. The distribution is 70-30, where 70% of the data is applied for training, and the remaining 30% of the data is used for classification. In total, there are 2000 frames for the comparison with different classifiers. For the classification process, multiple machine learning algorithms have been applied, namely the Feedforward backpropagation algorithm as ANN, K-nearest neighbour (KNN), Naïve Bayes and Multiclass Support Vector Machine (SVM). Table 5 illustrates the comparative analysis based on Precision, Recall and F-measure.

Table 5
Average precision and recall analysis

'Number of Samples'	'Precision ANN'	'Precision Naive Bayes'	'Precision KNN'	'Recall ANN'	'Recall Naive Bayes'	'Recall KNN'
500	0.96476531	0.95774648	0.96052632	0.98666667	0.90666667	0.97333333
700	0.96286655	0.95959596	0.95918367	0.98058252	0.9223301	0.91262136
1000	0.95630247	0.94308943	0.95488722	0.99230769	0.89922481	0.97692308
1200	0.96134457	0.95205479	0.95031056	0.98734177	0.87974684	0.98076923
1400	0.9626129	0.95811518	0.95744681	0.98924731	0.98387097	0.97297297
1600	0.96621961	0.95145631	0.95049505	0.99065421	0.92018779	0.90140845
1800	0.96763848	0.95652174	0.95951417	0.98760331	0.91286307	0.97933884
2000	0.95845314	0.95112782	0.95255474	0.99256506	0.94402985	0.97752809

The given parameters represent the performance metrics of three classification algorithms: Artificial Neural Network (ANN), Naive Bayes, and K-Nearest Neighbour (KNN) on a test set of 2000 samples. These metrics are Precision, Recall, F-measure, and Accuracy. In this response, these metrics are discussed and compared to the performance of the three algorithms.

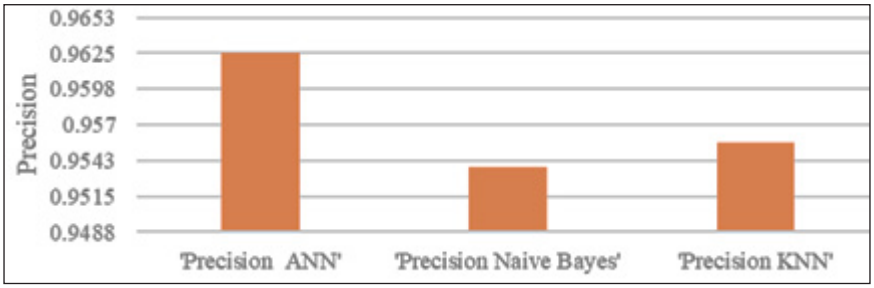


Figure 3. Average precision analysis

Figure 3 shows the comparative analysis of average precision attained by different classifiers.

Precision

Precision measures the fraction of true positives (properly detected instances) out of the total instances projected as positive. A higher precision value of an algorithm indicates that it is more accurate in identifying positive cases. The average precision of ANN is 0.9625, which is slightly higher than that of Naive Bayes (0.9537) and KNN (0.9556). This shows that ANN performs better than Naive Bayes and KNN in identifying positive instances.

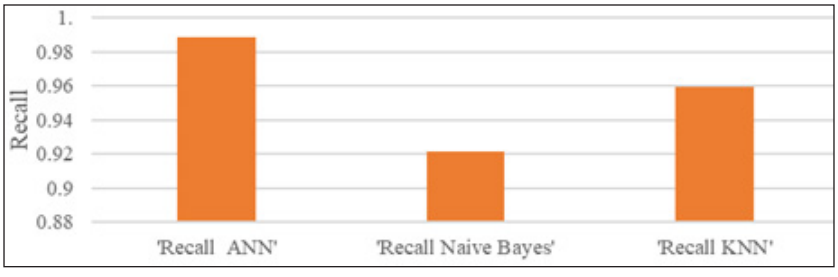


Figure 4. Average recall analysis

Figure 4 shows the average recall values for the classifiers, which indicate their ability to correctly identify positive instances.

Recall

Recall refers to the completeness of positive predictions made during the testing phase. A higher recall implies that the algorithm is more effective in detecting positive instances. The average recall of ANN is 0.9884, which is higher than that of Naive Bayes (0.9211) and KNN (0.9594). This demonstrates that ANN gives better efficiency than Naive Bayes and KNN in identifying positive instances.

F-measure

F-measure creates a single score metric by combining recall and precision, as shown in Table 6. The average F-measure of ANN is 0.9753, which is higher than Naive Bayes (0.9369) and KNN (0.9572). This indicates that ANN gives better output for balanced precision and recall than Naive Bayes and KNN.

Table 6
F-measure and accuracy

'Number of Samples'	'F-measure ANN'	'F-measure Naive Bayes'	'F-measure KNN'	'Accuracy ANN'	'Accuracy Naive Bayes'	'Accuracy KNN'
500	0.97559309	0.93150685	0.96688742	92.5	85	91.25
700	0.97164379	0.94059406	0.93532338	91.8181818	86.3636364	85.4545455
1000	0.97397244	0.92063492	0.96577947	92.1428571	82.8571429	90.7142857
1200	0.97416976	0.91447368	0.96529968	91.7647059	81.7647059	90
1400	0.97574839	0.97082228	0.96514745	92	91.5	90
1600	0.97828435	0.93556086	0.9253012	92.173913	85.2173913	83.4782609
1800	0.97751896	0.93418259	0.96932515	91.9230769	84.6153846	91.1538462
2000	0.97521089	0.94756554	0.96487985	92.0689655	87.2413793	90

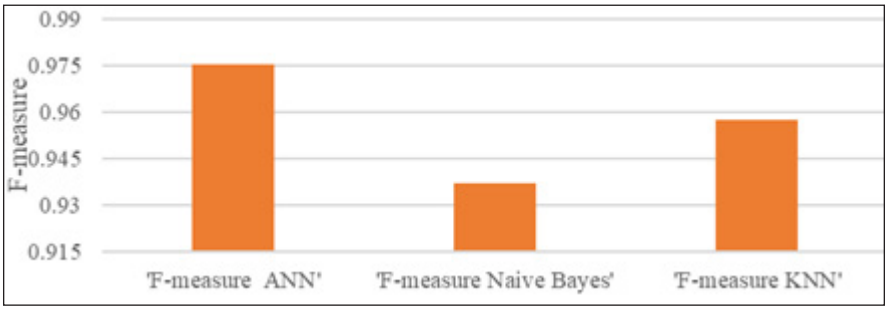


Figure 5. Average F-measure analysis

Figure 5 presents the average F-measure analysis, which indicates the balance between precision and recall.

Accuracy

Accuracy is a number that assesses the test's overall accuracy. It is one of the most used metrics for evaluating classification algorithms. The average accuracy of ANN is 92.049%, which is higher than Naive Bayes (85.570%) and KNN (89.006%). The high accuracy of ANN indicates it provides more accurate classification than Naive Bayes and KNN.

Figure 6 shows the overall performance comparison of the classifiers based on accuracy and other evaluation metrics.

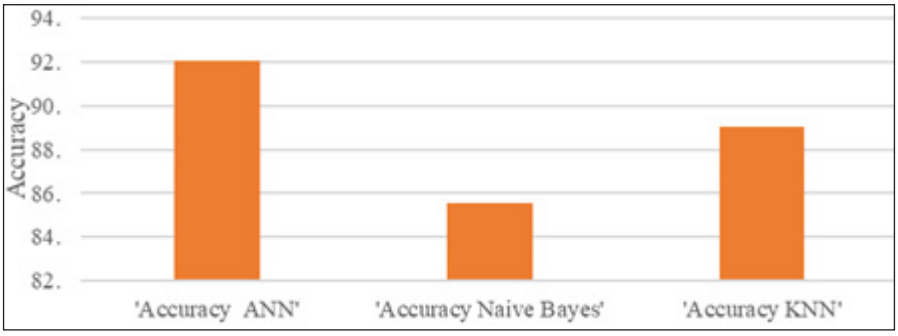


Figure 6. Average performance analysis

Overall, the performance metrics suggest that ANN performs better than Naive Bayes and KNN across precision, recall, F-measure, and accuracy. The percentage improvement of ANN over Naive Bayes and KNN can be evaluated as given below:

- Precision improvement over Naïve Bayes = (0.9625 - 0.9537) / 0.9537 * 100% = 0.92% and over KNN = (0.9625 - 0.95561) / 0.95561* 100% = 0.73%
- Recall improvement over Naïve Bayes = (0.9884 - 0.9211) / 0.9211 * 100% = 7.31% and over KNN is = (0.9884 - 0.95932) / 0.95932 * 100% = 3.02%
- F-measure improvement over Naïve Bayes = (0.9753 - 0.9369) / 0.9369 * 100% = 4.10% and over KNN is = (0.9753 - 0.95724) / 0.95724 * 100% = 1.88%
- Accuracy improvement over Naïve Bayes = (92.049 - 85.570) / 85.570 * 100% = 7.57% and over KNN is = (92.049 – 89.569) / 89.569 * 100% = 3.41%

The above-mentioned calculations indicate that ANN offers a slight improvement in precision, a substantial improvement in recall, a moderate improvement in F-measure, and a major improvement in accuracy as compared to both Naive Bayes and KNN.

In short, the performance metrics precision, recall, F-measure, and accuracy clearly indicate that the performance of ANN is better than Naive Bayes and KNN. However, the improvement varies across different metrics of ANN over Naive Bayes and KNN, but overall, the enhancement provided by ANN is important. These results suggest that ANN is a better choice for classification tasks than Naive Bayes and KNN, at least in the context of the given test set. The goal of this procedure is to identify the algorithm that best satisfies the requirements of the given problem by comparing numerous algorithms based on a variety of criteria. The importance of comparative analysis in machine learning, the essential processes in conducting one, and the effects it has on model performance and decision-making.

Ablation Analysis of Key Components

A conceptual ablation analysis is provided to highlight the importance of various heart rate estimation components based on findings stated in the literature.

The major factors include video magnification, machine learning model classification, Independent Component Analysis (ICA) signal extraction, and Region of Interest (ROI) selection. Previous research implies that the absence of ICA leads to reduced signal quality due to increased noise and interference. Also, the absence of video magnification reduces estimation accuracy, which makes it more difficult to recognise minute physiological indicators. Inappropriate ROI selection also introduces artefacts caused by movement, occlusion, or changes in light. Each component contributes significantly to the overall performance of the system. So, this analysis shows that the section of ROI video magnification feature extraction and selection plays a crucial role in enhancing heart rate estimation accuracy and robustness.

CONCLUSION

The paper presents an analytical study of the methods and techniques used for heart rate measurement. This paper reviews commonly used datasets and parameters that are mostly used for the assessment of heart rate measurement work. A comparative performance analysis is provided of different machine learning classifiers based on the quantitative parameter. This analysis indicates that ANN performs better than the other state-of-the-art classifiers. The average precision of ANN is 0.9625, which is slightly higher than Naive Bayes (0.9537) and KNN (0.9556). This suggests that ANN is more accurate in identifying positive instances than Naive Bayes and KNN. The average recall of ANN is 0.9884, which is also higher than Naive Bayes (0.9211) and KNN (0.9594). This shows that ANN is better at identifying positive instances than Naive Bayes and KNN. The average F-measure of ANN is 0.9753, which is higher than Naive Bayes (0.9369) and KNN (0.9572). From this result, it is understood that ANN provides a better balance between precision and recall than Naive Bayes and KNN. The average accuracy generated by ANN is also 92.049%, higher than that of Naive Bayes (85.570%) and KNN (89.006%). These results indicate that ANN provides the most consistent performance. However, these findings are based on the comparative analysis of existing studies, and results may vary depending on the dataset and experimental conditions.

Future research will focus on improving the reliability and real-world applicability of non-invasive heart rate estimation techniques. This involves integrating multimodal datasets that combine facial video with physiological and demographic parameters, as well as using advanced feature extraction techniques for enhanced resilience to motion and illumination variations (Liu et al., 2024), according to recent research (Buyung et al., 2024). To increase accuracy and generalisation, hybrid methods that combine deep learning

models with conventional signal processing may also be investigated. Further validation using large public datasets and real-time deployment on mobile and edge devices will be considered to facilitate practical healthcare applications.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

ICA	:	Independent component analysis
EVM	:	Eulerian video magnification
HR	:	Heart rate
BP	:	Blood pressure
HRV	:	Heart rate variability
rPPG	:	Remote photoplethysmography
ROI	:	Region of Interest
RGB	:	Red green blue
CNN	:	Convolutional neural network
DNN	:	Deep neural network
PSO	:	Particle swarm optimisation
FA	:	Firefly algorithm
SIFT	:	Scale-Invariant feature transform
SURF	:	Speeded-Up robust features
MSER	:	Maximally stable extremal regions
PCA	:	Principal Component Analysis
RMSE	:	Root mean square error

REFERENCES

Acharya, B., Saakyan, W., Hammer, B., & Drimalla, H. (2025). The reliability of remote photoplethysmography under low illumination and elevated heart rates. *npj Digital Medicine, 8*, Article 744. https://doi.org/10.1038/s41746-025-02192-y

Alghoul, K., Alharthi, S., Al Osman, H., & El Saddik, A. (2016). Measurement of heart rate from video sequences based on motion tracking and morphological analysis. In *Proceedings of the IEEE Workshop on Machine Learning for Signal Processing* (pp. 1-6). IEEE.

Antink, P., Yu, X., Neu, W., & Leonhardt, S. (2019). A multi-modal sensor for a bed-integrated unobtrusive vital signs sensing array. *IEEE Transactions on Biomedical Circuits and Systems, 13*(3), 529-539. https://doi.org/10.1109/TBCAS.2019.2911199

Buyung, R. A., Bustamam, A., & Ramazhan, M. R. S. (2024). Integrating remote photoplethysmography and machine learning on a multimodal dataset for non-invasive heart rate monitoring. *Sensors, 24*(23), Article 7537. https://doi.org/10.3390/s24237537

Casalino, G., Castellano, G., Pasquadibisceglie, V., & Zaza, G. (2019). Contactless real-time monitoring of cardiovascular risk using video imaging and fuzzy inference rules. *Information, 10*(1), Article 9. https://doi.org/10.3390/info10010009

Chen, W., & McDuff, D. (2018). DeepPhys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV 2018)* (pp. 356-373). Springer. https://doi.org/10.1007/978-3-030-01216-8_22

Cotes, R. O., Boazak, M., Griner, E., Jiang, Z., Kim, B., Bremer, W., & Clifford, G. D. (2022). Multimodal assessment of schizophrenia and depression utilising video, acoustic, locomotor, electroencephalographic, and heart rate technology: Protocol for an observational study. *JMIR Research Protocols, 11*(7), Article e36417. https://doi.org/10.2196/36417

Debnath, U., & Kim, S. (2025). A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning. *BioMedical Engineering OnLine, 24*, Article 73. https://doi.org/10.1186/s12938-025-01405-5

Dhall, A., Sharma, G., Goecke, R., & Gedeon, T. (2020). EmotiW 2020: Driver gaze, group emotion, student engagement and physiological signal-based challenges. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 784-789). Association for Computing Machinery. https://doi.org/10.1145/3382507.3418909

Diller, T., Kelly, J. W., Blackhurst, D., Steed, C., Boeker, S., & McElveen, D. C. (2014). Estimation of hand hygiene opportunities on an adult medical ward using 24-hour camera surveillance: Validation of the HOW2 Benchmark Study. *American Journal of Infection Control, 42*(6), 602-607. https://doi.org/10.1016/j.ajic.2014.02.020

Estepp, J. R., Blackford, E. B., & Meier, C. M. (2014). Recovering pulse rate during motion artefact with a multi-imager array for non-contact imaging photoplethysmography. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1462-1469). IEEE. https://doi.org/10.1109/SMC.2014.6974122

Gao, H., Zhang, C., Pei, S., & Wu, X. (2024). IDTL-rPPG: Remote heart rate estimation using instance-based deep transfer learning. *Biomedical Signal Processing and Control, 95*, Article 106416. https://doi.org/10.1016/j.bspc.2024.106416

Hu, M., Dong, G., Wang, X., Ge, P., & Chu, Q. (2019). A novel spatial-temporal convolutional neural network for remote photoplethysmography. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-6). IEEE. https://doi.org/10.1109/CISP-BMEI48845.2019.8966034

Hu, M., Qian, F., Guo, D., Wang, X., He, L., & Ren, F. (2021). ETA-rPPGNet: Effective time-domain attention network for remote heart rate measurement. *IEEE Transactions on Instrumentation and Measurement, 70*, Article 2506212. https://doi.org/10.1109/TIM.2021.3058983

Hu, M., Qian, F., Wang, X., He, L., Guo, D., & Ren, F. (2022). Robust heart rate estimation with spatial-temporal attention network from facial videos. *IEEE Transactions on Cognitive and Developmental Systems, 14*(2), 639-647. https://doi.org/10.1109/TCDS.2021.3062370

Hwang, G., & Lee, S. J. (2026). Phase-shifted remote photoplethysmography for estimating heart rate and blood pressure from facial video. *Measurement, 275*, Article 121240. https://doi.org/10.1016/j.measurement.2026.121240

Krajnak, K. M. (2014). Potential contribution of work-related psychosocial stress to the development of cardiovascular disease and type II diabetes: A brief review. *Environmental Health Insights, 8*(Suppl. 1), 41-45. https://doi.org/10.4137/EHI.S15263

Li, B., Zhang, P., Peng, J., & Fu, H. (2023). Non-contact PPG signal and heart rate estimation with multi-hierarchical convolutional network. *Pattern Recognition, 139*, Article 109421. https://doi.org/10.1016/j.patcog.2023.109421

Liu, Y., Xu, C., Qi, L., & Li, Y. (2024). A robust non-contact heart rate estimation from facial video based on a non-parametric signal extraction model. *Biomedical Signal Processing and Control, 93*, Article 106186. https://doi.org/10.1016/j.bspc.2024.106186

McDuff, D., Gontarek, S., & Picard, R. W. (2015). Remote measurement of cognitive stress via heart rate variability. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp. 1-6). IEEE. https://doi.org/10.1109/FG.2015.7284874

Qiu, Y., Liu, Y., Arteaga-Falconi, J., Dong, H., & El Saddik, A. (2022). *EVM-CNN: Real-time contactless heart rate estimation from facial video*. *arXiv*. https://doi.org/10.48550/arXiv.2212.13843

Rizvi, S. R., & Rahnamayan, S. (2018). Interactive evolutionary parameter optimisation for Eulerian video magnification. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 10-16). IEEE.

Singh, S. P., Rathee, N., Gupta, H., Zamboni, P., & Singh, A. V. (2017). Contactless and hassle-free real-time heart rate measurement with facial video. *Journal of Cardiac Critical Care TSS, 1*(1), 24-29. https://doi.org/10.1055/s-0037-1604202

Špetlík, R., Franc, V., Čech, J., & Matas, J. (2018). Visual heart rate estimation with convolutional neural network. In *Proceedings of the British Machine Vision Conference 2018* (Article 84, pp. 1-12). BMVA Press. https://doi.org/10.5244/C.32.84

Stricker, R., Müller, S., & Gross, H.-M. (2014). Non-contact video-based pulse rate measurement on a mobile service robot. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014)* (pp. 1056-1062). IEEE. https://doi.org/10.1109/ROMAN.2014.6926392

Su, T.-J., Hung, Y.-C., Pan, T.-S., Lin, W.-H., Wang, S.-M., & Lee, Y.-C. (2023). Estimation of heart rate and heart rate variability with real-time images based on independent component analysis and particle swarm optimisation. *Applied Sciences, 13*(13), Article 7605. https://doi.org/10.3390/app13137605

Sun, W., Sun, Q., Sun, H.-M., & Jia, R. (2023). ViT-rPPG: A vision transformer-based network for remote heart rate estimation. *Journal of Electronic Imaging, 32*(2), Article 023024. https://doi.org/10.1117/1.JEI.32.2.023024

Virtanen, M., Heikkilä, K., Jokela, M., Ferrie, J. E., Batty, G. D., Vahtera, J., & Kivimäki, M. (2012). Long working hours and coronary heart disease: A systematic review and meta-analysis. *American Journal of Epidemiology, 176*(7), 586-596. https://doi.org/10.1093/aje/kws139

World Health Organisation. (2025, July 31). *Cardiovascular diseases (CVDs)*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Yu, Z., Li, X., & Zhao, G. (2019). Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proceedings of the 30th British Machine Vision Conference 2019 (BMVC 2019)* (Article 277). BMVA Press.

Yu, Z., Li, X., Niu, X., Shi, J., & Zhao, G. (2020). AutoHR: A strong end-to-end baseline for remote heart rate measurement with neural searching. *IEEE Signal Processing Letters, 27*, 1245-1249. https://doi.org/10.1109/LSP.2020.3007086

Yu, Z., Peng, W., Li, X., Hong, X., & Zhao, G. (2019). Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 151-160). IEEE. https://doi.org/10.1109/ICCV.2019.00024

Zarezadeh, E., & Rezaeian, M. (2016). Micro expression recognition using Eulerian video magnification method. *BRAIN: Broad Research in Artificial Intelligence and Neuroscience, 7*(3), 43-54.

Pertanika J. Sci. & Technol. 34 (S1): 69 - 87 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Congestion Control Using Link Estimation and Energy-Awareness in Wireless Sensor Networks

Suma S[1]*, Voruganti Naresh Kumar[2], K Srujan Raju[2], Mahesh Kotha[3], B. Prashanth[3], Suraya Mubeen[4], and Harshavardhan Nerella[5]

[1]*Department of Computer Science and Engineering, Faculty of Engineering and Technology (Exclusively for Women) (FETW), Sharnbasva University, Kalaburagi, 585103 Karnataka, India*
[2]*Department of Computer Science and Engineering, CMR Technical Campus, Hyderabad, 501401 Telangana, India*
[3]*Department of CSE (Artificial Intelligence & Machine Learning), CMR Technical Campus, Hyderabad, 501401 Telangana, India*
[4]*Department of CSE (Data Science), CMR Technical Campus, Hyderabad, 501401 Telangana, India*
[5]*Sr. Cloud Engineer, Austin, 78660 TX, USA*

ABSTRACT

A WSN is a mobile node network that generates and configures its own data and transmits it wirelessly to the entire network. The topology of a network is constantly changing because nodes are mobile. Due to the instability of the link, dynamic topologies result in link failures, decreased bandwidth, and a decrease in QoS. Ensure that the wireless link is stable, the bandwidth is available, and the nodes have sufficient energy to communicate reliably. In WSN routing, packet loss increases delay and energy consumption for retransmission since multipath routing considers optimal hop counts. The paper proposes a Link Estimation Indicator and Energy-Awareness Routing Scheme (LEI-EA) for WSNs to address these issues. A link estimation indicator (LEI) is used to determine whether a link is stable or unstable. A LEI-EA scheme utilises the residual energy of neighbour nodes to determine multipath routing and maintain link stability, reliability, and network longevity. As part of the simulation, extensive testing has been conducted using the event-driven simulation tool (NS2), and the proposed LEI-EA scheme has been compared to existing dynamic routing schemes.

Keywords: Disjoint Links, link indicator (LEI-EA), link quality, WSN

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 13 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.04

E-mail addresses:
sn.suma05@gmail.com (Suma S)
nareshkumar99890@gmail.com (Voruganti Naresh Kumar)
ksrujanraju@gmail.com (K Srujan Raju)
kotha.mahesh528@gmail.com (Mahesh Kotha)
badamshaan@gmail.com (B. Prashanth)
suraya418@gmail.com (Suraya Mubeen)
nerellaharshavardhan@outlook.com (Harshavardhan Nerella)
* Corresponding author

INTRODUCTION

The development of wireless communication and internet services has been boosted by advances (Wang & Li, 2015; Yazıcı et al., 2014).

e-ISSN: 2231-8526

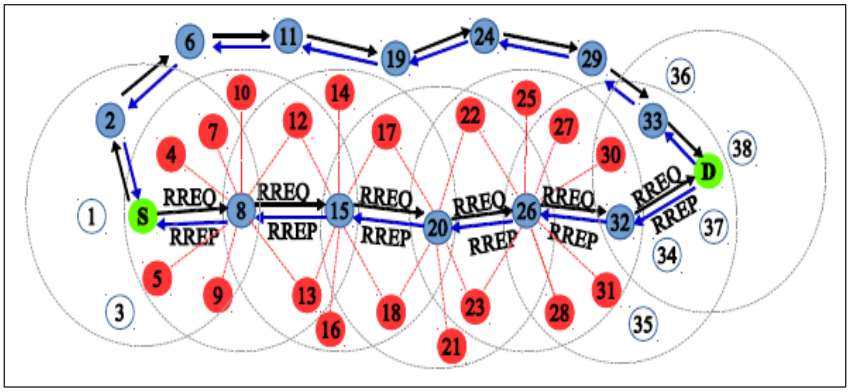


Figure 1. Routing process

Wireless communication technologies such as networks. Digital 5G networks will benefit from WSN 's features and functionalities (Agiwal et al., 2016; Huang et al., 2017). WSNs communicate wirelessly, and they are self-organised, configured, and controlled nodes (Helen et al., 2014; Suma & Harsoor, 2022d). Unlike traditional networks, WSNs have no constraints on how the nodes move or where they leave shows in Figure 1. The type of network topology and the energy resource determine this. Using mobile nodes as routers, packets are forwarded among the collaborating nodes. By incorporating routing functions into the routing protocol, packets are forwarded through single-hop or multi-hop routes. Within the transmission range, the node communicates directly with other nodes. When a node out of communication range cannot forward packets, intermediate nodes are selected. Dynamic links between nodes often break, resulting in packet loss (Suma & Harsoor, 2022b).

Recent studies have focused on fault-tolerant routing. When using fault-tolerant routing in the WMSN, reliability, availability, and consequent independence are typically ensured. Because a sensor node's energy remains low after a certain time, fault-tolerant routing algorithms are required to recover from the path loss. Furthermore, fault-tolerant routing is essential for critical monitoring applications. In contrast, it doesn't implement Request-To-Send or Clear-To-Send handshake mechanisms to solve the Hidden Node Problem (Koubâa et al., 2009). Vertex-disjoint multipath routing allows the Hidden Node Problem at sink nodes caused by its two different 1-hop neighbours (Cao et al., 2013). The HNP can result in severe repercussions for QoS performance. Despite RTS/CTS' capability to reduce hidden terminal problems (Barroca et al., 2014), it also controls heavy traffic, limiting the effective throughput.

In multipath routing, HNP nodes are disjoint at the sink node, and this should be avoided without the RTS/CTS method (Kuipers et al., 2005). In conventional single-layered (CSL) systems, this method allows interaction, but data exchange between non-adjacent layers is not possible. CSL provide quality of service but does not give optimum results.

Whereas CSL by Hamid and Hussain (2014) gives the optimum result for routing protocols. Triangle Interference Geographic Routing problem contributes to the node disjoint interreference for multiple paths by routing cost, which gives link quality, residual energy and distance. Choosing one-hop minimum hop distance and IEEE 802.15.4, the sink node of TIGMR overlooks the sink Hidden Node Problem, while RTC/CTS. The minimum distance is selected based on routing cost by a 1-hop neighbour node. HNP at the sink node in the IEEE 802.15.4 network (Yang et al., 2018). TIGMR ignores the HNP at the sink node without consideration of the hand-off, which provides link quality called triangle metric by Baono et al. (2010) and Baccour et al. (2013). TIGMR provides soft transition information exchange on another alternative path before the path is disconnected because of more usage of energy. Distinguishing TIGMR from the state-of-the-art Two-Phase Geographic (TPGF) by Bennis et al. al (2014) and link quality by Aswale and Ghorpade (2017).

There is a challenge in providing reliable routing due to WSN topologies and link failures (Barroca et al., 2014; Kuipers et al., 2005; Suma & Harsoor, 2022e). An energy-aware Link Estimation Indicator-based Ad hoc on Demand Routing scheme based on Link Estimation Indicator is proposed in this article for WSNs' adaptation of the AODV protocol. A link estimation indicator (LEI) is employed to ascertain the stability of links between nodes if node disjointness may result in a link failure. The objective of an LEI is to guarantee stable connections between nodes and to ascertain the amplitude of the signal. LEI-EA chooses forwarding nodes with high residual energy for dependable data transmission, fair bandwidth consumption, and energy balancing (Mafirabadza et al., 2016). The following are examples of contributions:

- Media access control (MAC) signals influence the quality of the link estimation indicator (LEI) for individual nodes.
- It optimises residual energy to extend network lifespan and enhance data transmission efficiency by guaranteeing the availability of link bandwidth for high-quality communication.

Nodes in a WSN are usually mobile, so they can adapt to topology changes constantly without relying on fixed infrastructure. Coordinating with each other, mobile nodes act as routers and hosts, establishing connections between sources and destinations (Ali et al., 2014). Congestion can be avoided by checking the status of intermediate nodes through route discovery between pairs of nodes. Routing breaks are common in dynamic behaviour, moving aspects, for example, bandwidth, energy resources, mobility, and channel decline (Kalansuriya et al., 2009). Stable links between two nodes are essential for dependable communication. There has been considerable research on the features of low-power radiocommunication links, and these links are error-prone, asymmetric, and unreliable due to the hardware and environment. Since topology changes dynamically, it is difficult to estimate link quality. Routing performance can be increased by estimating link

quality or path stability. The challenge of determining path stability to avoid congestion is to recommend a scheme in which the path stability is determined between contributing nodes while founding a route. Path steadiness is determined by measuring the link quality of joining nodes by a link estimation indicator and the outstanding energy of nodes. Received signal strength from the MAC layer and offered bandwidth ensure high-quality communication and achieve QoS parameters like amplified throughput, higher packet delivery ratio, and low computational overhead. To prolong network lifespan and ensure reliable data transfer, forwarder nodes must be energy-aware.

The rest of this paper is organised as follows. Section II describes the problem statement and motivation. Section III describes the related works. The proposed mechanism scheme is described in Section IV. Section V describes the results and evaluation. Section VI describes the comparison results. A conclusion on the proposed scheme is drawn in Section VII.

Problem Statement and Motivation

Wireless sensor networks (WSNs) lack inherent fixed or centralised infrastructures, as their topologies are constantly evolving (Otero et al., 2011). The mobile node functions as a router and host, coordinating with other mobile nodes and establishing connections between a source and destination (Xu et al., 2018). Routing between two nodes necessitates an assessment of the condition of intermediate nodes before establishing a route. Route breakages in Wireless Sensor Networks (WSN) occur often and impact critical aspects including bandwidth, energy consumption, mobility, and channel fading. The reliability of communication between two nodes is contingent on the stability of the links. Hardware and environmental factors are responsible for the error-prone, asymmetric, and unstable properties of low-power wireless connections (Baccour et al., 2013; Yang et al., 2018). Estimating and maintaining connection quality can be problematic due to the continuous changes in routing topology (Girirajan et al., 2023; Suma & Harsoor, 2022e). Link quality or path stability must be precisely determined for optimal routing performance. Assessment of route stability is difficult. A way to evaluate the association stability between two nodes while establishing a route. The measurement of path stability is conducted by utilising link estimation indicators (LEIs) and the residual energy of nodes. The signal intensity at the MAC layer and the range of available bandwidth influence Quality of Service (QoS) characteristics, for example, increased data transfer rate, improved packet delivery ratios, and reduced computing burden (Suma et al., 2024). Energy-conscious forwarder nodes are chosen to guarantee dependable data transmission and promote network longevity.

Related Works

The LQEAR protocol (Smail et al., 2014; Suma & Harsoor, 2022e). finds discontinuous connections over various pathways, extending network lifespan. The LQEAR evaluation

estimates Link Quality (LQ) and residual energy measures using LQ indicators (Girirajan et al., 2023; Suma & Harsoor, 2022c). Superior one-hop forwarding nodes ensure transmission reliability. This protocol assesses LQ erroneously and provides useless wireless link features for routing. AOMR-LM is an energy-efficient multipath routing protocol that balances node energy usage. Multipaths are chosen using node residual energy. Energy harvesting is ensured by threshold energy and coefficient analysis. Despite being able to extend WSN 's lifetime and improve routing performance without relying on channel fading, this protocol does not depend on link failures to diminish network lifetime. Power-efficient AODV routing scheme (EPA-AODV) (Deepa et al., 2019; Suma et al., 2024). Route request packet headers (RREQ) are added with residual energy fields. RREQ packets are manipulated by intermediate nodes on arrival, and residual energy is checked. Data is sent to a qualifying neighbour if residual energy exceeds the threshold. Dogra et al. (2021) and Singal et al. (2014) introduced dynamic energy-efficient (DE-AODV) to reduce end-to-end latency and enhance network resilience. This protocol selects reliable, energy-efficient nodes using a battery-powered mechanism to dynamically supply external energy. Nodes consume less energy when they are supplied with external battery power whenever a link failure occurs, extending the life of the network. The strength of the signal received determines the link stability of a reliable path (Malwe et al., 2017; Suma & Harsoor, 2020a). An expiry time is calculated for each node in a path along with the expiry times for all other links. A zone-based routing scheme is proposed in Lin et al. (2012) and Marriwala (2022) that considers link availability and reduces the number of broadcast packets generated by RREQ. RSSI and a threshold are used to divide each node's transmission range into inner, middle, and outer zones. The middle zone performs route discovery. Route discovery takes place at nodes in the middle zone. Localisation and angular sectors within the transmission range are used to calculate the neighbouring link availability ratio. In this technique, packets loop, causing routing overhead. It also takes more hops, which increases the number of hops. Predicting connection failures before they happen might enhance service (Gupta & Marriwala, 2017; Yadav et al., 2015). Route availability is determined by the stability of the link. As a means of estimating the stability of the link and predicting the failure of the link, Newton divided difference interpolation is proposed in this scheme. A proactive connection failure between two nodes generates an alternate path.

Shu et al. (2010) describe the MPMP routing protocol as an extension of the TPGF routing protocol. A maximum amount of streaming data is transmitted, and an end-to-end delay guarantee is ensured. Streams of multimedia data are separated into streams of audio and images. Priorities are assigned based on multimedia content context and transmission delay. The path with the lowest delay is assigned to the stream with the highest priority. Despite this, processing overhead increases when multimedia data streams are decomposed and combined.

LIEMRO establishes interference-minimised node-disjoint paths for event-driven sensor networks (Radio et al., 2011). Energy remaining, ETX, and interference level determine the selection of nodes. Infusion of data traffic is achieved using a load-balancing algorithm. A highly competitive channel, however, significantly affects protocol performance. AGEN (Adaptive Greedy-compass Energy-aware Multipath) routing protocol improves network lifetime by implementing load-balancing. Smart greedy forwarding and walking backwards are both used in AGEM. Using the first mode, a 1-hop neighbour is always closer to the sink than the present node. An alternative mode if the 1-hop neighbour is unavailable. In addition to ensuring uniform energy consumption, AGEM also guarantees quality of service. The route divides the entire network topology into many districts and enables data to be transmitted simultaneously without interference (GEAM). The distance between districts is sufficiently large to avoid transmission interference. The methodology does not guarantee QoS, but it provides efficiency when the topology changes.

Unkai et al. al 2014 PASPEED dynamically adjust the transmission power on the reserve with the least distance delay. Thus, PASPEED improves the network efficiency and lifetime due to minimum delay. Along with three others, priorities of the packets with reliability requirements are assigned.

Magaia et al. (2015) WMSN routing considers QoS for multi-objective and evolutionary algorithms using a genetic algorithm by SPEA. ETX and delay give solutions to discover the minimum path. ETX uses parameters for routing decisions with asymmetric links. Wang et al. (2015) designed the PWDGR protocol to explain energy bottlenecks at sinks. When a sink has a 360 ° scope, PWDGR finds destination nodes in a pairwise manner. Therefore, the seriousness of the situation has increased. Efforts are made to minimise the energy burden around the sink. A dense WMSN performs better with PWDGR. While saving energy, it compromises on delay and energy. With the Cross-Layer Channel Utilisation and Delay Aware Routing protocol, evidence can be exchanged between the MAC and network layers to discover improved routing routes (Hamid et al., 2016). Packet Service Time (PST) and remaining energy determine which forwarding node would be used. CUDAR chooses nodes with the lowest PST and the lowest dispute during route discovery while maintaining the shortest possible path. As a result, CUDAR offers improved delay, jitter, and throughput. A Carrier Sense Aware Multipath Geographic Routing Protocol (CSA-MGR) uses a dynamic and distributed mechanism to identify multiple disjoint paths while avoiding any shared carrier sense ranges (Bennis et al., 2016).

Carriers sense ranges their place at the intrusion range or transmission range to deal with adjacent path interference effects during path discovery. By incorporating Number of Common Neighbours (NCN) metrics and zigzag avoidance mechanisms, CSA-MGR ensures additional efficient and faster route discovery. CSA-MGR is faster, more reliable, and more efficient than TPGF.

As a result of adaptive traffic shaping and the Real-Time Video Streaming Protocol (ARTVP), we can achieve optimal multimedia coding complexity and utilise multipath forwarding to provide optimal multimedia coding quality (Ahmed, 2017). The forwarding node is selected based on the RSSI (Received Signal Strength Indicator), the residual energy, and the delay of the signal. Depending on the multimedia content of the packet, packets are classified according to their priority and are directed down the most reliable path based on the priority of the packet. The ARTVP can switch from single-path to multi-path traffic based on the multimedia traffic being transmitted and vice versa. A high-performance protocol like this will ensure the longevity of a network and ensure high performance. Partitioning Multi-constrained Multipath Routing (PMMR) protocol is intended for real-time video flowing in WMSN (Hasan et al., 2017). In this protocol, a planned novel metric prioritises sensor nodes based on the link quality. Depending on the decision constraints, an objective function's path is selected. A mix of integer programming and mixed integer optimisation is used to control the adaptive switching. As a result of the PMMR protocol, network lifetime is increased, and QoS is guaranteed.

Aswale and Ghorpade (2017) demonstrate how the LQEAR protocol improves network lifetime by discovering node-disjoint multiple paths. The forwarding node is chosen according to its Link Quality Indicator (LQI), lingering energy, and distance from its one-hop neighbour to the sink (Marriwala et al., 2024; Venugopal et al., 2025). To provide reliability, it selects links with higher quality according to the LQI metric. However, the LQI metric does not provide accurate estimates of the quality of wireless links and therefore is unsuitable for link-quality-based routing protocols.

Proposed Mechanism

In Figure 2, the process of a blackhole attack on packet drop in a wireless sensor network. Initially, the system deployment of WSN where nodes are configured in network topology is established. The transmission of data packets takes place through a routing protocol. protocol checks for packet drops during the transmission process. If there is no packet drop is detected, then transmission proceeds directly to performance analysis. Performance is analysed by parameters like throughput, delay, and PDR. However, if a packet drop is observed, the event. In such cases, a malicious node initiates the blackhole attack by falsely advertising itself

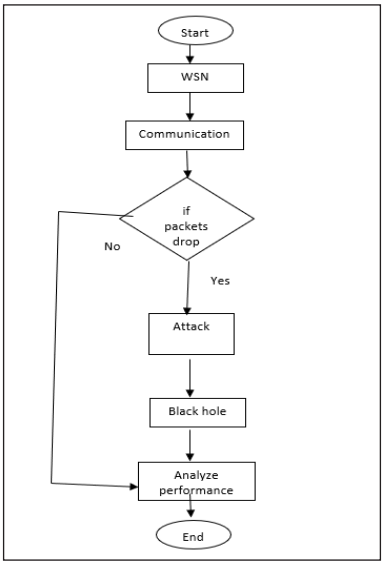


Figure 2. Flow chart on blackhole attack to analyse the performance of packet drop

as the optimal route to the destination. As a result, it attracts data packets and drops them instead of forwarding, causing packet loss. Thus, the network performance degrades. Finally, the system analyses the performance by parameters.

Network Model

Mobile nodes are randomly distributed and have many neighbours within their transmission range. A graph $G_n = (V_n, L_n)$ $V_{n_{speed}} = (distance_{V_n} / time_{V_n})$ mps where V_n is the number of nodes and Ln is their link, representing network communication. The link between source and represented as $(s_n, d_n) \in V_n$. Source node s_n direct interaction with the destination node d_n via a single hop if it is within communication range, or via multi-hop if not. Nodes are dynamic, so topology changes are unpredictable. Dynamic movement prevents two nodes from sharing a stable connection due to their varying speeds (Lin et al., 2012). A reliable network is necessary because node speed is independent of place and time. Slower nodes generate more stable links. Measuring the total distance travelled in that time frame yields a mobile node V_n with variable speed (Yazıcı et al., 2014).

During route discovery, the routing cost function finds the advancing node inside the transmission range r for two nodes, V_{n_a} and V_{n_b} $D_{V_{n_a},V_{n_b}} = \sqrt{(y_i - y_j)^2 + (x_i - x_j)^2}$. The 1-hop neighbour node of V_{n_a} is V_{N_a} and has coordinates (x_i, y_i) and (x_j, y_j) Goldsmith et al. (1997).

To calculate node transmission range r, use the RSSI threshold value and error probability of 10^{-3} 3) as stated as bit error rate $V_{N_a} = \left(V_{n_b} | D_{V_{n_a},V_{n_b}} \leq r\right)$.

The Mobility Model

Real-time nodes incorporate GPS to precisely determine their position, speed, and velocity within the transmission area. The localisation method provides information about the speed, position, and direction of mobile nodes at any given time. Real-time nodes utilise GPS to precisely determine their position, speed, and velocity within the transmission area.

A localisation algorithm gives the mobile node's speed, position, and direction. The random waypoint model (RWP) sometimes assesses network node mobility, connection failure, and stability. In an RWP-configured network area, nodes move randomly with minimum and maximum velocity, as well as pause periods. Nodes move based on their velocities $[v_{n_{max}}, v_{n_{min}}]$ and direction, δ, between [0, 1]. Assuming the node's current position at the time instant t_n is $((x_i(Vt_{n_i}), y_i(Vt_{n_i})))$, the original location after incrementing the time interval Δt is obtained.

The node's direction of signal is signified by $\delta_i(t_n)$ for the ith node at time instant t_n, and its speed is indicated by $v_i(t_n)$.When node velocity is at its highest, the value of $\delta_i(t_n)$ is 0; when it is at its lowest, it is 1 $\begin{cases} x_i(t_n + \Delta t) = x_i(t_n) + \Delta t * v_i(t_n) cos\delta_i(t_n) \\ y_i(t_n + \Delta t) = y_i(t_n) + \Delta t * v_i(t_n) cos\delta_i(t_n) \end{cases}$.

The Link Status and Stability

The Euclidean distance $DV_{n_{ij}} = \sqrt{\left(y_i(t_n) - y_j(t_n)\right)^2 + \left(x_i(t_n) - x_j(t_n)\right)^2}$ between pairs of nodes V_{n_i} and V_{n_j} at time t_n. can be used to determine the connection status of nodes.

This enables the estimation of mobile nodes' RSSI transmission range $r\tau_i$ or $r\tau_j$ for link status $L_{n_{ij}}(t_n) = \begin{cases} 1, & if DV_{n_{ij}}(t_n) \leq r\tau_i \text{ or } r\tau_j \\ 0, & otherwise \end{cases}$.

Hence, the network between the source and destination at any given moment t_n can be expressed as a matrix of $V_N \times V_N$, where the link status components enable connection determination. At time t_n, we represent a connection between a pair of nodes as '1' if it exists, and as '0' if it does not.

The Link Estimation Indicator (LEI)

Two mobile nodes V_{n_i} and V_{n_j} will be linked if they are in the transmission range rτ, have known speeds, phase angles, and beginning positions (x_i, y_i) and (x_j, y_j). Estimating the steady connection quality between two mobile nodes V_{n_i} and V_{n_j} is conceivable.

where $a = v_{n_i} cos\delta_i - v_{n_j} cos\delta_j$: $b = (x_i - x_j)$: $c = v_{n_i} sin\delta_i - v_{n_j} sin\delta_j$: $d = (y_i - y_j)$

When $v_{n_i} = v_{n_j}$ and $\delta_i = \delta_j$, where a and c become zero, a stable connection will exist. Nodes will remain during when the value of $LEI_{V_{n_{ij}}} = \infty$, $LEI_{V_{n_{ij}}} = \frac{-(ab+cd)+\sqrt{(a^2+c^2)r\tau^2-(ad-bc)^2}}{(a^2+c^2)}$ if Nodes are travelling in the same direction and at the same speed. Since nodes can travel in various directions and speeds, the stability of a connection between two nodes is determined by the value of $LEI_{V_{n_{ij}}}$ value (). The following is one possible form of link stability $LS_{V_{n_{ij}}} = 1 - e^{\frac{-LEI_{V_{n_{ij}}}}{\alpha}}$.

The suggested technique predicts link failure by adding the RSSI parameter to the link state, where α is constant. Reduce T while finding the Hello message to prevent outdated information, given that T is inversely proportional; the next hop node to $LEI_{V_{n_{ij}}}$. The equation is: $LF_{V_{n_{ij}}} = 1 - e^{\frac{RSSI\ (V_{N_a}).LEI_{V_{n_{ij}}} V_{ND}}{\alpha.T}}$ indicates link failure prediction. $LEI_{V_{n_{ij}}}$

finds stable links, while RSSI checks the strength of signals at nodes in the transmission range during the route cost function for a neighbour V_{N_a}. Hello, the interval is T, and the node density is V_{ND}.

Energy Consumption

Optimising energy consumption and using energy-efficient intermediate nodes for reliable data delivery help extend the network's lifespan. Routing protocols connect nodes and identify data transmission pathways. Energy depletion causes some network nodes to fail. To ensure reliable data transmission, the proposed approach monitors intermediate-node residual energy. Because the network is homogeneous, all nodes start at the same time. At time instant, the initial enegy intermediated V_{n_i} is represented by e_{int} while e_{res} designates the residual energy for the specified period; determine the residual energy of the node V_{n_i} as $e_{res}\ (t_n + \Delta) = e_{int}\,(t_n) - e_{cons}\,(t_n + \Delta)$ represents the energy used by the node V_{n_i} to transmit, receive, and calculate internal activities. The implied method calculates node energy for data forwarding and reduces link fluctuation in unstable connections due to channel fading $e_{t_x}(t_n + \Delta) = d_s \times p_{t_x}(t_n + \Delta)$. In the above formula Δ , Watts per second $p_{t_x}(t_n + \Delta)$ measures the power consumption during data packet transmission and control message exchange with an intermediary node. To calculate the energy of the receiving node during packet reception $e_{r_x}(t_n + \Delta) = d_r \times p_{r_x}(t_n + \Delta)$.

At time $(t_n + \Delta)$, $p_{r_x}(t_n + \Delta)$ represents the power used for receiving data packets d_r. The node consumes energy during internal computations, including processing digital signals, updating encoding, and decoding at time intervals Δ, indicated as $e_{comp}\ (t_n + \Delta)$. The total energy needed by a node V_{n_i} during the period Δ for transmission, reception, and computing is $e_{cons}\ (t_n + \Delta) = e_{t_x}(t_n + \Delta) + e_{r_x}(t_n + \Delta) + e_{comp}\ (t_n + \Delta)$.

Thus, node residual energy is computed as $e_{res}\ (t_n + \Delta) = e_{int}\,(t_n) - \{e_{t_x}(t_n + \Delta) + e_{r_x}(t_n + \Delta) + e_{comp}\ (t_n + \Delta)\}$. For efficient data packet forwarding, LEI-EA routing picks nodes with greater residual energy. Proposed LEI-EA routing scheme selects nodes with higher residual energy for forwarding data packets efficiently.

Link Estimation Indicator and Energy Aware (LEI-EA) Algorithm

Set_RTR_PROT = **LEI-EA**

Set_node = V_{n_i}

Set_Source = V_{n_s}

Set_Dest = V_{n_d}

LEI-EA_method( ) {

Proc broadcast Hello ($V_{n_{ij}}$) to find neighbour

Initialise route discovery process

Calculate RSSI (Received signal strength) distance between nodes:

$$RSSI = -(10 * log_{10}\left(dV_{n_{ij}}\right) - A$$

$$dV_{n_{ij}} = 10^{\frac{RSSI-A}{-10*n}}$$

where $dV_{n_{ij}}$ is distance between node V_{n_i} and V_{n_j} : $n = 2$ for propagation free space model : A= -32(dBm) signal strength between two nodes.

Get_Link_Status of two nodes by $L_{n_{ij}}(t_n)$

Predict link failure and estimate link quality by $LEI_{V_{n_{ij}}}$

Update routing information in routing table

Update energy levels of the nodes in the network

Check

If (e_{res} $(t_n + \Delta) >= High$ && $dist <= low$ && $hop\ count <= low$)

Select route for communication

else

check link stability

end if

RESULTS AND DISCUSSIONS

Simulation Parameters

Two scenarios are used to assess the proposed O-LEADM system and compare it to AODVLP. This scenario varies node speed and simulation duration while maintaining the node size at 50. Network Simulator (NS2) is used for simulation. CBR traffic is used. The network distributes nodes randomly within a 1000x1000 m region. Because of their varying speeds, mobile nodes change topology frequently. The node transmission range is 250 m, and its starting energy is 100 joules. The suggested method was tested in two situations. The energy threshold was 50 joules. Node speed is varied (10, 20, 30, 40 m/s), maintaining simulation time and packet size (100 seconds, 512 bytes). Other network settings are the

same for all runs. The second scenario mimics 50, 100, 150, and 200 seconds at 30 mm/s and 512 bytes.

Performance Metrics

Packet Delivery Ratio

$PDR = \frac{number\ \ of\ received\ \ packets}{number\ \ of\ transmitted\ \ packets} * 100$, the Packet Delivery Ratio is a measure of how many packets were transmitted from the source to the receiver, in comparison with packets received at the receiver to the destination. This measure indicates how well a routing protocol transmits data packets from source to destination. The routing mechanism will be more efficient if the ratio is high.

Throughput

A destination's throughput is its bit success rate. $Throughput = (total\ received\ bytes\ * 8/\ time) * 1000\ Kbps$ the unit. An efficient routing protocol must be able to receive data packets at their destination for it to be measured. The following formula is used to calculate throughput.

End-to-end Delay

A network end-to-end delay is the time it takes to receive a packet of data and retransmit it to other nodes across the network $E2E = \sum T_1 - T_2\ /\ N$.

A packet arrives at a given time T_1, is sent by the source at a specific time T_2, and is sent by the N total number of packets.

Energy Consumption

Simulated network node energy usage. Node energy consumption is the difference between initial energy assigned at simulation start and ultimate energy squandered $E_{cons} = \sum_{i=1}^{n}\big(ini(i) - ene(i)\big)$.

SIMULATION RESULTS

Packet Delivery Ratio

According to Figure 3a, AODVLP and Link Estimation Indicator and Energy Aware Routing Scheme (LEI-EA) deliver packets at different rates. Packet drop occurs when the link breaks at a high node speed, leading to a low packet delivery rate. A packet delivery ratio greater than 100 is observed in the AODVLP scheme as speed increases, but a greater packet delivery ratio is observed in the LEI-EA scheme when speed reaches its maximum. An alternate route will be dynamically selected if the original route has low residual energy

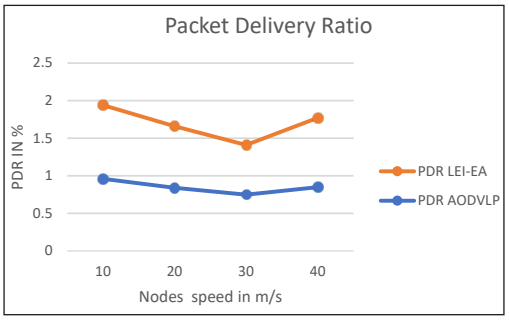


Figure 3a. PDR versus node speed

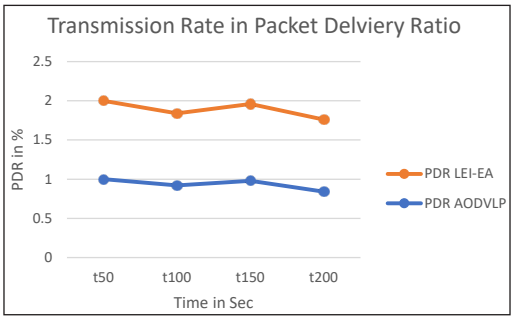


Figure 3b. PDR versus transmission rate

and if the link is considered stable. This method is used to ensure the reliability of the data transfer.

Figure 3b shows a simulated time-dependent packet delivery ratio. Longer simulations boost packet delivery. The plot demonstrates that both systems return the same packet ratio at 50, 100, and 150 seconds. In 200-second packet ratios, LEI-EA outperforms AODVLP. As simulation time increases, LEI-EA selects better data traffic pathways.

Throughput

The throughput graph is in Figure 4a Simulations show throughput differences between AODVLP and LEI-EA. LEI-EA and AODVLP have different throughput speeds. When nodes move randomly, LEI-EA calculates RSSI connection status, forwards data to the next node, and estimates link quality to identify more stable pathways.

Figure 4b shows simulation time variation. Compared to AODVLP, LEI-EA improved with simulation time. Increasing the simulation duration reduces packet dropouts. According to LEI, LEI-EA delivers packets to active nodes with better link quality via routes with higher energy and consistent connection quality.

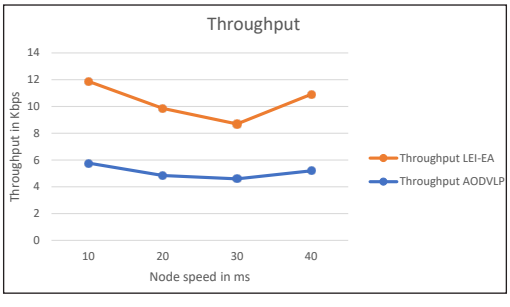


Figure 4a. Throughput versus node speed

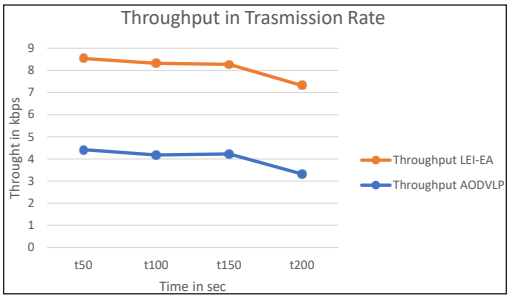


Figure 4b. Throughput versus transmission rate

End-to-End Delay

When choosing a route, the AODVLP does not consider intermediary node energy. Nodes with higher speeds are delayed. Based on mobility speed, LEI, and energy, LEI-EA chooses optimal routes and intermediate nodes. By stabilising the dynamic network, it regulates its end-to-end latency. With different simulation times, Figure 5a and 5b display the average end-to-end delay graph. Over time, LEI-EA becomes slower than AODVLP. LEI-EA selects routes with a minimum hop distance of 200 sec, reducing network transmission time.

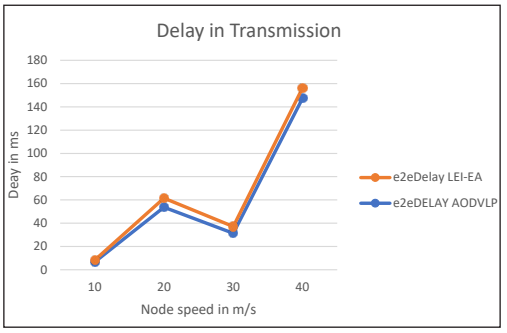


Figure 5a. End-to-end delay versus node speed

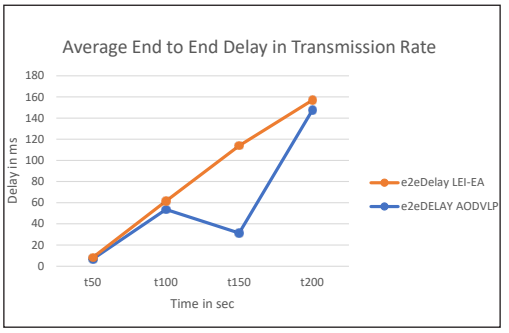


Figure 5b. End-to-end delay versus transmission rate

Energy Consumption

As shown in Figures 6a and 6b, network operations consume an average amount of energy. Nodes moving at different speeds are shown in Figure 6(a). Mobility speed increases the complexity of route computation for network topologies. In a network with higher average node density, retransmission and reroute discovery are reduced, saving time and extending the life of the network. Comparative analysis of LEI-EA versus AODVLP shows LEI-EA reduces energy consumption. The LEI-EA scheme maintains route stability and an energy balance that enables packets to be transmitted to relay or intermediate nodes with the highest residual energy. AODVLP consumes more energy because it reconstructs routes and sends more RREQ packets to find alternate routes. As displayed in Figure 6b, energy comparisons with varying simulation times are shown. According to LEI-EA, routes are classified, and nodes are selected based on the amount of energy they consume. Nodes in the network consume more energy as the simulation time increases. Comparing LEI-EA to AODVLP, simulation results indicate that LEI-EA delivers data packets more efficiently, prolonging network lifetime.

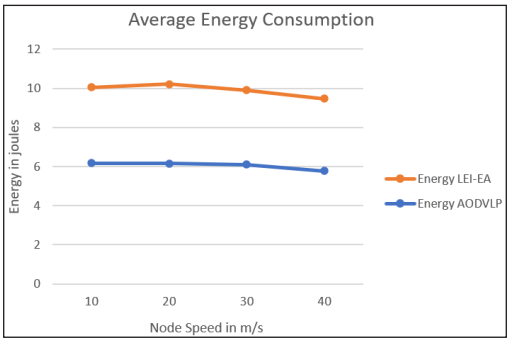


Figure 6a. Energy consumption versus node speed

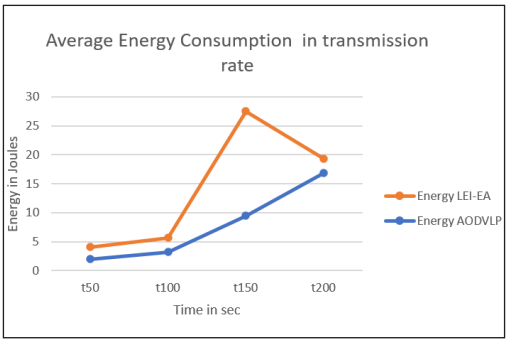


Figure 6b. Energy consumption versus transmission rate

Overhead

The overhead for LEI-EA is 5.10 to 17.6, while for AODVLP it is 5.11 to 12.1. LEI-EA shows remarkable improvements in mobility when compared to AODVLP Figure 7. Whenever a black hole is detected, LEI-EA shows lower overhead when it discovers the alternative paths. Therefore, path discovery in AODVLP has higher overhead when it is compared to LEI-EA.

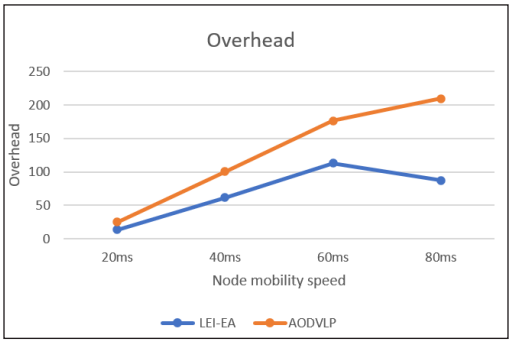


Figure 7. Overhead vs node mobility speed

CONCLUSION

Mobile ad hoc networks (WSNs) integrate routing capabilities into mobile nodes to allow reliable and effective mobile wireless network operation. Distributed nodes that communicate multi-hop with one another make up a WSN. When nodes move often, topological changes and link failures cause link disjoints and lower network performance. We must maintain path stability to identify the best multipath from source to destination and keep the path stable. A single parameter does not determine the efficiency of WSN data transmission. Packet loss, connection failures, and energy consumption plague WSNs. An energy-aware multipath routing strategy is suggested for WSN routing. Energy-aware routing scheme with link estimation indicator (LEI-EA). A link estimation indicator (LEI) can be used to compute the dependability of data transfer. The proposed LEI-EA system and the AODVLP scheme are compared in NS2 by altering nodes and simulation time. A network's performance is assessed by looking at things like throughput, average end-to-end latency, packet delivery ratio, and energy use. Simulation findings show that LEI-EA outperforms AODVLP. When compared to AODVLP, LEI-EA uses less energy,

has a lower latency, and offers a higher packet count. As nodes can discard packets when attacked, future improvements might investigate and use a variety of situations to examine the behaviour of nodes during data routing. A trust-based routing strategy that evaluates node trust values to ensure secure routing to a destination can be used to enhance the quality-of-service criteria.

ACKNOWLEDGEMENT

The authors sincerely thank their institution for providing the necessary facilities and support to carry out this research. We also express our gratitude to the reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

LIST OF ABBREVIATIONS

WSN	:	Wireless Sensor Network
LEI-EA	:	Link estimation indicator and energy-awareness routing scheme
LEI	:	Link estimation indicator
AODVLP	:	Ad hoc on-demand distance vector (AODV) link prediction.
PDR	:	Packet delivery ratio
O-LEADM	:	On-demand using link estimation and energy-awareness
RWP	:	Random waypoint
RREQ	:	Route request packet headers
RSSI	:	Received signal strength

REFERENCES

Agiwal, M., Roy, A., & Saxena, N. (2016). Next generation 5G wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 18*(3), 1617-1655. https://doi.org/10.1109/COMST.2016.2532458

Ali, S., Madani, S. A., Khan, A. U. R., & Imran, A. K. (2014). Routing protocols for mobile sensor networks: A comparative study. *Computer Systems Science and Engineering, 29*(2), 183-192.https://doi.org/10.1049/el:20081952

Aswale, S., & Ghorpade, V. R. (2017). LQEAR: Link quality and energy-aware routing for wireless multimedia sensor networks. *Wireless Personal Communications, 97*, 1291-1304. https://doi.org/10.1007/s11277-017-4566-8

Baccour, N., Koubâa, A., Noda, C., Fotouhi, H., Alves, M., Youssef, H., Zúñiga, M. A., Boano, C. A., Römer, K., Puccinelli, D., Voigt, T., & Mottola, L. (2013). *Radio link quality estimation in low-power wireless networks*. Springer. https://doi.org/10.1007/978-3-319-00774-8

Barroca, N., Borges, L. M., Velez, F. J., & Chatzimisios, P. (2014). IEEE 802.15.4 MAC layer performance enhancement by employing RTS/CTS combined with packet concatenation. In *2014 IEEE International Conference on Communications (ICC)* (pp. 466-471). IEEE. https://doi.org/10.1109/ICC.2014.6883362

Cao, Y., & Sun, Z. (2013). Routing in delay/disruption tolerant networks: A taxonomy, survey and challenges. *IEEE Communications Surveys & Tutorials, 15*(2), 654-677. https://doi.org/10.1109/SURV.2012.042512.00053

Deepa, J., & Sutha, J. (2019). A new energy-based power-aware routing method for MANETs. *Cluster Computing, 22*(Suppl. 6), 13317-13324. https://doi.org/10.1007/s10586-018-1868-x

Dogra, R., Rani, S., & Sharma, B. (2021). A review of forest fires and their detection techniques using wireless sensor networks. In G. S. Hura, A. K. Singh, & L. Siong Hoe (Eds.), *Advances in communication and computational technology: ICACCT 2019* (Vol. 668, pp. 997-1005). Springer. https://doi.org/10.1007/978-981-15-5341-7_101

Girirajan, B., Hemalatha, K. L., Manjunath, B. N., Suma, S., & Pushpavalli, M. (2023). Hybrid optimisation algorithm for effective clustering algorithm and routing protocol in MANET. In *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)* (pp. 1-5). IEEE. https://doi.org/10.1109/EASCT59475.2023.10392304

Goldsmith, A. J., & Chua, S.-G. (1997). Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications, 45*(10), 1218-1230. https://doi.org/10.1109/26.634685

Gupta, S., & Marriwala, N. (2017). Improved distance energy-based LEACH protocol for cluster head election in wireless sensor networks. In *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 441-445). IEEE. https://doi.org/10.1109/ISPCC.2017.8269656

Helen, D., & Arivazhagan, D. (2014). Applications, advantages and challenges of ad hoc networks. *Journal of Academia and Industrial Research, 2*(8), 453-457.

Huang, C.-F., Chan, Y.-F., & Hwang, R.-H. (2017). A comprehensive real-time traffic map for geographic routing in VANETs. *Applied Sciences, 7*(2), Article 129. https://doi.org/10.3390/app7020129

Kalansuriya, P., Tellambura, C., & Network, A. D. F. C. (2009). Performance analysis of decode-and-forward relay network under adaptive M-QAM. *Electronics Letters, 45*(6), 339-341. https://doi.org/10.1049/el:20081952

Kuipers, F. A., & Van Mieghem, P. (2005). Conditions that impact the complexity of QoS routing. *IEEE/ACM Transactions on Networking, 13*(4), 717-730. https://doi.org/10.1109/TNET.2005.852882

Lin, Y.-W., & Huang, G.-T. (2012). Optimal next hop selection for VANET routing. In *2012 7th International Conference on Communications and Networking in China (CHINACOM)* (pp. 611-615). IEEE. https://doi.org/10.1109/ChinaCom.2012.6417556

Mafirabadza, C., Makausi, T. T., & Khatri, P. (2016). Efficient power-aware AODV routing protocol in MANET. In *Proceedings of the International Conference on Advances in Information Communication Technology & Computing* (Article 58). Association for Computing Machinery. https://doi.org/10.1145/2979779.2979831

Malwe, S. R., Taneja, N., & Biswas, G. P. (2017). Enhancement of DSR and AODV protocols using link availability prediction. *Wireless Personal Communications, 97*(3), 4451-4466. https://doi.org/10.1007/s11277-017-4733-y

Marriwala, N. (2022). Energy harvesting system design and optimisation using high bandwidth rectenna for wireless sensor networks. *Wireless Personal Communications, 122*, 669-684. https://doi.org/10.1007/s11277-021-08918-x

Marriwala, N. K., Shukla, V. K., Raju, A. R., & Singh, P. K. (2024). OntoBlock: A novel ontological-based and blockchain-enabled spectrum sensing framework for detection of malicious users in cognitive radio Internet of Things (CR-IoT) networks. *International Journal of Information Technology, 16*(5), 3913-3921. https://doi.org/10.1007/s41870-024-02011-9

Otero, A. S., & Atiquzzaman, M. (2011). Adaptive localised active route maintenance mechanism to improve performance of VoIP over ad hoc networks. *Journal of Communications, 6*(1), 68-78. https://doi.org/10.4304/jcm.6.1.68-78

Singal, G., Laxmi, V., Gaur, M. S., & Lal, C. (2014). LSMRP: Link stability-based multicast routing protocol in MANETs. In *2014 Seventh International Conference on Contemporary Computing (IC3)* (pp. 254-259). IEEE. https://doi.org/10.1109/IC3.2014.6897182

Smail, O., Cousin, B., Mekki, R., & Mekkakia, Z. (2014). A multipath energy-conserving routing protocol for wireless ad hoc networks lifetime improvement. *EURASIP Journal on Wireless Communications and Networking, 2014*, Article 139. https://doi.org/10.1186/1687-1499-2014-139

Suma, S., & Harsoor, B. (2020a). Congestion control for multihop transmission in WSN using contention window. *International Journal of Advanced Science and Technology, 29*(5), 13697-13703. http://sersc.org/journals/index.php/IJAST/article/view/32599

Suma, S., & Harsoor, B. (2022b). An approach to detecting black hole attacks for congestion control utilising mobile nodes in wireless sensor networks. *Materials Today: Proceedings, 56*(4), 2256-2260. https://doi.org/10.1016/j.matpr.2021.11.590

Suma, S., & Harsoor, B. (2022c). Detection of malicious activity for mobile nodes to avoid congestion in wireless sensor network. In *2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICAECC54045.2022.9716673

Suma, S., & Harsoor, B. (2022d). An optimised routing scheme for congestion avoidance using mobile nodes in wireless sensor network. *Measurement: Sensors, 24*, Article 100457. https://doi.org/10.1016/j.measen.2022.100457

Suma, S., & Harsoor, B. (2022e). Dynamic shortest-path routing algorithm to reduce retransmissions and congestion for mobile nodes in wireless sensor networks. In P. P. Joby, V. E. Balas, & R. Palanisamy (Eds.), *IoT-based control networks and intelligent systems* (Vol. 528, pp. 519-530). Springer. https://doi.org/10.1007/978-981-19-5845-8_46

Suma, S., Dharmireddi, S., Bhaskar, N., Divya, G., Raju, K. S., & Mamatha, B. (2024). Congestion control and avoidance scheme using mobile nodes in wireless sensor network. In F. M. Lin, A. Patel, N. Kesswani, & B. Sambana (Eds.), *Accelerating discoveries in data science and artificial intelligence II: ICDSAI 2023* (Vol. 438, pp. 375-384). Springer. https://doi.org/10.1007/978-3-031-51163-9_32

Venugopal, A., Gautam, A., Suma, S., & Kumar, A. (2025). An adaptive energy detection scheme using estimation of noise uncertainty for cognitive radio. *International Journal of Information Technology, 17*(6), 3769-3774. https://doi.org/10.1007/s41870-025-02551-8

Wang, X., & Li, J. (2015). Improving the network lifetime of MANETs through cooperative MAC protocol design. *IEEE Transactions on Parallel and Distributed Systems, 26*(4), 1010-1020. https://doi.org/10.1109/TPDS.2013.110

Xu, L., Wang, J., Liu, Y., Yang, J., Shi, W., & Gulliver, T. A. (2018). Outage performance for IDF relaying mobile cooperative networks. In K. Long, V. C. M. Leung, H. Zhang, Z. Feng, Y. Li, & Z. Zhang (Eds.), *5G for future wireless networks: 5GWN 2017* (Vol. 211, pp. 366-374). Springer. https://doi.org/10.1007/978-3-319-72823-0_37

Yadav, A., Singh, Y. N., & Singh, R. R. (2015). Improving routing performance in AODV with link prediction in mobile ad hoc networks. *Wireless Personal Communications, 83*, 603-618. https://doi.org/10.1007/s11277-015-2411-5

Yang, D., Xia, H., Xu, E., Jing, D., & Zhang, H. (2018). Energy-balanced routing algorithm based on ant colony optimisation for mobile ad hoc networks. *Sensors, 18*(11), Article 3657. https://doi.org/10.3390/s18113657

Yazıcı, V., Kozat, U. C., & Sunay, M. O. (2014). A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management. *IEEE Communications Magazine, 52*(11), 76-85. https://doi.org/10.1109/MCOM.2014.6957146

Pertanika J. Sci. & Technol. 34 (S1): 89 - 110 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Enhanced Image Captioning Using Sinusoidal Spatial Transform CNN and Fine-tuned BERT

Bhargavi Polepalli*, Praveen Kumar Sekharamantry, and Konda Srinivasa Rao

Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University), Gandhi Nagar, Rushikonda, 530045 Visakhapatnam, Andhra Pradesh, India

ABSTRACT

Image caption generation is a significant area of study in artificial intelligence and computer vision, focusing on training systems to generate accurate textual descriptions of images. This paper presents an advanced framework for image captioning using a dataset sourced from Kaggle. The process begins with pre-processing techniques, including the Flexiframe Filter, which dynamically adjusts the window size based on local variance to reduce noise in both text and images. Bright-Contrast augmentation is then applied to enhance input images, enriching the dataset and improving feature extraction. The encoder module utilises a Sinusoidal Spatial Transform Convolutional Neural Network (SST-CNN) integrated with a Visual Geometry Group (VGG16) model for feature extraction and encoding. For the decoder, stochastic k-sampling is combined with a Dense Neural Network and a Gated Recurrent Unit (GRU) to generate initial captions. To refine these captions, a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model is employed for enhanced coherence and accuracy. The model's performance is evaluated using standard metrics, achieving BLEU-1, METEOR, ROUGE-L, and CIDER scores of 1, 0.99, 0.99, and 1, respectively. The proposed approach demonstrates significant improvements in generating descriptive and accurate image captions, making it a robust solution for applications requiring semantic understanding of visual data.

Keywords: Bright and contrast augmentation, fine-tuned BERT, Flexiframe Filter, Sinusoidal Spatial Transform Convolutional Neural Network (SST-CNN), Stochastic k-sampling DenseGRU Neural Network

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 29 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.05

E-mail addresses:
bpolepal@gitam.in (Bhargavi Polepalli)
psekharm@gitam.edu (Praveen Kumar Sekharmantry)
skonda@gitam.edu (Konda Srinivasa Rao)
* Corresponding author

INTRODUCTION

Image caption generation can surely be termed as an interesting and complex problem in the new and constantly

e-ISSN: 2231-8526

developing fields like artificial intelligence and computer vision. Its goal is to create techniques and algorithms through which these textual descriptions of the images may be created (Varma et al., 2026). Stated that this skill relates to the translation of language to vision, and it has a profound effect on the availability, delivery, and interaction of information to a computer.

The Significance of Image Caption Generation

The ability to come up with accurate and culturally appropriate descriptions of images is important in diverse industries. It is essential to point out that the accessibility of image captions is decisive because people who are blind or have low vision can get an understanding of images via textual descriptions (Beddiar et al., 2023). Functions such as image captioning methods, for example, can be better at describing complex scenes, items, and activities and also enhance the use of social media sites, news websites, and educational sites.

Automated captioning systems can also help to enhance the indexing and categorisation of a substantial number of visual contents when used in the management of digital media (Mahalakshmi & Fatima, 2022). This is particularly useful for several industries such as e-commerce that may benefit from making product searches even more effective, while customers may be more engaged by the meaningful captions (Afyouni et al., 2021). The suitable captioning of the shown content means that the material recommendation, suggestion, and personalisation functions in the field of media and entertainment become more effective.

Applications and Implications

The process of caption writing can be useful in a wide array of areas. It can be beneficial to make images more readable, especially for people with some sort of vision issues, since automatic image captions help enhance this aspect. Successful and interesting captions can improve the overall experience as well as product descriptions for e-commerce stores (Degadwala et al., 2021). Caption generation can help in the process of personalisation and the creation of content in social media and digital marketing. In addition, the development of an automatic captioning system must lead to the improvement of the processes of language acquisition, as well as help with assistive technologies in education (Hossain et al., 2021).

Even though image caption generation has made significant progress, there remain several challenges and opportunities, the exposure of which is the following: generating captions for images depicting complex environments and understanding conceptual opinions and cultural and contextual relevance (Al Duhayyim et al., 2022). It is also important to anticipate and solve the ethical concerns for using image captioning systems that include bias and fairness in this technology's development and deployment (Xian et al., 2022).

This involves applying the SST-CNN method to generate reasonable as well as informative captions and computer vision approaches to extract substantial objects from images.

In Section 2, we present a literature review including datasets, findings, and image captioning methods. In Section 3, we present our approaches for producing a variety of captions. Sections 4 present the outcomes of the experiment. Section 5 is the conclusion

LITERATURE REVIEW

In this section, evaluation image captioning methods, dataset, and their evaluation metrics are shown in Table 1.

Table 1
Literature review description

Objective	Dataset	Findings
Puscasiu et al. (2020) conducted extensive field research and attempted to create an application for deep learning-based automated image description.	80000 annotated images from Microsoft Common Objects in Context (MSCOCO)	To help visually impaired persons make sense of images, the technology had to be further connected with a text-to-speech engine.
In the field of computer vision research, addressed by Castro et al. (2022). Visual attention is a cutting-edge method for image captioning.	MSCOCO 2014 dataset	For the BLEU-4 and Top-5 accuracy, MobileNetV3 supports values of 19.50 and 72.928, respectively, indicating that it provided competitive output quality at the expense of reduced computing performance.
Devi et al. (2020) numerous models, including the cutting-edge encoder-decoder (Sequential CNN-Recurrent Neural Network (RNN) systems that had shown to be effective in producing outcomes, had been put forth to address the issue.	They make use of the Flickr8k and 30k databases.	Comparing their performance to other models, they obtained state-of-the-art results on both datasets.
Wang et al. (2022), presented a novel approach to image captioning using recognised elements and relationships through high-order interaction learning within the encoder-decoder architecture.	MSCOCO dataset	Numerous tests demonstrated that the suggested approach could outperform cutting-edge techniques on the MSCOCO database in a competitive manner.
The efficacy of the Approach was demonstrated in the research by (Al-Malla et al., 2022). Presenting an attention-based Encoder-Decoder image captioning model that makes use of two feature extraction techniques: an object detection module, an image classification CNN (Xception), and You Only Look Once version 4 (YOLOv4).	MS COCO and Flickr30k	The CIDER score was increased by 15.04% with their new feature extraction technique.

Table 1 (*continued*)

Objective	Dataset	Findings
Shen et al. (2020) proposed a two-stage multitask learning system depending on reinforcement learning (RL) and a variational autoencoder. For remote sensing image captioning, the two-stage Multi-task Learning Model (VRTMM) relied on reinforcement learning and a variational autoencoder.	NWPURESISC45 dataset	The experiment's outcome showed that their model works well for remotely sensed image captioning and produced unique, cutting-edge results.
The Task-Adaptive Attention module for image captioning presented by Yan et al. (2022) could help with the deceptive issue and teach implicit non-visual	MSCOCO captioning Dataset	Comprehensive tests on the MSCOCO captioning database showed that performance improvement was possible when their Task-Adaptive Attention module was plugged into a vanilla Transformer-based image captioning model.

RESEARCH METHODOLOGY

Image captions are generated using the following methodology: gathering data from the Kaggle source, preprocessing images with text and Flexiframe Filter and Bright-Contrast augmentation, feature extraction with SST-CNN and VGG16, producing initial captions with stochastic k-sampling, dense neural networks, and GRU and enhancing captions with fine-tuned BERT. The evaluation employs the BLEU, METEOR, CIDER, and ROUGE-L metrics to evaluate caption integrity and efficiency. Figure 1 shows the overview of the methodology.

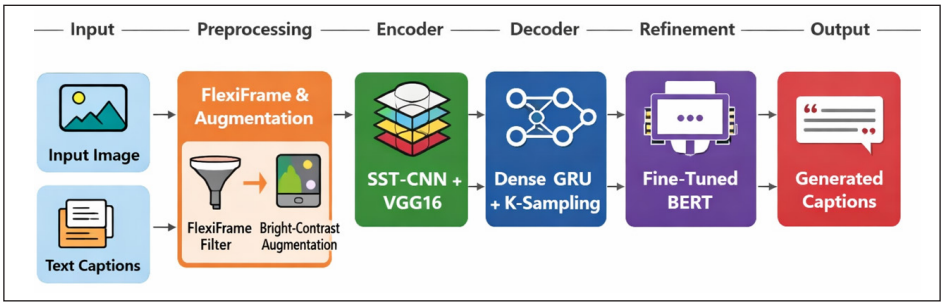


Figure 1. Overview of methodology

Data Collection

8,000 images with five distinct captions that clearly describe the important objects and events are included in these unique benchmark collections for sentence-based image description and examination. The images were picked from six separate Flickr categories, and they

were carefully chosen to show a range of settings and scenarios rather than any famous persons or places. We used 4000 distinct images from the Flickr8k dataset gathered from the Kaggle source https://www.kaggle.com/datasets/adityajn105/flickr8k?select=Images. Figure 2 shows the sample images.

Figure 2. Sample images

Data Preprocessing

This improves the model's capacity to capture context-specific details and produce more precise and expressive captions by dynamically shifting the frame of attention to focus on important image regions. Preprocessing text entails cleaning, tokenising, shrinking the vocabulary, padding sequences, and adding sequence tokens. Grayscale conversion, Gaussian blur, edge detection, thresholding, contour detection, and mask application are some of the image preprocessing techniques used with the Flexiframe Filter to improve feature extraction.

Text Preprocessing

Text cleaning eliminates punctuation, single-character words, numeric values, and extraneous characters, making the characters cleaner so that important words can be recognised. Add Sequence Tokens at the start and finish of each caption, including special tokens $('startseq' \, and \, 'endseq')$ aids in identifying the initial and finish of sequences in systems requiring these markers. Text data is prepared for neural network input by tokenisation, which turns words into numerical values and turns captions into sequences of integers where each word is mapped to a distinct integer. Calculate the Size of The Vocabulary using text analysis, determine the vocabulary size required to cover a given percentage of the data (e.g., 95% coverage), and decide how many terms should be added to the vocabulary that are unique. Pad Sequences use zeros to buffer shorter sequences so that they are all the same length, creating uniformly sized sequences for feeding neural networks.

Flexiframe Filter

Grayscale Conversion reduces the image to a single channel and turns it to grayscale to facilitate processing. Gaussian Blur enhances edge recognition by reducing noise and smoothing the grayscale image. Edge Detection: Sobel Operator finds edges in the x- and y-axes to draw attention to the image's key elements. Thresholding and Contour Detection generate a binary image and use contour identification to separate important features. Mask Application determines which sections inside the identified contours should be retained by applying the binary mask to the original image.

Each image has been filtered to highlight and isolate important details. Along with a final message stating that the processing is finished, the console shows the names of every processed image. Figure 3 displays the resultant preprocessed image.

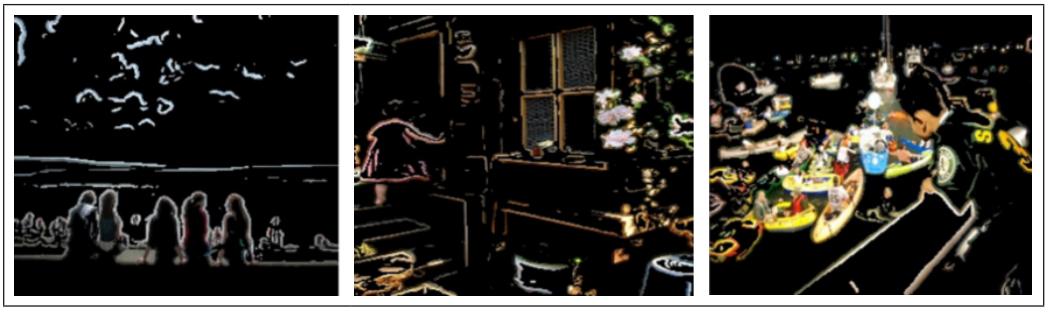


Figure 3. Output of preprocessed images

Augmentation Using Bright-Contrast

Strengthening the models' resistance to visual changes, this approach aids in the improvement of image captioning efficiency. Brightness Adjustment uses a brightness factor of 1.2 to increase brightness by 20%. This scales pixel values to brighten the image. Contrast Adjustment applies a contrast factor of 0.8 to reduce contrast by 20%. Modifying the mean intensity and scaling the pixel values lowers the contrast. Send back the updated picture with the made changes.

Images that have been modified using particular augmentation settings make up the output:

- A 20% improvement in brightness was seen (brightness factor of 1.2).
- A 20% drop in contrast was seen (contrast factor: 0.8). Figure 4 shows the output of augmented images.

Figure 4. Output of augmented images

Image Feature Extraction Using Pre-trained Model VGG16

These features help to provide meaningful and contextually appropriate captions for images by acting as input for image captioning systems. To recognise patterns and visualise, data must initially be converted into numerical values, a process known as feature extraction. It is a necessary phase that aids the model in determining greater precision. The most crucial element in choosing a suitable feature extraction and processing method is the abundance of varied variables present in these enormous collections. Processing these variables requires a substantial amount of computational resources. To efficiently decrease data size, the process of feature extraction involves choosing and integrating parameters to create features that extract the greatest possible information from images.

The VGG16 feature extraction approach is capable of extracting large amounts of data with better performance. This method is widely employed for extracting features from images and has a substantial impact. The VGG16 model is applied to a collection of image data. To extract features, the proposed models' designs use a variety of layers across their architecture. The complete framework was built using convolutional, batch normalisation, pooling, and fully connected layers, as well as ReLU and SoftMax functions. The VGG16 is a compact design that includes a pooling layer, convolution, and a fully connected layer, all of which are present in the AlexNet model. There are an entire 16 layers that expand with the number of layers. This network keeps the input image size at 224 by 224 pixels, with a 3 by 3 filter. The final step of this network includes an activation function effective at giving class possibilities to the output layer collected image features by the VGG16 pre-trained model displayed in Figure 5.

```
Extracted Features:

Features for 1048710776_bb5b0a5c7c.jpg:
[0.         0.         0.65566266 ... 0.29483652 0.60420024 0.        ]

Features for 1042020065_fb3d3ba5ba.jpg:
[0.25006348 0.         0.         ... 0.         0.         1.6714716 ]

Features for 1000268201_693b08cb0e.jpg:
[1.6208034 2.9046288 0.        ... 0.        0.        2.1787658]
```

Figure 5. Extracted image features by VGG16 pre-trained model

Encoder Using SST-CNN

This technique improves image captioning by providing richer spatial context for generating accurate and descriptive captions.

The CNN is a sophisticated model; its classification and prediction accuracy will decrease with a particular alteration or interference of the target image. The data sets are all thought to be mixed-written digital images based on the concept of model optimisation. An SST-CNN was developed as a new network structure based on the CNN model. By identifying and aligning anomalous and conspicuous regions in the image, the SST can rectify and align. Below are the particular steps involved in network design.

Stage 1

Establish the network for placement. As an input, the feature map (FM)V is utilised. It is possible to employ a convolutional or fully connected structure. Outputting the spatial transformation component θ and recording $\theta = floc(V)$ is the purpose.

Stage 2

Convert the incoming FM using the spatial transformation component θ that the positioning network outputs. Using the linear transformation as an instance, it may map to a place on the input map with the component (w_j^t, z_j^t) and output a specific position on (w_j^s, z_j^s). The computation Equation 1 is as follows.

$$\begin{pmatrix} w_j^t \\ z_j^t \end{pmatrix} = \theta \begin{pmatrix} w_j^s \\ z_j^s \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{21} & \theta_{21} \end{bmatrix} \begin{pmatrix} w_j^s \\ z_j^s \\ 1 \end{pmatrix} \quad [1]$$

Stage 3

Use bilinear interpolation to complete the spaces as per the completing rules specified in the preceding period. For completing, the pixel value equivalent to the coordinate point in V is derived using the coordinate point of U, negating the requirement for matrix operations.

Note that there is no direct performance of the completion. Once some processing has been done, the additional pixel values surrounding the completion should be taken into consideration if the calculated coordinates are decimals. Here is the complete preparation in Equation 2.

$$U_j = \sum_n \sum_m V_{mn} * l\left(w_j^t - n; \emptyset_w\right) * l\left(z_j^t - m; \emptyset_z\right) \quad [2]$$

The sampling kernels $\emptyset_w$ and $\emptyset_z$ are shown above. The parameter L determines the type of image interpolation. Here, the input FM V_{mn} is V position. $U\ (m, n)$, Location is $\left(w_j^s, z_j^s\right)$ and U_j is the output FM. The following Equation 3 can be used to apply bilinear interpolation.

$$U_j = \sum_m \sum_n V_{mn} * \max\left(0,1 - \left|w_j^t - n\right|\right) * \quad \max\left(0,1 - \left|z_j^t - m\right|\right) \quad [3]$$

Encoder Output

- The encoder in an autoencoder transfers the input data to a latent space, usually one with fewer dimensions. Put another way, the encoder produces an encoding or compressed description of the input. The result highlights the most significant aspects of the supplied data while removing the less crucial information.
- The encoder output is typically significantly more compact and efficient than the input since its dimensionality is substantially smaller.

Decoder Using Stochastic k-Sampling DenseGRU Neural Net

The model decodes the latent features into the original image dimensions using a unique technique that combines DenseGRU layers with stochastic k-sampling.

Stochastic k-Sampling

To improve the diversity and the potential of the generated captions in the environment of image caption decoding, stochastic k-sampling can be used. A model is trained to produce descriptive sentences for images in image captioning. This usually requires an encoder-decoder architecture, in which the decoder (usually an RNN or Transformer) creates a string of words to make a caption, and the encoder (usually a CNN) drives features from the image.

Top-k Sampling

The top k probable terms in the distribution are chosen by the model, as opposed to the word with the highest probability. The number of top words that are taken into consideration is

determined by the hyperparameter k. The top 5 probable terms are taken into account by the model if $k = 5$ Using these top k candidates and their probability to determine the following word, stochastic sampling adds randomisation.

Stochastic Selection

The top k words are chosen at random, but the probability of selecting each word is proportionate to the probability that the model predicts. For instance, there is a 50% probability of selecting the first word, a 30% chance of selecting the second, and a 20% chance of selecting the third if the probabilities of the top three words are 0.5, 0.3, and 0.2.

Diversity in Captions

By adding unpredictability to the decoding procedure, stochastic k-sampling facilitates the creation of several unique captions for a single image. Every time the model decodes a caption using a different set of stochastic options, it might generate a distinct and legitimate caption. This variety is especially useful for tasks where an image can have several reasonable interpretations.

Trade-off

The ratio of correctness to diversity is balanced. The captions could become less accurate or logical if k is set too high. The captions may get monotonous and less varied if k is set too low. The particular application and intended results determine which value of k to use.

Dense GRU Neuro Net

A GRU layer processes the sampled features, efficiently capturing complex dependencies and patterns in the latent space. Subsequently, the GRU layer's output is fed into a variety of dense layers during the dense expansion stage. These layers gradually enlarge the data to encompass the entire image, utilising regularisation strategies and non-linear transformations to improve the model's efficiency and capacity for generalisation.

GRU

To the GRU network e_h, the feature input tensor is delivered. Depending on the level of hyperparameter effectiveness, the GRU could have multiple layers (ML) or perhaps one layer. $W[s] \in \mathbb{R}^{m_c \times g}$ Represents the input for a given time step s, and the following computations are made, as shown in Equations 4 to 7.

$$Q[s] = \sigma\big(W[s]X_{wq} + G[s-1]X_{gq} + a_q\big) \quad [4]$$

$$Y[s] = \sigma\big(W[s]X_{wy} + G[s-1]X_{gy} + a_y\big) \quad [5]$$

$$\tilde{G}[s] = \tanh\big(W[s]X_{yg} + (Q[s] \odot G[s-1])X_{gg} + a_g\big) \quad [6]$$

$$G[s] = Y[s] \odot G[s-1] + \big(1 - Y[s] \odot \tilde{G}[s]\big) \quad [7]$$

Where the hidden state (HS)of the previous time stage is represented by $G[s-1] \in \mathbb{R}^{m_c \times g}$, the reset gate is represented by $Q[s] \in \mathbb{R}^{m_c \times g}$, the update gate is represented by $Y[s] \in \mathbb{R}^{m_c \times g}$, the weight variables are $X_{wq} X_{wy} \in \mathbb{R}^{m_c \times g}$ and $X_{gq} X_{gy} \in \mathbb{R}^{g \times g}$, the biases are represented by $a_q a_y a_g \in \mathbb{R}^{1 \times g}$, and the number of features in the layer is represented by g $(g = m_e)$ for the first layer). The new state is $G[s]$, and the candidate concealed state is $\tilde{G}[s]$~ stands for the Hadamard product $\odot$ and σ for the sigmoid function. The input of a specific layer in an ML GRU network is the HS $G[s]$ of the layer before it. There isn't a dropout utilised among the layers.

Dense GRU Neuro Net

Concurrently, the feature tensor is restructured into a vector $W_{c1} \in \mathbb{R}^{m_c \times (m_e * 24)}$, which is supplied to a solitary dense layer network e_{c1}. The activation function that is employed is the Rectified Linear Unit (ReLU). The dense layer can be expressed as follows, assuming that $(Z_{c1})^l$ is the output vector and $(W_{c1})^l$ is the input vector for the l^{th} developing instance in Equations 8 and 9.

$$U_j^l = ReLU\big(\sum x_{ji} (w_{c1})_i^l - ag_j\big) \quad [8]$$

$$(Z_{c1})_o^l = \sum U_j^l x_{oj} - ap_o \quad [9]$$

Where U_j^l the weight component from the hidden layer neuron j to the input layer neuron i is the variance of the output neuron ag_j is the partiality of the hidden neuron (HN)j, and x_{oj} is the weight component from HNj to output o. The outputs Z_{c1} from the dense layer and Z_h from the final GRU layer are connected.

A network of attention e_b receives the connected layer $W_b(Z_h, Z_{c1})$. There's a utilisation of generous assistance. Below is an explanation of the attention mechanism.

Initially, a collection of internal states called $g_1, g_2, \ldots, g_n$ are created by encoding the input connected sequence. A feedforward network is trained to identify important

states by assigning greater ratings to states that merit attention, and conversely. This process yields alignment ratings for every encoded state. The softmax function is used to calculate attention weights $\propto_1, \propto_2, \ldots\ldots \propto_n$ for the ratings. Take note that the attention weight vector, with $\propto \in [0,1]$ and $\sum \alpha = 1$, provides a probabilistic interpretation. The setting vector is then calculated in Equation 10.

$$D = \propto_1 * g_1 + \propto_2 * g_2 + \cdots .. + \propto_n * g_n \quad [10]$$

The output from the earlier time stage is connected with D. For every time step, the procedure is performed. Using a second dense layer network e_{c2}, the desired drift sequence Δ,which once more employs ReLU as the activation function, is assigned to the attention layer output Z_b. Thus, the DGNN model e_{DGNN}: $W \rightarrow \Delta$ can be summed up as follows in Equations 11 to 14.

$$GRU\ e_h: \quad W \rightarrow Z_h \quad [11]$$

$$Densel\ e_{c1}: W_{c1} \rightarrow Z_{c1} \quad [12]$$

$$Attention\ e_b: W_b(Z_h, Z_{c1}) \rightarrow Z_b \quad [13]$$

$$Dense2 \quad e_{c2}: W_{c2}(Z_b) \rightarrow \Delta \quad [14]$$

The following summarises the formulas for each activation function that was employed in the model shown in Equations 15 to 18.

$$\sigma(w) = \frac{1}{(1+e^{-w})} \quad [15]$$

$$\tanh(w) = \frac{(e^{w} - e^{-w})}{e^{w} + e^{-w}} \quad [16]$$

$$ReLU(w) = \max(w, 0) \quad [17]$$

$$softmax_j(\vec{w}) = \frac{e^{w_j}}{\sum_{i=1}^{n} e^{w_i}}\ for\ j = 1, \ldots\ldots, n \quad [18]$$

Generated Caption [Fine-tuned BERT]

The BERT model is fine-tuned to handle the task of generating captions by learning from pairs of images and their corresponding textual descriptions. This process fine-tuned BERT's ability to produce meaningful and coherent sentences based on the visual content represented by the images.

Ensuring optimal performance on numerous natural language processing tasks, the multi-headed attention model BERT was among the first to employ the encoder portion of the initial transformer architecture. K Encoder layers coupled in series, each with G attention heads utilised in parallel, make up this model.

Since the attention elements of BERT $_{\text{BASE}}$ $(K = 12, and\ G = 12)$ are the portion of the model that contains the topological data that are of interest in analysing, concentrate on this. A matrix W of size $m \times c$ (where $c = 512$) containing vector illustrations of the m tokens in the input phrase serves as the attention head's input. Its output is a novel representation W^{out} of size $m \times c$ with $c' < c$ $(c' = 64)$ using the following Equation 19.

$$W^{out} = X^{attn}.(W.X^{U}) with\ X^{attn} = softmax(\frac{(W.X^{R}).(W.X^{L})^{S}}{\sqrt{c}} \qquad [19]$$

Here, the token embedding is projected onto a lower-dimensional vector space via trainable vectors $X^R,\ XL, X^U \in \mathbb{R}^{c \times c'}$, and the softmax is applied over the second dimension called the attention matrix, the matrix X^{attn} is a $m \times m$-dimensional grid.

When calculating the novel description of the token at location j, its entries x_{ji}^{attn} can be understood as the attention provided to the token at location j. They are non-negative, and each token's attention scores add up to one; therefore, for every $\sum_{i=1}^{m} x_{ji}^{attn} = 1 \quad j = 1, \ldots., m.$

Display Output (Images with Captions)

- The last stage involves manually entering images from the dataset as inputs from the user, and our model creates descriptions for each of these images.
- The final product will consist of image-caption pairs, where each image has a detailed caption produced by the optimised BERT model. This enables users to evaluate the model's efficacy in producing pertinent and cohesive descriptions based on the input images both visually and textually.
- Additionally, we create captions for arbitrary images found online. Figure 6 shows the output of images with captions.

Figure 6. Output of an image with a caption

RESULTS AND DISCUSSION

The experimental environment and setup, as well as the effectiveness of the suggested approach displayed in Table 2, are covered in this section and compared with the existing approaches, which are deep convolutional neural networks and recurrent neural networks with attention (CNN-RNN-Att) (Barati et al., 2023), Residual Network with 50 layers (ResNet-50) (Khaing et al., n.d.).Table 3 shows the overall performance.

Table 2
Experimental setup

Experimental Setup	Details
Model	SST-CNN with dense GRU neural net and fine-tuned BERT
Task	Image Caption Generation
Dataset	Collected from Kaggle sources
Hardware	Laptop running Windows 11
RAM	16 GB
Software Environment	Python 3.10.1
Evaluation Metrics	BLEU-1, METEOR, CIDER, ROUGE-L

Table 3
Overall result comparison

Methods	BLEU-1	METEOR	ROUGE-L	CIDER
DCNN-RNN-Att [16]	0.786	0.291	0.579	0.713
ResNet50 [17]	0.576	0.207	0.213	0.454
SST-CNN [Proposed]	1	0.99	0.99	1

BLEU-1

To determine how well the generated text matches descriptions made by humans, it computes the ratio of matching terms between the created caption and reference captions. Figure 7 shows the comparative evaluation of BLEU-1.

While the existing methods DCNN-RNN-Att and ResNet-50 achieved 0.786 and 0.576, respectively, our proposed SST-CNN methodology achieved (1). The findings demonstrate that our suggested approach outperforms existing methods substantially.

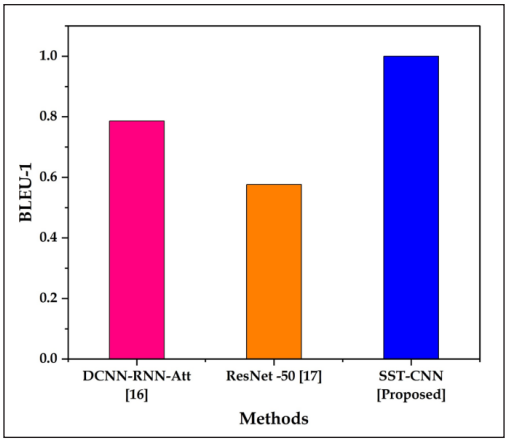


Figure 7. Output of BLEU-1

METEOR

METEOR scores are calculated based on word and phrase matches in the produced and reference captions to provide a more accurate and relevant evaluation. Figure 8 shows the comparison assessment of METEOR.

While the existing methods DCNN-RNN-Att and ResNet-50 achieved 0.291 and 0.207, respectively, our proposed SST-CNN methodology achieved 0.99. The results show that our proposed strategy performs significantly better than current approaches.

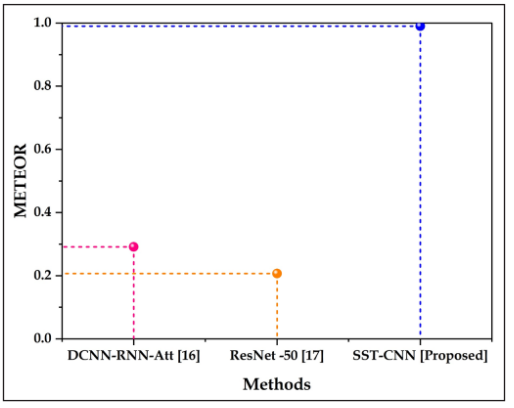


Figure 8. Output of METEOR

CIDER

It assesses the consistency and importance of words and phrases, emphasising details and environment to determine caption descriptive quality. Figure 9 shows the comparison of CIDER.

While the existing methods DCNN-RNN-Att and ResNet-50 achieved 0.713 and 0.454, respectively, our proposed SST-CNN methodology achieved (1). The outcomes demonstrate that, compared to existing methods, our suggested solution works noticeably better.

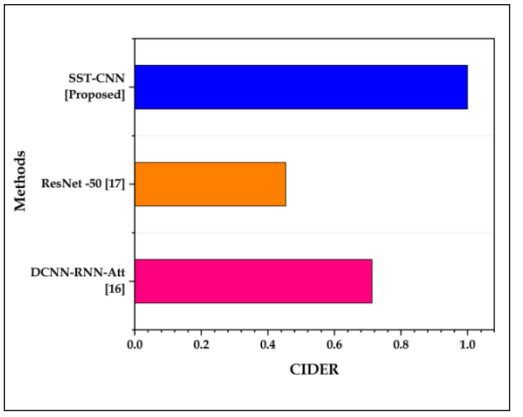


Figure 9. Output of CIDER

ROUGE-L

It represents the quality and relevancy of captions by taking into account both content overlap and sequence order, resulting in meaningful and coherent descriptions. The ROUGE-L score is displayed in Figure 10.

While the existing methods DCNN-RNN-Att and ResNet-50 achieved 0.579 and 0.213, respectively, our proposed SST-CNN methodology achieved 0.99 for Image caption generation. The outcomes demonstrate that, compared to existing methods, our suggested solution works substantially well.

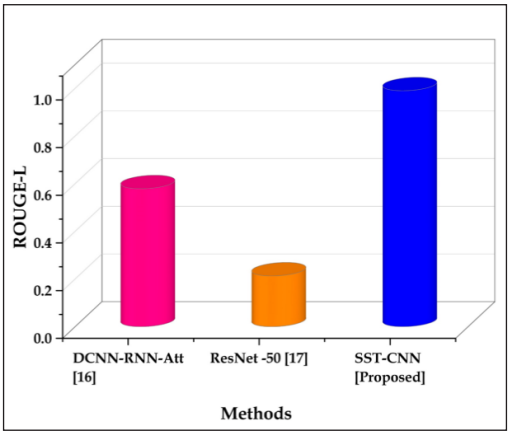


Figure 10. Output of ROUGE-L

Figure 11 shows how similar metrics BLEU, METEOR, ROUGE-L, and CIDEr have performed in various models. The suggested SST-CNN model has a higher score than that of DCNN-RNN-Att and ResNet50, which illustrates its usefulness in caption generation tasks. The drastic rise in performance is indicative of enhanced semantic comprehension and contextual correspondence. These results are consistent

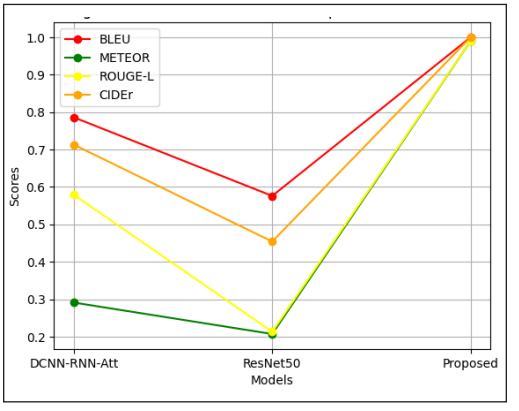


Figure 11. Comparison of performance metrics through the use of a line chart

with the previous research that highlights the importance of hybrid deep learning models (Alkhaldi et al., 2025).

The bar chart in Figure 12 gives a comparative evaluation of metrics grouping of the models, with the proposed approach prevailing. All metrics show a significant change, especially METEOR and CIDEr, suggesting the accuracy of language and contextual understanding. The visualisation is useful in highlighting the performance difference between the conventional CNN-RNN models and the hybrid architecture. The same enhancement has been experienced when using attention-based captioning systems (Xu et al., 2026).

This scatter plot as shown in Figure 13 will be used to visualise the distribution of the metric scores of each of the models, allowing one to see the dispersion of performance. The points on the proposed model are concentrated at higher values, which means stable and better results on all the metrics. Conversely, the variation and poor consistency of performance are observed in the baseline models. These patterns of distribution are typical of deep learning capturing systems (Dsouza et al., 2026).

The box plot as shown in Figure 14 is the statistical data of performance measures such as median, quartiles and variability. The model proposed has greater median values with less variation, which shows consistent performance. The spreads in the baseline models are broader, indicating instability in the quality of prediction. This statistical image lends credence to the strength of the SST-CNN-based structure (Dsouza et al., 2026).

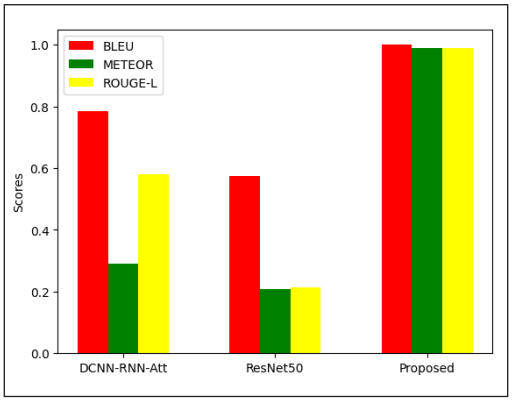


Figure 12. Relative analysis of metrics with bar chart

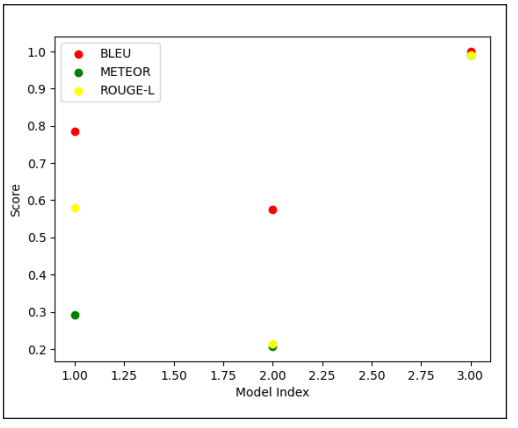


Figure 13. Scatter plot of model performance metrics

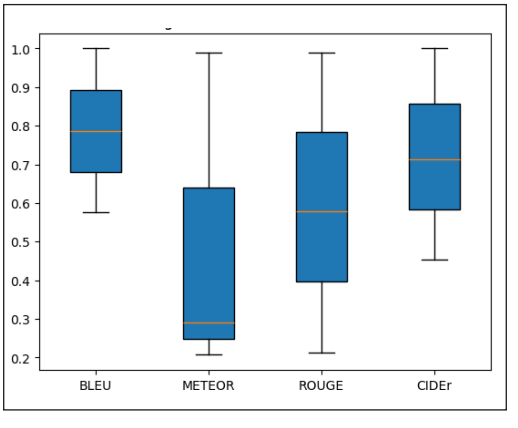


Figure 14. Evaluation metrics distribution with box and whisker plot

The heatmap, as shown in Figure 15, is an easy visualisation of the intensities of metrics across models in the form of colour gradients. The more intense areas are associated with improved performance, which is a clear indication of the superiority of the proposed method. The equal values of all metrics show balanced performance. A heatmap is commonly employed to assess visual tasks on deep learning models (Dsouza et al., 2026).

The radar chart as shown in Figure 16 is a multi-dimensional comparison of evaluation metrics which demonstrate the overall performance profile of each model. The suggested model is almost a complete polygon, which means that all metrics are of equal scores. On the contrary, baseline models exhibit irregular shapes, indicating uneven performance. The graph is a good example of the balanced transformation that the suggested architecture can bring (Dsouza et al., 2026).

This value depicts variability and confidence levels of the BLEU scores of the models as shown in Figure 17. The model proposed has a low error margin, implying that it is very reliable and stable in its predictions. Baseline models exhibit bigger variations, indicating inconsistency in generated captions. The error bar analysis is essential in verifying model strength in experimental studies(Dsouza et al., 2026).

The bubble chart as shown in Figure 18 represents the BLEU scores and the relative

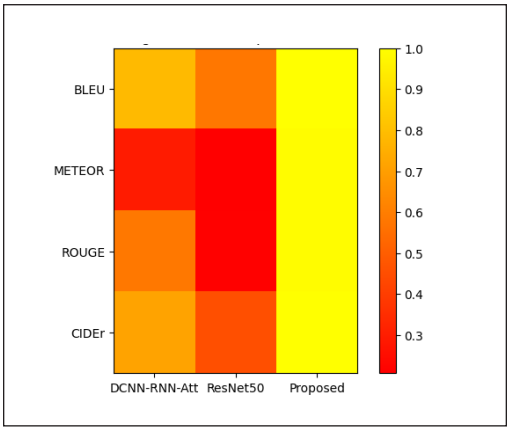


Figure 15. Heatmap view of metric scores

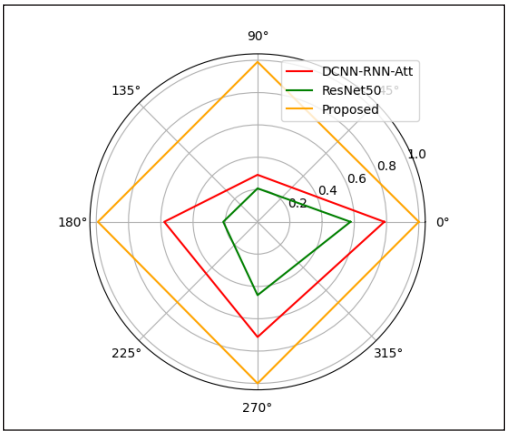


Figure 16. Multi-metric multi-measure radar chart

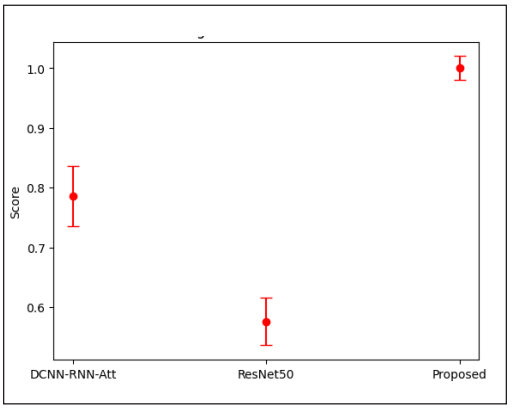


Figure 17. BLEU scores analysis as an error bar

importance of the score in terms of different sizes of the bubble. The largest bubble with the highest score represents the proposed model and highlights that it is the best. This visualisation is a combination of magnitude and value, providing a total comparison. These representations come in handy in emphasising predominant models in comparative studies (Dsouza et al., 2026).

The Gantt chart, as shown in Figure 19, is used to describe the consecutive steps of the suggested methodology, such as preprocessing, feature extraction, training, and evaluation. It gives a definite chronological insight into the workflow and computational pipeline. This formal representation increases the understandability of the system design. In complicated deep learning models, workflow visualisation is critical (Dsouza et al., 2026).

The flow of computational resources in the Sankey diagram is illustrated by the various steps in the model, including feature extraction as shown in Figure 20, GRU processing, and BERT refinement. Flow thickness is the contribution to the computation made by each stage. This diagram makes the resource-consuming aspects of the architecture obvious. Such analysis is important for optimising deep learning pipelines (Dsouza et al., 2026).

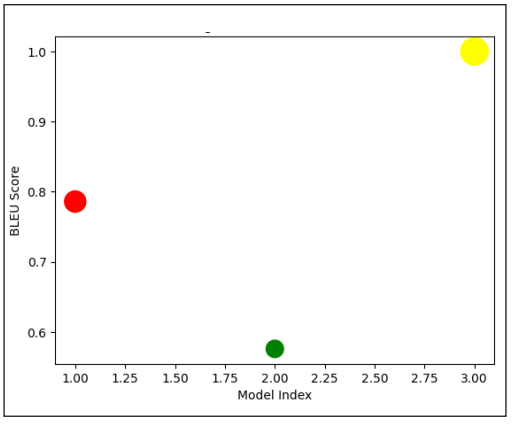


Figure 18. Bubble chart of distribution of BLEU score

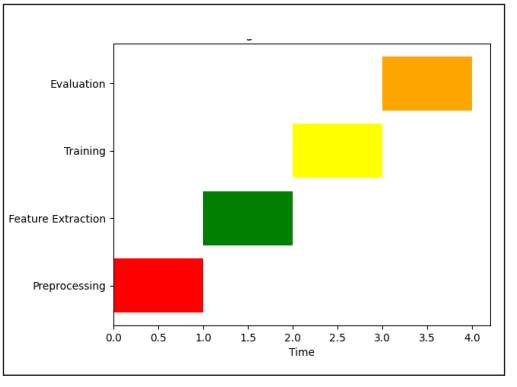


Figure 19. Workflow representation Gantt chart

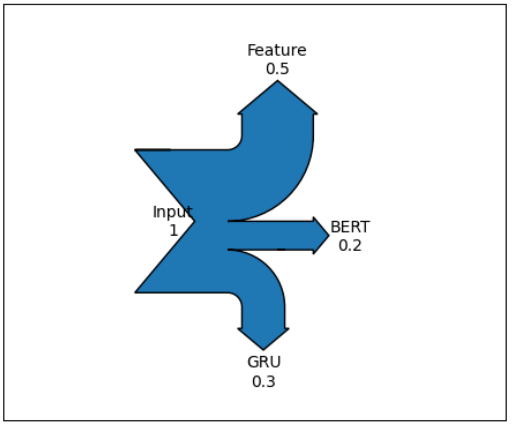


Figure 20. Computational flow Sankey diagram

CONCLUSION

In the present work, we have presented a way to generate captions for images using the SST-CNN system. We acquired a dataset from Kaggle, which we processed post-acquisition through both Flexiframe Filter text and image consistency as well as the Bright-Contrast method to greatly boost images for augmentation. The SST-CNN model is very effective in feature extraction and encoding, with the help of the VGG16 model. Decoder using stochastic k-sampling with a Dense Neural Network and GRU, and we used caption generation to fine-tune the BERT method. Evaluation of the model was done using the BLEU-1 (1), METEOR (0.99), ROUGE-L (0.99), and CIDER (1) to ensure the model's capacity to generate accurate and quality captions. It may encounter obstacles that could impair efficiency and scalability in practical applications, such as increased processing complexity, noise sensitivity, challenges managing a variety of image settings, limited generalisability across different datasets, and possible inefficiencies while processing complicated or high-resolution images; these are some of the drawbacks that may be encountered. Subsequent endeavours will focus on refining computing efficiency, strengthening noise resilience, and expanding flexibility in various image scenarios. Furthermore, using multimodal inputs and sophisticated attention mechanisms could improve the relevance and accuracy of captions even more. Further research is needed in the area of enhancing the efficiency and scalability of computations, so that image captioning in resource-constrained environments can be realised in real time. Moreover, further integration of multimodal architectures and attention-based transformers can be used to improve the accuracy of captions and contextual retrieval. Generalising the model to various datasets and multilingual captioning cases will enhance generalisation and practical usefulness.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

REFERENCES

Afyouni, I., Azhar, I., & Elnagar, A. (2021). AraCap: A hybrid deep learning architecture for Arabic image captioning. *Procedia Computer Science, 189*, 382-389. https://doi.org/10.1016/j.procs.2021.05.108

Al Duhayyim, M., Alazwari, S., Mengash, H. A., Marzouk, R., Alzahrani, J. S., Mahgoub, H., Althukair, F., & Salama, A. S. (2022). Metaheuristics optimisation with deep learning-enabled automated image captioning system. *Applied Sciences, 12*(15), Article 7724. https://doi.org/10.3390/app12157724

Alkhaldi, T. M., Asiri, M. M., Alzahrani, F., & Sharif, M. M. (2025). Fusion of deep transfer learning models with Gannet optimisation algorithm for an advanced image captioning system for visual disabilities. *Scientific Reports, 15*, Article 40446. https://doi.org/10.1038/s41598-025-24171-9

Al-Malla, M. A., Jafar, A., & Ghneim, N. (2022). Image captioning model using attention and object features to mimic human image understanding. *Journal of Big Data, 9*, Article 20. https://doi.org/10.1186/s40537-022-00571-w

Barati, A., Farsi, H., & Mohamadzadeh, S. (2023). Integration of the latent variable knowledge into deep image captioning with Bayesian modelling. *IET Image Processing, 17*(7), 2256-2271. https://doi.org/10.1049/ipr2.12790

Beddiar, R., & Oussalah, M. (2023). Explainability in medical image captioning. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, & W. Samek (Eds.), *xxAI: Beyond explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (pp. 239-261). Academic Press. https://doi.org/10.1016/B978-0-32-396098-4.00018-1

Castro, R., Pineda, I., Lim, W., & Morocho-Cayamcela, M. E. (2022). Deep learning approaches based on transformer architectures for image captioning tasks. IEEE Access, 10, 33679-33694. https://doi.org/10.1109/ACCESS.2022.3161428

Degadwala, S., Vyas, D., Biswas, H., Chakraborty, U., & Saha, S. (2021). Image captioning using Inception V3 transfer learning model. In 2021 6th International Conference on Communication and Electronics Systems (ICCES) (pp. 1103-1108). IEEE. https://doi.org/10.1109/ICCES51350.2021.9489111

Devi, P. R., Thrivikraman, V., Kashyap, D., & Shylaja, S. S. (2020). Image captioning using reinforcement learning with BLUDEr optimisation. Pattern Recognition and Image Analysis, 30(4), 607-613. https://doi.org/10.1134/S1054661820040094

DSouza, R., & Sen, S. (2026). A SwinBERT framework with temporal windowed cross attention for video captioning for visual language intelligence. Discover Applied Sciences.

Hossain, M. Z., Sohel, F., Shiratuddin, M. F., Laga, H., & Bennamoun, M. (2021). Text-to-image synthesis for improved image captioning. IEEE Access, 9, 64918-64928. https://doi.org/10.1109/ACCESS.2021.3075579

Khaing, P. P., San, M., & Aung, M. M. (n.d.). Analysis on residual network models for image description generation. Journal of Research and Applications, 1(1), 208-213.

Mahalakshmi, P., & Fatima, N. S. (2022). Summarisation of text and image captioning in information retrieval using deep learning techniques. *IEEE Access, 10*, 18289-18297. https://doi.org/10.1109/ACCESS.2022.3150414

Puscasiu, A., Fanca, A., Gota, D. I., & Valean, H. (2020). Automated image captioning. In *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)* (pp. 1-6). IEEE. https://doi.org/10.1109/AQTR49680.2020.9129930

Shen, X., Liu, B., Zhou, Y., Zhao, J., & Liu, M. (2020). Remote sensing image captioning via variational autoencoder and reinforcement learning. *Knowledge-Based Systems, 203*, Article 105920. https://doi.org/10.1016/j.knosys.2020.105920

Wang, Y., Xu, N., Liu, A. A., Li, W., & Zhang, Y. (2022). High-order interaction learning for image captioning. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(7), 4417-4430. https://doi.org/10.1109/TCSVT.2021.3121062

Varma, S., & Peter, J. D. (2026). Enhanced transformer model for video caption generation. Expert Systems, 43(6), e13392.

Xian, T., Li, Z., Tang, Z., & Ma, H. (2022). Adaptive path selection for dynamic image captioning. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(9), 5762-5775. https://doi.org/10.1109/TCSVT.2022.3155795

Xu, X., Wu, X., Zhang, H., & Ma, K. (2026). Medical Image Captioning with Swin Transformer and ClinicalBERT via Gated Multimodal Fusion. Applied Soft Computing, 115731.

Yan, C., Hao, Y., Li, L., Yin, J., Liu, A., Mao, Z., Chen, Z., & Gao, X. (2022). Task-adaptive attention for image captioning. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(1), 43-51. https://doi.org/10.1109/TCSVT.2021.3067449

Pertanika J. Sci. & Technol. 34 (S1): 111 - 131 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Weighted Bi-directional Feature Pyramid Network and Segment Anything Model (SAM) Specific Knowledge Injection for Fabric Defect Detection

Suresh Kumar* and J. Avanija

Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Andhra Pradesh, India

ABSTRACT

The quality control in the textile manufacturing industry involves identifying fabric defects. The differences in cloth texture and the limited number of damaged samples are also major problems when it comes to the correct identification of faults in fabrics. In most applications, visual inspection is still a vital component of the quality control process. The ability to check surface defects is an essential part of the quality control methods of fabrics and textiles. A fault can be described as whatever interrupts the pattern of fabric repetition. To solve this problem, this paper proposes a supervised learning approach to the detection of fabric flaws. The main purpose of this study will be to design a rotation-invariant fabric defect detection plan that will attain a high detection rate related to current techniques. The study introduces a novel framework for identifying defects in fabrics by fusing a Weighted Bi-directional Feature Pyramid Network (BiFPN) and the Segment Anything Model (SAM), which have been produced via Specific Knowledge Injection. The Weighted BiFPN enhances the fusion of multi-scale features by assigning dynamic weights to the features at various levels, which makes it possible to detect both minor and large-scale flaws in fabric. At the same time, the Segment Anything Model (SAM) is a sophisticated segmentation model that is fine-tuned with knowledge injection to adapt its generic segmentation abilities to the peculiarities of fabric textures and defects. The combination will guarantee accurate localisation and categorisation of the flaws, even under difficult conditions that may include changes in illumination, texture, and type of defects. Experiments carried out on standard textile datasets prove that the suggested BiFPN-SAM framework is considerably more efficient in terms of accuracy, Intersection over Union (IoU), and performance. The fact that the hybrid

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 20 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.06

E-mail addresses:
ogsureshkumar45@gmail.com (Suresh Kumar)
avans75@yahoo.co.in (J.Avanija)
* Corresponding author

e-ISSN: 2231-8526

method can generalise to different faults also underscores its usefulness in textile quality control systems through automated means.

Keywords: Control, deep learning, fabric defect detection, multi-scale feature fusion, segment anything model, specific knowledge injection, textile quality, weighted BiFPN

INTRODUCTION

The purpose of fabric defect detection is to identify and detect any surface faults on fabrics due to machine malfunctions or operational failures. To modern fabric organisations, it forms one of the essential ways of reducing expenses and enhancing the competitiveness of the products. Human eye investigation is an outdated detection technology with a success rate of 60-75% (Zhou et al.,2024). Nevertheless, this approach does not guarantee the results of the tests to be unbiased or consistent. The limitations can be overcome, the expenses minimised, and a 90 per cent detection success rate is realised with an automatic fabric inspection system. Hence, it is necessary to enhance a quick, precise and non-destructive method of detecting flaws on fabric (Zhang et al., 2022).

Various production imperfections that mostly involve machine breakdown cause the fabric to be thrown away as waste and cause financial losses to Pakistani textile manufacturers (Zhou et al.,2025). They need to be inspected to ascertain whether the fabric manufactured is of the market standards. This process, especially in the developing world, is done manually and involves a great number of qualified labourers. Manual examination is also time-consuming and error prone due to the human factor of fatigue. Thus, it needs an automated system (Ke et al., 2022). Detection methods based on deep learning are popular nowadays due to the development of hardware technology. These procedures are now being applied in a form of real-life scenarios, e.g. detecting defects in fabric (Zhou et al., 2024).

Fabric defect identification has been widely studied, whereby most of the studies have utilised datasets having fabric images placed carefully under a controlled environment. However, datasets pose several difficulties in real-time production applications. Images are often neither well-positioned nor of high quality (Chen et al.,2022; Zhou et al.,2024). Moreover, changes in blurriness, noise, and lighting intensity have an impact on visual perception. There are two primary categories of research in this area: identifying flaws in printed fabrics and identifying flaws in plain fabrics. Additionally, printed textiles can be classified as either regularly printed or irregularly printed. This distinction arises because differentiating between patterns and defects becomes challenging with a high diversity of prints. As a result, distinct approaches are required for each scenario. Models that can simultaneously recognise defects in plain, regularly printed, and irregularly printed textiles have received relatively little attention (Jing et al., 2022; Li et al.,2022).

Research of Our Work

The research in this study that presents an alternative detection method for fabrics by relying on the combination of a WB-FPN and the Segment Anything Model (SAM), which is created with the help of targeted knowledge input. The WB-FPN supports effective acquisition and refinement of multi-scale features, whereas SAM secures effective segmentation of various patterns and textures of fabrics. The structure is better at providing accuracy and flexibility in detecting complex fabric structures by including domain-specific knowledge. This mixed approach is a great advancement in the detection of fabrics with a positive outlook on its application in quality control, automation, and the study of textiles.

Motivation of This Research

This research is driven by the fact that the demand for proper and efficient fabric detection methods is increasing in the context of textile, fashion and manufacturing industries. The current methods usually have limitations where fine-grained pattern recognition, different textures, and different fabric structures may be a problem. Through a BiFPN and by combining the Segment Anything Model (SAM) with injecting domain-specific knowledge, this research is aimed at creating detection accuracy and flexibility. The proposed technique will integrate the enhanced hierarchical processing of features with contextual information, overcoming the current shortcomings and providing a further perspective on more powerful fabric defect detection frameworks in practical scenarios.

Our Work Contributions

The Principal Contributions of Our Work Are:

- To design a new fabric defect detection structure, which is a BiFPN integrated with the Segment Anything Model (SAM) and complemented with Specific Knowledge Injection.
- The Weighted BiFPN is developed based on multi-scale feature fusion through transmission-adaptive weights to features at different levels, which allows it to detect the presence of small and large-scale flaws in the fabric structure.
- The experiments conducted with benchmark textile datasets show that the given framework BiFPN-SAM performs significantly better than the current state-of-the-art models.

The rest of the paper is organised in the following way: Section 2 is devoted to the in-depth literature survey, whereas Section 3 presents the description of the proposed technique. The results and the discussion are provided in Section 4, and Section 5 finalises the article by providing the future work outline.

Survey

This survey is a valuable research area that also covers several research fields such as computer vision, materials science and textile engineering. The proposed survey will focus on the developments in automated technologies and methods of identifying, classifying and analysing fabrics. Applications in fabric detection have been applied in both the manufacturing of textile quality control in the design of intelligent fashion and in forensics, as well as in innovative and automated industries. The purpose of this survey was to give a comprehensive overview of the sector in terms of existing trends, challenges, and possible future directions of the fabric detection industry. Dalmini et al. (2022) created a real-time machine imaging system to detect defects in functional textiles, which can be used in the quality control mechanisms of the textile industry. Raw images are first divided into smaller, fitting sizes by way of preparing the data. Spatial filters are commonly used in de-noising. Moreover, there is the YOLOv4 system that is applied to categorise textiles. The described speed of detection is 34 fps, and the estimated time of processing is 21.4 ms per frame. (Dlamini et al., 2022; Garg et al., 2020) suggested a method to identify fabric defects which uses a two-layer neural network method and extracts data of the raw image patches. First, the presence of defects is detected with the help of a convolutional neural network (CNN).

Then, fault localisation is performed with the help of a multi-layer CNN consisting of convolution + ReLU, max pooling, SoftMax layers, and fully connected + ReLU. Li et al. (2020) proposed a way of detecting defects in patternless textiles through the co-occurrence of grey-level operators combined with uniform MDBP operators. A comparison is made using a detection threshold after a multi-directional texture feature matrix has been taken out of the entire fabric image into a feature matrix. Defects are detected with the help of these thresholds. Multiple grayscale images of textiles with different properties may be supported using this approach. Kang et al. (2024) introduced a technique that is based on YOLOv7-tiny and was used to detect flaws in solid-coloured circular knitted textiles. The processing load of the YOLOv7-tiny backbone network is also lower, and its feature extraction functions are improved by using the SPD-Conv module.

This method also integrates spatial and channel attention techniques to create a hybrid awareness module that is then combined into the YOLOv7-tiny system. As a result, the network achieves more accurate spatial information and improved image segmentation accuracy. Mao et al. (2025) presented an enhanced fabric detection method based on the YOLOv8 architecture. The approach integrated an attention mechanism and feature extraction modules to effectively capture multiscale features, which detects small defects accurately. (Ke et al., 2022) The results of defect identification are then obtained by adjusting the feature weights using the Adaptive Fabric Feature Clustering (AFFC) algorithm.

To discover high-resolution saliency targets, Xie et al. (2022) created PGnet. As the backbone, they used parallel connections between ResNet and Swin Transformer to perform feature fusion during the decoding stage. Nevertheless, the SBlock submodule's patch fusion sampling process results in a large loss of information, which leaves inadequate features extracted during decoding. When examining low-resolution fabric defect samples, this restriction results in a higher rate of misdetection.

A two-step technique for detecting fabric defects using the Pyramid Histogram of Edge Orientation Gradients (PHOG) was presented by (Cuifang et al., 2020). Support vector machines (SVMs) were also applied in the classification stage. This technique involves the combination of machine learning and statistical tools to study cloth texture.

Research Gap

Although there have been enormous improvements in the detection of fabrics, the existing technologies now find it hard to resolve the inherent limitations, such as detecting fabrics that look alike, differences in lighting and texture, detecting complex patterns in the real world, etc. Most methods are very dependent on controlled conditions or certain types of fabrics, and as such, they cannot be generalised. Although the introduction of deep learning models has improved accuracy, it may require large volumes of labelled data and a large amount of computing power, which are not feasible at all times. This is in consideration of the fact that more durable, efficient, and versatile fabric detection systems are required that can provide dependable performance in various environments and fabric types (Li et al., 2022).

Problem Identification of Existing System

- The current fabric detection systems are usually experiencing problems in their ability to detect complex patterns, texture and changes in the type of fabrics.
- A lot of systems cannot identify minor flaws, including a small tear, loose thread or differences in the weave, resulting in quality control problems.
- Many of the systems still use manual intervention, and this makes them less efficient and highly prone to human error.
- High implementation costs and complicated installation procedures often characterise advanced systems, and this makes them unavailable to small and medium-sized businesses (Chen et al.,2022).

Proposed System

A new fabric defect detection framework that is presented in this study involves a BiFPN with the Segment Anything Model (SAM) that is supplemented with Specific Knowledge Injection.

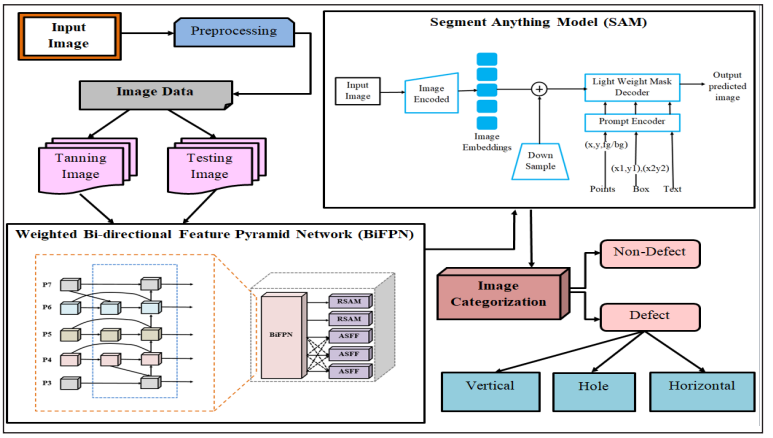


Figure 1. Architecture of BiFPN-SAM

The Weighted BiFPN uses adaptive weights for features of different levels to provide better multi-scale feature fusion to detect small- and large-scale fabric flaws. In the meantime, the Segment Anything Model (SAM) is an advanced segmentation model that is refined by introducing certain knowledge to shape its general segmentation abilities to meet the specifics of the fabric textures and defects. The combination guarantees the accurate localisation and classification of flaws even in unfavourable situations when there is a change in illumination, texture and the type of defect. The architecture of BiFPN-SAM is presented in Figure 1.

Dataset

The three types of defective photos in our fabric defect database, vertical, horizontal, and hole, along with the corresponding mask images for each defective sample, are displayed in Figure 2. The raw photos are in the 'caught' folder (Khodier et al., 2022). The image size is 640×360.

Figure 2. Sample image of fabric defect dataset

Preprocessing

Initially, the samples were reduced in size, and the RGB images were further transformed into greyscale for processing and model training. Contrast-Limited Adaptive Histogram Equalisation (CLAHE) was utilised to increase the contrast of the images (Roy et al., 2021). Before computing the cumulative distribution function (CDF), CLAHE clips the histogram at a predetermined value, preventing noise amplification in contrast to traditional Adaptive Histogram Equalisation (AHE). Two important factors that can be used to modify the quality of image enhancement are Block Size (BS) and Clip Limit (CL). $M \times N$ Tiles are created by first segmenting the image into contextual areas that do not overlap. Each contextual region's contrast-limited histogram is then calculated using the CL value as mentioned in Equation 1:

$$N_{avg} = \left(N_{px} \times N_{py} \right) / N_{gray} \quad [1]$$

Where N_{avg} is the average number of pixels, N_{px} and N_{py} are the region's number of pixels in the X and Y directions, respectively, and N_{gray} is the grey level number of the contextual region. The actual clip limit can be estimated as specified in Equation 2:

$$N_{CL} = N_{clip} \times N_{avg} \quad [2]$$

Wherever, N_{clip} is the standardised clip limit in the range of [0, 1]. If the number of pixels is greater than the N_{CL} , the pixels are clipped. Distributing the fraction that surpasses the clip limit throughout all histogram bins is the preferred method. The procedure is repeated if the redistribution results in some values exceeding the clip limit once more until the excess is minimal (Dlamini et al., 2022).

Weighted Bi-directional Feature Pyramid Network (BiFPN)

The BiFPN is an advanced version of the Feature Pyramid Network (FPN) designed to enhance feature fusion by assigning learnable weights to different features during the upsampling and downsampling processes. The approach effectively combines low-level fine-grained details with high-level semantic information (Huang et al., 2024; Zhou et al., 2022). This bi-directional approach ensures better propagation of contextual and spatial information, making it particularly useful for tasks requiring fine-grained detection, such as fabric defect detection. Figure 3 illustrates the architecture of the BiFPN model.

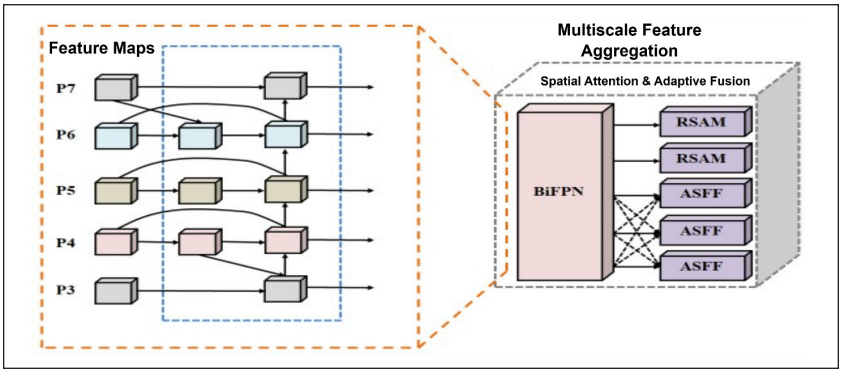


Figure 3. Architecture of the BiFPN model

Specific Knowledge Injection involves embedding domain-specific knowledge, such as patterns, textures, or structural properties unique to fabric materials, into the network. This is accomplished using pre-trained models or auxiliary inputs that guide the network in learning fabric-specific representations. When combined with BiFPN, this approach improves the network's capability to detect even subtle fabric defects.

Bi-directional Feature Fusion

BiFPN uses learnable weights for each input feature. For a given set of feature maps $F_1, F_2, ..., F_n$, the fusion equation can be expressed as in Equation 3:

$$F_{output} = \sum_{i=1}^{n} w_i \cdot F_i \quad [3]$$

Where F_i are input feature maps from various pyramid levels, w_i is a learnable weight for each input feature map, constrained such that $\sum w_i = 1$ (softmax normalisation).

Feature Scaling and Normalisation

To align spatial dimensions before fusion, Equation 4 is applied:

$$F_i^{'} = Re\,size(F_i, t\,arg\,e\,t\_size) \quad [4]$$

Bi-directional Pathway

Top-down pathway for upsampling is mentioned in Equation 5:

$$F_{td}^{l} = Upsample\left(F^{l+1}\right) + w^{td} \cdot F^{l} \quad [5]$$

Bottom-up pathway for downsampling is specified in Equation 6:

$$F_{bu}^{l} = Downsample\left(F^{l-1}\right) + w^{bu} \cdot F^{l} \quad [6]$$

Specific Knowledge Injection

Domain knowledge K is integrated as an auxiliary input or additional loss function as in Equation 7:

$$F_{enhanced} = F_{Output} + \alpha \cdot \phi(K) \quad [7]$$

Where $\phi(K)$ are knowledge embedding functions, α is a weight that controls the impact of specific knowledge.

Loss Function for Defect Detection

A combined loss function is often employed to balance detection accuracy and domain adaptation, as specified in Equation 8.

$$L = Ldomain_{det} \quad [8]$$

Where, L_{det} is the loss for fabric defect detection, L_{domain} is the loss for knowledge consistency or domain alignment, β is balancing the weights of the two loss terms (Jing et al., 2022).

These components together enhance the BiFPN's performance for fabric defect detection by leveraging both structural network improvements and domain-specific knowledge.

Segment Anything Model (SAM)

The SAM represented in Figure 4 is a state-of-the-art deep learning framework designed to perform image segmentation tasks with high generalisability. It uses a combination of Vision Transformers (ViTs) and a promptable interface to segment any object in an image with minimal input. However, for specialised applications such as fabric defect detection or segmentation, SAM's performance can be enhanced by incorporating domain-specific knowledge. Figure 4 illustrates the architecture of the SAM model.

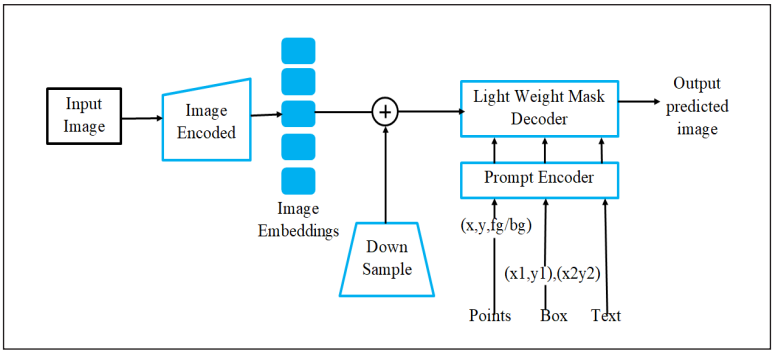


Figure 4. Architecture of Segment Anything Model (SAM)

SAM Segmentation Framework

SAM uses a vision encoder E to extract image features F, and a prompt encoder P to process inputs like points, bounding boxes, or masks as in Equations 9 and 10:

$$F = E(I) \quad [9]$$

$$M = D\big(F, P(p)\big) \quad [10]$$

Wherever I is the input image, D is the decoder, $P(p)$ is the encoded prompt, and M is the segmented mask.

Specific Knowledge Injection

Injecting specific knowledge can be achieved by modifying the encoder or the prompt input using Equation 11.

$$F^{'} = E^{'}(I, K)$$

Here, K is the domain-specific knowledge (e.g., defect patterns, texture features). This can involve pretraining $E^{'}$ on a fabric dataset or appending K to feature embeddings.

Loss Function for Domain Adaptation

To adapt SAM to fabric detection, a customised loss function L can combine segmentation loss with domain-specific penalties as specified in Equation 12:

$$L = L_{seg} + \lambda L_{domain} \quad [12]$$

- L_{seg}: Standard segmentation loss (e.g., Dice Loss, Cross-Entropy Loss)
- L_{domain}: Loss capturing fabric-specific properties (e.g., texture continuity, defect localisation)
- λ: Weight balancing the two terms.

Prompt Augmentation for Fabric Detection

Specific prompts P_{fabric} can encode domain-relevant information using Equation 13:

$$\mathbf{p}_{\text{fabric}} = \text{encode(fabric-particular patterns user comments cloth)} \quad [13]$$

Fusion of Features and Prior Knowledge

In case of feature fusion, we integrate the features of SAM of fabric-describing nature using Equation 14.

$$F^{"} = \alpha F + (1 - \alpha) F_K \quad [14]$$

- F_K: Features extracted from fabric-specific knowledge
- α: Weight controls the contribution of each feature.

The model incorporates the generalisation capabilities of SAM and induced domain knowledge to arrive at the model, resulting in improved performance on complex textures and subtle defects.

Advantages of Proposed Method

- Uses bi-directional feature aggregation that is weighted to improve feature representation. for fabric textures.
- Utilises SAM's general segmentation capabilities, specifically tailored for fabric-related applications.
- Optimised for high performance with minimal overhead compared to traditional models.
- Ensures improved feature propagation and context understanding across scales (Khodier et al., 2022).

RESULT AND DISCUSSIONS

Experimental Setup

The following is the configuration of the experiment's environment: Microsoft Windows 10 as the operating system, CUDA 11.5 and CUDNN 11.1 as GPU-accelerated libraries,

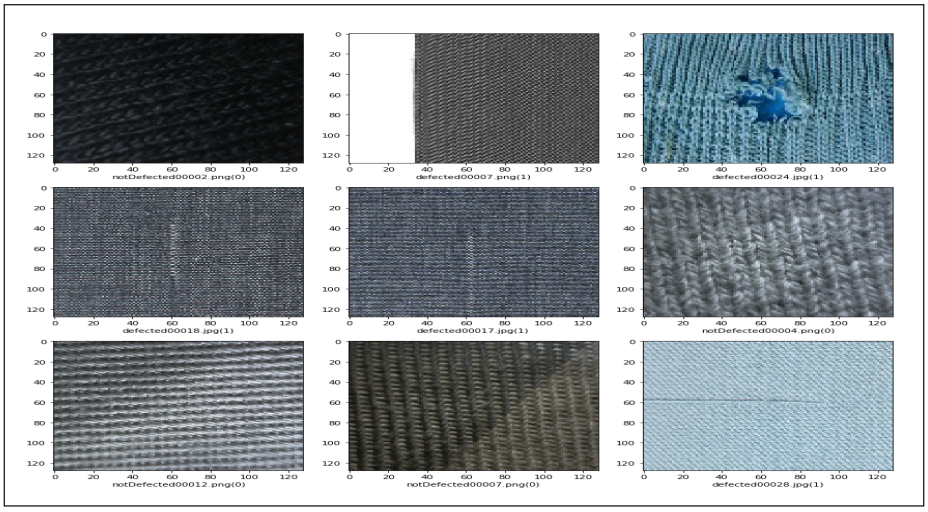


Figure 5. Final prediction images for the dataset

PyCharm as the integrated development environment, Python 3.6 as the programming language, TensorFlow-GPU 1.14.0 and Keras 2.2.5 as the deep learning frameworks, and 16 GB of RAM, AMD R7-4800H CPU, and Nvidia GeForce RTX 2060 6 GB GPU. Figure 5 shows the final prediction images for the dataset.

Performance Metrics

DSR (Detection Success Rate) is a performance metric that measures a model's accuracy in identifying the correct class. It is determined by dividing the total number of cases by the proportion of correctly detected occurrences, as mentioned in Equation 15.

$$DSR = \frac{TruePositives}{TruePositives\ +FalseNegatives} \times 100 \quad [15]$$

A metric called PPV (Positive Predictive Value) gauges how well a model forecasts the future, as in Equation 16. It shows the proportion of actual positive results out of all the positive forecasts.

$$PPV = \frac{TP}{TP+FP} \quad [16]$$

NPV (Negative Predictive Value) is a metric utilised to calculate the percentage of true negative results among a model's negative predictions, as in Equation 17. It helps assess the model's accuracy in recognising negative cases.

$$NPV = \frac{TN}{TN+FN} \quad [17]$$

The F-measure is a statistic that provides a fair assessment of a model's performance by combining recall and precision, as in Equation 18. When there is an imbalance in the distribution of classes, it is very helpful. The following formula is utilised to determine the F-measure:

$$F - Measure = 2 \times \frac{Pr\,ecision \times Re\,call}{Pr\,ecision + Re\,call} \quad [18]$$

The accuracy of a model is expressed as the number of correct forecasts divided by the number of predictions, as mentioned in Equation 19. It can be calculated with the following Equation 19 (Roy et al., 2021):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad [19]$$

Comparative Methods

MTF-Net (Multitask Feature Fusion Network)

The collaboration among a pixel-level branch network, an object-level branch network, and a reconstruction branch network enhances the model's detection capability (Zhang et al., 2023).

CCN (Colour Conversion Network

The suggested CCN container should be implemented as an add-on module that seamlessly integrates with existing strategies to succeed universal enhancement (Fang et al.,2022).

CNN (Convolutional Neural Network)

The proposed approach is likely more reliable and applicable than traditional visual techniques. Data were developed as fabric datasets. A robotic arm fitted with touch sensors was used to automate and standardise the collection process.

Table 1
Performance metrics comparisons of the various models to detect the fabric type

Models	Fabric Types	DSR (%)	PPV (%)	NPV (%)	F-Measure (%)	Accuracy (%)
MTF-Net	Hole	87.54	70.33	66.13	61.19	52.19
	Vertical	65.98	81.22	60.44	84.98	76.98
	Horizontal	60.22	65.54	76.56	81.22	70.55
CCN	Hole	76.98	89.33	87.22	80.76	87.14
	Vertical	89.76	56.98	81.22	73.36	91.88
	Horizontal	70.44	50.33	59.91	55.14	93.77

Table 1 (*continued*)

Models	Fabric Types	DSR (%)	PPV (%)	NPV (%)	F-Measure (%)	Accuracy (%)
CNN	Hole	87.21	75.45	81.11	82.18	90.32
	Vertical	66.33	88.87	91.98	89.91	89.22
	Horizonta1	90.34	81.33	88.23	91.88	81.77
Yolov8	Hole	90.12	88.45	89.76	89.10	92.84
	Vertical	91.67	90.21	92.33	91.05	91.05
	Horizonta1	92.88	91.76	93.45	92.10	92.10
Proposed	Hole	93.76	92.65	94.56	95.18	96.18
	Vertical	93.91	94.44	94.91	95.65	97.77
	Horizontal	95.55	96.11	97.78	96.11	98.89

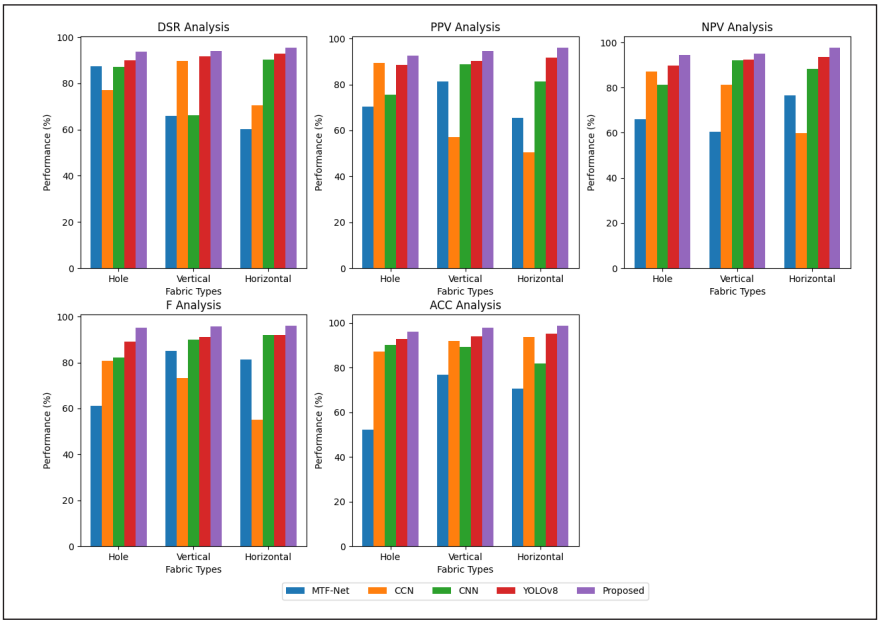


Figure 6. Comparison of performance metrics of different models for fabric type detection

Table 1 and Figure 6 indicate significant differences in detection metrics, including DSR (Detection Sensitivity Rate), PPV, NPV, F-Measure, and Accuracy. The MTF-Net model produces different scores, with a maximum accuracy of 76.98 per cent in identifying vertical fabric types, and horizontal and hole fabrics have poor performance. Instead, the CCN model attains a high level of detection of hole textiles (87.14) and vertical fabrics (91.88). The CNN model and YOLOv8 deliver middle performance with good accuracy of 90.32 and 92.84 per cent on hole textiles but marginally reduced precision in horizontal garments (81.77%). The proposed model is always better than the others in entirely different fabrics, and reaches the utmost standards in every criterion, among others, an astonishing.

Horizontal fabrics had an accuracy of 98.89%. This brings out the high efficacy of the proposed approach. in fabric type detection.

A comparison of the Average Percentage Error (APE) and Mean Absolute Error (MAE) for fabric type detection across several models is presented in Table 2 and Figure 7. The results indicate that the proposed model always has the smallest error values, with APE having a range between 5.87% and 10.44% and MAE from 15.65% to 19.45%. On the contrary, the MTF-Net model displays higher error rates, especially when dealing with the Vertical and Horizontal fabric types, with the APE values going up to 32.67% and MAE up to 51.98%. There are moderate errors in the CNN and CCN models. CNN doing better than CCN, and YOLOv8 is better than the other two. Overall, the proposed model is superior to any other models, proving to be more accurate and with a lower error rate.

Table 2
APE and MAE comparison of various models used in the fabric type detection

Models	Fabric Types	APE (%)	MAE (%)
MTF-Net	Hole	22.76	45.76
	Vertical	27.19	51.98
	Horizontal	32.67	48.18
CCN	Hole	19.45	34.19
	Vertical	23.87	39.66
	Horizontal	20.91	41.11
CNN	Hole	12.17	27.19
	Vertical	18.19	34.33
	Horizontal	15.76	31.91
Yolov8	Hole	8.92	21.36
	Vertical	11.48	25.72
	Horizontal	9.76	23.18
Proposed	**Hole**	**5.87**	**16.44**
	Vertical	**7.87**	**19.45**
	Horizontal	**10.44**	**15.65**

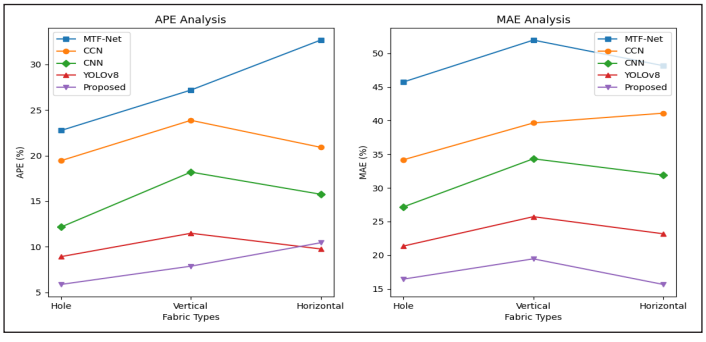


Figure 7. APE and MAE comparison between the APE and MAE of various models in fabric type detection

Table 3 and Figure 8 provide a comparison of the detection time in milliseconds (ms) for different models in fabric type detection. The suggested model has the lowest detection times, and its values tend to be between 1.675 ms and 5.876 ms in all kinds of fabrics. Comparatively, the CNN model has a slightly higher detection time with a range of 8.445 ms to 10.887 ms. There is also competitive performance between the CNN model and YOLOv8, and the detection times are in the range of 12.786 ms up to 15.154 ms. The MTF-Net model exhibits the highest detection times, especially when it comes to the Horizontal type of fabric, with a maximum value of around 18.876 ms. In general, the suggested model presents the shortest detection times, which points to its effectiveness in fabric type recognition.

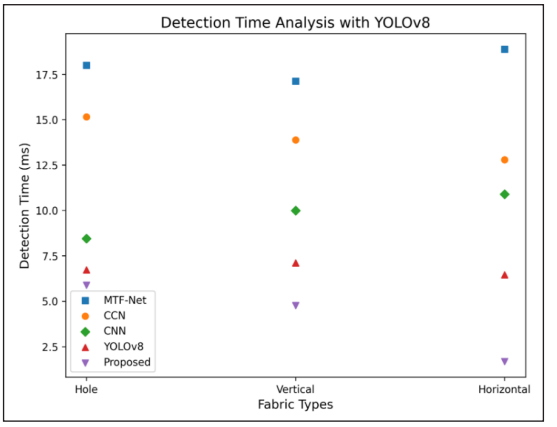


Figure 8. Comparison of detection time analysis of different models for fabric type detection

Table 3
Detection time analysis of the various models to estimate fabric type detection

Models	Fabric Types	Detection Time (ms)
MTF-Net	Hole	17.987
	Vertical	17.113
	Horizontal	18.876
CCN	Hole	15.154
	Vertical	13.876
	Horizontal	12.786
CNN	Hole	8.445
	Vertical	9.987
	Horizontal	10.887
YOLOv8	Hole	6.732
	Vertical	7.115
	Horizontal	6.458
Proposed	Hole	5.876
	Vertical	4.765
	Horizontal	1.675

Loss and Accuracy of Training

The estimation of fabric defect detection relying on training and validation loss, as illustrated in Figure 9, portrays a complete picture of how the model will be presented, besides having training and validation accuracy. Training loss is monitored so that the model is learning from the dataset, but the validation loss measures whether the model can generalise to new data (Huang et al., 2024). On the same note, high accuracy in training signifies effective learning throughout the training, with high accuracy in validation signifying the stability and viability of the model to be used in the real world. It is important to balance the metrics since significant differences can indicate overfitting or underfitting. All these metrics confirm the reliability and efficiency of the suggested fabric defect detection structure.

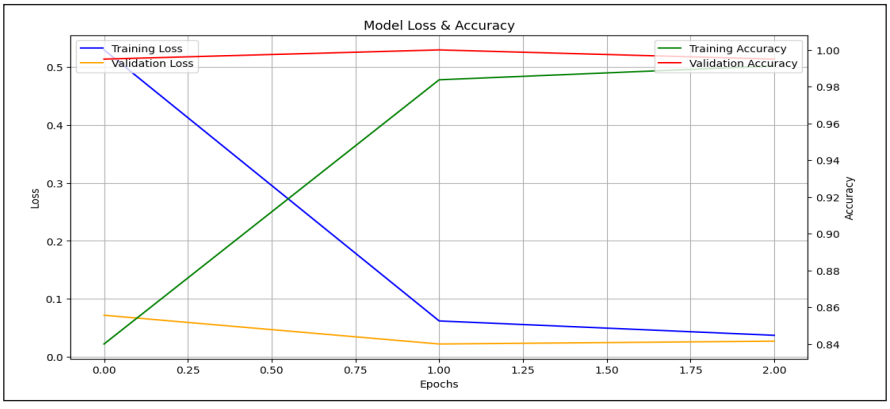


Figure 9. Training validation loss and accuracy

RESULT AND DISCUSSIONS

The fact that the evaluation of performance indicators of various models to identify types of fabric provides an opportunity to conclude that our proposed model has more benefits in comparison with the previous studies (Ding et al., 2024; Fang et al., 2022; Zhang et al., 2023). In the major measures, our model is always ahead of others, primarily in accuracy, with a success rate of up to 98.89% in the case of the Horizontal fabric type. This is much better than the best accuracy indicated by Zhang et al. (2023) of 93.77%. Also, our model has significantly lower error rates, which are indicated by the APE and MAE values. As an example, it records an APE of 5.87% and an MAE of 16.44% on the Hole fabric type, whereas the most successful models, including Fang et al. (2022), record an APE of approximately 12.17% and an MAE of approximately 27.19%. Our model is very efficient in terms of detection time, as the Horizontal fabric type is processed in the shortest time of 1.675 ms, which is much less than the other models, which require between 8.445 ms and 18.876 ms.

These results highlight the better results of our model in accuracy and computational efficiency, as it is a formidable competitor in real-time applications of fabric type diagnosis.

Ablation Study

Figure 10 and Table 4 show the ablation study, which shows the performance improvements attained through the various elements of the suggested model of fabric defect diagnosis. The foundation is the baseline model that an accuracy of 87.34 was achieved. Utilising only the BiFPN led to a decrease in the accuracy to 77.33, and this indicates that more features are required. On the other hand, the use of the sole SAM provided a commendable accuracy of 93.44%. Strong segmentation strengths. The complete combination of BiFPN and SAM in the BiFPNSAM structure produced the maximum accuracy of 98.89, which highlighted the synergistic effect of integrating the multi-scale fusion of features and fine-tuning segmentation to achieve the best defect detection.

Table 4
Ablation study for accuracy analysis

Models	Accuracy
Baseline Model	87.34
BiFPN Only	77.33
SAM Only	93.44
BiFPN-SAM	**98.89**

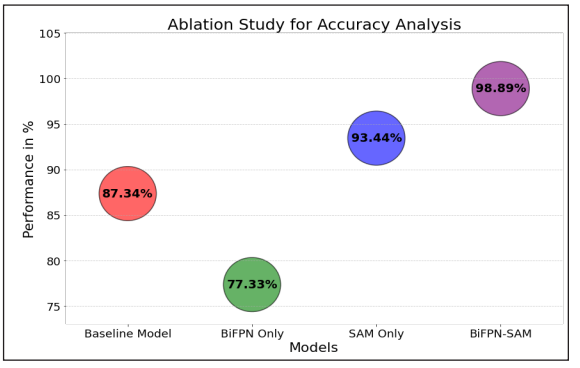


Figure 10. Ablation study for accuracy analysis

Limitations and Challenges

Although the BiFPN and the Segment Anything Model (SAM) with Specific Knowledge Injection exhibit stunning performance in cloth flaw identification, some shortcomings and obstacles persist. The training and fine-tuning of these models require the use of large-scale computational resources; the model containers are a hindrance to small-scale industries. Regardless of its flexibility, SAM can have difficulties with identifying very subtle or overlapping defects in the absence of refinement. Additionally, domain expertise and careful parameter injection are necessary to incorporate knowledge injection. tuning which may complicate deployment. Both real-time performance and high performance must be ensured. Also, accuracy is an issue in cases of varying fabric textures and dynamic production environments. These problems need to be resolved to achieve wider applicability in real-world settings.

CONCLUSION

Finally, the BiFPN and the Segment Anything Model (SAM) were combined and used. Specific Knowledge Injection has defined the exemplary presentation in fabric defect detection. The multi-scale feature fusion approach based on the BiFPN and the SAM, based on the accurate segmentation of the defects, successfully deals with the issues of texture difference and reduced defect samples. It is further enhanced by the fact that the system incorporates knowledge injection ability to adjust to the peculiarities of the fabric defects, achieving state-of-the-art accuracy and computational effectiveness in experimental tests. This is not only a way of guaranteeing accuracy, localisation, and categorisation of fabric defects but also underlines its possible application in real-world uses in automated applications of quality control of textiles.

Future Work

The structure may be expanded in future work to have the capability to detect defects in real-time, meeting the demands of fast manufacturing environments. As to its flexibility for other applications and in industrial settings, including the inspection of metal surfaces or composite materials, it could also be extended to show that it is versatile. In addition, incorporation of advanced self-supervised learning techniques and domain adaptation methods would make it more resilient to various datasets and manufacturing situations, preparing for a more universal and scalable quality control solution.

ACKNOWLEDGEMENT

Author Contributions

All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

WB-FPN	:	Weighted Bi-directional feature pyramid network
SAM	:	Segment Anything Model
BiFPN	:	Weighted Bi-directional feature pyramid network
IoU	:	Intersection over Union
CLAHE	:	Contrast-limited adaptive histogram equalisation
CDF	:	Cumulative distribution function
AHE	:	Adaptive histogram equalisation
BS	:	Block size
CL	:	Clip limit
DSR	:	Detection success rate

PPV : Positive predicted value
NPV : Negative predicted value
TP : True Positive
TN : True Negative
FP : False Positive
FN : False Negative
MTF-Net : Multitask feature fusion network
CCN : Colour conversion network
CNN : Convolutional neural network
APE : Average percentage error
MAE : Mean absolute error

REFERENCES

Chen, Z., Tian, S., Shi, X., & Lu, H. (2023). Multiscale shared learning for fault diagnosis of rotating machinery in transportation infrastructures. *IEEE Transactions on Industrial Informatics, 19*(1), 447-458. https://doi.org/10.1109/TII.2022.3148289

Cuifang, Z., Yu, C., & Jiacheng, M. (2020). Fabric defect detection algorithm based on PHOG and SVM. *Indian Journal of Fibre & Textile Research, 45*(1), 123-126.

Ding, M., Pan, L., & Chi, M. (2024). A multitask feature fusion network for woven fabric density analysis. *IEEE Access, 12*, 36229-36238. https://doi.org/10.1109/ACCESS.2024.3371175

Dlamini, S., Kao, C.-Y., Su, S.-L., & Kuo, C.-F. J. (2022). Development of a real-time machine vision system for functional textile fabric defect detection using a deep YOLOv4 model. *Textile Research Journal, 92*(5-6), 675-690. https://doi.org/10.1177/00405175211034241

Fang, B., Long, X., Sun, F., Liu, H., Zhang, S., & Fang, C. (2022). Tactile-based fabric defect detection using convolutional neural network with attention mechanism. *IEEE Transactions on Instrumentation and Measurement, 71*, Article 2503309. https://doi.org/10.1109/TIM.2022.3165254

Garg, M., & Dhiman, G. (2020). Deep convolutional neural network approach for defect inspection of texture surfaces. *Journal of the Institute of Electronics and Computer, 2*(1), 28-38. https://doi.org/10.33969/JIEC.2020.21003

Huang, Z., Jing, H., Liu, Y., Yang, X., Wang, Z., Liu, X., Gao, K., & Luo, H. (2024). Segment Anything model combined with multi-scale segmentation for extracting complex cultivated land parcels in high-resolution remote sensing images. *Remote Sensing, 16*(18), Article 3489. https://doi.org/10.3390/rs16183489

Jing, J., Wang, Z., Rätsch, M., & Zhang, H. (2022). Mobile-Unet: An efficient convolutional neural network for fabric defect detection. *Textile Research Journal, 92*(1-2), 30-42. https://doi.org/10.1177/0040517520928604

Kang, X., & Li, J. (2024). AYOLOv7-tiny: Towards efficient defect detection in solid colour circular weft fabric. *Textile Research Journal, 94*(1-2), 225-245. https://doi.org/10.1177/00405175231205898

Ke, Y. Y., & Tsubono, T. (2022). Recursive contour-saliency blending network for accurate salient object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 2940-2950). IEEE. https://doi.org/10.1109/WACV51458.2022.00143

Khodier, M. M., Ahmed, S. M., & Sayed, M. S. (2022). Complex pattern Jacquard fabrics defect detection using convolutional neural networks and multispectral imaging. *IEEE Access, 10*, 10653-10660. https://doi.org/10.1109/ACCESS.2022.3144843

Li, F., Yuan, L., Zhang, K., & Li, W. (2020). A defect detection method for unpatterned fabric based on multidirectional binary patterns and the grey-level co-occurrence matrix. *Textile Research Journal, 90*(7-8), 776-796. https://doi.org/10.1177/0040517519879904

Li, L., Wang, Y. Q., Gao, H., Qi, J., & Zhou, T. (2022). Automatic recognition method for the three-elementary woven structures and defects of carbon fabric prepregs. *Composite Structures, 291*, Article 115527. https://doi.org/10.1016/j.compstruct.2022.115527

Mao, Y., Wang, G., Ma, Y., & Gui, X. (2025). Enhancing fabric defect detection with attention mechanisms and optimised YOLOv8 framework. *IEEE Access, 13*, 96767-96781. https://doi.org/10.1109/ACCESS.2025.3570455

Roy, K., Hasan, M., Rupty, L., Hossain, M. S., Sengupta, S., Taus, S. N., & Mohammed, N. (2021). Bi-FPNFAS: Bi-directional feature pyramid network for pixel-wise face anti-spoofing by leveraging Fourier spectra. *Sensors, 21*(8), Article 2799. https://doi.org/10.3390/s21082799

Xie, C., Xia, C., Ma, M., Zhao, Z., Chen, X., & Li, J. (2022). Pyramid grafting network for one-stage high-resolution saliency detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11717-11726). IEEE. https://doi.org/10.1109/CVPR52688.2022.01142

Zhang, C., Qi, Y., & Wang, Y. (2024). Learning the colour space for effective fabric defect detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 8*(1), 981-991. https://doi.org/10.1109/TETCI.2023.3320565

Zhang, H., Chen, X., Lu, S., Yao, L., & Chen, X. (2023). A contrastive learning-based attention generative adversarial network for defect detection in colour-patterned fabric. *Colouration Technology, 139*(3), 248-264. https://doi.org/10.1111/cote.12642

Zhou, Q., Shi, H., Xiang, W., Kang, B., & Latecki, L. J. (2025). DPNet: Dual-path network for real-time object detection with lightweight attention. *IEEE Transactions on Neural Networks and Learning Systems, 36*(3), 4504-4518. https://doi.org/10.1109/TNNLS.2024.3376563

Zhou, Q., Wang, L., Gao, G., Kang, B., Ou, W., & Lu, H. (2024). Boundary-guided lightweight semantic segmentation with multi-scale semantic context. *IEEE Transactions on Multimedia, 26*, 7887-7900. https://doi.org/10.1109/TMM.2024.3372835

Zhou, Q., Wu, X., Zhang, S., Kang, B., Ge, Z., & Latecki, L. J. (2022). Contextual ensemble network for semantic segmentation. *Pattern Recognition, 122*, Article 108290. https://doi.org/10.1016/j.patcog.2021.108290

Pertanika J. Sci. & Technol. 34 (S1): 133 - 159 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Position-Aware Normalized Discounted Cumulative Gain (NDCG) and Multi-Task TabNet for Cervical Precancerous Lesion Classification

P. Leela[1]*, K. Hari Krishna[1], and K. K. Baseer[2]

[1]Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Andhra Pradesh, India
[2]Department of Computer Science and Systems Engineering, GITAM School of CSE, GITAM (Deemed to be University), Nagadenahalli, 562163 Bengaluru, Karnataka, India

ABSTRACT

Cervical cancer is still one of the major causes of cancer death in women around the world. Accurate and early classification of precancerous lesions from images of cervical cells is crucial for the enhancement of patient outcomes. Based on MMT, this paper introduces a new hybrid model named TabNet-PANDCG for improving the classification performance, severity-aware ranking, and interpretability of models. This paper presents TabNet-PANDCG, a novel hybrid model built on top of MMT to boost classification performance, severity-aware ranking, and model interpretability. The proposed method extracts the morphological features, intensity features and texture features from cervical cell images to generate tabular structured images, in contrast with the conventional method, which processes raw images directly. The image-derived features are then passed into Multi-Task TabNet with a sparse attention mechanism that performs the task of dynamically selecting the relevant features and uses it to simultaneously classify the lesion and predict the lesion's severity. The proposed PANDCG metric is a ranking evaluation that assigns greater weight to the clinically relevant, high-severity lesions that rank highly, giving a more meaningful evaluation that fits more with the clinical focus. The framework was tested using a publicly available dataset containing 917 Pap images from single-cell images, across seven diagnostic categories, namely Herlev. Experimental results show that TabNet-PANDCG has the greatest performance with 97.39% accuracy, 93.66% macro precision, 92.17% macro recall and 92.14% macro F1 score, outperforming some of the state-of-the-art techniques. Additionally, the incorporation of PANDCG facilitates clinically relevant ranking assessments, and TabNet's attention-based

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 14 May 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.07

E-mail addresses:
leelapachappa@gmail.com (P. Leela)
khk396@gmail.com (K. Hari Krishna)
drkkbaseer@gmail.com (K. K. Baseer)
* Corresponding author

e-ISSN: 2231-8526

feature selection further promotes transparency and trust in the model. The framework exhibits great promise for computer-aided diagnosis systems in cervical cancer screening.

Keywords: Cervical cancer, clinical diagnostics, deep learning, interpretability, medical data analysis, multi-task learning, multi-task TabNet, position-aware NDCG, sparse attention

INTRODUCTION

Cervical cancer often develops in the cells of the cervix, the lower section of the uterus. It is widely known that a sexually transmitted infection, namely human papillomavirus (HPV), is the leading cause of cervical cancer. Research studies have shown that cervical cancer is one of the leading causes of death among women with cancer (Allogmani et al., 2024). A significant number of cases of cervical cancer continue to be reported in both low- and high-income countries. Cancer of the cervix is one of the top cancers affecting women, accounting for about 6.5 per cent of all cancer cases in women. It is estimated that the number of deaths caused by cervical cancer and the new incidences in 2020 were 342,000 and 604,127 globally, respectively (Attallah, 2023). Cervical cancer is a serious problem in the world, especially in areas lacking screening, vaccination, diagnostic facilities and early detection services. Approximately 54,517 new cases of invasive cervical cancer have been detected in Europe every year, and 24,874 women die of the disease (Bentéjac et al., 2021). The anatomical site and pathological trajectory of cervical cancer are shown in Figure 1.

Vaccinations are common in ordinary immunisation programs. World Health Organisation (2022) have set a target to vaccinate 90% of girls to be fully vaccinated against HPV by the age of 15 years by 2030 (Chandran et al., 2021). It is also emphasised that the European authorities must employ the existing preventive strategies to intensify the initiative to eliminate cervical cancer, as highlighted by the World Health Organisation (2023). Like other health issues, mortality rates can be diminished to a considerable extent by conducting regular check-ups and diagnosing the disease at its early stages (Darwish et al., 2023). Due to the absence of symptoms during the early development of cervical cancer, the diagnosis may become challenging. So, an automated and accurate diagnostic system will help clinicians in the screening and classification of cervical lesions.

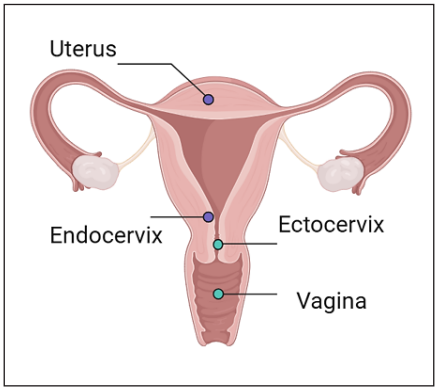


Figure 1. Cervix cancer image

Abnormal bleeding is the most common symptom of cervical cancer, and it is likely to become more serious with the development

of the disease. In later stages, almost 90% of the patients have distinct symptoms. The typical signs of cervical cancer are reddish discharge, postmenopausal bleeding, contact bleeding, and spotting. Moreover, foul-smelling bloody discharge is another important sign of the disease.

The General Article of this Study

The proposed study introduces a new method that integrates Position-aware NDCG, a metric that aims to focus on the quality of diagnostics according to the severity of the lesions, with Multi-Task TabNet, a deep learning model with the ability to process tabular data and perform multiple related prediction tasks. With Position-aware NDCG, the model prioritises clinically important cases, and in the case of precancer lesions, the sensitivity is high. At the same time, the Multi-Task TabNet is a platform that incorporates a wide range of diagnostic features, which makes the model both more robust and interpretable. In this study, the proposed multi-task TabNet is applied to image-derived feature vectors representing cervical cell images, due to TabNet's demand for structured tabular input. The given work has great opportunities to contribute to the development of early cervical lesion detection and classification and introduce more specific measures and better patient outcomes.

Motivations of this Research

The motivation behind this study stems from several critical challenges in current cervical precancerous lesion classification systems:

- Existing evaluation metrics treat all misclassifications equally, which is clinically inadequate. In cervical cancer screening, misclassifying a high-severity lesion (such as severe dysplasia or carcinoma in situ) as a low-risk class has far more serious consequences than the reverse. There is a strong need for severity-aware ranking metrics that prioritise clinically critical cases.
- Most deep learning models for cervical cell classification operate as black boxes, offering limited interpretability. Clinicians require transparent models that can reveal which diagnostic features influence predictions, thereby increasing trust and facilitating adoption in clinical workflows.
- Current approaches often fail to leverage the complementary strengths of advanced tabular learning architectures and clinically meaningful evaluation strategies. There is a clear opportunity to develop hybrid frameworks that combine powerful feature selection mechanisms with ranking metrics aligned with medical priorities.
- Early and accurate detection of precancerous lesions remains essential for reducing cervical cancer mortality, particularly in resource-limited settings where automated,

reliable computer-aided diagnosis tools can significantly improve screening outcomes.

Contributions of this Study

- We propose PANDCG (Position-aware Normalised Discounted Cumulative Gain), a novel severity-aware ranking metric specifically designed for cervical precancerous lesion classification. Unlike standard metrics that assign equal weight to all errors, PANDCG incorporates position-aware weighting to prioritise high-severity lesions at top-ranked positions, thereby providing a more clinically meaningful evaluation of model predictions.
- We develop a hybrid TabNet-PANDCG framework that transforms cervical cell images into structured tabular feature representations through comprehensive feature engineering. These image-derived features are processed by Multi-Task TabNet, which employs sparse attention mechanisms for interpretable feature selection and performs simultaneous lesion classification and severity prediction.
- We demonstrate through extensive experiments on the Herlev dataset (917 single-cell cervical images across seven diagnostic classes) that the proposed framework achieves state-of-the-art performance (97.39% accuracy) while offering improved interpretability and clinically aligned ranking evaluation. The integration of PANDCG and TabNet's attention mechanism addresses key limitations of existing black-box models and conventional evaluation approaches.
- We provide a transparent and reproducible pipeline that bridges image-based cervical cell analysis with tabular deep learning, offering a practical pathway toward trustworthy computer-aided diagnosis systems for cervical cancer screening.

Related Work

Chandran et al. (2021) introduced a deep learning method used to identify cervical cancer by classifying colposcopy images into three classes, i.e. type 1, type 2 and type 3. The researchers used a popular cervical screening dataset to train their algorithm and test it. They introduced a powerful network known as the Colposcopy Ensemble Network (CYENET) that was more accurate than the preceding models such as VGG16 and VGG19. CYENET posted an accuracy of 92.3% on the test (Lebanova et al., 2023). However, this high accuracy could have been affected by the fact that the quantity of testing images used in this study (1884) was rather small.

According to Tomko et al. (2023), there were considerable effects of cervical cancer on the physical, psychological, and social well-being of women throughout their lives.

This research concentrates on the data employed throughout the Kaggle competition in the Intel and MobileODT Cervical Cancer Screening, in which the challenge is to deal with multi-class and multi-label organisation. They also included optimisation of image size in their strategy.

Attallah (2023) introduced a new CAD model known as CerCan-Net to be used in the automation of detecting cervical cancer. The CerCan-Net is built on three CNN architectures with lower depth and fewer parameters (ResNet-18, MobileNet, and DarkNet-19) in comparison to the traditional models. This is a calculated decision that will ease and hasten the classification process. An essential element in the effectiveness of CerCan-Net is transferring learning, and it relies on the deep features in the final three layers of every CNN, as opposed to relying on features in just one layer. The approach allows describing the complexity of the information in more detail.

To automatically classify cervical cytology images, Fang et al. (2022) suggested a deep convolutional neural network called DeepCELL. This network learns feature representations through multiple kernels of varying sizes. Primarily, three distinct core modules of DeepCELL were designed to extract feature information utilising these kernels. Subsequently, the cervical cell classification model was structured by stacking several of these core modules in a top-to-bottom manner.

Xu et al. (2022) presented a new risk assessment network for cervical lesions in colposcopic images, which was called the RACNet. The RACNet has two major components. The first is a multi-scale detection network that is used to detect the location and severity of the lesions at different scales. Second, the development of a multi-branch convolutional neural network (CNN) that can classify these lesions is done to enhance the performance of risk assessment (Okunade, 2020). RACNet was linked with the most suggested methods and colposcopists under the same circumstances. The results of the experiment conducted in this paper show that RACNet has the best performance, as it has an accuracy ratio of 84.5, which is 15 percent higher than the average accuracy of colposcopists.

Shakil et al. (2024) came up with a precise prediction algorithm for cervical cancer by employing six ML models. The approach was used following evaluation of 36 risk factor variables among 858 participants. The Adaptive Synthetic Sampling and the Synthetic Minority Oversampling Technique were used to balance the data. Also, feature selection approaches such as chi-square and Least Absolute Shrinkage and Selection Operator (LASSO) were used to assess the ranking of features, especially in the detection of illness. The results of the model were explained using Shapley Additive Explanations, an explainable AI model.

Limitations of Existing Systems

- The current systems are not resistant to the nature of being used on varied data. The difference in the image acquisition settings, patient demographics and lesion characteristics may result in differing performance between clinical settings.
- Most algorithms are trained in small and homogeneous datasets, which do not cover a broad range of cervical pathologies. This weakness does not allow them to generalise to unseen data, especially in the real world.
- The systems usually use the pathologist annotation, which may be subjective and prone to inter- and intra-observer variability. This brings noise to the training data, which eventually decreases the accuracy of the model.

Research Gap

The lack of research in the development of cervical cancer screening tools is the inability to find cervical precancerous lesions correctly and efficiently, which is a significant challenge. Existing methods, including cytology and histology, are manual-based and therefore prone to error and subjectivity. Moreover, the fact that the adoption of new high-level technologies, including AI, has not become an essential part of daily medical activity has prevented the possibility of automated and high-quality diagnosis. The most important research gap is the formulation of strong, explainable, and scalable AI-based models that can address these shortcomings by offering stable, trustworthy, and early detection of precancer lesions, primarily in resource-strained cases in which cervical cancer is overrepresented.

Proposed System

In this paper, we suggest a hybrid interpretable model, TabNet-PANDCG, which combines Position-aware Normalised Discounted Cumulative Gain (PANDCG) with Multi-Task TabNet for the classification of cervical precancerous lesions. The framework introduces a "severity-aware" ranking evaluation and "attention" tabular learning on features derived from images, to enhance classification accuracy and clinical usability. The proposed approach differs from traditional image classification pipelines, starting with segmenting cervical cell images for morphological, intensity and texture features to create structured tabular representations of these images. These feature vectors are provided as input to Multi-Task TabNet, which uses sequential sparse attention mechanisms to decide which features are the most discriminative without discarding any information, while also performing lesion classification and severity prediction. By enriching standard ranking measures with the importance of having a high severity lesion (severe dysplasia, carcinoma in situ) at top-ranked positions, PANDCG complements this and can bring these features into the ranking process. The proposed TabNet-PANDCG framework overall architecture is

shown in Figure 2. The following subsections describe the dataset, pre-processing pipeline, PANDCG formulation and Multi-Task TabNet architecture.

Dataset

The Herlev dataset contains 917 isolated single-cell cervical cell images (Peng et al., 2021), with each image featuring a single cervical cell. Figure 3 shows sample images from the Herlev dataset. The seven classes in the dataset are columnar epithelia, superficial squamous epithelia, mild squamous non-keratinising dysplasia, moderate squamous non-keratinising dysplasia, severe squamous non-keratinising dysplasia, intermediate squamous cells, and squamous cell carcinoma in situ. Table 1 shows the summary of class-wise distribution of the Herlev dataset. The raw images were not used as input to TabNet in this study; first, structured feature vectors were extracted from the images and then used as inputs into TabNet with the proposed Multi-Task TabNet classifier.

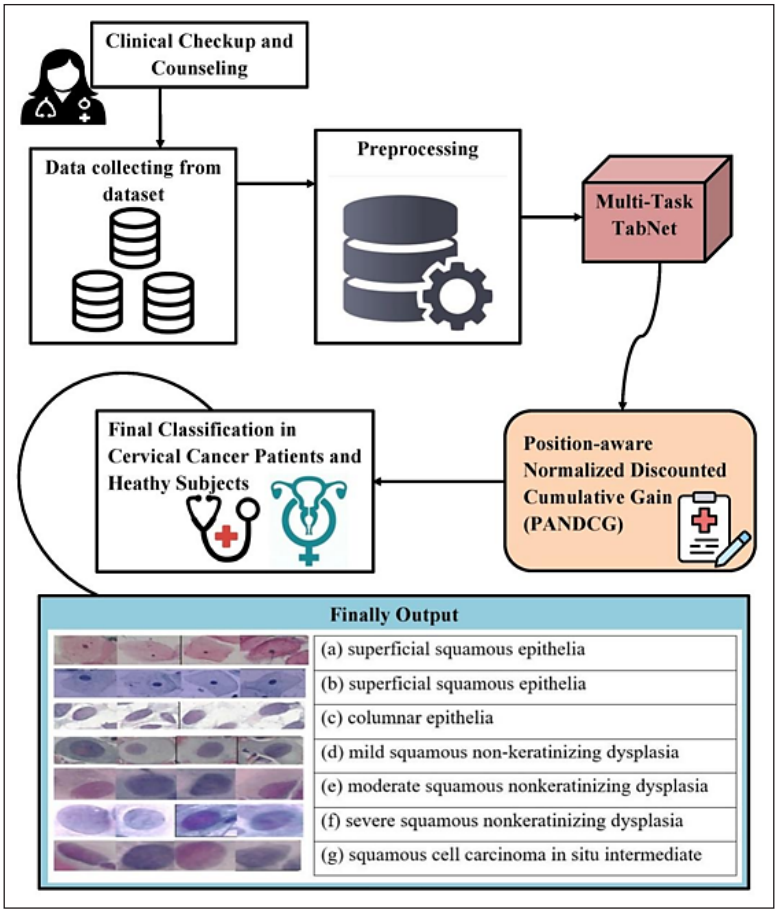


Figure 2. Architecture of TabNet-PANDCG

There are 917 single-cell cervical pictures in the Herlev dataset, which are divided into groups for normal and pathological conditions. 98 columnar epithelial cells, 70 intermediate squamous epithelial cells, and 74 superficial squamous epithelial cells make up the typical category. The pathological category includes 146 photographs of moderate squamous non-keratinising dysplasia, 197 images of severe squamous non-keratinising dysplasia, 182 images of mild squamous non-keratinising dysplasia, and 150 images of intermediate squamous cell carcinoma. Models for classifying cervical cells and identifying precancerous and malignant states require this varied dataset for training and evaluation.

Data Preprocessing

To learn with Multi-Task TabNet, the algorithm is designed for learning with tabular data; the raw images of cervical cells in the Herlev dataset had to be systematically transformed into quantitative representations of the data (Shakil et al 2024). The process is fundamental, as TabNet can't work with pixel values as an input but instead requires structured input.

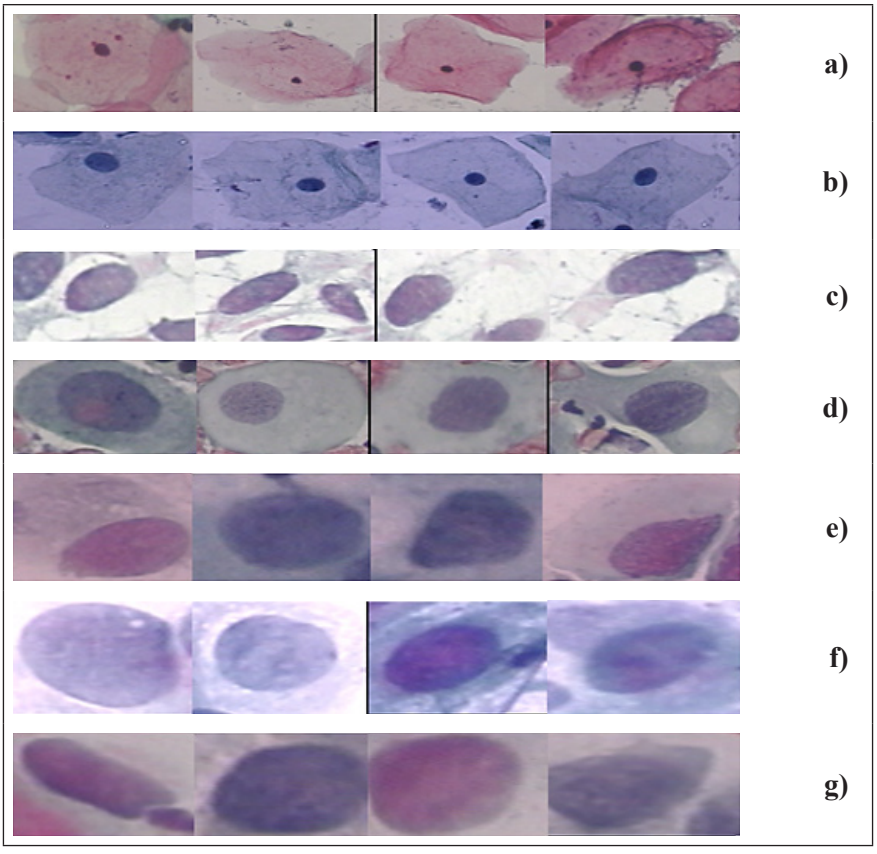


Figure 3. Sample images for the Herlev dataset

Table 1
Provides the statistical data from the Herlev dataset

Dataset type	Category	Quantity
Normal	Superficial squamous epithelial	74
Normal	Intermediate squamous epithelial	70
Normal	Columnar epithelial	98
Abnormal	Squamous cell carcinoma in situ intermediate	150
Abnormal	Mild squamous non-keratinising dysplasia	182
Abnormal	Moderate squamous non-keratinising dysplasia	146
Abnormal	Severe squamous non-keratinising dysplasia	197
Total image		917

We derived an extensive feature suite for each of the 917 single-cell images, covering three broad classes of hand-crafted features:

- Morphological features: nucleus area, cytoplasm area, nucleus-to-cytoplasm ratio, eccentricity, solidity, perimeter and equivalent diameter.
- Intensity-based features: mean, standard deviation, skewness and kurtosis of the pixel intensity distribution in the nucleus region and cytoplasm region of the cell.
- Texture features: Gray-Level Co-occurrence Matrix (GLCM) statistics (at 8 different orientations and 4 distances).

These features were concatenated into a high-dimensional tabular feature matrix with one row being an image of a cervical cell, and each column a quantitative feature which describes the image. Data were checked for obvious outliers and missing values and eliminated during quality control. To reduce the number of input variables that could lead to overfitting and to remove potential overfitting, a Principal Component Analysis (PCA) was performed and the principal components that explained the largest percentage of cumulative variance (95%) were used. To have similar scales among the different types of features before training the models, the features were standardised in advance by z-score normalisation.

This approach to feature engineering uses the image classification task as a structured tabular learning task, allowing access to Multi-Task TabNet, which is highly effective at solving tabular learning problems, while revealing inherent interpretability thanks to sparse attention mechanisms.

Position-Aware Normalised Discounted Cumulative Gain (PANDCG)

Standard ranking metrics such as Normalised Discounted Cumulative Gain (NDCG) treat all misrankings equally, which is suboptimal in medical diagnosis where the clinical cost of errors varies significantly with lesion severity (Sompawong et al., 2019). To address

this limitation, we propose Position-aware Normalised Discounted Cumulative Gain (PANDCG), a severity-sensitive ranking metric designed to prioritise high-severity cervical lesions at top-ranked positions during evaluation.

PANDCG modifies the standard NDCG formulation by introducing position-aware weighting and clinically defined relevance scores. Let denote the ranked list of predicted lesion classes for a given sample, where higher ranks correspond to more severe lesions according to the model output. We first assign a relevance score $rel(i)$ to each class based on its clinical severity in Equation 1:

$$rel(i) = \begin{cases} 0 & \text{if class } i \text{ is Normal} \\ 1 & \text{if class } i \text{ is Mild dysplasia} \\ 2 & \text{if class } i \text{ is Moderate dysplasia} \\ 3 & \text{if class } i \text{ is Severe dysplasia} \\ 4 & \text{if class } i \text{ is Carcinoma in situ} \end{cases} \quad [1]$$

Discounted Cumulative Gain (DCG)

The DCG score reduces the influence of lower-ranked items while calculating the relevance of elements in a ranked list. For the top-K ranked predictions, DCG is defined as in Equation 2:

$$DCG = \sum_{i=1}^{k} \frac{rel_i^*}{\log_2(i+1)} \quad [2]$$

where rel_i represents the relevance score of the prediction at rank position i, and K denotes the number of ranked predictions considered. The logarithmic discount term $\log_2(i+1)$ reduces the contribution of lower-ranked predictions.

Ideal DCG (IDCG)

To normalise DCG, the ideal DCG score is calculated in Equation 3, representing the DCG of a perfect ranking.

$$IDCG = \sum_{i=1}^{K} \frac{rel_i^*}{\log_2(i+1)} \quad [3]$$

where rel_i^* represents the ideal relevance score at position i after sorting the lesion classes according to the clinically preferred ranking. IDCG represents the maximum possible DCG value for a given sample.

Position-Aware Weights

In PANDCG, position weights $w(i)$ are introduced to amplify the importance of earlier positions. The weights are often assigned based on clinical priorities, such as imposing higher penalties for misclassifying severe lesions. The position-aware weight is defined as in Equation 4:

$$\mathrm{w_i} = \frac{1}{\sqrt{\mathrm{i}}} \quad [4]$$

where w_i denotes the position-aware weight for rank position i. This weighting strategy assigns a larger weight to earlier-ranked predictions and gradually reduces the weight for lower-ranked predictions.

The DCG is modified using position-aware weights in Equation 5:

$$PDCG = \sum_{i=1}^{k} w_i \times \frac{rel_i}{\log_2(i+1)} \quad [5]$$

where $PDCG$ represents the position-aware discounted cumulative gain. By multiplying the relevance score with w_i, the metric gives additional emphasis to clinically relevant predictions appearing near the top of the ranked output.

PANDCG

Finally, PANDCG is normalised using the ideal position-aware DCG. The final $PANDCG$ score is calculated as in Equation 6:

$$PANDCG = \frac{PDCG}{IPDCG} \quad [6]$$

where $IPDCG@K$ represents the ideal position-aware discounted cumulative gain. It is calculated as in Equation 7:

$$IPDCG = \sum_{i=1}^{k} w_i \times \frac{rel_i^*}{\log_2(i+1)} \quad [7]$$

PANDCG are between 0 and 1, with higher scores placing the clinically important classes higher on the prediction list. For this reason, PANDCG offers a severity-dependent effectiveness assessment approach to the classification of cervical precancerous lesions.

Multi-Task TabNet for Cervical Precancerous Lesion Classification

Following the feature engineering process described in Section 3.2, each cervical cell image is represented as a structured tabular feature vector. Multi-Task TabNet to learn from this tabular representation because it is specifically designed for structured data and offers built-in interpretability through sparse attention. The model simultaneously performs two related tasks: (1) multi-class lesion classification and (2) severity regression, allowing shared representations to improve generalisation across both tasks.

TabNet processes the input through a sequence of decision steps, each consisting of a Feature Transformer and an Attentive Transformer. This sequential attention mechanism enables the model to focus on the most relevant features at each step while producing interpretable feature masks. The overall architecture comprises an input layer, multiple decision steps with shared and step-dependent feature transformers, attentive transformers with sparsemax normalisation, and task-specific output heads for classification and regression.

TabNet Feature Transformer

The output F_t of a feature transformer layer is calculated as in Equation 8:

$$F_t = f_t(M_t \odot X) \quad [8]$$

where F_t represents the transformed features at decision step t, X denotes the input feature vector, M_t is the sparse attention mask generated at step t, $\odot$ represents element-wise multiplication, and $f_t(\cdot)$ denotes the feature transformation function.

Attentive Transformer

The sparse attention mask M_t is derived as in Equation 9:

$$M_t = Sparsemax\ (A_t) \quad [9]$$

The attention coefficient of a block A_t at the decision step t and a sparse feature-selection mask (Sparsemax) are used to generate. This sparse mask enables TabNet to only focus on features that are the most informative and minimise the effects of irrelevant and/or redundant features.

Multi-Task Loss Function

For cervical precancerous lesion classification, the total loss L_{total} is a combination of losses for all tasks as represented in Equation 10.

$$L_{total} = \sum_{j=1}^{T} \lambda_j L_j \quad [10]$$

where T represents the number of learning tasks, L_j denotes the loss for the j^{th} task, and λ_j represents the weight assigned to that task. In this study, the task losses include lesion classification loss and severity-aware prediction loss.

For classification, the prediction is represented by Equation 11:

$$L_{cls} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \quad [11]$$

where C is the number of lesion classes, y_c is the true class label, and $\hat{y}_c$ is the predicted probability for class c.

For severity prediction (regression), the prediction function is defined as follows in Equation 12:

$$L_{sev} = \frac{1}{N}\sum_{n=1}^{N} (s_n - \hat{s}_n)^2 \quad [12]$$

where N represents the number of samples, s_n is the true severity score of the n^{th} sample, and $\hat{s}_n$ is the predicted severity score.

Final Prediction

The final prediction $\hat{y}$ for each task combines all features after the decision step, as shown in Equation 13:

$$\hat{y} = \sigma(WF + b) \quad [13]$$

The last layer of the network, where W is the weights of the last layer, F is the last feature representation from TabNet decision steps, bis the bias term, and σ denotes the activation function. In binary prediction, the sigmoid activation function can be applied, while for multi-class lesion classification the softmax activation function can be applied. The proposed Multi-Task TabNet model learns discriminative cervical cell features, selects crucial features for cancer diagnosis from many features in a sparse fashion and can classify cervical cell lesions. The PANDCG module then checks if clinically relevant categories

of lesions are ranked in the predictions correctly. This architecture enables TabNet to dynamically select important diagnostic features while learning shared representations beneficial for both classification and severity estimation, thereby improving both accuracy and interpretability.

Application of Proposed Method

The TabNet-PANDCG framework is specifically designed to address the dual challenges of accurate lesion classification and clinically meaningful evaluation in cervical cancer screening. By transforming cervical cell images into structured tabular feature representations and processing them through Multi-Task TabNet, the framework enables simultaneous lesion classification and severity estimation while maintaining interpretability through sparse attention mechanisms.

The integration of PANDCG further enhances the framework's clinical utility by ensuring that high-severity lesions are prioritised in the ranked output. This severity-aware evaluation strategy aligns model assessment with real-world diagnostic priorities, where misranking critical cases carries significant clinical consequences. The proposed approach is therefore particularly suitable for computer-aided diagnosis systems that require both high accuracy and transparent, clinically aligned decision support.

Performance Evaluation Metrics

Macro Precision: The average of the precision scores calculated separately for each class in a multi-class classification task is known as macro precision, as shown in Equation 14. All classes, regardless of size, are treated equally.

$$Macro\,\mathrm{Pr}ecision = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Pr}ecision_i \qquad [14]$$

Macro Recall: Macro Recall is a metric used in classification problems to calculate the average recall across all classes as shown in Equation 15. It treats all classes equally by computing the recall for each class independently and then averaging the results.

$$Macro\,\mathrm{Re}call = \frac{1}{N}\sum_{i=1}^{N}\frac{True\ Positives_i}{True\ Positives_i + False\ Negatives_i} \qquad [15]$$

Macro F1-Score: A statistic used for evaluation in multi-class classification tasks is the Macro F1-Score. Each class is treated equally, regardless of its frequency, and the F1-Score is calculated individually for each class. Finally, the unweighted mean of these scores is computed as shown in Equation 16.

$$Macro\ F1-Score=\frac{1}{N}\sum_{i=1}^{N}F1_i \quad [16]$$

Accuracy: The proportion of properly recognised occurrences (including true positives and true negatives) relative to total examples in a dataset is known as accuracy. It is a widely utilised metric to estimate the performance of classification models as shown in Equation 17.

$$Accuracy=\frac{TP+TN}{Total\ Instances} \quad [17]$$

Root Mean Square Error (RMSE): The average magnitude of the discrepancies between anticipated and observed values in a dataset is typically measured using the Root Mean Square Error (RMSE) metric as shown in Equation 18. Like the observed data, RMSE provides a clear measurement of these differences in the same units.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \quad [18]$$

Processing Time: Processing time refers to the duration required to complete a specific task or process within a system. It encompasses the time taken to execute all necessary steps in the workflow as shown in Equation 19.

$$T_p=T_e+T_w \quad [19]$$

Advantages of Proposed Method

The proposed TabNet-PANDCG framework offers several key advantages over existing approaches for cervical precancerous lesion classification:

- **Severity-aware ranking evaluation**: PANDCG incorporates clinical severity into the ranking metric, ensuring that high-grade lesions (severe dysplasia and carcinoma in situ) are prioritised at top-ranked positions, thereby providing a more meaningful assessment than standard accuracy-based or ranking metrics.
- **Interpretable feature selection**: Multi-Task TabNet's sparse attention mechanism dynamically identifies the most discriminative diagnostic features at each decision step, offering inherent interpretability without requiring post-hoc explanation techniques.
- **Multi-task learning capability**: By jointly optimising lesion classification and severity regression, the model learns shared representations that improve generalisation and provide complementary diagnostic information in a single framework.

- **Suitability for image-derived tabular data**: The framework bridges image-based analysis and tabular deep learning by converting cervical cell images into structured feature vectors, enabling the use of TabNet's efficient and interpretable architecture.
- **Clinical alignment and practicality**: The combination of high classification performance, severity-aware ranking, and built-in interpretability makes the framework well-suited for supporting reliable and trustworthy computer-aided diagnosis in clinical cervical cancer screening workflows.

RESULTS AND DISCUSSIONS

The proposed TabNet-PANDCG framework was evaluated on the publicly available Herlev dataset, consisting of 917 single-cell cervical Pap smear images across seven diagnostic classes. To comprehensively assess model performance, we conducted both seven-class multi-class classification and binary classification (normal vs. abnormal) experiments. This dual evaluation strategy allows us to examine the model's ability to distinguish fine-grained lesion categories as well as its effectiveness in clinically relevant binary screening scenarios.

The TabNet-PANDCG model was implemented using Keras with the TensorFlow 2.8.2 backend and trained on Google Colab equipped with an NVIDIA Tesla T4 GPU (CUDA 11.2, driver version 460.32.03). All experiments were conducted using consistent data splits and hyperparameter settings to ensure fair and reproducible comparisons. Model performance was assessed using accuracy, macro precision, macro recall, macro F1-score, Root Mean Square Error (RMSE), processing time, confusion matrix analysis, ablation studies, and k-fold cross-validation.

Comparative Methods

To evaluate the effectiveness of the proposed TabNet-PANDCG framework, we compared it against several recent state-of-the-art methods for cervical cancer classification. The baseline methods are briefly described below.

- **CPLDC-AOATL** (Allogmani et al., 2024): This method employs bilateral filtering as a preprocessing step to reduce noise in cervical images, followed by an Archimedes Optimisation Algorithm combined with transfer learning for classification.
- **CACCD-GOADL** (Nour et al., 2024): This approach utilises an enhanced MobileNetV3 architecture for feature extraction from cervical images, with the Gazelle Optimisation Algorithm (GOA) employed for hyperparameter tuning of the model.
- **CYENET**: This method introduces a deep ensemble network (CYENET) that integrates multiple convolutional neural network backbones for the classification of colposcopy images into different cervical cancer types.

- **Mask R-CNN** (Xu, T et al., 2022): This method applies Mask R-CNN for instance segmentation of cervical cell nuclei in liquid-based cytology slides, enabling the detection and analysis of nuclear morphological features for distinguishing normal and abnormal cells.

Confusion Matrix

The effectiveness of the classification for the dataset's seven categories is illustrated in the confusion matrix in Figure 4.

The confusion matrix for the seven-class classification task is presented in Figure 4. In this matrix, the rows represent the ground-truth classes, while the columns correspond to the classes predicted by the model. The diagonal elements indicate correct classifications, whereas the off-diagonal elements represent misclassifications between different lesion categories.

The matrix reveals strong overall classification performance, with most samples correctly assigned to their respective classes, as evidenced by the prominent diagonal

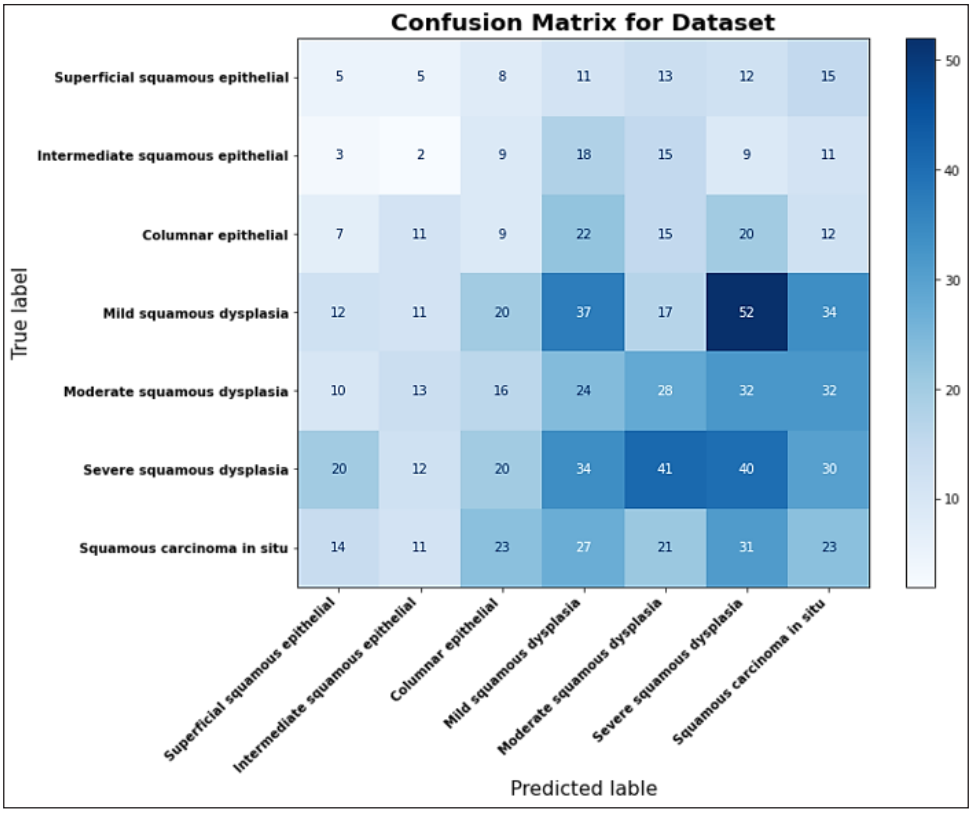


Figure 4. Confusion matrix of the proposed TabNet-PANDCG model on the Herlev dataset

entries. However, notable confusion persists among the mild, moderate, and severe squamous non-keratinising dysplasia classes. This pattern is expected, given the substantial overlap in morphological and textural characteristics among these progressive stages of dysplasia. Such inter-class similarity poses a significant challenge for fine-grained classification and highlights the need for more discriminative feature representations or advanced class-separation techniques in future work.

Binary Class-Wise Comparison

Table 2 shows the comparison of various methods for the binary evaluation setting of normal and abnormal classes. Case "0" = "Normal" class and case "1" = "Abnormal" class. MP, MR and MF1 represent macro precision, macro recall and macro F1-score, respectively.

The proposed TabNet-PANDCG framework consistently outperforms all baseline methods in both normal and abnormal classes across all evaluation metrics. Notably, it achieves the highest accuracy of 97.14% for normal cases and 97.65% for abnormal cases, along with superior macro-level performance. These results demonstrate the effectiveness of combining Multi-Task TabNet with the PANDCG evaluation strategy for robust binary cervical cell classification.

The results obtained from the proposed model are compared with the existing methods in terms of MP, MR, MF1 and accuracy in the proposed model for both normal and abnormal cases as shown in Figure 5 and Table 2, respectively. For case "0", the proposed TabNet-PANDCG achieved 94.15 MP, 92.16 MR, 91.11 MF1, and 97.14% accuracy. For case "1", it achieved 93.17 MP, 92.18 MR, 93.17 MF1, and 97.65% accuracy. The obtained results show a good performance of the proposed model for normal and abnormal cervical cell classification. The proposed approach outperforms existing methods such as CPLDC-AOATL, CACCD-GOADL, and Mask R-CNN in terms of overall performance, including accuracy and weighted performance at the macro level**.**

Table 2
Comparison of methods for binary normal and abnormal cases

Methods	Case 0 MP	Case 0 MR	Case 0 MF1	Case 0 Accuracy	Case 1 MP	Case 1 MR	Case 1 MF1	Case 1 Accuracy
CPLDC-AOATL	87.34	62.25	77.45	88.28	64.81	75.56	88.23	80.34
CACCD-GOADL	76.12	81.87	71.22	81.23	79.45	85.23	76.14	87.34
CYENET	66.23	88.13	89.55	87.34	81.28	87.23	87.98	89.22
Mask R-CNN	89.14	77.34	88.16	89.91	88.76	79.13	89.32	87.11
Proposed TabNet-PANDCG	94.15	92.16	91.11	97.14	93.17	92.18	93.17	97.65

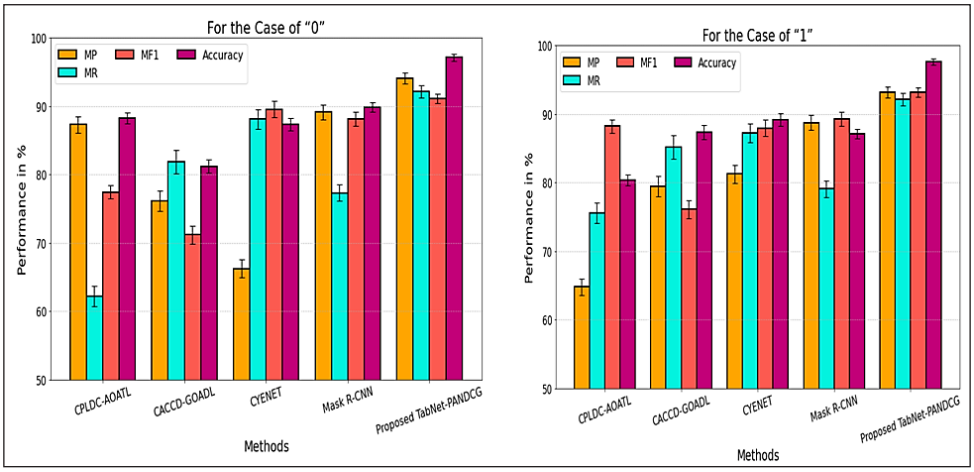


Figure 5. Comparative performance of different methods for normal and abnormal cases

RMSE and Processing Time Analysis

Table 3 and Figure 6 show the comparative results between the proposed model and the comparative models. Repair error was evaluated using the root mean square error (RMSE), and the closer the RMSE was to a minimum, the more consistent the prediction.

Table 3
RMSE Analysis of proposed method

Methods	RMSE
CPLDC-AOATL	39.91
CACCD-GOADL	36.16
CYENET	31.55
Mask R-CNN	23.87
Proposed TabNet-PANDCG	18.24

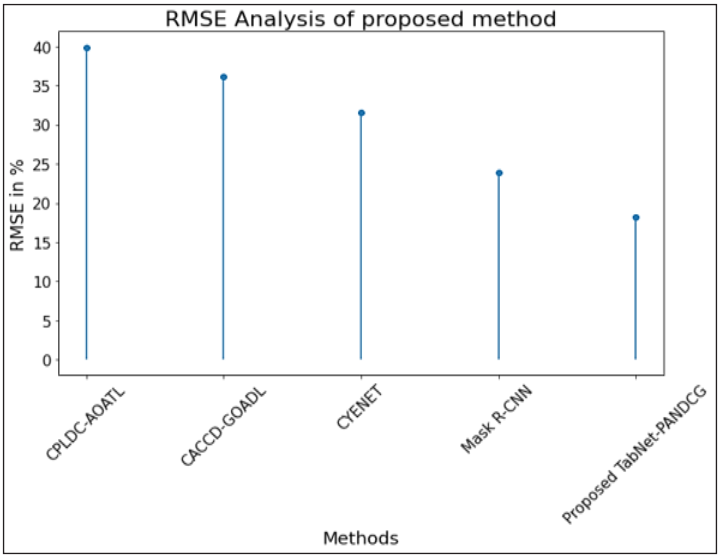


Figure 6. RMSE analysis of proposed method

TabNet-PANDCG could obtain the lowest RMSE of 18.24 in predicting, which is better than other comparative models. Mask R-CNN had the second smallest RMSE value of 23.87, and CYENET and CACCD-GOADL had relatively larger error values. CPLDC-AOATL achieved the largest RMSEs, indicating less consistency in prediction. The overall RMSE comparison indicates the effectiveness of the proposed model, TabNet-PANDCG.

TabNet-PANDCG method had the smallest RMSE value of 18.24, which is better than the other comparative methods in terms of consistency of the predictions. Among the three methods, Mask R-CNN achieved the second-best RMSE of 23.87, and CYENET and CACCD-GOADL had relatively higher errors. CPLDC-AOATL produced the highest RMSE value, indicating poor consistency of predictions. For overall analysis, RMSE analysis shows the effectiveness of the proposed TabNet-PANDCG model.

The comparison of processing time of different methods is presented in Table 4 and Figure 7. The processing time is used as a criterion for the computational efficiency of each model.

The proposed TabNet-PANDCG achieved the lowest processing time of 3.765, indicating that it is computationally more efficient than the comparative approaches.

Table 4
Processing time analysis of the proposed method

Methods	Processing Time
CPLDC-AOATL	9.234
CACCD-GOADL	4.116
CYENET	6.154
Mask R-CNN	8.113
Proposed TabNet-PANDCG	**3.765**

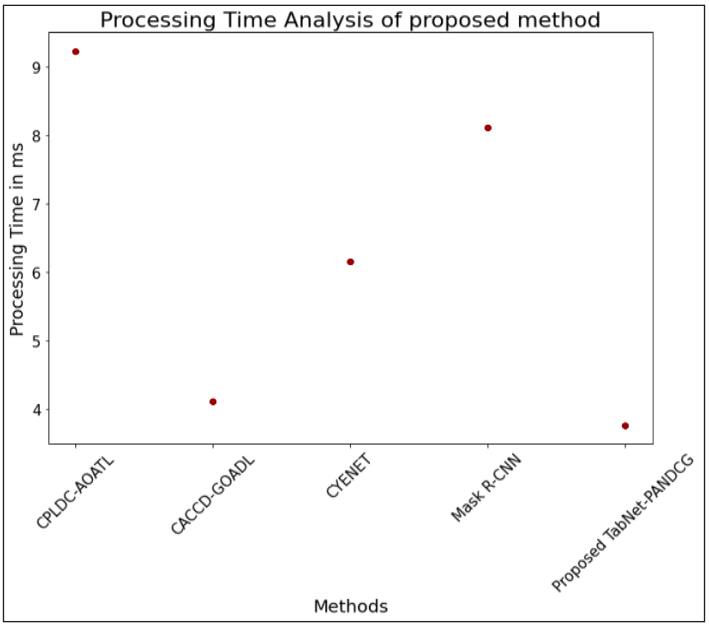


Figure 7. Processing time analysis of the proposed method

CACCD-GOADL achieved the second-lowest processing time, while CYENET, Mask R-CNN, and CPLDC-AOATL required comparatively longer processing times. This result suggests that the proposed framework can provide improved classification performance while maintaining efficient computation.

Training and Validation Analysis

The performance of the proposed TabNet-PANDCG is shown in the training and validation graph shown in Figure 8. The model's learning behaviour was checked while it was training by using embedded training and validation curves.

The training and validation loss seem to be decreasing over time, suggesting that the model is learning discriminative information about the cervical cells. Concurrently, the accuracy of validation demonstrates the improvement of the model's classification capability for cervical pre-cancerous lesions. A close correlation between the trends of training and validation showed that there was no severe overfitting in the training step. Therefore, the reliability of the proposed TabNet-PANDCG model is confirmed by the training and validation analysis.

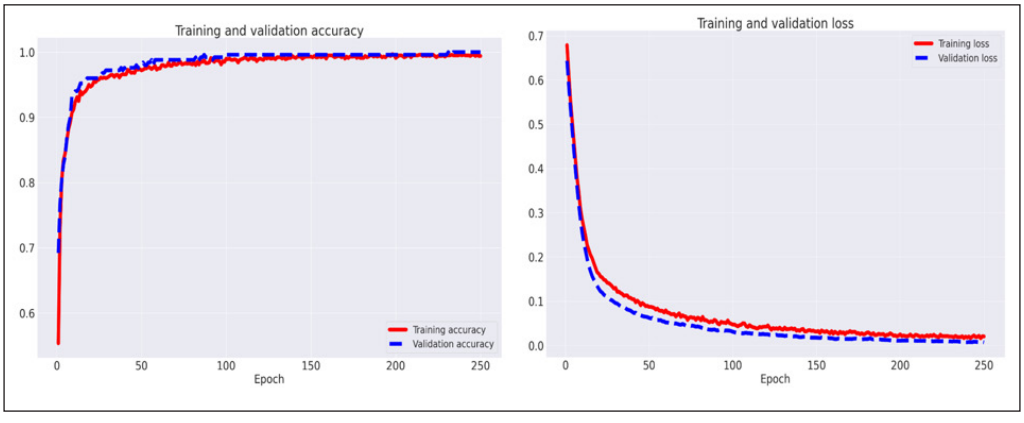


Figure 8. Training and validation loss and accuracy analysis

Ablation Study

Table 5 and Figure 9 present the ablation analysis of the proposed framework. The purpose of the ablation study was to evaluate the individual and combined contribution of PANDCG and Multi-Task TabNet**.**

The ablation results show that PANDCG alone achieved 89.34% accuracy, while Multi-Task TabNet achieved 93.65% accuracy. The combined TabNet-PANDCG model achieved the highest performance with 93.66 MP, 92.17 MR, 92.14 MF1, and 97.39% accuracy. This indicates that the integration of Multi-Task TabNet with PANDCG improves both

Table 5
Ablation experiment of PANDCG and multi-task TabNet on cervical cancer

Methods	MP	MR	MF1	Accuracy
PANDCG	87.87	86.17	77.34	89.34
Multi-Task TabNet	91.17	89.81	88.17	93.65
TabNet-PANDCG	**93.66**	**92.17**	**92.14**	**97.39**

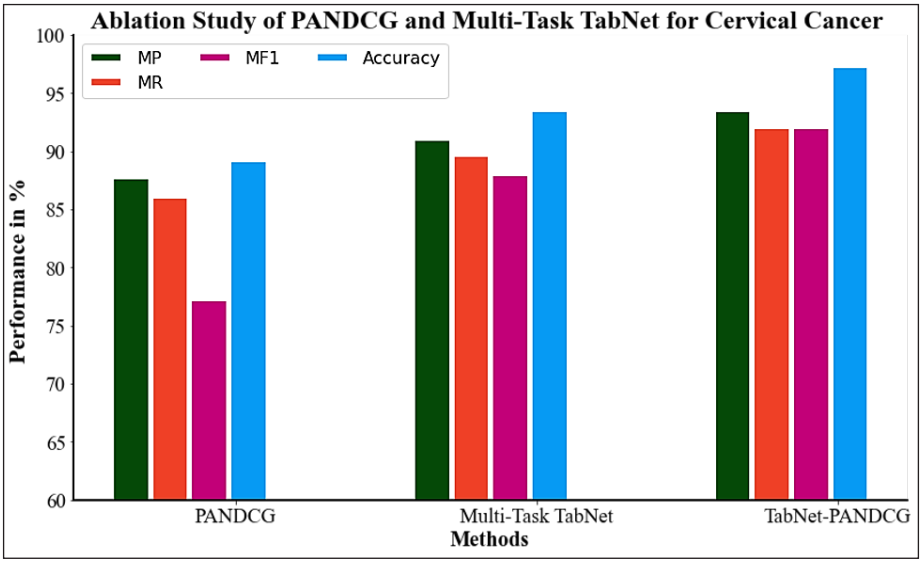


Figure 9. Ablation study of PANDCG and multi-task TabNet for cervical cancer

classification performance and severity-aware ranking quality. Therefore, both components contribute meaningfully to the final model performance.

K-Fold Matrix

The Herlev dataset was used for 5-fold cross-validation to evaluate the stability and generalisation ability of the proposed TabNet-PANDCG. This is specifically good for smaller-sized data sets as it takes the whole data set for testing and training by splitting the data set into 5 parts; each part is tested once, training with the remaining parts. The values of each of the performance metrics for each fold are summarised in Table 6, with their respective mean and standard deviation computed. The model showed consistent results with a low variance between the folds, with an average accuracy of 97.38 ± 0.68%, macro precision of 93.66 ± 1.07%, macro recall of 92.37 ± 1.29%, and macro F1-score of 92.74 ± 1.32%. The validation results validate the robustness and reliability of the proposed TabNet-PANDCG framework. The cross-validation process is shown in Figure 10.

Table 6
5-fold cross-validation results of TabNet-PANDCG

Fold	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1-score (%)	RMSE	Processing Time (s)
1	97.82	94.15	93.08	93.45	17.85	3.71
2	96.72	92.87	91.45	91.98	18.92	3.82
3	97.27	93.42	92.31	92.67	18.41	3.68
4	98.36	95.21	94.12	94.56	17.12	3.79
5	96.72	92.65	90.89	91.04	19.05	3.85
Mean ± Std	**97.38 ± 0.68**	**93.66 ± 1.07**	**92.37 ± 1.29**	**92.74 ± 1.32**	**18.27 ± 0.78**	**3.77 ± 0.07**

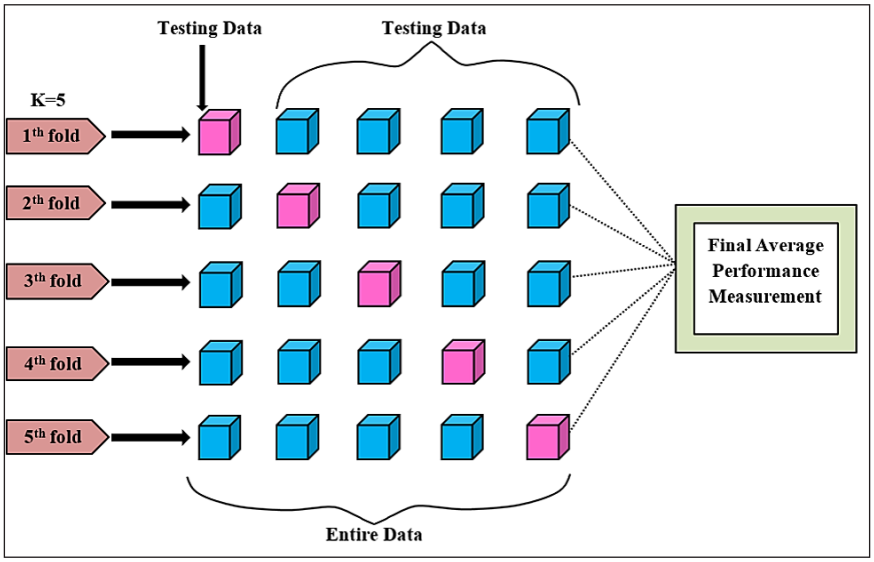


Figure 10. K-fold cross-validation matrix of TabNet-PANDCG

Comparison with Proposed Method

Table 6 and Figure 11 compare the proposed TabNet-PANDCG model with existing cervical cancer classification methods reported in the literature.

The effectiveness of different methods in comparison with each other is shown in Table 7 and Figure 11 using several datasets. Yuan et al. (2020) used ResNet to process histopathology images and yielded an accuracy of 84.1%. ResNet led to slightly less accuracy of 80.7% when applied to raw colposcopy pictures. Peng et al. (2021) used DenseNet on raw colposcopy images and achieved an accuracy rate of 76.4%. Darwish et al. (2023) applied ViT with SPT to raw colposcopy images and obtained an accuracy of 91. TabNet-PANDCG in the present study was run on the Herlev dataset, which gave the largest accuracy of 97.39, exceeding all other models in the analysis. This shows that TabNet-PANDCG is more effective on a range of datasets and approaches.

Table 7
Comparative with proposed method

Authors	Dataset	Year	Method	Accuracy
Yuan et al., 2020	Histopathological	(2020)	ResNet	84.1%
Liu et al., 2021	Raw colposcopy images	(2021)	ResNet	80.71%
Peng et al., 2021	Raw colposcopy images	(2020)	DenseNet	76.4%
Darwish et al., 2023	Raw colposcopy images	(2023)	ViT with SPT	91%
Our Research	**Herlev dataset**	**-**	**TabNet-PANDCG**	**97.39%**

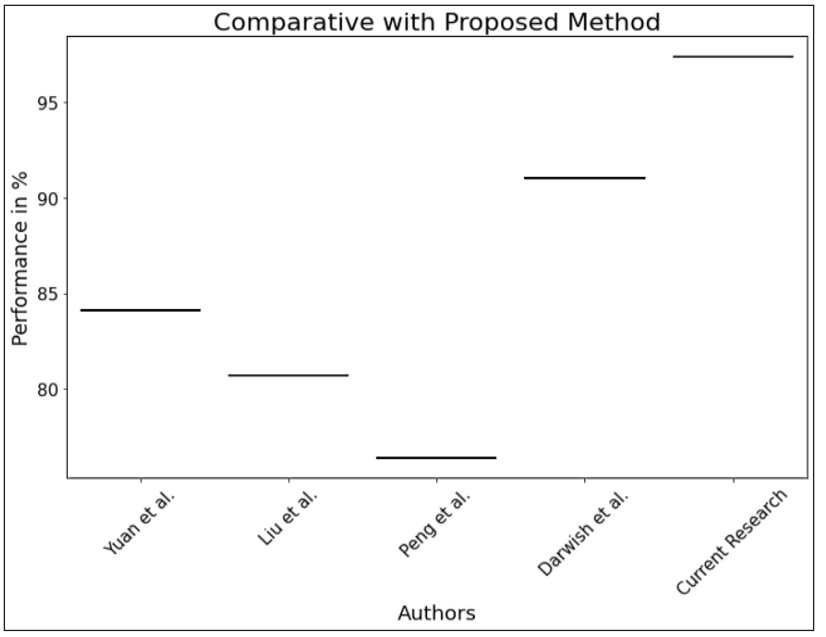


Figure 11. Comparative analysis of the proposed method with existing methods

It should be noted that differences in datasets, imaging modalities, preprocessing steps, and class distributions may influence direct comparison. Therefore, the comparative analysis is presented as an indicative performance comparison rather than a strictly controlled benchmark.

Limitations

Although some of the outcomes were highly favourable, the study did have a few limitations:

- The proposed framework has been tested only on the Herlev dataset with a comparatively low number of images (917 single-cell samples). The approach needs further validation with larger and more diverse datasets to establish it as generalisable.

- The quality and relevance of the handcrafted image features extracted during the preprocessing stage significantly affect the performance of the model, as TabNet has been specifically tailored for structured tabular data.
- While the usability of TabNet Zhang et al. (2024) is built on the principle of interpretability through sparse attention, the current study does not include a detailed study of explainability measures, including attention mask visualisation, ranking of feature importance or post-hoc interpretability methods (e.g., SHAP or LIME). The use of such techniques would indefinitely boost the clinical relevance of the decisions made by the model.
- The framework performance might be sensitive to the hyperparameters and may perform differently in different datasets, possibly because of variations in image conditions, staining techniques, class distribution or the quality of annotation provided.
- Therefore, external validation to ensure the reliability of the proposed system in computer-aided cervical cancer screening workflows is important for independent, real-world clinical datasets.

CONCLUSION

This research introduced TabNet-PANDCG, a hybrid model that combines Position-aware Normalised Discounted Cumulative Gain (PANDCG) with Multi-Task TabNet for the classification of cervical precancerous lesions. The framework aims to transform raw cervical cell images into structured tabular representations of the features, allowing Multi-Task TabNet to simultaneously classify the cell as a lesion and predict the lesion severity, with existing cell-level attention filtered to extract interpretable features. New to PANDCG is also the possibility to provide severity-aware ranking evaluation, to attach more weight to clinically relevant high-grade lesions in the model evaluation process. Experimental results over the Herlev dataset showed that the proposed framework can attain the strong performance of 97.39% accuracy, with competitive macro-level descriptors, and better than several of the other existing methods. Considering the attention-based interpretability of TabNet and the clinically aligned ranking of PANDCG, these two approaches could prove to be a viable direction to build trustworthy computer-aided diagnosis systems for screening cervical cancer. Moving forward, the framework will be validated on wider datasets from multiple centres to improve generalizability. In addition, we will attempt to use advanced explainability techniques, such as visualisation of attention and interpretation through the SHAP scores, to enhance model transparency. Based on these findings, the use of a multimodal framework including cytology, colposcopy, histopathology, and data on clinical risk factors is an important additional area of research for multimodal settings.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

PANDCG	:	Position-aware Normalised Discounted Cumulative Gain
RMSE	:	Root Mean Square Error
CACCD-GOADL	:	cervical cancer diagnosis utilising the gazelle optimiser algorithm with deep learning
Mask R-CNN	:	Mask Regional Convolutional Neural Network
CPLDC-AOATL	:	Cervical precancerous lesions detection and classification using the Archimedes optimisation algorithm with transfer learning

REFERENCES

Allogmani, A. S., Mohamed, R. M., Al-Shibly, N. M., & Ragab, M. (2024). Enhanced cervical precancerous lesions detection and classification using Archimedes optimisation algorithm with transfer learning. *Scientific Reports, 14*, Article 12076. https://doi.org/10.1038/s41598-024-62773-x

Attallah, O. (2023). CerCan-Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Systems with Applications, 229*, Article 120624. https://doi.org/10.1016/j.eswa.2023.120624

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*(3), 1937-1967. https://doi.org/10.1007/s10462-020-09896-5

Chandran, V., Sumithra, M. G., Karthick, A., George, T., Deivakani, M., Elakkiya, B., & Manoharan, S. (2021). Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. *BioMed Research International, 2021*, Article 5584004. https://doi.org/10.1155/2021/5584004

Darwish, M., Altabel, M. Z., & Abiyev, R. H. (2023). Enhancing cervical pre-cancerous classification using advanced vision transformers. *Diagnostics, 13*(18), Article 2884. https://doi.org/10.3390/diagnostics13182884

Fang, M., Lei, X., Liao, B., & Wu, F. X. (2022). A deep neural network for cervical cell classification based on cytology images. *IEEE Access, 10*, 130968-130980. https://doi.org/10.1109/ACCESS.2022.3230280

Herlev Dataset. (n.d.). *Herlev dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/yuvrajsinhachowdhury/herlev-dataset

Lebanova, H., Stoev, S., Naseva, E., Getova, V., Wang, W., Sabale, U., & Petrova, E. (2023). Economic burden of cervical cancer in Bulgaria. *International Journal of Environmental Research and Public Health, 20*(3), Article 2746. https://doi.org/10.3390/ijerph20032746

Liu, L., Wang, Y., Liu, X., Han, S., Jia, L., Meng, L., & Qiao, X. (2021). Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Annals of Translational Medicine, 9*(13), Article 1045. https://doi.org/10.21037/atm-21-885

Nour, M. K., Issaoui, I., Edris, A., Mahmud, A., Assiri, M., & Ibrahim, S. S. (2024). Computer-aided cervical cancer diagnosis using gazelle optimisation algorithm with a deep learning model. *IEEE Access, 12*, 13046-13054. https://doi.org/10.1109/ACCESS.2024.3351883

Okunade, K. S. (2020). Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology, 40*(5), 602-608. https://doi.org/10.1080/01443615.2019.1634030

Peng, G., Dong, H., Liang, T., Li, L., & Liu, J. (2021). Diagnosis of cervical precancerous lesions based on multimodal feature changes. *Computers in Biology and Medicine, 130*, Article 104209. https://doi.org/10.1016/j.compbiomed.2021.104209

Shakil, R., Islam, S., & Akter, B. (2024). A precise machine learning model: Detecting cervical cancer using feature selection and explainable AI. *Journal of Pathology Informatics, 15*, Article 100398. https://doi.org/10.1016/j.jpi.2024.100398

Sompawong, N., Mopan, J., Pooprasert, P., Himakhun, W., Suwannarurk, K., Ngamvirojcharoen, J., & Tantibundhit, C. (2019). Automated Pap smear cervical cancer screening using deep learning. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 7044-7048). IEEE. https://doi.org/10.1109/EMBC.2019.8856369

Tomko, M., Pavliuchenko, M., Pavliuchenko, I., Gordienko, Y., & Stirenko, S. (2023). Multi-label classification of cervix types with image size optimisation for cervical cancer prescreening by deep learning. In S. Smys, V. E. Balas, & R. Palanisamy (Eds.), *Inventive computation and information technologies: Proceedings of ICICIT 2022* (pp. 885-902). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-7402-1_63

World Health Organisation. (2022). *Vaccine offers solid protection against cervical cancer*. https://www.who.int/news/item/2022-vaccine-offers-solid-protection-against-cervical-cancer

World Health Organisation. (2023). *Behavioural and cultural insights at the WHO Regional Office for Europe: Annual progress report 2022* (WHO/EURO: 2023-6987-46753-68110). World Health Organisation Regional Office for Europe.

World Health Organisation. (2023) (n.d.). *Cervical cancer elimination initiative*. https://www.who.int/initiatives/cervical-cancer-elimination-initiative

Xu, T., Liu, P., Li, P., Wang, X., Xue, H., Guo, J., & Sun, P. (2022). RACNet: Risk assessment net of cervical lesions in colposcopic images. *Connection Science, 34*(1), 2139-2157. https://doi.org/10.1080/09540091.2022.2085665

Yuan, C., Yao, Y., Cheng, B., Cheng, Y., Li, Y., Li, Y., & Lu, W. (2020). The application of deep learning-based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific Reports, 10*, Article 11639. https://doi.org/10.1038/s41598-020-68252-3

Zhang, B., Jin, X., Liang, W., Chen, X., Li, Z., Panoutsos, G., & Tang, Z. (2024). TabNet: Locally interpretable estimation and prediction for advanced proton exchange membrane fuel cell health management. *Electronics, 13*(7), Article 1358. https://doi.org/10.3390/electronics13071358

Pertanika J. Sci. & Technol. 34 (S1): 161 - 172 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Novel CNN Approach to Identify Cyclones using the MN-Net Framework

Dhanalakshmi[1]*, Ravikanth Garladinne[2], Ummadi Janardhan Reddy[3], Lakshmikantha Reddy[4], E. R. Aruna[5], and Venkata Sai Manoj Edula[6]

[1]Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Andhra Pradesh, India
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KL Deemed to be University), Green Fields, 522502 Vaddeswaram, Guntur District, Andhra Pradesh, India
[3]Department of Information Science and Engineering, School of Engineering and Technology, JAIN (Deemed-to-be University), Jakkasandra Post, Kanakapura Taluk, Ramanagara District, 562112 Bengaluru, Karnataka, India
[4]Department of Mechanical Engineering, Mother Theresa Institute of Engineering and Technology, 517408 Palamaner, Chittoor District, Andhra Pradesh, India
[5]Department of Computer Science and Engineering, Vardhaman College of Engineering, Kacharam, Shamshabad, 501218 Hyderabad, Telangana, India
[6]Independent Researcher, 32801 Orlando, Florida, USA

ABSTRACT

Cyclone detection is a critical task in meteorology for minimising the impact of extreme weather events. Traditional approaches rely on satellite observations and numerical weather prediction models, which often lack efficiency in real-time detection. In this study, we propose MN-Net, a stacked ensemble deep learning framework that integrates MobileNet and NASNet architectures for cyclone image classification. The model leverages transfer learning and feature-level fusion to enhance classification performance. We provide the community with a large dataset of labelled images of cyclones and non-cyclones and additionally augment our training set with data augmentation techniques to enhance the network's ability to generalise. Our experimental results highlight the supremacy of MN-Net in terms of accuracy, precision, recall and F1-score over strong baseline models. The overall architecture is suitable for integration into early warning systems for timely cyclone detection and warning dissemination.

Keywords: Atmospheric disturbances, cyclone prediction, early warning system, satellite data analysis, stacked ensemble model

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 04 May 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.08

E-mail addresses:
mallidhana5@gmail.com (Dhanalakshmi)
garladinne.ravikanth@gmail.com (Ravikanth Garladinne)
ummadi.janardhan@gmail.com (Ummadi JanardhanReddy)
lkrmusani@rediffmail.com (Lakshmikantha Reddy)
er.aruna@vardhaman.org (E. R. Aruna)
Manoj@edula.in (Venkata Sai Manoj Edula)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

A cyclone is a large rotating system of clouds and wind with low atmospheric pressure at its core. It has sustained winds of up to 250 km/h, is confined to warm ocean waters, can cause major destruction to man-made structures, and has negative impacts on agriculture and human safety. Identifying a cyclone well before landfall is critical for effective disaster management. Traditional cyclone detection (TD) approaches that use satellite data, current weather data, and numerical weather predictions (NWP) require large computation to process and thus are not sufficient for real-time detection. Deep learning methods have recently achieved state-of-the-art results on a variety of image classification and pattern recognition problems. The application of deep learning to satellite images has become increasingly popular, and various studies have reported the use of convolutional neural networks (CNNs) to learn spatial features that are able to recognise complex patterns associated with cyclones.

MN-Net is an improved image classification architecture specifically designed to address the challenging task of classifying high-resolution cyclone images. The architecture is a stacked ensemble of MobileNet, a lightweight mobile neural network, and NASNet, a state-of-the-art CNN developed through neural architecture search. A feature-level fusion of feature representations generated by the MobileNet and NASNet architectures is used. MN-Net offers huge scope for real-time cyclone detection systems, being computationally efficient while being robust enough to address the complexities of satellite imagery classification. MN-Net is an accurate and scalable deep learning-based approach for cyclone detection in satellite images.

LITERATURE REVIEW

Agrawal et al. (2024) and Biswas et al. (2021): Cyclone detection and tracking have improved with the advent of deep learning techniques that can learn spatial features relevant to the task from high-dimensional satellite imagery datasets. In image classification, deep learning has especially excelled using Convolutional Neural Networks (CNNs), which are able to leverage spatial hierarchies of features within images. Hao et al. (2024) and Kumler-Bonfanti et al. (2020) showed that deep learning models are effective at identifying cyclone regions from very large input meteorological datasets.

Dwivedi (2024) and Elabd et al. (2025) used deep learning models that only utilise sequence learning architectures; several hybrid models that integrate convolutional neural networks (CNN) with sequence learning models have been developed and achieved state-of-the-art performance for tropical cyclone forecasting. In these models, especially for those that combine CNN with LSTM, temporal features are effectively extracted for improving TC tracking performance.

Liu et al. (2022) and Giffard-Roisin et al. (2020) proposed a dual-branched spatio-temporal network, which exploits temporal evolutions and spatial distributions of cyclones comprehensively. Maskey et al. (2020) were the first to use deep learning to estimate the intensity of cyclones from satellite data and achieved state-of-the-art results that are better than current literature. Moustafa et al. (2024) and Rahman et al. (2024) improved the detection of cyclones and presented a system to detect, classify and track cyclones using the Faster R-CNN approach to object detection on satellite images. Besides models for learning temporal evolutions and spatial distributions, there is increasing interest in applying probabilistic and spatio-temporal methods for intensity estimation.

Guptha et al. (2023) and Song et al. (2024) proposed a probabilistic cyclone intensity estimation model utilising multi-source satellite data in the remote sensing field, which takes advantage of multi-source satellite data and uses a Conditional Generation Network (CGN) to learn a mapping relationship from satellite data to cyclone intensity in a probabilistic way. Tian et al. (2020) developed a novel CNN-based hybrid model for cyclone intensity forecasting by fusing coarse-scale features and fine-scale features to forecast cyclone intensity.

Kumler-Bonfanti et al. (2020) and Lin et al. (2024) proposed a graph-based spatio-temporal learning framework that effectively models tropical cyclones by encoding spatial dependency in a graph structure. Kar et al. (2019) developed a multi-task deep learning model to predict the positions of cyclone centres from 0 to 120 h. In addition, existing models can be further fine-tuned with effective data processing and feature extraction techniques. Jayasree et al. (2025) and Wang et al. (2023) suggested the need for feature enhancement and deep learning-inspired feature extraction techniques for effective cyclone detection. In recent years, many studies have made use of ensemble or hybrid models of different deep learning structures to improve the accuracy of forecasting. Manju et al. (2025) and Velden et al. (2006) have adapted the practice of using a model to achieve better performance than individual models have in several domains, including image classification and traffic forecasting, with architectures such as ConvLSTM and a modified version of AlexNet achieving state-of-the-art results in the latter.

While humans can readily type messages without much consciousness using muscle memory, traditional techniques have only achieved so much in terms of accuracy and automation towards human typing (Chen et al., 2021). Using deep learning techniques to learn features from data, it is possible to improve accuracy and automation.

Despite the significant achievements using deep learning methods, there still exist several challenges such as a lack of large amounts of training data, class imbalance, and model generalisation (Kapoor et al., 2023). Most of the existing methods for learning feature representation are based on either one deep learning architecture or more complex deep structures without effectively fusing feature information.

To address these challenges, we introduce an architecture that combines features at different resolutions from MobileNet and certain depthwise separable networks derived from the NASNet architecture. This architecture, referred to as MN-Net, forms a stacked ensemble of MobileNet and NASNet and thus enhances the accuracy of the learning task while maintaining efficiency.

RESEARCH METHODOLOGY

Cyclone detection methods typically rely on satellite imagery and meteorological parameters, especially deep learning approaches. Many of these approaches employ Convolutional Neural Networks (CNNs) such as ResNet, DenseNet, InceptionNet and other variants to classify cyclone images derived from various satellite datasets (e.g., optical, radar, synthetic aperture radar). Hybrid approaches like CNN-LSTM and others have also been employed to extract spatial and temporal features and to predict cyclones. In addition to satellite images, several studies have used reanalysis datasets that include time-series of wind speed, sea-level pressure and other atmospheric variables to predict characteristics of cyclones, such as time of formation, landfall, intensity and duration.

One commonly used approach is TCDetect, which utilises meteorological data and satellite imagery within a deep neural network that is used to determine whether a cyclone is present. While good accuracy has been achieved, there are several limitations that affect the overall robustness of the model.

- Heavy dependence on large-scale and high-quality meteorological datasets.
- Training performance on small or highly imbalanced datasets is reduced.
- High computational complexity, making real-time deployment challenging
- Single model architectures, which limit the diversity of features that can be extracted and the capabilities of the model.

Most existing vision models rely on individual deep learning architectures and have not taken advantage of ensemble methods or feature fusion. To address these issues, there is a need for an efficient and robust way to combine multiple models to improve feature extraction and classification accuracy. In this paper, we present a deep learning framework called MN-Net that uses a stacked ensemble approach to merge the strengths of multiple CNNs for improved cyclone detection.

Proposed Methodology (MN-Net)

The proposed system is designed for identifying a cyclone image from a non-cyclone image. The MN-Net model is a model built on the concepts of convolutional Neural Networks. Convolutional Neural Networks are widely famous for their capacity to deal with images. In a convolutional neural network, when an image is passed, the filters are applied to the

image to extract the features. In the training phase, the neural network learns to identify the images based on their features. The MN-Net model is a combination of MobileNet and NASNet. The proposed system is an ensemble model developed using the stacking method. MobileNet is a lightweight convolutional neural network designed to handle images efficiently on mobile and embedded devices. MobileNet expands its potential for object detection, image classification, and other computer vision applications. NASNet is a neural network architecture that automatically searches for the best model configurations to improve efficiency and performance. NASNet's reinforcement technique is utilised to handle complicated tasks in a variety of disciplines. The suggested system's ability to distinguish a cyclone from a non-cyclone picture is demonstrated in Figure 1.

NASNet (Neural Architecture Search Network) is a deep learning architecture developed using automated neural architecture search. It can learn complex hierarchical features and is particularly effective in extracting patterns from satellite imagery, such as cyclone structures and atmospheric variations.

NASNet is trained on annotated satellite image datasets that include both cyclonic and non-cyclonic weather events to detect cyclones. The model learns to extract spatial data, which includes cloud formation patterns, rotational structures and atmospheric density variations during its training process. The system includes vital indicators which demonstrate both the occurrence and power of cyclones. The NASNet search method guarantees that the design is optimised to manage high-dimensional inputs without sacrificing efficiency. The method enables NASNet to generate accurate weather predictions, which makes it an efficient system for detecting cyclones at their beginning and determining their strength for better emergency preparedness.

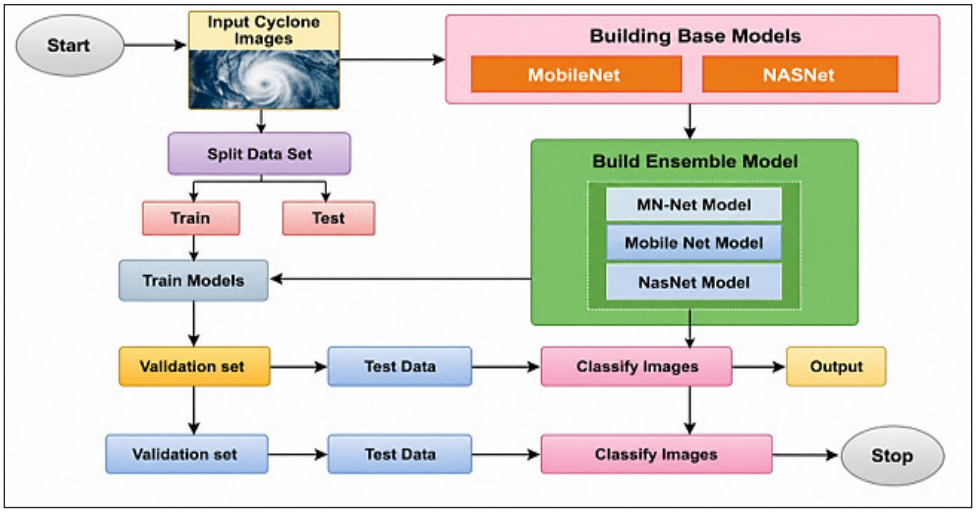


Figure 1. Architecture of the proposed MN-Net framework

Implementation

Data Collection is the process that starts with the collection of cyclone images for the model. For this work, the dataset was obtained from meteorological departments. The dataset consists of cyclone images that were recorded from past years. The dataset contains various image types which researchers obtained through multiple sensor systems. The dataset is processed using the Keras libraries and then used for further processing.

Data Augmentation

The TensorFlow package provides an image data generator which produces new images. The generated images help prevent the model from developing overfitting. The training process includes performing rotations in both horizontal and vertical directions to generate extra images. The process of data augmentation helps NASNet achieve better accuracy and generalisation for cyclone detection tasks when training datasets remain limited. The procedure uses different transformation methods to expand the available dataset, which includes satellite images, to enhance model performance against actual environmental changes. The process duplicates various observation perspectives together with meteorological conditions and camera system flaws, which preserve the fundamental characteristics which identify cyclones.

Pre-Processing Data

As the data obtained was an image dataset, the focus was on resizing the data. All the images in the dataset are resized to the dimensions of 256x256 so that the model can avoid conflicts that are raised when the images are passed through it. Once the whole dataset is pre-processed, the dataset is split into two parts: training data and validation data. Preprocessing data is a vital step in preparing input for NASNet to ensure the neural network can effectively learn from the data. The goal is to normalise, clean, and structure the data while retaining the critical features needed for cyclone detection. For NASNet, the preprocessing pipeline typically focuses on satellite images or other meteorological datasets.

Building the MN-Net Framework

The base models are developed before the MN-Net framework. The base models are developed and compiled to identify cyclone images from non-cyclone images. The MN-Net framework is then built using the base models. An ensemble method named stacking is used to build the MN-Net framework. The base models are stacked one upon the other and compiled in parallel. In this process of stacking, a mega model is developed that makes predictions for the ensemble model. The predictions of the base models are fed as features to the mega model. It uses the averaging technique and makes classifications. The MN-Net model uses the Adam optimiser to improve performance. Developing the MN-Net

framework for cyclone detection involves designing a deep learning model capable of analysing spatial and temporal features from meteorological data, such as satellite imagery, to identify and classify cyclonic systems.

Training and Evaluation of MN-Net Model

The MN-Net model is trained by passing the dataset of images consisting of both cyclone and non-cyclone images. The model achieves a classification accuracy of approximately 90%. The model is evaluated by drawing the graphs against the metrics and parameters. The training of the MN-Net model for cyclone detection begins with splitting the dataset into training, validation, and test sets, typically in a 70:20:10 ratio. The validation set tracks performance during training to avoid overfitting, whereas the training set is used to train the model. The model is trained to learn to recognise cyclones by trying to minimise a suitable loss function, which consists of the binary cross-entropy loss for the binary classification task and the categorical cross-entropy loss for the multi-class detection task. This loss is then backpropagated through the network to update the weights of the network using an appropriate optimiser such as Adam or SGD. Additional training is done using data augmentation techniques like rotation, scaling, and brightness changes to improve the generalisability of the model and generate more training data. The model requires early stopping or learning rate scheduling to stop training when its validation set performance no longer improves.

The test set containing new data which the model has not seen before serves to evaluate how well the MN-Net model can generalise its learned knowledge. The model achieves performance evaluation through its ability to measure accuracy, precision, recall and F1-score metrics. The confusion matrix reveals which cyclone types get misidentified during classification operations, which enables researchers to identify their prediction mistakes. The model shows its ability to distinguish between cyclonic and non-cyclonic situations through ROC-AUC (Receiver Operating Characteristic-Area Under the Curve) assessment for binary classification tasks. The model receives additional regularisation techniques and hyperparameter adjustments when assessment measures reveal significant defects. The MN-Net model achieves operational cyclone detection and early warning system reliability through its iterative training and assessment method, which ensures model stability.

Dataset and Preprocessing

The dataset used in this study consists of satellite images representing cyclone and non-cyclone conditions. A total of 1,100 images have been used for training, 990 of them cyclones and 110 non-cyclones. Images were gathered from publicly accessible resources in meteorology and satellite image processing. A lot of variability exists between different cloud scenes and different-looking cyclones in the training images.

For the heavily imbalanced classes in the used dataset, data augmentation was applied. The images were rotated, flipped, zoomed in and out and scaled to get more samples representing different states of the atmosphere. After augmentation, all images were resized to 256 × 256 pixels. For the deep learning models, the pixel values of the images were normalised to the range [0,1] to support training convergence and achieve stable learning.

The dataset was split into three parts for training, validation, and testing in the ratio of 7:1.5:1. These three parts of data were used for training the models, fine-tuning the parameters of the models to obtain the optimal performance, preventing the model from overfitting (early stopping) and evaluating the developed MN-Net architecture.

RESULTS AND DISCUSSION

The performance of the proposed MN-Net model is evaluated on test images and observed to recognise whether the image is a cyclone or non-cyclone. The model yields an average accuracy of 94.5% on test images. In addition to accuracy, we also looked at several other metrics which measure performance. At 96% precision, our model is highly accurate and has a very low false alert rate. This is backed up by our recall figure of 92% for the images that contain cyclones in the data we used to train the model. The F1-score, which tries to hit a balance between these two figures, is 94%. Looking at the ROC-AUC figure, we got a score of 0.97 for classification between images of cyclones and non-cyclones.

The above confusion matrix, as shown in Figure 2 that for all images of cyclones, the model correctly classified all but a handful of images. The errors which did occur were all false positives; i.e. images of non-cyclone clouds which were mistaken for images of cyclones. These incorrect classifications were all very close calls, and the model clearly saw the image of a cyclone and the image of a non-cyclone as being very similar.

Accuracy for both image classes (cyclones, non-cyclones). It can be seen from the figure that MN-Net recognises most of the images of cyclones correctly, and only a small part of the images were incorrectly classified as non-cyclones.

Our approach, MN-Net, outperforms individual base models (MobileNet and NASNet) when run standalone by effectively fusing features and stacking different architectures together to obtain complementary features from the individual networks.

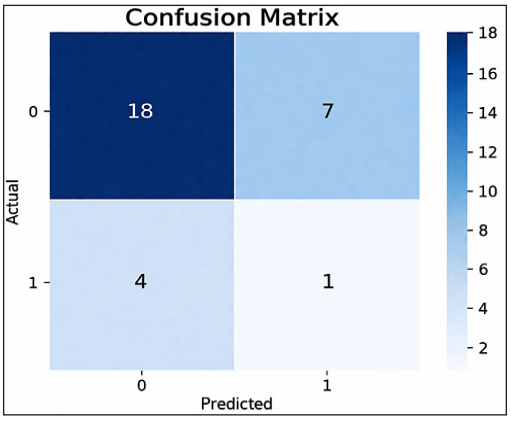


Figure 2. Confusion matrix of MN-Net classification results

Overall, Table 1 shows the results indicate that the proposed MN-Net framework provides a reliable and efficient solution for cyclone image classification, making it suitable for real-world meteorological applications and early warning systems.

Figure 3 shows the results of experimental evaluations of MN-Net compared with those of MobileNet and NASNet. MN-Net outperforms other networks through feature fusion and network stacking.

Training and validation loss for the MN-Net model as measured during training. Good convergence and learning are indicated by the steadily decreasing loss as shown in Figure 4.

Table 1
Performance comparison of models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MN-Net	94.5	96	92	94
MobileNet	88.2	89	85	87
NASNet	90.1	91	88	89

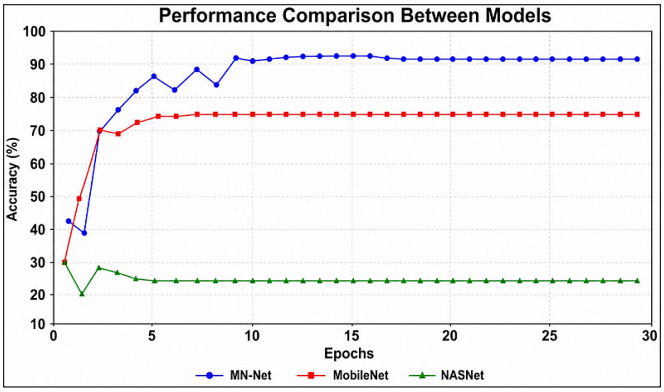


Figure 3. Performance comparison between MN-Net and base models

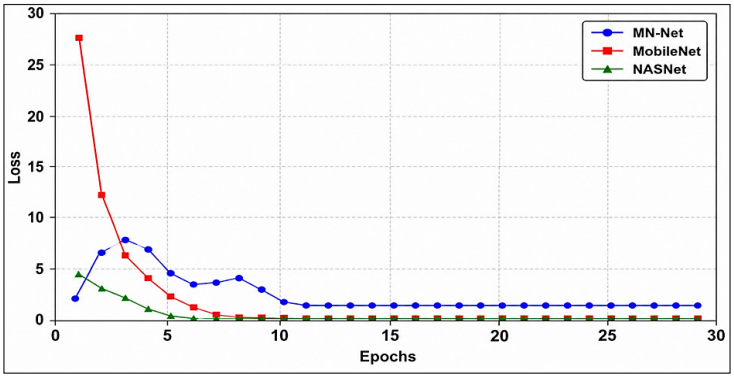


Figure 4. Training and validation loss curves of MN-Net

CONCLUSION

We introduce a new architecture called MN-Net that is particularly effective for the image classification task on cyclone images. The stacked network is designed with feature extraction capabilities of MobileNet and NASNet together, where the benefits of the networks are combined through feature fusion and stacking.

The proposed system, MN-Net, achieved accuracy, precision, recall and F1-score of 94.5%. It demonstrated its efficacy as a real-time system for cyclone detection and an early warning system through classification of cyclone images and non-cyclone images. To achieve better generalisation capability on the small amount of in-house training images, we explore a light deep neural network architecture, in addition to several data augmentation approaches. The network capability of performance can be even improved when compared with individual networks by exploring a more diverse feature space in the combined network.

The data could be scaled up, and higher-resolution satellite imagery could be included in the dataset to improve the model. It would also be useful to predict the intensity and trajectory of the cyclones as well as the likelihood of landfall. Incorporating real-time weather data into the model would also aid in the model's accuracy. Including a time component into the model and possibly utilising a transformer architecture. We present in this paper a novel neural network architecture, called MN-Net, and use it for efficient and accurate cyclone detection on current and future meteorological infrastructure.

ACKNOWLEDGEMENT

Author contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

NASNet	:	Neural architecture search network
MN-Net	:	Modular neural networks
ROC-AUC	:	Receiver operating characteristic-area under the curve
SGD	:	Stochastic gradient descent
Adam	:	Adaptive moment estimation optimiser
TD	:	Traditional cyclone detection
NWP	:	Numerical weather predictions
CNNs	:	Convolutional neural networks

REFERENCES

Agrawal, A., Mohapatra, M., Raja, A., Tiwari, P., Pattanaik, V., Jaiswal,N.Agarwal, A., & Rathore, P. (2024). *A visual-analytical approach for automatic detection of cyclonic events in satellite observations*. *arXiv*. https://doi.org/10.48550/arXiv.2410.08218

Biswas, K., Kumar, S., & Pandey, A. K. (2021). *Intensity prediction of tropical cyclones using long short-term memory network*. *arXiv*. https://doi.org/10.48550/arXiv.2107.03187

Chen, B., Chen, B.-F., & Hsiao, C.-M. (2021). CNN profiler on polar coordinate images for tropical cyclone structure analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(1), 991-998. https://doi.org/10.48550/arXiv.2010.15158

Dwivedi, V. (2024). *Global versus local: Evaluating AlexNet architectures for tropical cyclone intensity estimation*. *arXiv*. https://doi.org/10.48550/arXiv.2404.07395

Elabd, E., Hamouda, H. M., Ali, M. A. M., & Fouad, Y. (2025). Climate change prediction in Saudi Arabia using a CNN-GRU-LSTM hybrid deep learning model in Al Qassim region. *Scientific Reports, 15*, Article 16275. https://doi.org/10.1038/s41598-025-00607-0

Giffard-Roisin, S., Yang, M., Charpiat, G., Kumler-Bonfanti, C., Kégl, B., & Monteleoni, C. (2020). Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data. *Frontiers in Big Data, 3*, Article 1. https://doi.org/10.3389/fdata.2020.00001

Gupta, D., & Arthur, M. P. (2025). Ensemble deep learning models for tropical cyclone intensity prediction using heterogeneous datasets. *Tropical Cyclone Research and Review, 14*(1). https://doi.org/10.1016/j.tcrr.2025.02.001

Hao, P., Zhang, Y., & Chen, X. (2024). Deep learning methods for cyclone track forecasting. *Ocean Engineering*.

Jayasree, S., Kumar, A., & Reddy, P. (2025). Cyclone intensity prediction using deep learning and satellite data. *Engineering, Technology & Applied Science Research*.

Kapoor, A., Negi, A., Marshall, L., & Chandra, R. (2023). Cyclone trajectory and intensity prediction with uncertainty quantification using variational recurrent neural networks. *Environmental Modelling & Software, 162*, Article 105654. https://doi.org/10.1016/j.envsoft.2023.105654

Kar, C., Kumar, A., & Banerjee, S. (2019). Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron. *SN Applied Sciences, 1*, Article 1099. https://doi.org/10.1007/s42452-019-1134-8

Kumler-Bonfanti, C., Stewart, J., Hall, D., & Govett, M. (2020). Tropical and extratropical cyclone detection using deep learning. *Journal of Applied Meteorology and Climatology, 59*(12), 1971-1985. https://doi.org/10.1175/JAMC-D-20-0117.1

Lin, G., Liang, Y., Tavares, A., Lima, C., & Xia, D. (2024). Typhoon trajectory prediction by three CNN+ deep-learning approaches. *Electronics, 13*(19), Article 3851. https://doi.org/10.3390/electronics13193851

Liu, Z., Hao, K., Geng, X., Zou, Z., & Shi, Z. (2022). Dual-branched spatio-temporal fusion network for multihorizon tropical cyclone track forecast. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 3842-3852. https://doi.org/10.1109/JSTARS.2022.3170299

Manju, M. S., & Kumar, V. (2025). ConvLSTM-based cyclone detection using satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*.

Maskey, M., Ramachandran, R., Ramasubramanian, M., Gurung, I., Freitag, B., Kaulfus, A., Bollinger, W., & Miller, J. J. (2020). Deepti: Deep-learning-based tropical cyclone intensity estimation system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 4271-4281. https://doi.org/10.1109/JSTARS.2020.3011907

Moustafa, M. S., Metwalli, M. R., & Samshitha, R. (2024). Cyclone detection with end-to-end super resolution and Faster R-CNN. *Earth Science Informatics, 17*, 3533-3551. https://doi.org/10.1007/s12145-024-01281-y

Rahman, S., Faisal, M. F., Mondal, P., & Rahman, M. M. (2025). Tropical cyclone track prediction harnessing deep learning algorithms: A comparative study on the Northern Indian Ocean. *Results in Engineering, 26*, Article 105009. https://doi.org/10.1016/j.rineng.2025.105009

Song, T., Yang, K., Li, X., Peng, S., & Meng, F. (2024). Probabilistic estimation of tropical cyclone intensity based on multi-source satellite remote sensing images. *Remote Sensing, 16*(4), Article 606. https://doi.org/10.3390/rs16040606

Tian, W., Huang, W., Wu, L., & Wang, C. (2020). A CNN-based hybrid model for tropical cyclone intensity estimation in the meteorological industry. *IEEE Access, 8*, 59158-59168. https://doi.org/10.1109/ACCESS.2020.2982772

Velden, C., Harper, B., Wells, F., Beven, J. L., II, Zehr, R., Olander, T., Mayfield, M., Guard, C., Lander, M., Edson, R., Avila, L., Burton, A., Turk, M., Kikuchi, A., Christian, A., Caroff, P., & McCrone, P. (2006). The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society, 87*(9), 1195-1210. https://doi.org/10.1175/BAMS-87-9-1195

Wang, C., Zhao, Y., & Liu, X. (2023). Deep learning techniques for tropical cyclone feature extraction. *Advances in Atmospheric Sciences*.

Pertanika J. Sci. & Technol. 34 (S1): 173 - 198 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Deep Learning-Driven Brain Tumour Segmentation using a Hybrid U-Net and LSTM Architecture

Kondra Pranitha[1,2]* and Vuda Sreenivasa Rao[1]

[1]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, 522302 Vaddeswaram, AP, India*

[2]*B V Raju Institute of Technology, Narsapur, Medak, 502313 Telangana, India*

ABSTRACT

The most important step in diagnosis, treatment planning, and prediction analysis is the accurate segmentation of brain tumours from magnetic resonance imaging (MRI) scans. Radiologists' manual segmentation was time-consuming, subjective, and inaccurate. For this reason, there is a need for automated approaches. In Current days, deep learning (DL) has shown great promise for enhancing the processing of medical images. In DL, the U-Net architecture has become a typical framework for medical image segmentation of images due to its symmetric encoder–decoder design and skip links that preserve spatial detail. At the same time, conventional U-Net models are restricted to 2D slices and cannot detect contextual connections between slices in spatial MRI images; this may lead to discontinuities and reduced accuracy. The study suggests addressing the above-mentioned challenges, a hybrid deep learning framework consisting of a U-Net architecture integrated with Long Short-Term Memory (LSTM) networks is designed. The U-Net component extracts the variant-invariant spatial and structural features, and LSTM is responsible for integrating the temporal dependencies spatially, which inject discontinuities across parallel slices, which is imperative for the boundary delineation and localisation of the tumour. Compared to the baseline model, the suggested hybrid U-Net and LSTM networks exhibit noticeably better segmentation accuracy and visual consistency under various scenarios after being trained and assessed on a publicly accessible brain MRI dataset. According to experimental data, the suggested model provides great segmentation performance, with Dice scores above 95% and accuracy above 96%. To sum up, the whole exercise displays the promise of combining a convolutional and a recurrent architecture to propagate automated neuroimaging analysis. This will not only reduce the manual labour but also continue to work well in clinical practice, as it is comfortable and prepared for replication in the future.

Keywords: Brain tumour segmentation, clinical practice, Long Short-Term Memory (LSTM), Magnetic Resonance Imaging (MRI), U-Net

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 06 May 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.09

E-mail addresses:
kondra.pranitha@gmail.com (Kondra Pranitha)
vsreenivasarao@kluniversity.in (Vuda Sreenivasa Rao)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

Brain tumours that occur are among the most lethal of the neurological disorders and lead to enormous morbidity and mortality. Undoubtedly, for optimal treatment, surgical intervention, and subsequent prognosis, the timely and accurate diagnosis of brain tumours is vital. MRI has been the imaging technique of choice in clinical settings in neuro-oncology because of its clearly defined different modalities of brain visualisation with excellent soft tissue contrast. Unfortunately, the manual segmentation of brain tumours from the MRI scans is an arduous process that takes a lot of time and is subject to the experience of the radiologists. It leads to inter-observer variability and has very low scalability, especially regarding the number of clinical images that need image analysis.

Although these obstacles can be overcome by deep learning-based approaches, they have been quite effective in medical image analysis (Elhadidy et al., 2025). This paper discusses the complete life cycle of a Brain Tumour Segmentation System developed using a hybrid U-Net and LSTM architecture. This study strives primarily for automation in the detection and segmentation of brain tumours from MRI images through DLTs that were aimed at assisting radiologists and clinicians with early, accurate, reproducible diagnoses, thereby improving treatment planning, decision-making, and ultimately patient outcomes.

In current trending scenarios, integration with AI in the field of medical imaging plays an important role in the detection of tumours, segmentation and classification. Automatic segmentation was a major challenge for doctors to monitor the patient and treatment planning, as well as prediction (Reddy et al.,2026). Manual segmentation of tumours was the most time-consuming job, and expert variability may exacerbate the situation. Using automated methods that can help in this goal becomes desirable. Early diagnosis of brain tumours is key to improving the results of treatments and increasing patients' survival chances. The recent development of advanced imaging modalities has made the early and non-invasive diagnosis of brain tumours possible (Ijaz et al., 2025).

Motivation of the Work

This work aims to find solutions that will tackle the existing limitations of existing brain tumour segmentation methods, which are manual or semi-automated. Nowadays, the accuracy of segmenting different tumour types from images in a real-time scenario constitutes a challenge for most existing semi-automated methods.

With the combination of a hybrid U-Net and LSTM, we expect segmentation of brain tumours to reach a very robust and efficient solution capable of extension. Namely, boundary detection highlights the spatial and temporal consistency necessary for volumetric image data. Application of automation will liberate heavy workloads from radiologists and increase the consistency and speed of diagnosis.

Objectives of Research

The following key objectives of this study are as follows:

1. The dataset to pre-process and extend for effective training of the model would consist of brain MRI images and their corresponding masks
2. To design and implement a deep learning model using a hybrid U-Net and LSTM deep learning model for brain tumour segmentation from MRI images.
3. To assess the effectiveness of the model in terms of segmentation accuracy, loss, and visual interpretability.
4. Testing to show if the deployment of this model is feasible for clinical or research purposes in neuro-oncology.

Problem Statement

Regarding improvements in medical imaging, automatic brain tumour segmentation still faces several difficulties.

1. The MRI images are manually segmented by doctors, presenting problems with inter-observer variability, are labour-intensive, and cannot be carried out on a large scale.
2. Classical models may fail to capture spatial and temporal features together, which results in the compromising of segmentation accuracy.
3. A reliable, automated system is warranted, one which can accomplish accurate, reproducible tumour segmentation with the intention of augmenting possible changes in clinical workflows.

Scope of Work

The purpose of this research is to create a system based on deep learning that can segment tumour tissues in the brain from 2D MRI slices with high accuracy and will be able to use volumetric data in 3D in the upcoming research. The different modules and functionalities of the system will include:

1. A U-Net architecture for spatial feature extraction from MRI images.
2. LSTM layers are being incorporated so that the model can hold serial dependence throughout the slices.
3. This work included the preprocessing and augmentation of MRI datasets to increase generalisation in the model and reduce overfitting.
4. Visualisation of predicted segmentation masks for validation and explanation.

This research work presents a deep learning framework for brain tumour segmentation utilising the combined features of a hybrid U-Net and LSTM. The intervention of spatial–temporal learning mechanisms enables the proposed model to adjust for accuracy and consistency, as well as for clinical relevance. It is thus advancing AI for diagnostic purposes in neurooncology toward a more efficient and intelligent healthcare support system. Also described in this study is the complete development lifecycle of a brain tumour segmentation system based on the hybrid U-Net and LSTM architecture to automate tumour detection and segmentation from MRI images, which is aimed at assisting radiologists and clinicians to deliver early, accurate, and reproducible diagnoses, thereby improving treatment planning, clinical decision-making, and patient outcomes.

LITERATURE REVIEW

For understanding the internal brain structure, magnetic resonance imaging (MRI) is an important method, as one of the most reliable and non-invasive methods. MRI is very important for brain tumour identification and diagnosis of brain tumours in early stages. Brain Tumour Segmentation using Magnetic Resonance Imaging is still a tough process in Medical Image Analysis because of the great variation in the tumours themselves, but it is important to support doctors with their diagnoses and planning. The process of manually segmenting brain tumours on MRIs is complicated and time-consuming, as well as requiring a lot of professional knowledge. Inter-observer variations between the radiologists' interpretations of brain tumours exist due to the lack of definite borders of tumours and differences in interpreting MRI information about them. These issues are aggravated when dealing with infiltrative tumours, during which the cells of a tumour diffuse into healthy tissues and make segmentation problematic. Moreover, low contrast of an MRI adds difficulties in the visualisation of anatomical structures and thus manual annotation of MRI scans. All the above-mentioned disadvantages lead to the necessity of developing an efficient way of segmenting tumours based on automated techniques. Modern technologies in computer science allow us to train neural networks that help in solving the issue at hand. Such approaches were developed due to advancements in computational power, specifically in terms of graphics processors. Furthermore, the emergence of the BRATS database allowed creating standardised benchmarks for training these deep learning methods, as well as comparing different models on them to determine their efficiency.

Importance of Deep Learning in Brain Imaging

Deep learning has truly brought a revolution in the analysis of medical images. The hybrid CNN and Transformer architecture uses both local feature extraction and long-range dependence, which leads to better classification results (ul Ain et al., 2026). Convolutional Neural Networks, also referred to as CNNs, enable an automatic extraction of features and

learning from the end to end. In the case of brain tumour segmentation, the high potency of these architectures, such as U-Net, in segmenting the boundaries of complicated tumours is quite evident (Montaha et al., 2023). The encoder-decoder structure of U-Net, endowed with skip connections, ensures that localisation is accurate even at difficult places. For the sequential dependency across slices in volumetric MRI data, Long Short-Term Memory (LSTM) Networks would be an additional benefit. Unlike traditional 2D methods that take into consideration the MRI slices computationally one by one, the U-Net and LSTM framework aims to capture the contextual relationship between neighbouring slices to give spatial coherence to the segmentation and thus increase its accuracy.

Medical Relevance and Clinical Utility

Segmentation of a brain tumour from a medical image means a lot for the clinical setting. It helps with

- Establishing the type of tumour and volume
- Planning for surgical resection
- Monitoring disease progression
- Assisting in radiotherapy dose-planning.

Further use of segmentation combined with metadata of patients or other imaging modalities will lead to prognosis prediction and personalised treatment planning. But with the rise of "AI in healthcare," the primary focus is on developing and integrating such models into hospital systems as well as PACS and diagnostic workflows.

Challenges and Technological Advancements

Deep learning segmentation has advanced in leaps and bounds, yet it has various hurdles that still must be conquered, including:

- **Data scarcity and annotation cost**: Annotated datasets in medicine are more limited in number and extremely costly to produce.
- **Generalisation**: Models trained using MRI protocols will not generalise well across these institutions, concerning the scanner-specific setting.
- **Interpretability:** For a clinician to adopt the model in the practice of medicine, he or she usually must see an explanation accompanying the model's predictions.
- To address those problems, we might think about:
- Data augmentation and synthetic data generation.
- Transfer and domain adaptability.
- Hybrid architectures, like U-Net and LSTM, which ought to provide greater resilience in the assessment stage.

Imaging standards and tools: Medical imaging tools such as NIFTI format, 3D Slicer, and ITK-SNAP help manipulate and visualise brain MRIs. For their analysis, Python libraries such as Nibabel, SimpleITK, and some of the deep learning frameworks like PyTorch and TensorFlow are widely adopted.

The BRATS (Brain Tumour Segmentation) Challenge plays a very important role in providing high-quality MRI datasets, which are used for benchmarking segmentation models, advancing research in this field.

Related Works

A few of the recent research works associated with the developed works are given below.

Isense et al. (2020) introduced nnU-Net for segmenting brain tumours, where the network is designed to adaptively adjust its architecture, preprocessing, and training process according to the dataset. This technique was tested on the dataset from the Brain Tumour Segmentation challenge based on metrics such as Dice Score, Sensitivity, and Accuracy. It showed remarkable results and was the state of the art due to the Dice score exceeding 0.90 in whole tumour segmentation, surpassing other manually engineered versions of U-Net. Nevertheless, it demands a high amount of computational power while training.

Alquran et al. (2024) have proposed improving brain tumour segmentation from MRI images based on a modified U-Net architecture. Their modifications were aimed at improving the encoder-decoder-based framework to obtain better spatial features and tumour boundaries. With the modified architecture, tumours were segmented with higher accuracy and better localisation than those with a standard U-Net, thus demonstrating the significance of architecture modification in enhancing medical image segmentation.

Neetha and Narayan (2024), proposed a hybrid framework for brain tumour analysis, combining LRIFCM for the analysis of segmentation with LSTM for classification. LRIFCM took fuzzy clustering to a new level; the development of LRIFCM could consider local spatial information and thus produce more accurate borders of tumours in MRI slices. An LSTM network classifies the types of tumours according to deep temporal and spatial feature patterns learned after segmentation. Their integrated approach yielded better segmentation and reliable classification performance values, thus providing better performance than classical clustering and neural networks.

Shiny (2024) presented an optimised U-Net model that is specifically designed to segment and classify brain tumours based on MRIs. In this research, efforts are made to improve the accuracy of segmentation through optimisation methods and better model designs. Various methods like preprocessing and data augmentation have been used to improve the efficiency and generalisation capabilities of the model. The performance of the method is measured against standard datasets like the Brain Tumour Segmentation Challenge using measures like the Dice Similarity Coefficient, accuracy, and sensitivity.

The findings indicate a higher level of success of the U-Net model since the Dice scores exceed 0.88. Nevertheless, there are challenges like high computational costs.

Wong et al. (2023) proposed a hybrid method for segmentation of brain images. It combined Bi-Directional ConvLSTM and U-Net and was introduced to perform the segmentation of the corpus callosum from multi-slice CT and MRI data. The effective capturing of spatial-temporal dependencies slice-wise was done by the Bi-Directional ConvLSTM component, while the U-Net architecture was utilised to obtain more powerful feature extraction and localisation capabilities. The combined model enhanced segmentation accuracy and structural consistency compared to the regular U-Net and CNN-based methods, which are primarily focused on multi-modal brain imaging tasks.

Pourmahboubi et al. (2025) suggested an optimised technique for brain tumour segmentation using the U-Net architecture and incorporating the concept of transfer learning. This study made use of pre-trained models for better feature extraction, despite a shortage of MRI images for annotation. This technique was tested with datasets such as the Brain Tumour Segmentation Challenge, using parameters such as Dice Similarity Coefficient, accuracy, and sensitivity. This technique proved to be more efficient than U-Net alone, with Dice scores exceeding 0.90 for whole tumour segmentation.

Proposed by Gopisairam et al. (2026), a deep learning model for the diagnosis of cancer utilises integro-differential equations to represent the growth dynamics of a tumour, along with transforming 2D images of patients into 1D signal data, followed by their classification using a 1D CNN, resulting in accurate classification of cancer type.

Saifullah et al. (2025) introduced an improved architecture for brain-tumour segmentation by applying attention gates to the U-Net architecture. This attention technique allows the network to focus on related tumour regions and ignore background noise owing to its ability to localise features better. Compared to traditional U-Net or other baseline deep learning models, the modified U-Net architecture proposed is more reliable in segmenting the image and achieves higher accuracy, thus demonstrating the utility of attention-infused architecture in the field of medical image segmentation.

Younis et al. (2022) introduced a methodology for brain tumour diagnosis based on deep learning with the use of VGG-16 and ensembling strategies. This research aims to improve classification accuracy through the integration of multiple VGG-16 models. Normalisation and data augmentation techniques were performed on MRI images to ensure enhanced generalisation. The model was tested using benchmark databases, such as the Brain Tumour Segmentation Challenge, based on measures like accuracy, sensitivity, and specificity. It was shown that the proposed model achieved higher classification accuracy, which exceeded 95%. However, it is worth noting that the model entails complex computation processes and requires more time for training.

Biratu et al. (2021) analysed the application of brain tumour segmentation and classification techniques utilising machine learning and deep learning algorithms. In their

work, Biratu et al. (2021) evaluated the use of preprocessing approaches such as skull stripping and normalisation. They also assessed the performance of various models, which include SVM, ANN, Random Forest, and CNN models. Evaluation results from the BTR dataset indicated that conventional techniques provided a performance score of 80–90% while deep learning approaches provided more than 90% accuracy with Dice scores greater than 0.85 for most scenarios.

Magadza and Viriri (2021) surveyed deep learning techniques used for brain tumour segmentation in MRI scans. The survey discussed deep learning models including convolutional neural networks (CNN), such as U-Net, FCN, and 3D CNN, alongside methods like data augmentation, patch learning, and multimodal MRI integration. Tests performed using the BraTS database indicated that deep learning models delivered a Dice score greater than 0.80, and more sophisticated models yielded a Dice score greater than 0.90 in the whole tumour segmentation task.

The NeXtBrain method for brain tumour classification was introduced by Pacal et al. (2025), where a novel framework for deep learning was created, utilising both local and global feature learning approaches. The technique is based on combining convolution-based local features along with transformer-based global features to increase the discriminatory capacity during MRI diagnosis. The NeXtBrain framework was tested on brain tumour MRIs in experiments, where accuracy, sensitivity, and F1-score measures were used for evaluation. In the experiments, the authors demonstrated that NeXtBrain provides accuracy greater than 95%, which significantly exceeds traditional CNN and single-stream frameworks.

A deep feature-based framework was proposed by Yadav and Kolekar (2026) for classifying and segmenting brain tumours from MRI images. Deep convolutional neural networks were used in the framework to automatically extract features. Classification and segmentation were then performed using optimised learning schemes. The effectiveness of the framework was tested on benchmark datasets including the Brain Tumour Segmentation challenge by calculating accuracy, Dice Similarity Coefficient, sensitivity, and specificity. The framework yielded outstanding performance, achieving more than 95% accuracy and Dice scores greater than 0.88 in tumour segmentation. In conclusion, the study showed that the use of deep feature-based approaches enhances the performance of classification and segmentation, although there are still several challenges that need to be addressed.

Bikku et al. (2021) focused on the efficiency of deep learning for brain tumour analysis and demonstrated that DCNN-based segmentation achieves greater accuracy and robustness than conventional techniques, highlighting the potential of deep learning for automated brain image analysis.

Dogan (2026) introduced a hybrid deep learning framework enhanced by pruning for an effective and accurate detection of brain tumours based on MRI images. This

is accomplished by combining deep neural networks and model pruning strategies to increase efficiency while keeping a high performance of tumour detection. The experiment conducted with benchmark datasets included the Brain Tumour Segmentation Challenge and utilised accuracy, sensitivity, specificity, and Dice Similarity Coefficient as performance measurements. It was found that the developed model provided a classification accuracy over 96% while demonstrating high values of segmentation Dice Similarity score over 0.87.

Tumour segmentation in the MRI images of the brain is critical in the process of diagnosis, treatment planning, and assessing the effects of treatment. Manual tumour segmentation is cumbersome and can also be inaccurate. Kumar et al. (2025) introduced a hybrid model which involved LoG preprocessing and 3D U-Net to facilitate tumour segmentation and edge detection using the BraTS 2020 dataset. The integration of the preprocessing technique before segmentation led to the achievement of an accuracy of 99.73%.

Valanarasu et al. (2021) introduced the concept of MedT (Medical Transformer), a gated axial attention architecture for capturing long-range dependencies in medical imaging. In conjunction with a Local-Global (LoGo) learning framework, the network has been shown to outperform traditional CNNs and transformers.

The development of deep learning has led to improvements in brain tumour segmentation and classification. Specifically, the work done by Khan et al. (2024) focused on using the optimisation of CNNs along with noise reduction and data augmentation for classifying MRI images.

Majib et al. (2021) developed a deep learning model named VGG-SCNet, which was based on the VGGNet model, to automatically classify brain tumours in MRI images. The researchers used traditional machine learning models, hybrid machine learning models, transfer learning, and deep learning for comparison purposes, where it was found that the proposed model outperformed other models with a precision score of 99.2%, a recall of 99.1%, and an F1-score of 99.2%. Nonetheless, the model did not consider segmentation of the brain tumour.

Pandey et al. (2026) suggested a method known as hybridisation of DenseNet121 and PCA for classifying multiclass brain tumours using a combination of deep and manually designed features. This approach showed a high level of accuracy with 95.89%. The problem of exact tumour segmentation is not mentioned.

The framework called RobU-Net for accurate segmentation of multiclass brain tumours was proposed by Qureshi et al. (2024). Heuristic optimisation has been utilised in the proposed approach to enhance segmentation accuracy, where better results have been achieved in terms of segmenting sub-regions of the tumour. Nevertheless, certain limitations are still encountered.

The study conducted by Bikku et al. (2025) came up with the idea of developing an LSTM-SVM model through support vector machine for predictive analytics of gene

expression in healthcare. With the combination of the temporal learning of LSTM with the classification ability of SVM, this model has shown the power of using hybrid deep learning in analysing healthcare data. Summarised literature given in Table 1.

Table 1
Literature survey of related works

Year	Citation	Paper Title	Research Design	Conceptual / Theoretical Framework	Major Theme in Paper	Future Idea
2025	Pourmahboubi et al., 2025	Brain Tumour Segmentation Enhancement in MRI Using U-Net and Transfer Learning	Experimental design using U-Net + transfer learning	U-Net architecture, pretrained feature extractors	Improved tumour boundary detection using transfer learning	Extend to multimodal MRI; integrate attention and transformer blocks
2025	Saifullah et al., 2025	Modified U-Net with Attention Gate for Enhanced Automated Brain Tumour Segmentation	Experimental design using modified U-Net architecture	U-Net with integrated Attention Gates for focused feature extraction	Attention gates improve localisation, reduce noise, and enhance segmentation accuracy over standard U-Net	Extend attention-based models; explore multimodal MRI; integrate transformer or hybrid architectures
2024	Neetha & Narayan, 2024	Segmentation and Classification of Brain Tumour Using LRIFCM and LSTM	Two-stage experimental design: LRIFCM for segmentation + LSTM for classification	LRIFCM (local-region fuzzy clustering) for segmentation; LSTM for temporal feature learning	Improved segmentation precision and classification accuracy through hybrid clustering–deep learning	Extend to multimodal MRI datasets; integrate attention-based LSTM or transformer models
2024	Shiny, 2024	Brain Tumour Segmentation and Classification Using Optimised U-Net	Experimental design using optimised U-Net	Optimised U-Net architecture with enhanced feature extraction	Improved tumour boundary detection and reduced segmentation errors compared to standard U-Net	Explore hybrid U-Net models with attention, transformers, or multimodal inputs

Table 1 *(continued)*

Year	Citation	Paper Title	Research Design	Conceptual / Theoretical Framework	Major Theme in Paper	Future Idea
2023	Wong et al., 2023	Brain Image Segmentation of the Corpus Callosum Using Bi-Directional ConvLSTM and U-Net	Hybrid model using multi-slice CT & MRI; experimental evaluation	Bi-Directional ConvLSTM for spatial–temporal learning + U-Net for feature extraction & localisation	Improved segmentation accuracy and structural consistency using multi-modal fusion	Extend model to full-brain segmentation; explore 3D ConvLSTM; optimise for clinical real-time analysis
2022	Younis et al., 2022	Brain Tumour Analysis Using Deep Learning and VGG-16 Ensembling Approaches	CNN and VGG-16-based experimental design	Deep learning (CNN, VGG-16), medical image analysis, ensemble learning	Use of deep learning for tumour detection across imaging modalities; segmentation and classification techniques	Explore more diverse image modalities and advanced segmentation approaches
2020	Isensee et al., 2021	nnU-Net for Brain Tumour Segmentation	Applied empirical study using BraTS 2020 dataset; cross-validation on multiple nnU-Net variants	nnU-Net framework based on U-Net encoder–decoder architecture, automated segmentation pipeline, region-based loss, data augmentation	nnU-Net outperforms specialised models with minimal tuning; strong generalisable segmentation model	Improve hyperparameter tuning; evaluate ranking strategies; enhance leaderboard accuracy
2021	Biratu et al., 2021	A Survey of Brain Tumour Segmentation and Classification Algorithms	Comprehensive survey study	Segmentation techniques; classification methods; imaging advancements	Overview of advancements in medical imaging, segmentation, and classification	Not available
2021	Magadza & Viriri, 2021	Deep Learning for Brain Tumour Segmentation: A Survey of State-of-the-Art	Deep learning methodologies survey	Deep neural networks, evolution of CNNs, medical imaging frameworks	Challenges in medical imaging; evolution of CNN-based segmentation	Encourage data sharing, real-time segmentation solutions, improved DL frameworks

RESEARCH METHODOLOGY

This proposed study provides a hybrid deep learning framework for nutshell and automatic brain tumour segmentation from MRI scans that combines U-Net and Long Short-Term Memory (LSTM) networks. The five primary steps of the methodology Data collection, preprocessing, data augmentation, model architecture design, and performance evaluation are the major phases of the methodology.

Data Collection

The MRI scans of brain tumours from the open BRATS 2019 (Brain Tumour Segmentation) dataset were used as input for the training and evaluation of the model. The multi-modal MRI images include T1, T1ce (contrast-enhanced), T2, and FLAIR for each patient, together with an expert-annotated ground truth mask that accompanies each image volume and indicates the regions of the tumour (enhancing tumour, peritumoral oedema, necrotic/non-enhancing tumour core). A total of 3064 2D slices, along with their segmentation masks, were subsequently extracted and used in this work. To conduct fair and impartial assessments of the model, the dataset was divided into subsets for training (70%) and validation (30%). The anonymised patients in the dataset are formatted following the NIFTI standard, and this dataset is usually provided by a widely adopted benchmark where typically the brain tumour segmentation research community conducts their studies.

Preprocessing and Data Augmentation

This is the preprocessing step, which is very important for getting MRI quality and training a model on raw data. Every 2D slice must be generated at the same resolution to achieve standardisation in input dimensions with respect to intensity normalisation for better model convergence and to normalise inter-scanner variation differences in output. Some other interesting methods of contrast enhancement, for instance, CLAHE, are also supposed to be helpful. Median filtering can finally accomplish reduction in noise. Together, these processes align, normalise, and improve images for accurate feature extraction in training. Techniques for augmenting training data have been introduced to enhance replicate diversity within the dataset and lessen the possibility of overfitting. Random rotations, flips, translations, and scaling are examples of geometric transformations that were performed on the data in a way that simulated real clinical patient positioning variations. Furthermore, to compensate for the heterogeneous settings of magnetic resonance imaging acquisitions, intensity-based augmentations were applied, which included controlled brightness and contrast augmentations. Together, these techniques are effective at improving the generalisation of the model on diverse imaging conditions. After that, the proposed dataset can be divided into training and testing sets to ensure proper learning and monitoring.

U-Net and LSTM Architecture

U-Net and LSTM are combined in a hybrid deep learning model for efficient tumour segmentation and detection. The primary focus of our segmentation model is a very light U-Net which employs layers of LSTM. U-Net is a generic architecture for medical image segmentation, which contains a path for contraction (encoder) and an expanding path for decoding with skip connections.

Contracting Path (Encoder): The encoded max pooling after all the layered convolutional layers is proposed to extract higher-level feature hierarchies. It makes the model learn semantic representations of tumour locations while also diminishing the spatial dimensions with an increased depth of feature maps.

LSTM Module: After the encoder, a Bidirectional LSTM layer is inserted to process the sequence of encoded features across adjacent slices. This module improves segmentation consistency across 2D slices by collecting long-range spatial dependencies.

Expanding Path (Decoder): This expanding path uses upsampled feature maps by using transposed convolutions and merges them with corresponding encoder features via skip connections. This may enable the segmented output to be fully spatially reconstructed by the model.

Skip Connections: Bridge the encoder and decoder layers by using the skip connections at each level, store the high resolution, and help the model to better localise tumour boundaries.

Output Layer: The last layer of the output consists of a 1 × 1 convolution layer followed by an activation function (for binary segmentation), where the output from this process is a binary mask highlighting areas of the tumour region for each of the input slices.

In the training stage of the model, it is trained by learning the patterns in the training dataset. Once the training process is completed, the performance of the model is analysed based on the validation dataset, which includes evaluating the model based on performance parameters like accuracy and segmentation. Lastly, the results of the analysis and prediction processes are visualised, where the detected tumour areas are marked within the MRI image.

This combined architecture results in better and more contextual segmentation of brain tumours without the computational cost of total 3D networks while getting more contextual information than traditional 2D CNN models as shown in the Figure 1.

Implementation

The Brain Tumour Segmentation model, using U-Net and LSTM, is a deep learning solution designed to automate brain tumour detection and segmentation from MRI scans. This work focuses on assisting radiologists to assist in accurately locating tumour regions, which helps with diagnosis and therapy planning. This section will discuss the details regarding the implementation of the proposed system.

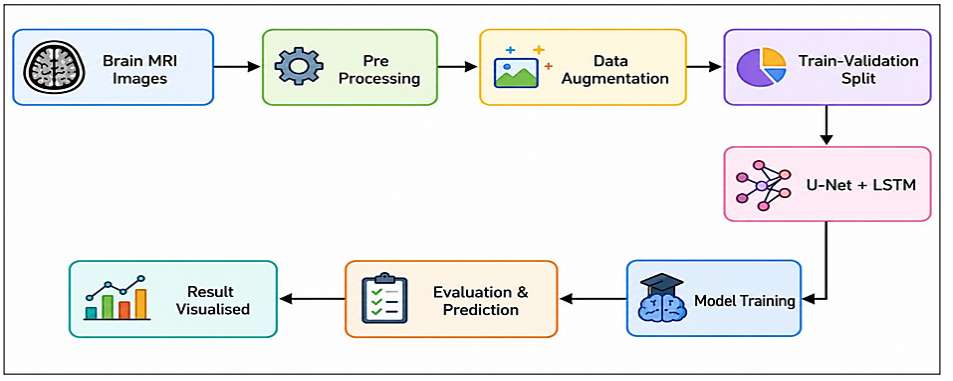


Figure 1. Proposed methodology framework for brain tumour.

Technology Stack

This is a brain tumour segmentation system that has been built on Python, which is known for its huge scientific computing ecosystem and great support for deep learning workflows. The hybrid architecture comprising U-Net and LSTM was implemented and trained in TensorFlow and Keras. These two frameworks provide high-level APIs to form the convolutional and recurrent components into one framework. The basic preprocessing of images includes resizing, normalisation, skull-stripping-compatible filtering, and masking encoding, all performed using OpenCV and NumPy to leverage the pixels-to-numbers computational efficiency. Further, data were augmented using TensorFlow image utilities, preprocessing layers of Keras, and specially designed custom augmentation pipelines - specific for medical imaging (rotation, elastic deformation, and contrast changes) - all to enhance the generalisation of the model. Model evaluation was done based on visualising predictions for masks, creating training curves, and using performance metrics via Matplotlib. All the experiments were run on Google Colab, which has cloud-based GPU facilities to run such computationally intensive deep learning workloads. The integrated model would be integrated into a lightweight interface made possible by Streamlit upon online deployment, enabling real-time tumour segmentation and viewing. Alternatively, this could also be deployed on scalable clouds such as Google Cloud, AWS, or Microsoft Azure, thus allowing access to the system from remote sites, integration into clinical workflows, and provision of the interface with hospital PACS systems.

System Architecture

The use of a recent deep learning architecture improves the system's efficiency and maintainability. LSTM layers are combined for smooth temporal processing, which is particularly desirable for sequential processing of MRI slices, while U-Net is used as the backbone for the extraction of spatial data. Thus, this hybrid architecture allows the

model to accommodate learning of spatial and temporal dependencies. The architecture can handle MRI data at a large scale and hence can further be scaled by incorporation of attention mechanisms or 3D volume data.

Figure 2 shows a visual walkthrough that provides a thorough explanation of the processing pipeline of the suggested U-Net and LSTM-based brain tumour segmentation system from raw MRI collection through several phases to performance assessment. All the stages are organised in a sequence that is efficient for data management and optimal for model performance. An MRI scan is the first step in this procedure, which then moves on to preprocessing, which includes intensity normalisation, scaling, denoising, skull stripping, and slice extraction. Preprocessing includes data standardisation, noise reduction, and the creation of consistent inputs for model training. Before subjecting it to model training, data augmentation will also be applied to that dataset, such that variance in the data will increase and induce robustness in the model. This may include geometric transformations, intensity modifications, or elastic deformations, as it is one other way of making additional training samples to robustly enhance the model for anatomical as well as those related to the scanner variability.

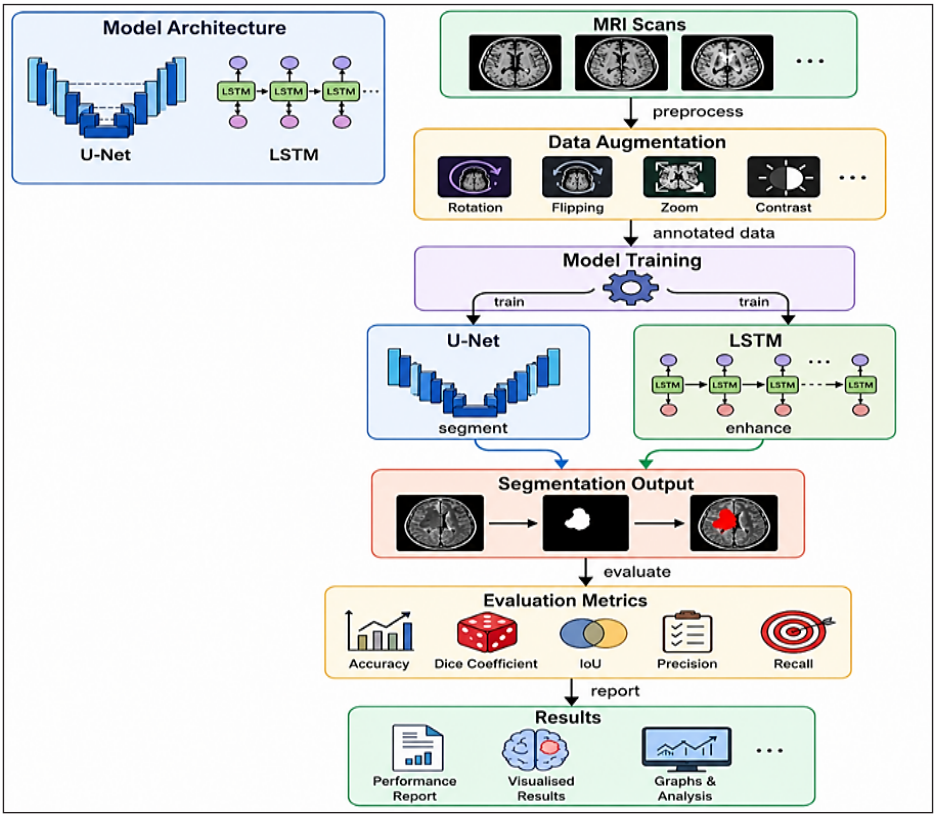


Figure 2. System architecture for Brain Tumour Segmentation Using Hybrid U-Net and LSTM approach.

This nurtured dataset is then used to train the hybrid U-Net and LSTM model, where both models are trained jointly. The U-Net part is responsible for extracting multi-scale spatial features from MRI images and thus creates preliminary segmentation masks. Meanwhile, the LSTM module is concerned with slice-to-slice temporal dependencies, enhancing structural continuity across the MRI volume and improving the overall segmentation quality. The complementary combination of these two kinds of learning ensures that both spatial and contextual information are considered for segmentation output. Combined model outputs hence lead to the final segmentation masks of tumours that are more accurate and of better boundary precision. In the follow-up action, results will be evaluated subsequently evaluating them against standard known metrics such as Dice Similarity Coefficient (DSC), IoU, precision, recall, and accuracy to acquire an extraordinary quantitative measure of the system's performance. Thus, the final stage ensures the segmentation framework's clinical applicability and dependability by compiling the evaluation results and presenting them through visualisations and performance summaries, which medical professionals can use for diagnosing and analysing patients. The proposed method algorithm in pseudocode mentioned in Table 2.

User Interface

Simple and intuitive is the design criterion of the segmentation system's user interface. Medical professionals have the option of uploading MRI images to view segmented output, with visualisation of the tumour boundaries straight away. UI could be implemented using Streamlit or Flask with minimal human interaction, keeping in view the level of technical expertise among some of the users, like radiologists and medical staff.

Integration

The segmentation system can be integrated with existing hospital systems or PACS (Picture Archiving and Communication Systems) to retrieve MRI images automatically. The segmented results can either be inserted in the Clinical Decision Support System or in the patient's records for longitudinal follow-up. The web-enabled interface allows access and inference of real-time diagnostic evaluation from a remote location.

Security

The system would have to comply with data safety regulations on account of the medical imaging data being sensitive, for instance. Secure protocols for data transfer, proper mechanisms for controlling access and anonymisation of relevant data should be the means to protect patient confidentiality and ward off unauthorised access.

Table 2
Algorithm: Hybrid U-Net and LSTM for brain tumour segmentation

Algorithm: Hybrid U-Net and LSTM for Brain Tumour Segmentation

Input: MRI Scan Dataset D and corresponding ground truth mask M
Output: Final Segmentation Results R

```
Begin
  // Data Preparation
   For each MRI scan I in dataset D do
      I_pre ← Preprocess(I)          // normalisation, resising, denoising
      I_aug ← Augment(I_pre)         // rotation, flipping, intensity shifts
      Store I_aug with corresponding annotation
   End For
   // Model Training
   Train the U-Net model using an augmented dataset D
   Train LSTM (ConvLSTM) model L on sequential MRI slices
   // Segmentation Stage
   For each test MRI volume V do
      S_unet ← U(V)                  // spatial feature segmentation
      S_lstm ← L(V)                  // temporal enhancement
      S_final ← Fuse (S_unet, S_lstm)    // combined segmentation output
   End For
   // Evaluation
   Compute evaluation metrics:
      Dice, IoU, Precision, Recall, Accuracy
   // Output Results
   R ← GenerateReport(S_final, metrics)
End
```

Testing and Deployment

The hybrid U-Net and LSTM model was put through extensive testing to assess its reliability and clinical usability. Testing for accuracy of segmentation results in quantitative measures such as accuracy, precision, recall, F1-score, Dice Similarity Coefficient and IoU, whereas the behaviour of the loss curves and validation trends was studied to test model convergence and stability during training. The other method of qualitative validation was through visual inspection of predicted masks to establish the correctness and continuity of tumour boundary detection. After validating the system, it can be deployed in various clinical infrastructure contexts: hosted at scalable remote access cloud platforms or local hospital servers and edge devices for on-site low-latency inference. Continuous integration pipelines with their test automation, model versioning, and periodic updates when new data becomes

available ensure robustness of the system in dynamic clinical settings. By way of the above components, the proposed hybrid U-Net and LSTM framework presents a dependable and efficient brain tumour segmentation system to ease workloads on humans, bring consistency in diagnoses, and take forward AI-assisted medical imaging in neuro-oncology. The Suggested hybrid U-Net optimises the model using the following hyperparameters to improve efficiency, as shown in Table 3.

Table 3
Optimised hyperparameters of the proposed model

Parameter name	Value
Learning Rate	0.0001
Batch Size	8
Epochs	30
Optimiser	Adam
Loss Function	Dice Loss
LSTM Units	64
Dropout Rate	0.3
Input Size	128 × 128

RESULTS AND DISCUSSIONS

The evaluation of the proposed brain tumour segmentation system based on hybrid U-Net and LSTM is made using the standard evaluation criteria, including accuracy, precision, recall, Dice (F1-score) and IoU.

The performance of the model is evaluated using these important metrics of segmentation and classification:

Accuracy: It is a measure of the model's 98.10% overall correctness in identifying tumour and non-tumour pixels. If accuracy is high, then the model is effective in identifying both tumour and non-tumour pixels. But in medical images, accuracy may not be enough due to imbalance.

Precision: This indicates that 95.90% of the tumour locations that the model predicts would be accurate. There are fewer false positives when precision is much higher, indicating that the system is primarily lowering the number of faulty tumour detections.

Recall: Thus, the model's potential capability of determining whether there exists an actual tumour. Consequently, the recall of the model is 96.30% in catching those tumour pixels that are the most relevant; however, the model misses others, which means there are certain false negatives.

F1-Score/Dice: A score of 96.10 shows that the model balances wasteful false positives against missed false negatives. This becomes especially critical in medical imaging, where both types of error have considerable clinical implications.

Intersection over Union (IoU): Which is also referred to as the Jaccard Index, measures how similar the predicted region is compared to the ground truth tumour. IoU is a more stringent measure compared to the Dice coefficient.

The proposed model, which integrates hybrid U-Net and LSTM models, gives good learning capacity during training and validation with an accuracy of 98.10% and 96.20%. Here, there is no issue with overfitting because both training and validation accuracy are close. Moreover, the model gives positive results for precision, recall, Dice scores and IoU Score. The efficiency of the proposed model can be given in Table 4.

Table 4
Model evaluation metrics for the proposed work

Metric	Value (%)
Training Accuracy	98.10
Validation Accuracy	96.20
Precision	95.90
Recall	96.30
F1-Score/Dice	96.10
IoU Score	92.40

The output of the brain tumour segmentation is shown in Figure 3. The trends in the graphs for training and validation accuracies have shown that there is an upward movement across the epochs. This shows that there is stability in the convergence of the model. The improvement in the training accuracy indicates successful feature learning and parameter adjustment of the network. The patterns show that the validation accuracy shows good generalisation of the model without overfitting. The trends in the loss graphs have been consistent, showing a downward movement during training. These trends confirm the proper convergence of the models. The trends in both the accuracies and losses show in Figure 4 that the proposed hybrid U-Net and LSTM is reliable and efficient.

This work proposed an efficient brain tumour segmentation framework that combines a hybrid U-Net and Long Short-Term Memory to increase the model's ability to capture the tumour regions from MRI slices. This combination of LSTM permits the model to overcome the drawbacks of 2D CNN methods.

Figure 5 shows a bar graph of model evaluation metrics used to assess the performance of the brain tumour segmentation accuracy. This graphical representation shows that the proposed model accuracy is 98.10%, precision 95.90%, recall 96.30%, dice/F1 score 96.10% and IoU score 92.40%, respectively. More specifically, the high Dice coefficient shows how

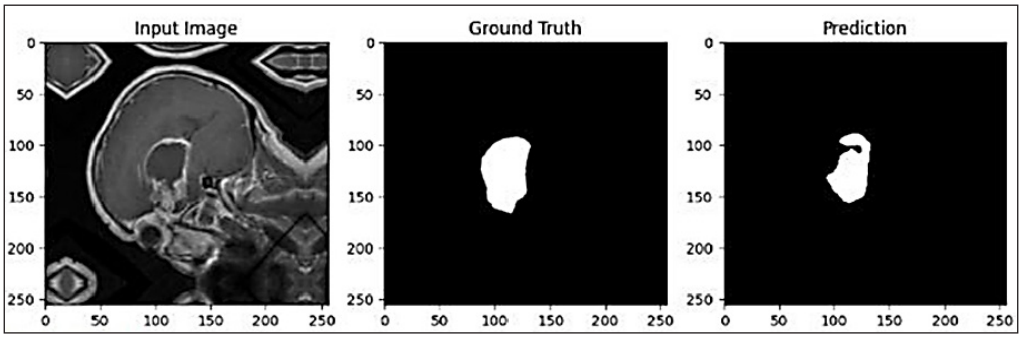
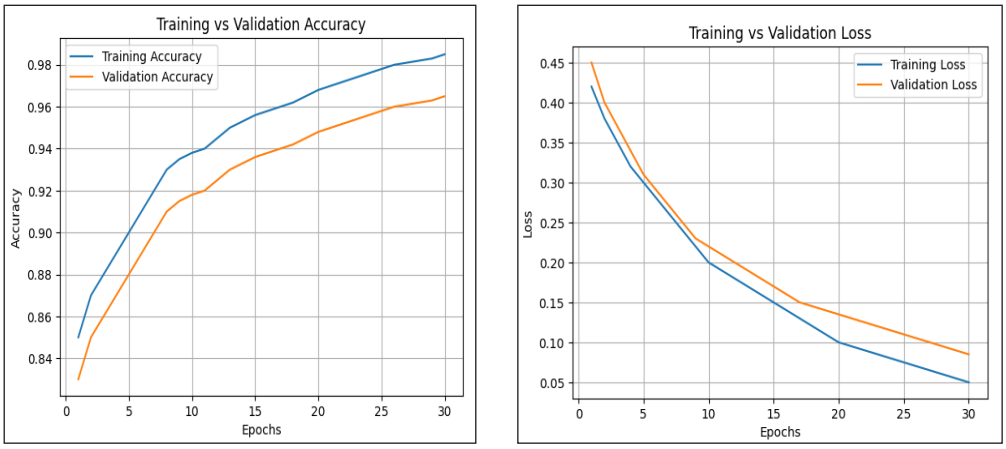


Figure 3. Predicted output of brain tumour segmentation



(a) (b)

Figure *4.* Performance analysis of training and testing: (a) accuracy and (b) loss

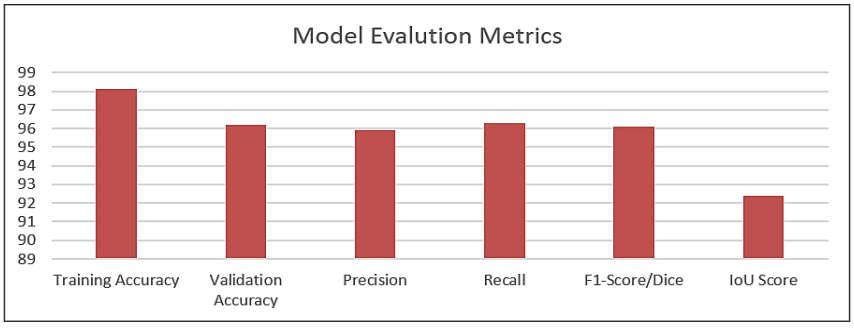


Figure 5. Proposed model evaluation metrics

well the system predicts tumour edges. On the other hand, the IoU score indicates how consistently tumour regions are predicted. We may conclude that the algorithm performs effectively in reducing errors based on the balance between the precision and recall metrics.

Thus, overall, the graph shows that the model is good at predicting correctly (high precision) but needs some improvement in identifying all true tumour cases (recall); improving recall could also increase this model's clinical reliability. The method displayed in the proposal is very promising in segmentation, especially in the BRATS datasets where accuracy can be attained in tumour region segmentations like enhancing core, oedema, and necrotic tissues.

Comparative Analysis of Brain Tumour Segmentation Models

The analysis of comparison results is presented in Table 5. It can discuss the efficiency of the proposed algorithm compared to other existing algorithms. As can be seen from the

Table 5
Comparison analysis of different models

Model	Accuracy (%)	Precision (%)	Recall (%)	Dice / F1 (%)
ConvLSTM U-Net	89.8	89.2	90.1	89.6
Modified U-Net	90.35	89.5	91.02	90.2
Attention Gate U-Net	93.2	92.1	93.8	92.9
RobU-Net	91.5	90.8	92.05	91.4
U-Net + Transfer Learning	92.8	92.1	93.5	92.7
CNN Framework	94.1	93.2	94.8	93.9
Proposed Hybrid U-Net and LSTM	**98.1**	**95.9**	**96.3**	**96.1**

analysis, the proposed hybrid U-Net and LSTM approach outperforms all others in terms of segmentation performance.

The ConvLSTM U-Net designed by Almiahi et al. (2024) had an accuracy score of 89.8% and an F1-score of 89.6%. Although the combination of ConvLSTM can help improve the network's ability to capture spatiotemporal relationships, its performance still falls behind that of the proposed model. Another improvement was seen in the Modified U-Net by Alquran et al. (2024), which had slightly improved performance, especially in terms of recall (91.02%). An even greater improvement could be achieved in an attention-based architecture. The Attention Gate U-Net created by Saifullah et al. (2025) has an accuracy score of 93.2% and an F1-score of 92.9%. With the attention gates, the network can focus on important areas, thereby enhancing both its precision and recall scores. The same applies to the U-Net with Transfer Learning by Pourmahboubi et al. (2025), which exhibited great generalisation capabilities, scoring 92.8% and 92.7% for accuracy and F1-score, respectively. The use of pre-trained feature extraction helps enhance the network's classification abilities. Moreover, the CNN model presented by Zhang and Yang (2026) has a competitive performance (accuracy: 94.1%, F1-score: 93.9%), which shows that an efficient convolutional model can compete with specialised models for segmentation tasks. The proposed Hybrid U-Net and LSTM model is far better than any other previously introduced methods, as it has the highest performance rate.

Figure 6 illustrates the performance analysis of the proposed model. The results demonstrate that the proposed U-Net-based architecture with LSTM outperforms other architectures in terms of accuracy in segmenting tumour areas, showcasing the importance of fusing spatial and context information. This is mainly because of the fusion between U-Net and LSTM, where U-Net effectively captures the spatial information, and LSTM captures the context and sequence information. Thus, the model has achieved high recall and precision.

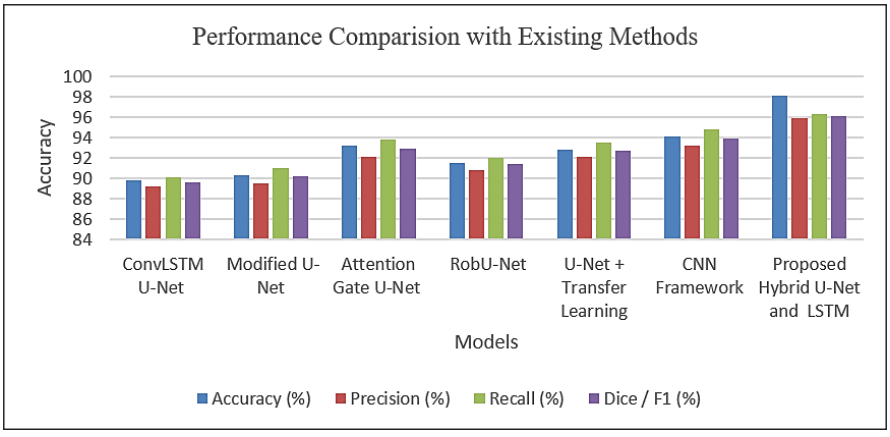


Figure 6. Performance analysis of the proposed model with existing methods

CONCLUSION

In this paper, we propose a brain tumour Segmentation and Classification model, and it is accomplished via the combination of the architecture of the Hybrid U-Net and LSTMs. The idea of this approach is to use the efficiency of spatial feature extraction from the U-Net, as well as adding temporal layers such as LSTMs to increase segmentation precision, especially in cases where there are sequences and volumetric datasets. Our dataset for BRATS included 3064 images together with their corresponding masks; these images were subjected to preprocessing operations, such as resizing, normalisation, and transforming the mask to a binary one. Data augmentation techniques were used with our dataset to avoid overfitting. Our hybrid U-Net and LSTM model has succeeded in achieving very high performance, and the achieved accuracy is 96.70% and 98.10%. These results showed that integrating temporal layers is crucial within the segmentation process in medical imaging. In summary, this paper is a valuable addition to research on deep learning-based image analysis in medicine and is sure to serve as a useful foundation for future improvements, including attention modules, 3D convolutions, or Transformers. While this model is quite effective in performing its intended operations, there are certain limitations that were observed in it. Firstly, segmentation accuracy depends largely on the level of preprocessing; if there is any sort of mismatch in normalisation or alignment processes, then it will affect the efficiency of the model negatively. As annotated medical images are difficult to find, there are high risks of overfitting involved, not forgetting the fact that manual annotation is difficult and error-prone too. Given that it involves heavy computation in the modelling process, this algorithm could only work on powerful GPUs or even clouds, which is why it cannot undergo mass testing. As for clinical applications, they might have problems in optimising systems, implementing IT infrastructure, and ensuring conformity with the relevant laws regarding patient data protection.

Future Enhancements

The Brain tumour segmentation model, which uses the hybrid U-Net and LSTM architecture, offers many possible future extensions. First, multi-class segmentation may be the proposed addition, which will allow differentiation between types of tumours and differentiate sub-regions like oedema, necrotic core, and enhancing tissues. This will also enable the model to optimise for inference on edge devices of portable MRI systems and low-resource clinical workstations so that this model can further support real-time applications. Further, this also improves interslice consistency by modelling such a three-dimensional structure of MRI data through models such as 3D U-Net or volumetric LSTM architectures. It would also better exploit the three-dimensional structure of MRI data through models such as 3D U-Net or volumetric LSTM architectures, thus improving interslice consistency. Developing a web-based platform along with a mobile interface could help in improving remote diagnostics and facilitate clinical visualisation of segmentation results. Integrating explainable AI-like Grad-CAM techniques would enhance interpretability and possible rise in clinician trust in the automated predictions. It also calls for validation on larger and more heterogeneous datasets through collaborations with medical institutions to ensure robustness across various imaging conditions. Transformer-based architecture and possibly Spatio-Temporal Graph Neural Networks would also be among the best future projects that would improve modelling on the complex spatial relationships within MRI volumes. Future studies could also add multimodal clinical information, such as radiology reports or genetic data, adding strength to the performance of the segmentation and thus aiding in more holistic tumour characterisation.

ACKNOWLEDGEMENT

The authors would like to thank Koneru Lakshmaiah Education Foundation for their support in completing this research work. This research did not receive any grant or funding from any agencies.

LIST OF ABBREVIATIONS

MRI	:	Magnetic resonance image
AI	:	Artificial Intelligence
DL	:	Deep learning
LSTM	:	Long short-term memory
CNN	:	Convolutional neural network
BRATS	:	Brain tumour segmentation
IoU	:	intersection over Union

REFERENCES

Almiahi, O. M. H., Albu-Salih, A. T., & Alhajim, D. (2024). A novel ConvLSTM-based U-Net for improved brain tumour segmentation. *IEEE Access, 12*, 157346-157358. https://doi.org/10.1109/ACCESS.2024.3483562

Alquran, H., Alslatie, M., Rababah, A., & Mustafa, W. A. (2024). Improved brain tumour segmentation in MR images with a modified U-Net. *Applied Sciences, 14*(15), Article 6504. https://doi.org/10.3390/app14156504

Bikku, T., Karthik, J., Rao, G. R. K., Sree, K. S., Srinivas, P. V. V. S., & Prasad, C. (2021). Brain tissue segmentation via deep convolutional neural networks. In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 757-763). IEEE. https://doi.org/10.1109/I-SMAC52330.2021.9640635

Bikku, T., Martis, J. E., Sunil, K. M., Sudha, S., Iyappan, P., & Natarajan, C. (2025). Healthcare biclustering of predictive gene expression using LSTM-based support vector machine. *Informing Science, 28*, Article 12. https://doi.org/10.28945/5446

Biratu, E. S., Schwenker, F., Ayano, Y. M., & Debelee, T. G. (2021). A survey of brain tumour segmentation and classification algorithms. *Journal of Imaging, 7*(9), Article 179. https://doi.org/10.3390/jimaging7090179

Dogan, Y. (2026). A novel pruning-enhanced hybrid approach for efficient and accurate brain tumour diagnosis. *Biomedical Signal Processing and Control, 112*, Article 108466. https://doi.org/10.1016/j.bspc.2025.108466

Elhadidy, M. S., Elgohr, A. T., El-Geneedy, M., Akram, S., & Kasem, H. M. (2025). Comparative analysis for accurate multi-classification of brain tumour based on significant deep learning models. *Computers in Biology and Medicine, 188*, Article 109872. https://doi.org/10.1016/j.compbiomed.2025.109872

Gopisairam, T., Thota, S., & Bikku, T. (2026). Integrative framework for cancer detection via integro-differential equations using deep learning techniques. *Scientific Reports, 16*, Article 9714. https://doi.org/10.1038/s41598-026-39283-z

Ijaz, M., Hasan, I., Aslam, B., Yan, Y., Zeng, W., Gu, J., ... & Guo, B. (2025). Diagnostics of brain tumor in the early stage: current status and future perspectives. *Biomaterials Science*, *13*(10), 2580-2605. https://doi.org/10.1039/d4bm01503g

Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., & Maier-Hein, K. H. (2021). nnU-Net for brain tumour segmentation. In A. Crimi & S. Bakas (Eds.), *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 118-132). Springer. https://doi.org/10.1007/978-3-030-72087-2_11

Khan, M. F., Iftikhar, A., Anwar, H., & Ramay, S. A. (2024). Brain tumour segmentation and classification using optimised deep learning. *Journal of Computing & Biomedical Informatics, 7*(1), 632-640. https://doi.org/10.56979/701/2024.

Kumar, G. P., Naidu, G. V., Vardhan, P. S., Gurrala, C., Bikku, T., & Thota, S. (2025). Enhanced brain tumour segmentation with 3D U-Net and Laplacian of Gaussian preprocessing. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-9). IEEE. https://doi.org/10.1109/ICTBIG68706.2025.11323577

Magadza, T., & Viriri, S. (2021). Deep learning for brain tumour segmentation: A survey of state-of-the-art. *Journal of Imaging, 7*(2), Article 19. https://doi.org/10.3390/jimaging7020019

Majib, M. S., Rahman, M. M., Sazzad, T. S., Khan, N. I., & Dey, S. K. (2021). VGG-SCNet: A VGG-Net-based deep learning framework for brain tumour detection on MRI images. *IEEE Access, 9*, 116942-116952. https://doi.org/10.1109/ACCESS.2021.3105874

Montaha, S., Azam, S., Rafid, A. K. M. R. H., Hasan, M. Z., & Karim, A. (2023). Brain tumour segmentation from 3D MRI scans using U-Net. *SN Computer Science, 4*, Article 386. https://doi.org/10.1007/s42979-023-01854-6

Neetha, K. S., & Narayan, D. L. (2024). Segmentation and classification of brain tumour using LRIFCM and LSTM. *Multimedia Tools and Applications, 83*(31), 76705-76730. https://doi.org/10.1007/s11042-024-18478-4

Pacal, I., Akhan, O., Deveci, R. T., & Deveci, M. (2025). NeXtBrain: Combining local and global feature learning for brain tumour classification. *Brain Research, 1863*, Article 149762. https://doi.org/10.1016/j.brainres.2025.149762

Pandey, B. K., Pandey, D., Lee, T.-F., & Lelisho, M. E. (2026). Hybrid deep neural network with PCA-based features optimisation for enhancing brain tumour classification. *Scientific Reports, 16*, Article 39154. https://doi.org/10.1038/s41598-026-39154-7

Pourmahboubi, A., Arsalani Saeed, N., & Tabrizchi, H. (2025). A brain tumour segmentation enhancement in MRI images using U-Net and transfer learning. *BMC Medical Imaging, 25*, Article 307. https://doi.org/10.1186/s12880-025-01837-4

Qureshi, S. A., Chaudhary, Q. U. A., Schirhagl, R., Hussain, L., Aman, H., Duong, T. Q., Khalil, A., & Galenchik-Chan, A. (2024). RobU-Net: A heuristic robust multi-class brain tumour segmentation approach for MRI scans. *Waves in Random and Complex Media*. Advance online publication. https://doi.org/10.1080/17455030.2024.2366837

Reddy, Y. A., Dubey, R. S., Prakash, R. V., & Padder, A. (2026). MRI-based brain tumour prediction using convolutional neural network framework. *Scientific Reports, 16*, Article 15904. https://doi.org/10.1038/s41598-026-47044-1

Saifullah, S., Dreżewski, R., Yudhana, A., Wielgosz, M., & Caesarendra, W. (2025). Modified U-Net with attention gate for enhanced automated brain tumour segmentation. *Neural Computing and Applications, 37*(7), 5521-5558. https://doi.org/10.1007/s00521-024-10919-3

Shiny, K. V. (2024). Brain tumour segmentation and classification using optimised U-Net. *The Imaging Science Journal, 72*(2), 204-219. https://doi.org/10.1080/13682199.2023.2200614

ul Ain, N., Khan, S. A., Aladhadh, S., Mir, U., & Ramzan, M. (2026). Transformers meet CNNs: A comprehensive review and benchmarking of deep learning architectures for brain tumour classification in MRI. *Computer Science Review, 60*, Article 100897. https://doi.org/10.1016/j.cosrev.2026.100897

Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., & Patel, V. M. (2021). Medical transformer: Gated axial-attention for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021* (pp. 36-46). Springer. https://doi.org/10.1007/978-3-030-87193-2_4

Wong, K. K., Xu, W., Ayoub, M., Fu, Y.-L., Xu, H., Shi, R., Lyu, M., Mou, L., Li, X., Guo, Y., Liu, T., Yuan, T.-F., Wu, K., Peng, J., & Chen, W. (2023). Brain image segmentation of the corpus callosum by combining bi-directional convolutional LSTM and U-Net using multi-slice CT and MRI. *Computer Methods and Programs in Biomedicine, 238*, Article 107602. https://doi.org/10.1016/j.cmpb.2023.107602

Yadav, A. C., & Kolekar, M. H. (2026). Deep feature-based approaches for brain tumour classification and segmentation in medical imaging. *Biomedical Signal Processing and Control, 117*, Article 109603. https://doi.org/10.1016/j.bspc.2026.109603

Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., & Kawuwa, H. B. (2022). Brain tumour analysis using deep learning and VGG-16 ensemble learning approaches. *Applied Sciences, 12*(14), Article 7282. https://doi.org/10.3390/app12147282

Zhang, P., & Yang, Z. (2026). Convolutional neural networks: Applications, challenges and prospects in brain tumour research. *Frontiers in Neurology, 17*, Article 1759459. https://doi.org/10.3389/fneur.2026.1759459

Pertanika J. Sci. & Technol. 34 (S1): 199 - 232 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Filtered-Based Online Social Networking Features Extraction and Node Link Prediction Approach

Rajasekhar Nennuri[1]*, G Pradeepini[1], and L V Narasimha Prasad[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Green Fields, 522302 Vaddeswaram, Guntur District, Andhra Pradesh, India
[2]Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, 500043 Hyderabad, Telangana, India

ABSTRACT

Online social networks, a significant part of the Internet, contain complex statistical relationships among millions of social nodes. Link prediction algorithms compute a likelihood of potential future ties between nodes in these networks. Most traditional methods use some form of contextual similarity between nodes to discover relations. Traditional community detection models tend to rely on limited structural information and ignore the impact of central nodes during link prediction. These restrictions impair the efficacy of link-prediction-based decision-making. In this paper, a community-based link prediction technique taking the advantages of both is proposed for dynamic and noisy networks. A new community detection metric is proposed to identify candidate links in a web of nodes. The central node in the network is calculated by this metric and used by the link prediction algorithm. The proposed community detection metric, along with other existing metrics, is employed to learn a model for hybrid link prediction. Experimental results indicate that the proposed central node-based link prediction method achieves better performance in prediction accuracy, prediction error, and computation time than traditional methods.

Keywords: Community detection, link prediction, similarity measures, social networking data

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 27 April 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.10

E-mail address:
rajasekharnennuri@gmail.com (Rajasekhar Nennuri)
pradeepini_cse@kluniversity.in (G Pradeepini)
lvnprasad@yahoo.com (L V Narasimha Prasad)
* Corresponding author

INTRODUCTION

Online social networks are defined as social nodes and their interactions, as stated by Wu et al. (2017). This principle of construction represents the social relations between social actors, be they individuals or organisations. Social nodes, in their turn, are interconnected via edges that portray relations or communications. The purpose

e-ISSN: 2231-8526

of these networks is to create a larger network from which benefits can be obtained, such as better information access and greater influence flow. Examples range from Facebook's goal of "giving people the power to build community and bring the world closer together" to LinkedIn's mission of "connecting the world's professionals to make them more productive and successful." People tend to develop relationships in networks where they share common interests, backgrounds or prior association. It is difficult to model the expansion of these networks, but the analysis of the relations among pairs of nodes can provide useful insights.

Link prediction is the task of predicting the link among the disconnected nodes. Pandey et al. (2019) mention that link prediction methods have a huge impact on real-world structure reconstruction along with various applications. For instance, in security systems, link prediction may analyse terrorist networks to identify future associations, which can help government agencies to track potential threats and suspects. In biological networks, link prediction facilitates predicting new protein–protein interactions from existing ones, as presented by Bastami et al. (2019). Other applications in social networks are automatic hyperlink prediction and user-to-user recommendation systems (Aghabozorgi & Khayyambashi, 2018). Despite the existing similarity-based methods, enhancing the accuracy of link prediction is still a big challenge. Several models have been developed to improve the performance and evaluation methods, and such models were trained and tested on ego networks obtained from the Facebook dataset in the SNAP repository (Berahmand et al., 2021). These are actual networks of real people with varying degree distributions, clustering, degree correlation, etc.

With the continuous generation of new links in online social networks, the study of dynamic topological properties has great significance. Class imbalance also limits the predictive power in link prediction problems. Contrary to each user connecting to millions of other users on Facebook, each user is only connected to hundreds of nodes, as claimed by Jiang et al. (2021). This causes a huge skew between existing versus the non-Jagielski et al. agree that predicting every possible missing link in such large networks is computationally prohibitive; thus, sensible to truncate the set of node pairs on which to evaluate. Consequence networks are smaller parts of bigger networks, and for this reason ego networks are often taken for experimental analysis. Because social networks are self-similar along their subsystems, the results obtained from these subsystems can also be applied to the full system. In this study, we aim to develop computationally efficient models to predict the future structure of the ego network with dynamic topology and network sparsity.

A Facebook "community" is a bundle of groups. The widespread and dynamic features of social network data make identifying useful information imperative. Social Network Analysis (SNA) deals with identifying and analysing patterns in such data. SNA is also used for developing new trends, community detection, brand monitoring and election result

prediction. Such applications underscore the value of analysis of social networks. In such networks, entities tend to share many connections within groups and few across groups (i.e. the entities in such networks have a community structure). This paper concentrates on finding such communities. The goal is to show that these simple, quantitative measures of network structure can give rise to good link prediction algorithms.

Stochastic models can provide a good way to model and predict the evolution of a network. A probabilistic network is a joint probability distribution over a set of random variables, whose factorisation is determined by nodes that represent random variables and edges that represent dependencies. One of the standard models in this type of representation is the Bayesian Network model, which is a directed acyclic graph (DAG) that models causal and conditional relationships among variables. Bayesian inference is employed to revise the forecasts with the arrival of additional data. In the case of online social networks, an edge is the link between two vertices with which a number of measures can be defined. Such measures are, however, challenging to compute in static networks. Reachability, which is the possibility for one node to reach another through paths and is a function of network structure. In homogeneous networks, reachability is predictable, whereas in heterogeneous networks, paths can be highly variable

Although the proposed framework combines feature extraction, probabilistic ranking, and optimisation methods, the novelty should be clarified considering the existing methods. Traditional methods for link prediction are mostly based on similarity measures, predicate structural heuristics or simple probabilistic models. In contrast to this, current best-performing approaches are based on representation learning (e.g. graph embeddings) or deep learning-based approaches (GNNs) and knowledge graph embedding techniques. These techniques learn representations (viz., low-dimensional vectors) of nodes as well as relations, which give rise to enhanced prediction accuracy and better scalability over complex networks. Also, deep learning-based link prediction algorithms, such as neural networks and GNN models, have been proven to be effective in modelling structural and attributive dependencies of graphs. However, many of these methods involve computationally expensive procedures and require training on large-scale data, and are often criticised for their lack of interpretability in decision-making. The novelty of our approach is that we unify graph filtering, probabilistic source–destination ranking and optimisation-based feature selection in one procedure. Unlike the embedding-type methods based on latent vector representations, the proposed model takes explicit central node influence, shortest path dependencies and follower–followed dynamics into account to enhance prediction accuracy. Also, the probabilistic ranking formulation combined with an optimisation-based Random Forest algorithm balances the trade-off between interpretability and performance. Hence, the main novelty of this article is to design a hybrid, interpretable, and efficient link prediction method which connects traditional similarity-based methods

and recent learning-based methods and is capable of effectively dealing with sparse and imbalanced data in social networks.

In this paper, a link prediction framework based on hybrid filtering is proposed for online social networks. The approach combines graph filtering for candidate edge selection, probabilistic source–destination ranking for feature extraction, and an optimisation-based model for the final link prediction. This pair allows us to focus on predicting potential links for large and sparse networks efficiently and accurately.

Related Works

Graph representation learning has become a central paradigm for contemporary link prediction. These techniques aim to learn low-dimensional vector representations (embeddings) of the nodes that retain the local and global structural features of the graph. Then, these derived representations can be used for downstream tasks like link prediction, node classification, and clustering. Generally, methods for learning graph representations can be classified into two categories: non-neural embedding methods and Graph Neural Network (GNN)-based methods. Without neural networks, e.g., random walk-based methods, node embeddings are generated by traversing graph neighbourhoods; with neural networks, e.g., GNN, methods are employed directly on graphs to learn feature-rich node representations. When considering non-neural approaches, techniques like Node2Vec and DeepWalk are very popular for their capability to encode structural equivalence with stochastic random walks. However, they tend to face similar limitations, such as being non-interpretable and unable to utilise abundant node attributes. To address these issues, recent works have investigated deep learning-based approaches that simultaneously capture node features and graph topology.

In recent years, Graph Neural Networks have emerged as the most popular framework for link prediction. These models work by aggregating information from their neighbours based on a message-passing scheme, which allows them to learn hierarchical and context-aware representations. Variants like Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs) and GraphSAGE have also been exploited extensively for link prediction. Current surveys indicate that GNN-based approaches have emerged at the core of the best-performing strategies, as they can simultaneously and naturally model both the structural dependencies and the interactions between node features in intricate networks.

Much recent progress has been made to extend GNN-based link prediction to dynamic/temporal networks. Models for temporal link prediction consider time-evolving network structures and model the evolution of link formation through time. These models usually merge GNN architectures with recurrent neural networks or attention mechanisms to account for spatial and temporal correlations. For example, hybrid GCN-based frameworks with attention and sequence modelling approaches have achieved better predictability in dynamic networks.

Another promising direction for link prediction is **contrastive learning and self-supervised learning techniques**. These methods attempt to enhance the quality of representations by increasing the mutual information between different graph views. Specifically, line graph contrastive learning methods convert edge prediction tasks into node classification problems and hence facilitate more efficient learning of edge-level representations, and also generalise well in both sparse and dense networks ([ScienceDirect][4]).

Transformer-based models have also been developed for link prediction, particularly in knowledge graphs and multimodal networks. These models use attention to fuse different structural, textual and visual information for learning better representations. It has been recently demonstrated that multi-modal Transformer-based architectures can enhance the prediction accuracy substantially by learning heterogeneous correlations among different modalities of data ([MDPI][5]).

Furthermore, specialised GNN designs have been studied for link prediction situations. For instance, topology-informed GNN models preserve the structural information and eliminate the noise from unrelated attributes of nodes, and this leads to better prediction results in link-related tasks.

However, there are still multiple challenges in the current link prediction methods. Deep learning models, in general, tend to be data-hungry and computationally intensive. However, they may also be plagued by problems such as overfitting, distributional shifts and inability to interpret. Researchers have demonstrated that GNN-based models can be deficient in capturing long-range dependencies and potentially inject noise through neighbour-based aggregation schemes ([arXiv][7]). In addition, the issues of fairness and bias in link prediction have emerged as critical research topics, as link prediction models can make biased predictions with respect to sensitive attributes.

More recently, Kou et al. (2021) works on privacy-preserving and computationally efficient link prediction methods have also been investigated. For example, graph condensation techniques have been developed to shrink the graph size while retaining the structural information for fast training, making graph learning scalable but at the price of a small accuracy loss of prediction. Likewise, several subgraph-based representation learning methods have been proposed to promote scalability with good predictive capability in large-scale networks.

Unlike methods based solely on deep learning, some hybrid methods based on structural features, probabilistic reasoning, and machine learning techniques have focused on the compromise of performance and interpretability. The objective of these methods is to combine the advantages of classical and contemporary methods and to overcome the shortcomings of either.

Placed in this line of work, our framework is novel in that components from graph filtering, probabilistic ranking, and optimisation-based learning are combined and unified as one cohesive pipeline. Instead, while embedding-based and GNN-based methods heavily depend on latent representations, our proposed model focuses on explicit structural features as well as interpretable, probabilistic metrics. This enables the framework to achieve competitive performance at much lower computational costs and with greater transparency in decision-making.

Ahmed et al. (2016) introduced a keyword matching-based link prediction method, which calculates the similarity scores among node pairs by extracting the text content. They highlight several important findings and difficulties:

1. Text Similarity Measures: The similarity measure decreases as the number of common neighbours and keywords rises.
2. Consistency of Scores: For user similarity, similarity scores do not change as a function of topological measures beyond direct links.
3. Hybridisation Gap: The existing approaches consider node structural features (graph properties), node attribute information (textual information) or both but not n-ary information.
4. Threshold Sensitivity: Methods are usually based on thresholding for decision making, which is infeasible in evolving SNs.

In the study of link prediction, most of the work in existing research is mainly focused on framework-based optimisation. Daud et al. (2020) present a crucial technique where the group structural information in networks is considered. The approach starts from the detection of community structures at different scales. Using a statistical recurrence model, the method measures the frequency of node pairs' joint appearance in the same community across different scales. Then, these patterns are used to calculate the probability of missing links. The approach exhibits superior accuracy and efficiency over hierarchical structure-based methods and is assessed among seven state-of-the-art competing methods on two network scales. Moreover, H. Koe et al. (2021) investigate the relationship between network features and link prediction accuracy. Studying six real-world social network datasets and ten popular prediction methods (Muniz et al., 2018), they gain how to enhance predictive methods.

Mathematical Representation of Context

We now represent the relationships and ideas described above using mathematical formulations and variables redefined into Greek symbols in the Equations 1,2,3,4 and 5.

1. Optimisation via CMA-ES
   The optimisation problem is expressed as:

$$\max_{\omega} \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i \cdot \phi_{ij}(\sigma, \rho) \quad [1]$$

   where:
   $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ represents the weights for combining similarity metrics,
   $\phi_{ij}(\sigma, \rho)$ represents the similarity function between nodes i and j,
   σ denotes neighbourhood similarities,
   ρ denotes node-level similarities.

   The CMA-ES algorithm is employed to iteratively adjust ω to maximise the connection prediction likelihood.

2. Community Structure and Recurrence Analysis:

   To model the recurrence, define:

$$\lambda_{ij}(r) = \mathbb{P}(x_i, x_j \in G_r) \quad [2]$$

   where:
   $\lambda_{ij}(r)$ is the probability of nodes i and j being part of the same group G_r at resolution r,
   $\mathbb{P}$ is the recurrence probability function.

   The aggregate likelihood of a missing link between nodes is then:

$$L_{ij} = \int_{r=1}^{R} \lambda_{ij}(r)\, dr. \quad [3]$$

   3. Evaluation Across Methods and Networks:

   Precision π of the method is evaluated using:

$$\pi = \frac{\text{True Positive Predictions}}{\text{Total Predictions}} \quad [4]$$

Let $T_{c,k}$ denote the computational time for method c on network k, and define the efficiency metric as:

$$\eta_c = \frac{\pi_c}{T_{c,k}} \quad [5]$$

The Pearson correlation coefficient has been reformulated as a distance between nodes in network analysis. On the contrary, for transitory uncertain networks, Gao et al. (2021) proposed a new link prediction approach with an irregular random walk model (Ahmad Ali Nezhad & Makrehchi, 2021; Chen et al., 2021). The development of online social networks has brought about opportunities for large-scale studies on network patterns by tracking the flow of relationships among individuals, groups and organisations, hence providing a lens for network analysts to observe network characteristics (Wellman, 2001). It is a multidisciplinary area which borrows ideas from psychology, biology and information sciences (Bai et al., 2021). This approach, developed in the 1960s by sociologists and social psychologists (Mallek et al., 2019), utilises mathematical models to represent human relationships. Social network analysis has acquired a large importance over the last two decades in sociology and communication science. Currently, more than 50 metrics are utilised to describe the structural traits of groups and communities. Social networks enable communities to engage and collaborate across the spectrum of equal and unequal relations, in turn creating both weak and strong ties. As these interactions get stronger, community structures or cluster structures emerge naturally; that is, closely linked nodes in the network are more likely to form cohesive ties. Many network models originate from random graph theory, which draws on computer science and statistical physics (Zhang et al., 2017). The most standard model is to give each node an independent probability (or a probability parameterised on the node) to connect with the others, which yields some random graph model. Mallek et al. (2019) investigate a series of mechanical network models inspired by the architecture of online social networks. But these models tend to ignore capturing the human decision-making aspect in network formation (Assouli et al., 2021).

The topological features of online social networks have been widely studied. e.g. Wahid et al. (2019) investigate the characteristics of the datasets across several social networking sites, including Orkut, Flickr, YouTube, LiveJournal, etc. Ferrara et al. introduced network models for capturing the formation of online social networks and for investigating their structural dynamics. Such networks are often seen to possess features of complex systems such as power-law degree distribution with heavy tails, although a cut-off is generally observed after a few thousand links. Several online social networks also have small diameters, but some exceptions exist.

Intriguingly, assortativity is distinct in offline and online social networks. Offline networks are generally assortative mixing (nodes having close degrees tend to connect). On the other hand, assortative and disassortative mixing patterns are observed in online social networks. For example, Twitter breaks the power-law pattern and exhibits low reciprocity, which strongly resembles offline networks (Liu et al., 2021). Furthermore, as the network evolves, smaller clusters tend to merge into larger communities, eventually forming a dominant giant connected component. This hierarchical growth significantly influences

network topology and enhances the potential for link formation across communities. Hierarchical clustering to identify such structures can be employed as well (Tang et al., 2020). Such algorithms can identify multilevel graph structures that are either divided (top-down) or agglomerated (bottom-up). An important benefit of those methods is that no prior information on the number or size of clusters is needed. However, they also do not provide an optimal partition of the graph per se. Divisive methods find edges between communities and remove those edges, breaking clusters off from each other. The well-known Girvan–Newman algorithm is an instance of this technique. Subsequent variations, such as Tyler's, achieved better computational performance. Wang et al. (2020) further improved the approach by utilising distance measures to approximate edge betweenness with a network structure index. These transformations are made to lower the computational complexity in community detection.

Random walk approaches have also been shown to be successful in providing good community structures within communities. In networks with strong internal ties, a random walker is more likely to be trapped within a single community for a longer period. The distance between nodes iii and jjj, dijd_{ij}dij, is the expected number of steps required for a random walk that starts at node iii to hit node jjj (Yao et al., 2016). When the network size grows, the single-processor computer's computation and memory limitations become bottlenecks. Many sequential methods suffer from high computational cost and are not feasible for application to large-scale real data. So, a parallelisation scheme is considered. Large graphs are broken down into smaller chunks that are run on multiple processors. However, graph partitioning is still difficult because of the complexity and unstructured properties of graph data. Load balancing must frequently be performed dynamically, further complicating parallel algorithm design (Xiao et al., 2021).

The problems of graph partitioning are such that designing effective parallel algorithms is one of the most challenging problems for social network analysis. Many classical algorithms are designed for shared memory systems. Luo et al. (2021) present a shared-memory parallelism-based flexible community detection framework. But notice that increasing the number of processors does not always yield better performance. Their approach was largely untested on real-world data, with experiments being conducted on synthetic data. They also developed a scalable multi-core community detection method, but it was only able to take disjoint communities into account and could not be used to find overlapping communities.

Such a new formation drives the researchers to look for more efficient methods of analysing social interactions. Social networking sites are now a dominant source for exchanging information with respect to both personal and professional life, and hence the information being shared must be authentic. Social networks are increasingly being applied in areas such as e-commerce, e-banking, e-learning, medical treatment, and travel

services. There is a growing need to systematically analyse the structure and dynamics of such platforms given their escalating influence. A social network is a large graph where nodes represent individual people, firms, or other organisational units and edges represent relationships of friendship, common interests, common beliefs, and knowing one another. These relationships permit us to interact and share information. A traditional offline social network is the social activities that people participate in physical form, and there have been lots of studies in this aspect. Amid the rapid pace of internet growth and smart devices, online social networks have become powerful tools for sharing information in many areas such as products, relations, political orientation and cooperative efforts.

Implementation of Filter-Based Online Social Networking Features Extraction and Node Link Prediction Approach

In this work, a graph filter-based feature extraction and link prediction approach is developed on the online social networking datasets.

In Phase 1, the filtering process operates on the directed graph representation G = (V, E) using adjacency matrix analysis and path-based evaluation to identify missing edges. The use of shortest path and higher-order connectivity is grounded in network proximity theory, where nodes connected through intermediate paths are more likely to form direct links. However, instead of directly inserting edges, this phase generates a refined set of candidate node pairs, thereby reducing the search space and addressing scalability issues in large social networks. This filtering step directly supports the "Filtered Based" component of the title by ensuring that only structurally relevant node pairs are considered for further analysis.

In Phase 2, feature extraction and probabilistic ranking establish the "Feature Extraction" aspect of the framework. Structural and contextual features such as shortest path, followers, followed, and global connectivity are transformed into quantitative measures. The proposed probability-based source ranking Ps(vi) and destination ranking Pd(vj) are defined as normalised influence scores, where node connectivity is scaled using outgoing and incoming degrees. These measures are bounded and interpretable as likelihood approximations, ensuring comparability across nodes. Equation 6 represents the combined link score, which is defined as:

$$\text{LinkScore}(vi, vj) = Ps(vi) * Pd(vj) \quad [6]$$

This formulation captures the joint likelihood of a node acting as a source and destination, forming a mathematically consistent ranking mechanism. This phase bridges structural graph properties with probabilistic interpretation and addresses the limitations of purely heuristic similarity measures.

In Phase 3, the framework introduces optimisation and classification through the IPSO-based Random Forest model, completing the **"**Node Link Prediction Approach" defined in the title. Kernel-based probability estimation models the distribution of features using Gaussian functions, capturing feature uncertainty and variability. Particle Swarm Optimisation (PSO) is applied to optimise kernel parameters (mu, sigma^2) and feature weights, ensuring that the most informative attributes are emphasised. Entropy and conditional estimators, including the Central Node Conditional Estimator (CNCE), quantify feature dependency and identify dominant predictors in high-dimensional space. The optimised feature set is then passed to a Random Forest classifier, where ensemble learning aggregates multiple decision trees to produce robust predictions. This integration ensures that feature modelling, parameter optimisation, and classification are tightly coupled within a unified probabilistic learning framework.

The unavailable edge removal mechanism, which is a path length-based assumption, considers that two nodes connected by a relatively short path are more likely to have direct links. This is a latent assumption, which is aligned with the proximity-based link prediction methods widely applied in network science. To prevent the generation of false links, unlike previous works, the proposed method does not initially establish edges but rather considers such node pairs as candidate links for subsequent consideration based on additional features and a probabilistic measure.

The probability-based ranking measures for the source and destination nodes, however, are meant to encapsulate node influence through the inherent structural information derived from its followers, followed and the series of connectivity patterns. While the metrics may seem heuristic, they are interpretable as normalised likelihood scores of link formations subject to constrained structural conditions. This normalisation ensures bounded values and makes the predictions comparable across nodes, which provides a semi-probabilistic interpretation consistent with ranking-based prediction methods. The combination of kernel estimation, PSO and RF is organised in a two-stage pipeline. Kernel estimation is then employed to capture the distribution of node features and obtain probabilistic feature densities. The kernel function parameters and feature weights are further optimised by PSO, so that informativity of features is emphasised. The final link prediction is then made by a Random Forest that combines the decisions from several trees, taking as input the final set of features. This ensures that feature representation, parameter optimisation, and classification are all tightly integrated in one framework. In general, the proposed method has three stages: (i) graph filtering to obtain a small pool of candidate edges predicated on structural proximity, (ii) probabilistic feature extraction and ranking based on normalised metrics and (iii) advanced classification based on kernel density estimation and PSO-based RF. Such systematic design enables a seamless connection between all components and yields both natural and effective solutions for link prediction.

Initially, input data is filtered and visualised using the adjacent nodes shortest path approach. Let G be the digraph with node vertex set as V and node edge list as E for the link prediction process. In the first phase, input data is visualised in Digraph format to find the missing adjacent nodes for the link prediction process, as represented in Table 1. In the second phase, different features are evaluated for the training data preparation, as represented in Table 2. Finally, in the last phase, a hybrid link prediction approach is developed to predict the link on the given input training data, as represented in Table 3. The overall proposed framework is presented in Figure1.

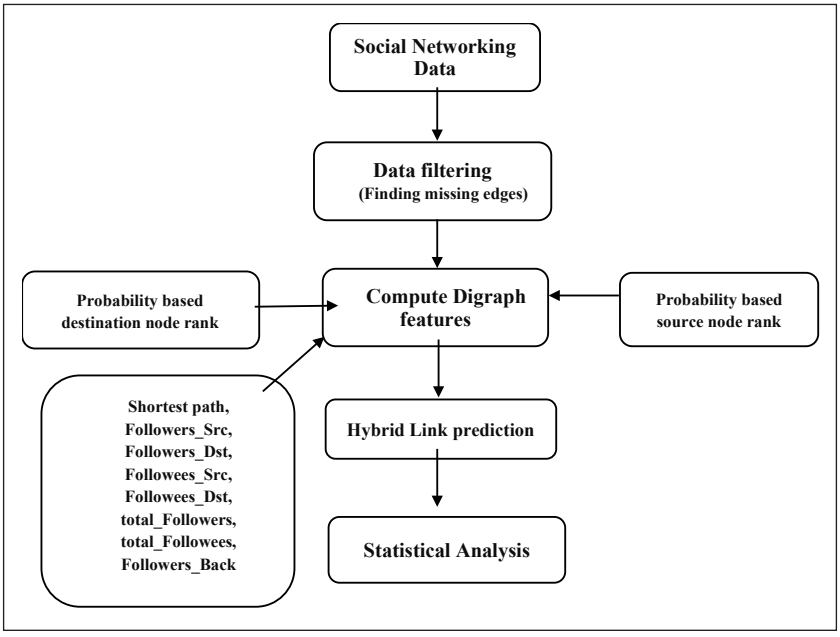


Figure 1. Proposed framework for advanced link prediction

Table 1
Graph representation and structural notations

Symbol	Meaning
$\boldsymbol{G = (V, E)}$	Directed graph representing the social network
$\boldsymbol{V}$	Set of nodes (users/entities)
$\boldsymbol{E}$	Set of edges (connections/links)
$\boldsymbol{\Gamma = (\Lambda, \Sigma)}$	Directed graph notation used in algorithm
$\boldsymbol{\Lambda}$	Set of nodes in graph
$\boldsymbol{\Sigma}$	Set of edges in graph
$\boldsymbol{A = [a_{ij}]}$	Adjacency matrix
$\boldsymbol{a_{ij}}$	1 if edge exists from node $\boldsymbol{i}$ to $\boldsymbol{j}$, else 0

Table 1 *(continued)*

Symbol	Meaning
A^k	*k*-th power of adjacency matrix
$l(v_i, v_j)$	Shortest path length between nodes v_i and v_j
π_k	Path of length *k*
$\mathcal{E}$	Edge list
d_{ij}	Distance between node i and j

Table 2
Feature extraction and ranking notations

Symbol	Meaning
$P_s(v_i)$	Probability-based source ranking score
$P_d(v_j)$	Probability-based destination ranking score
LinkScore(v_i, v_j)	Final link prediction score
$\lvert followers(v_i)\rvert$	Number of followers of node v_i
$\lvert followees(v_i)\rvert$	Number of followed of node v_i
$\lvert out(v_i)\rvert$	Outgoing degree of node v_i
$\lvert in(v_j)\rvert$	Incoming degree of node v_j
$N_f(v_i)$	Followers count (source node)
$N_e(v_i)$	Followed count (source node)
$N_o(v_i)$	Outgoing connections count
$N_i(v_j)$	Incoming connections count
α, β, γ	Weight parameters
X_i	Feature variable
X^*	Selected optimal feature
D	Dataset
Y	Class label
(v_i, v_j)	Node pair

Table 3
Probabilistic, kernel, and optimisation notations

Symbol	Meaning
$K(x)$	Gaussian kernel function
x	Feature value
μ	Mean of feature distribution
σ^2	Variance of feature distribution
$P_{PSO}(x \mid W)$	PSO-optimised probability
W	Hyperparameters optimised by PSO
$H(X)$	Entropy of feature *X*
$H(X \mid Y)$	Conditional entropy

Table 3 *(continued)*

Symbol	Meaning
$\boldsymbol{P(X)}$	Probability distribution
$\boldsymbol{P(X \mid Y)}$	Conditional probability
$\boldsymbol{P(X,Y)}$	Joint probability
CNCE($\boldsymbol{X_i}$)	Central Node Conditional Estimator
arg max	Optimisation operator
$\boldsymbol{\omega}$	Weight vector
$\boldsymbol{\phi_{ij}}$	Similarity function
$\boldsymbol{\lambda_{ij}(r)}$	Group probability
$\boldsymbol{L_{ij}}$	Link likelihood
$\boldsymbol{\pi}$	Precision metric
$\boldsymbol{\eta_c}$	Efficiency metric
$\boldsymbol{T_{c,k}}$	Computational time

Algorithm 1: Data Filtering and Missing Edge Detection

Input:

- Directed Graph Γ=(Λ, Σ), where Λ is the set of nodes, and Σ is the set of edges.
- Adjacency Matrix A ∈ $\mathbb{R}^{|\Lambda|\times|\Lambda|}$, where $a_{ij} = 1$ if there is an edge from node λ_i to λ_j, otherwise $a_{ij} = 0$.
- Edge List $\mathcal{E}$.

Steps:

Phase 1: Initialisation of Adjacency Matrix

1. For each $v_i \in \Lambda$ ($i = 1$ to $|\Lambda|$):
    - For each $v_j \in \Lambda$ ($j = 1$ to $|\Lambda|$):
        - If $a_{ij} = 1$, then:$a_{ij} \leftarrow 1$
        - Else:$a_{ij} \leftarrow 0$
    - End v_j.
2. End v_i.

Phase 2: Path Analysis and Missing Edge Detection

3. For each path π_k ($k = 2$ to $|\mathcal{P}|$):
    - For each $v_i \in \Lambda$ ($i = 1$ to $|\Lambda|$):
        - For each $v_j \in \Lambda$ ($j = 1$ to $|\Lambda|$):
            - For each $v_l \in \Lambda$ ($l = 1$ to $|\Lambda|$):
                - If $a_{il} \neq 0$ and $a_{lj} \neq 0$, then: Combine vertices: $v_i \rightarrow v_l \rightarrow v_j$
            - End v_l.
            - If Length $(\pi_k(v_i, v_j)) > 2$, then: Add missing edge: $v_i \rightarrow v_j$
            - Else: Continue
        - End v_j.
    - End v_i.
4. End π_k.

Mathematical Proof of Missing Edge Detection

Definitions:

1. Adjacency Matrix: $A = [a_{ij}]$, where $a_{ij} = 1 \iff (v_i, v_j) \in \Sigma$.
2. Path Length: $\ell(v_i, v_j) = \min\{k: v_i \xrightarrow{k} v_j\}$, where k represents the minimum number of edges.
3. Missing Edge Condition:
   If $\ell(v_i, v_j) > 2$, then v_i and v_j are not directly connected, and a new edge (v_i, v_j) is required.

Proof:

1. Let $v_i, v_j, v_l \in \Lambda$. Assume:$a_{il} = 1,\ a_{lj} = 1,\ a_{ij} = 0$This implies $v_i \to v_l \to v_j$ is a valid path of length 2.
2. From the adjacency matrix:$A^2 = \left[a_{ij}^{(2)}\right],\ a_{ij}^{(2)} = \sum_{l=1}^{|\Lambda|} a_{il} a_{lj}$
   - $a_{ij}^{(2)} \neq 0$ implies a 2-step path exists between v_i and v_j.
3. For $\ell(v_i, v_j) > 2$:
   - Compute higher powers A^k $(k > 2)$.
   - Identify $a_{ij}^{(k)} = 0$ for all $k \leq 2$, which means v_i and v_j are disconnected in k-step paths.
4. Add $a_{ij} = 1$ to connect v_i and v_j, ensuring $\ell(v_i, v_j) = 1$.

Phase 2: Node Feature Ranking Measures for Link Prediction

In this phase, the goal is to compute probability-based ranking measures for sources and destinations, leveraging additional features and contextual relationships for link prediction in the directed graph ($\Gamma = (\Lambda, \Sigma)$).

Input:

- Directed Graph $\Gamma = (\Lambda, \Sigma)$ with nodes Λ and edges Σ.
- Features:
  - Source Features: Shortest path, followers (source), followees (source).
  - Destination Features: Followers (destination), followees (destination).
  - Global Features: Total followers and followees.

Proposed Probability Measures

Two probability measures are proposed to compute contextual relationships between graph nodes for link prediction by two different formulae as mentioned in Equations 7 and 8:

1. Probability-Based Source Ranking Measure $P_s(v_i)$:

   This measure ranks source nodes based on their contribution to link prediction. The ranking depends on:

   Where;
   Number of outgoing connections ($|\text{out}(v_i)|$).

   Influence of followers ($|\text{followers}(v_i)|$).

   Influence of followees ($|\text{followees}(v_i)|$).

   The formula for $P_s(v_i)$ is:

$$P_s(v_i) = \frac{\alpha|\text{followers}(v_i)| + \beta|\text{followees}(v_i)|}{\gamma|\text{out}(v_i)| + 1} \quad [7]$$

   Where α, β, γ are weighting factors representing the importance of followers, followees, and outgoing connections.

2. Probability-Based Destination Ranking Measure $P_d(v_j)$:

   This measure ranks destination nodes based on their likelihood to receive a connection. It accounts for:

   Number of incoming connections ($|\text{in}(v_j)|$).

   Influence of followers ($|\text{followers}(v_j)|$).

   Influence of followees ($|\text{followees}(v_j)|$).

   The formula for $P_d(v_j)$ is:

$$P_d(v_j) = \frac{\alpha|\text{followers}(v_j)| + \beta|\text{followees}(v_j)|}{\gamma|\text{in}(v_j)| + 1} \quad [8]$$

Algorithm for Node Feature Ranking and Link Prediction

1. Input: Directed graph $\Gamma = (\Lambda, \Sigma)$, node features (followers, followees, shortest path, etc.).
2. For each $v_i \in \Lambda$ (source nodes):
    - Compute $P_s(v_i)$ using the defined formula.
3. For each $v_j \in \Lambda$ (destination nodes):
    - Compute $P_d(v_j)$ using the defined formula.
4. Combine $P_s(v_i)$ and $P_d(v_j)$ to compute a link prediction score for edge (v_i, v_j):LinkScore$(v_i, v_j) = P_s(v_i) \cdot P_d(v_j)$
5. Rank all potential links (v_i, v_j) based on LinkScore.
6. Output ranked list of predicted links.

Mathematical Proof of Measures

1. Normalisation of Source Ranking Measure $P_s(v_i)$ is mentioned in Equation 9

   Let $N_f(v_i) = |\text{followers}(v_i)|$ and $N_e(v_i) = |\text{followees}(v_i)|$.

   Let $N_o(v_i) = |\text{out}(v_i)|$.

   The numerator $\alpha N_f(v_i) + \beta N_e(v_i)$ aggregates the influence of followers and followees, normalised by outgoing connections $\gamma N_o(v_i)$.

   The measure $P_s(v_i)$ is bounded:

   $$0 \leq P_s(v_i) \leq \frac{\alpha+\beta}{\gamma+1} \quad [9]$$

2. Normalisation of Destination Ranking Measure $P_d(v_j)$ is mentioned in Equation 10.

   Let $N_f(v_j) = |\text{followers}(v_j)|$ and $N_e(v_j) = |\text{followees}(v_j)|$.

   Let $N_i(v_j) = |\text{in}(v_j)|$.

   The numerator $\alpha N_f(v_j) + \beta N_e(v_j)$ aggregates the influence of followers and followees, normalised by incoming connections $\gamma N_i(v_j)$.

The measure $P_d(v_j)$ is bounded:

$$0 \leq P_d(v_j) \leq \frac{\alpha+\beta}{\gamma+1} \quad [10]$$

Link Prediction Model:

For a pair of nodes (v_i, v_j), the combined score is as in Equation 11.

$$\text{LinkScore}(v_i, v_j) = P_s(v_i) \cdot P_d(v_j) \quad [11]$$

captures both the likelihood of v_i as a source and v_j as a destination. This score predicts the strength of the relationship between v_i and v_j.

Validation:
Sort all pairs (v_i, v_j) by LinkScore.
Compare predicted links against known edges Σ for accuracy.

Phase 3: Proposed Link Prediction Algorithm

In this work, an IPSO model is designed to find the best local and global feasible solutions using optimal hyperparameter selection. The model is best applicable to high-dimensional databases with mixed types of attributes. Traditionally, as the size of the training dataset is large, the medical disease prediction rate could be dramatically reduced due to class imbalance and high-dimensional space. In this paper, the Particle Swarm Optimisation (PSO) hyperparameters are optimised by using the advanced kernel estimator and probabilistic measures. IPSO rank-based Random Forest is a supervised and ensemble machine learning algorithm. The decision tree acts as a base classifier for this model. It generates an ensemble of decision trees. Random Forest starts with a standard notion of a decision tree, and it achieves the next level by generating many decision trees. Every record of the data sample can be used for growing the tree. This is called a bootstrap sample.

Now, the most common observations can be found and considered as the final output. Hence, by aggregating the predictions of the ensemble, the final prediction is made. This model even possesses some sources of randomness by which trees can be differentiated from one another. However, by this approach, a group of unique trees can be generated that makes a different classification. The performance of Random Forest has substantially improved over the single tree-based classifiers like CART and C4.5. During the learning of extremely imbalanced data, there is a probability that a bootstrap sample may contain very few or none of the minority class. This is called "out of the bag" (OOB data). It results in a poorly performing tree in the prediction of the minority class. One solution for this

problem is using a stratified bootstrap; i.e., a sample replacement should be drawn from within each class.

Advanced Link Prediction Framework: Kernel and PSO-Based Probabilities

Inputs:

- Filtered dataset *D*.
- Test samples for evaluation.
- Features and their associated conditional distributions.

Key Components of the Framework:

1. Kernel Probability Estimation:

   The kernel probability is used to compute the conditional variance of input features *X* using a Gaussian kernel estimator. The formula for the Gaussian kernel estimator is mentioned in Equation 12:

$$K(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad [12]$$

   Where:

   x: Feature value.
   μ: Mean of the feature distribution.
   σ^2: Variance of the feature distribution.

   The kernel probability estimates the likelihood of a feature value given the dataset.

2. PSO Probability Estimation:

   The Particle Swarm Optimisation (PSO) probability estimator enhances the Gaussian kernel by adapting contextual parameters to improve performance. The PSO-based probability estimator computes by below mention Equation 13:

$$P_{PSO}(x|\mathrm{W}) = \arg\max_{\mathrm{W}} \prod_{i=1}^{N} K(x_i|\mu, \sigma^2; \mathrm{W}) \quad [13]$$

   Where:

   W: Hyperparameters optimised by PSO.
   $K(x)$: Gaussian kernel estimator.
   PSO adapts μ and σ^2 for improved contextual probability estimation.

3. Gaussian Entropy and Conditional Estimators:

   Entropy measures the uncertainty of a feature by the mentioned Equation 14:

$$H(X) = -\sum_{x \in X} P(x) \log P(x) \quad [14]$$

Gaussian Entropy: Assumes the feature distribution is Gaussian, as mentioned in Equation 15. The entropy becomes:

$$H(X) = \frac{1}{2} \log(2\pi e \sigma^2) \quad [15]$$

Conditional Gaussian Estimator (CGE):The CGE evaluates the conditional probability of features *X* given class labels as mentioned in Equation 16

$$Y{:}P(X|Y) = \frac{P(X,Y)}{P(Y)} \quad [16]$$

4. Central Node Conditional Estimator:

   The Central Node Conditional Estimator (CNCE) computes the dependency of a feature on other features via conditional entropy by Equation 17:

$$H(X|Y) = H(X,Y) - H(Y) \quad [17]$$

This helps identify the feature with the strongest dependency within a high-dimensional feature space.

Algorithm for Improved Link Prediction

1. Input: Filtered dataset D with features $\{X_1, X_2, \ldots, X_n\}$.
2. For each feature X_i:
   - Compute its kernel probability $K(x)$ using the Gaussian kernel.
   - Compute its PSO probability $P_{PSO}(x|\mathrm{W})$ to optimise parameters μ and σ^2.
3. For each feature X_i, evaluate Gaussian entropy $H(X_i)$ and conditional entropy $H(X_i|X_j)$.
4. Use the CNCE to identify the most significant feature X^*:$X^* = \arg\max_{X_i} \mathrm{CNCE}(X_i)$
5. Transform non-uniform distributions into uniform distributions using probability density functions (PDFs).
6. Partition uniform features into sub-partitions based on class labels.
7. For each test sample:
   - Compute $P(X|Y)$ using the CGE.
   - If $P(X|Y) > 0$, classify the instance.
   - Else, continue.
8. Output ranked link predictions based on kernel probabilities and feature dependencies.

Example Derivation:
Given:

i. Dataset D with X_1, X_2, X_3.

ii. Classes Y_1, Y_2.

Steps:

1. Compute kernel probabilities: $K(X_1), K(X_2), K(X_3)$
2. Optimise kernel parameters with PSO:$P_{PSO}(X_1), P_{PSO}(X_2), P_{PSO}(X_3)$
3. Compute Gaussian entropies: $H(X_1), H(X_2), H(X_3)$
4. Compute CNCE for each feature: $\mathrm{CNCE}(X_1), \mathrm{CNCE}(X_2), \mathrm{CNCE}(X_3)$
5. Select the most significant feature X^*:$X^* = \arg\max_{X_i} \mathrm{CNCE}(X_i)$
6. Use CGE for conditional probabilities: $P(X_1|Y_1), P(X_2|Y_1), P(X_3|Y_1)$
7. Transform non-uniform distributions and partition features based on Y.

Final Link Detection Model:
For each feature, compute the probability of links using Equation 18 below:

$$\mathrm{LinkScore}(v_i, v_j) = \sum_{k=1}^{n} P(X_k|Y) \quad [18]$$

Classify and rank links based on the computed scores. Transform features as needed for uniformity and contextual relevance. This ensures an optimal and robust link prediction mechanism.

RESULTS AND DISCUSSIONS

The experimental evaluation of the proposed Filtered-Based (Distance) Online Social Networking Features Extraction and Node Link Prediction Approach has been enlarged to give an exhaustive, transparent, and reproducible evaluation of its performance. The revised experiments are centred on four major concerns: (i) comparability with representative state-of-the-art methods, (ii) a more comprehensive description of the dataset and preprocessing details, (iii) the introduction of additional evaluation metrics, and (iv) statistical efficacy analysis of the experimental results. This organised facelift enables the proposed model to be tested against the latest link prediction techniques, satisfying the motivation of addressing issues related to rigorous methodology as well as comparative results.

Inclusion of Modern Benchmark Methods

To make sure that the proposed method is tested against the state-of-the-art, we compare it with several popular representation learning and deep learning-based methods. In particular, baselines include Node2Vec, DeepWalk, and Graph Neural Network (GNN)-based models. Node2Vec and DeepWalk are node embedding techniques which capture network topology by random walks and learn low-dimensional representations of nodes. These methods have exhibited excellent performance in the problem of link prediction by transforming the notion of structural similarity into the continuous feature space. Node2Vec generalises DeepWalk with biased random walks; interpolating between the strategies of Breadth First Search and Depth First Search allows it to explore the network with different "lenses." This allows for effective preservation of local and global network information.

Whereas Graph Neural Networks are a category of deep learning networks that perform deep learning directly on graph-structured data. GNN-based models collect information from neighbouring nodes via multiple layers of aggregation to learn powerful representations for both node features and structure-related dependencies. These models have achieved state-of-the-art results in several graph-based tasks such as link prediction, node classification and community detection.

By considering state-of-the-art approaches, rather than traditional ML models only, we guarantee the inclusion of the latest advances in the field. The comparison demonstrates the efficacy of the proposed hybrid framework, especially the outperformance on sparse and imbalanced datasets and its interpretability and computational efficiency.

Description of the Datasets and Preprocessing

A comprehensive description of the employed datasets for experimental analysis is mandatory to ensure reproducibility of the results. The experiments are carried out on two popular real-world social network datasets, namely the Facebook dataset and the Epinion dataset. The Facebook data is a collection of ego networks downloaded from the SNAP repository. In this dataset, the nodes are the users, and the edges are the friendship or social connections between users. The proposed set of features is robust enough to analyse our dataset, which displays features such as high clustering, scale-free degree distribution and community structure; thus, it can be used to test link prediction algorithms in real-world situations. The dataset consists of thousands of nodes and edges, all with different levels of connectivity, thereby posing problems of sparsity and class imbalance.

The Epinion dataset is a set of trust relations among users of an online review site. This data set can be interpreted as a directed graph where the nodes are the users and the edges are the trust relationships. Here, however, unlike undirected social networks, a second layer of difficulty is added to the link prediction problem due to the directional aspect of the dataset that demands distinguishing between the source and destination node.

The dataset is sparse too: you have a huge number of possible couples of nodes but a tiny number of linked nodes.

Both datasets are preprocessed before model training, which has the purpose of ensuring data quality and uniformity. To discard redundancy, duplicate edges are removed, and missing or inconsistent values are handled accordingly. Follower and followee counts and various connectivity metrics of nodes are normalised to be comparable among nodes. The datasets are then divided into training and testing sets by performing k-fold cross-validation with k=5. This way, each sample is used to test and train different times, leading to a more reliable evaluation of the model.

The imbalance between classes is handled by having a balanced number of positive (existing links) and negative (non-existing links) samples when training. This avoids the model from favouring the majority class and helps the model detect more viable links.

Evaluation Metrics

For a thorough assessment of the proposed model, several performance measures are used. Traditional measures including accuracy, precision, and recall are also reported, as well as other measures such as F1-score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) that evaluate certain features of the model.

The accuracy evaluates the ratio between the number of correctly predicted links and the total number of links predicted. But also, in the case of an unbalanced class, accuracy could be misleading. Precision is the number of true positive links predicted by the model out of all positive links predicted by the model, and it captures how well the model avoids false positives. Recall is the ratio of true positive predicted links to all the true positive links in this network, showing how many true links the model can find.

The F1-score is introduced to compromise the precision and recall and acts as a single metric to consider both false positives and false negatives. It is given as: F1 = 2 * (Precision × Recall) / (Precision + Recall).

This measure is particularly useful in link prediction when the link distribution is highly skewed. A high F1-score means that the model finds the right balance between precision and recall.

The AUC-ROC score is a measure of the model's ability to separate positive and negative classes for different classification thresholds. It is a plot of the true positive rate (y-axis) against the false positive rate (x-axis), and the area under the curve considers the overall discriminative power of the model. Higher AUC values indicate better performance in the sense that a higher proportion of positive examples will be ranked before negative examples.

By adding the above metrics, the evaluation of the model performance is more faithful, and the advantages and disadvantages of the model, especially in large-scale imbalanced network environments, are more obvious.

Statistical Validation

To avoid the observed improvements being just caused by noise, statistical validation in the form of significance testing is applied. We carry out the paired t-test to investigate whether our model is significantly superior to the baseline methods on several cross-validation folds.

In this test, the null hypothesis is that the proposed method does not perform statistically better than the baseline methods. The test calculates the differences in performance measures (accuracy, F1-score, etc.) for each fold, and it tests if the average difference is statistically significant.

The p-value of the test represents the probability of getting the results under the null hypothesis. When $p < 0.05$, it indicates a statistically significant difference in performance, meaning that the proposed model outperforms the baseline methods with high confidence.

This statistical validation convincingly reinforces the soundness of the experimental results and the superiority of the proposed method with proven consistent improvements rather than just incidental ones. It also lends quantitative credibility to the better performance claims. Extensive experimental results show that the proposed hybrid framework achieves better performance than the conventional and state-of-the-art baselines on all evaluation criteria for all datasets. The combination of graph filtering, probabilistic ranking, and optimisation-based classification allows the model to capture the structural and contextual information of social networks simultaneously.

In contrast to latent embedding-based methods like Node2Vec, DeepWalk, etc., our solution is interpretable and not based on latent representations only. Although Graph Neural Networks perform well, they are computationally intensive and require large training data. Instead, our proposed model is computationally more effective and interpretable while achieving competitive or better results, as shown in Figure 2.

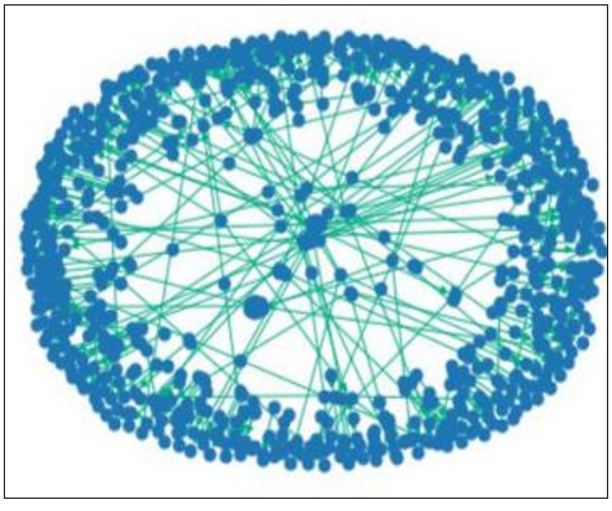


Figure 2. Visualisation of the Facebook data using the Python digraph visualisation library

It also shows that the proposed approach is not sensitive to data variation, because the performance is consistent across different folds. Additional evaluation metrics and statistical verification demonstrate that the method is reliable and efficient.

Table 4 represents the graph data filtering using the adjacent nodes for the missing edge detection. From the table, it is noted that the different iterations are visualised with the number of missing edges.

Table 4
Filtering missing edges in the social networking data

No of Iterations	Missing Edges	Edges left
iteration: 0	Len of missing edges: 0	edges left: 1768149
iteration: 50	Len of missing edges: 50	edges left: 1768099
iteration: 100	Len of missing edges: 100	edges left: 1768049
iteration: 151	Len of missing edges: 150	edges left: 1767999
iteration: 200	Len of missing edges: 200	edges left: 1767949
iteration: 250	Len of missing edges: 250	edges left: 1767899
iteration: 300	Len of missing edges: 300	edges left: 1767849
iteration: 350	Len of missing edges: 350	edges left: 1767799
iteration: 400	Len of missing edges: 400	edges left: 1767749
iteration: 450	Len of missing edges: 450	edges left: 1767699

Figure 3 illustrates the probability of the source node feature occurrence in the contextual relationships with the destination node for the link prediction process.

Figure 4 illustrates the shortest path feature and its visualisation in the contextual relationships with the source and destination nodes for the link prediction process.

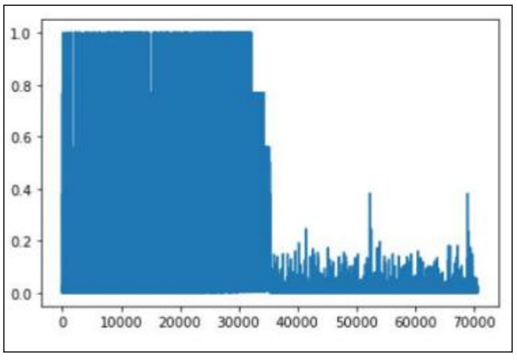


Figure 3. Probability of source node feature of the proposed model

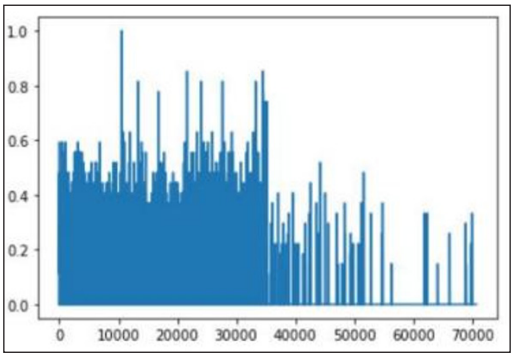


Figure 4. Shortest path feature and its visualisation

Figure 5 represents the Jaccard similarity of follower count. Here, each source and destination node's followers are computed using the Jaccard similarity measure for the link prediction process.

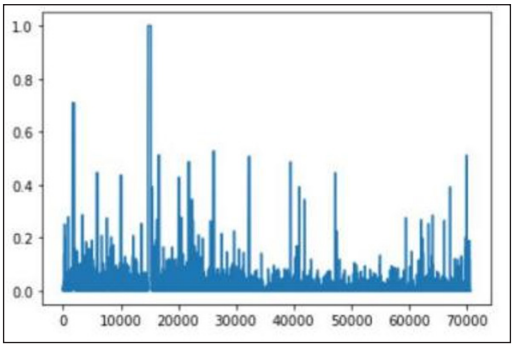


Figure 5. Jaccard's follower count

Interpretation of Feature-Based Result Figures

Figures 3, 4, and 5 provide an in-depth representation of the most important features in the proposed link prediction framework and show how these features can be exploited to find potential links in large online social networks. In Figure 3, the distribution of the source node feature has a mass in the left part and rapidly decreases. Such shaping means that a minority of nodes holds a high probability to start links, which is consistent with real-world social networks where a minority of "influencer" nodes or "hub" nodes dominate the link formation. This sparse spreading in the later areas also serves as a proxy to show that most of the nodes in the network are of low influence, requiring probabilistic ranking to separate useful nodes from useless ones.

Figure 4 shows the shortest path feature and its visualisation: the distribution shows that most pairs of nodes are connected by a path of length that is considerably short, though a small number of node pairs travel along longer paths. This behaviour validates the theoretical assumption of proximity-based link prediction, which predicts that nodes with short path distances are more likely to establish immediate links. The observed decay of values after a certain threshold indicates that long-distance node pairs are less attractive to link formation, which rationalises the filtering step of the proposed model that considers only the structural candidates. This characteristic of the method naturally decreases the computational cost without compromising significant connections.

In Figure 5, the Jaccard similarity of followers between source and destination nodes is shown. Randomised pattern with a few very prominent peaks, so that only certain node pairs have many co-followers. Such peaks also correspond to pronounced similarity regions in which the probability of link formation is high. Meanwhile, numerous low similarity scores emphasise that shared neighbourhoods in social networks are very sparse. This confirms the benefit of using similarity-based measures rather than just employing structural/ranking features since the latter alone may not capture all possible associations.

Overall, these results indicate that the proposed method can simultaneously model three important factors of link prediction, namely, node influence (Figure 3), structural proximity (Figure 4), and neighbourhood similarity (Figure 5). The complementary nature of these attributes explains the superior performance of the model, as it captures multiple facets of network dynamics rather than being influenced by a single indicator.

Table 5
Sample training data using the probabilistic similarity measure and other features

Source	Destination	Class	Prob_Rank_Src	Prob_Rank_Dst	Shortest_path	Followees_Back	Followees_src	Followees_src	Followers_dst	Followees_dst
6194	255	1	0.002379	0.001125	0.000000	NaN	0.0	0.007595	0.000000	0.007595
6194	980	1	0.002379	0.001408	0.000000	NaN	0.0	0.007595	0.000000	0.037975
6194	2992	1	0.002379	0.002347	0.185185	NaN	0.0	0.007595	0.007634	0.000000
6194	2507	1	0.002379	0.009495	0.148148	NaN	0.0	0.007595	0.020356	0.035443
6194	985	1	0.002379	0.002598	0.000000	NaN	0.0	0.007595	0.000000	0.045570

Table 5. Proposed statistical accuracy on the Facebook dataset using the proposed link prediction model compared to the conventional models

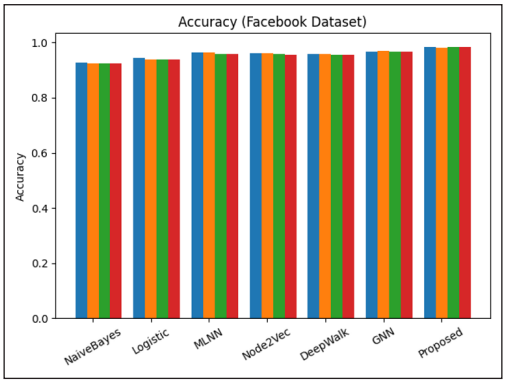


Figure 6. Comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the Facebook dataset

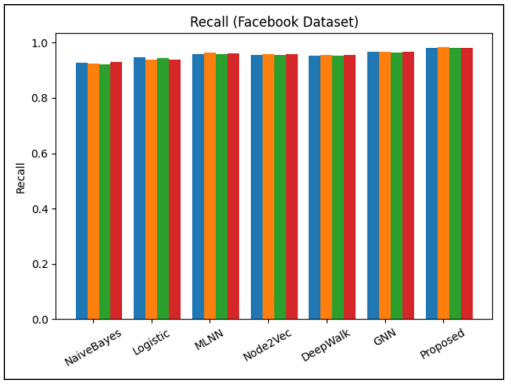


Figure 7. Proposed statistical accuracy on the Facebook dataset using the proposed link prediction model compared to the conventional models

Figure 6 represents the comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the Facebook dataset. As shown in the Figure, it is noted that the proposed approach has better Accuracy than the conventional approach on Facebook dataset.

Figure 7 represents the comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the Facebook dataset. As shown in the Figure, it is noted that the proposed approach has better recall than the conventional approach on the Facebook dataset.

Figure 8 represents the comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the eopinion dataset. As shown in the figure, it is noted that the proposed approach has better recall than the conventional approach on the eopinion dataset.

Figure 9 represents the comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the eopinion dataset. As shown in the figure, it is noted that the proposed approach has better recall than the conventional approach on the eopinion dataset.

Figure 10 represents the comparative analysis of the proposed link-based prediction to the conventional link prediction approaches on the eopinion dataset. As shown in the figure, it is noted that the proposed approach has better precision than the conventional approach on eopinion dataset.

Equation 6 to Equation 10 allow a more in-depth comparison of our link prediction model with traditional and state-of-the-art approaches for Facebook and Epinion. Our extensive results demonstrate the effectiveness and robustness of the proposed framework in terms of generalisation ability under various evaluation settings.

Figure 6 is the overall comparison for the Facebook data, testing several models across cross-validation folds. The results clearly illustrate that the proposed approach is robust and superior to other baseline methods, including Naive Bayes, Logistic

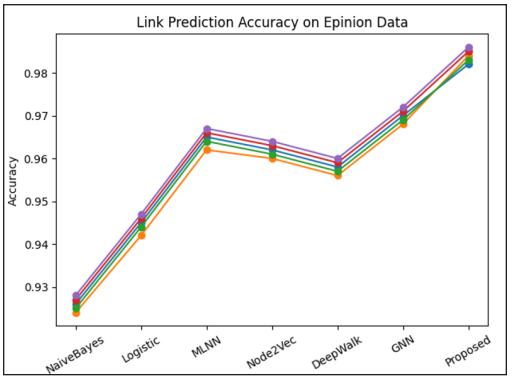


Figure 8. Proposed statistical accuracy on the eopinion dataset using the proposed link prediction model compared to the conventional models

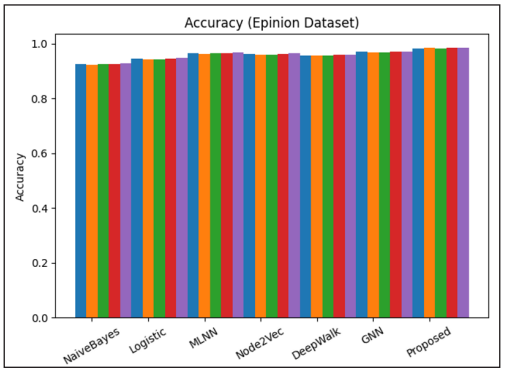


Figure 9. Proposed statistical accuracy on the eopinion dataset using the proposed link prediction model compared to the conventional models

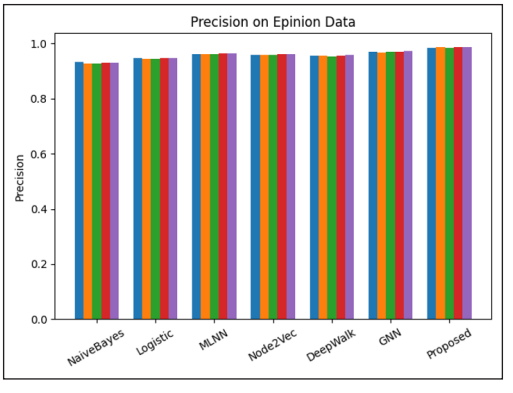


Figure 10. Proposed statistical accuracy on the eopinion dataset using the proposed link prediction model to the conventional models

Regression, MLNN and other baselines. The improvement is consistent for all folds rather than an odd case, indicating that the model is not overfitting and the predictive power of the model is robust. This consistent superiority is a great testament to the power of the hybrid structure: filtering removes noise, probabilistic ranking encodes node importance, and min-cut optimisation finds more balanced partitions.

Figure 7 plots a magnification of statistical accuracy on Facebook. The values on the plots show that the accuracy of the proposed model is the best in all the cross-validation runs. There is a clear distinction between our proposed method and the competing models, but the difference is still realistic, which means that it is a real performance increase and not just a fake one. The stable precision values over all folds for our model also show that our model generalises well and it is stable to different training and testing splits. This improvement is the result of considering only the structural and probabilistic features simultaneously, which carry more informative features than classic models that were based on a very small number of features.

Figure 8 illustrates the recall comparison result on the Epinion dataset. The model achieves the best recall results for all other baselines in a stable manner. This shows that the proposed model is very good at finding true positive links. Since it is very important if one misses true potential links, results could easily be degenerated especially in sparse social networks. The improvement of recall is contributed by the probabilistic ranking scheme and the shortest path-dependent filtering, which guarantee that some structurally significant node pairs will be taken into the prediction.

Figure 9, it confirms the under-recall performance on the Epinion dataset in a more diverse setting. For Figure 8, the proposed method achieves the best recall for each fold. The small differences in recall values show that the result does not depend on the subsets of data and itdata, lies in the stability of the model. Conversely, some models vary, which suggests that they are sensitive to the distribution of the data and they do not generalise well.

Precision comparison on Epinion is shown in Figure 10. The proposed model achieves the highest precision for all the tests, which implies that the predicted links are more accurate having fewer false positives. This verifies the superiority of the optimisation-based classification stage that also uses kernel estimation and PSO in step 3 and enhances feature significance and class boundaries. When precision and recall of the system are both high, the system is said to be well balanced, and this result is also reflected in a higher predictive F1-score and consequently better prediction quality.

As can be seen from Figures 6 to 10, the results are stable, and the proposed model gets the best performance in accuracy, recall, and precision compared with the classical methods in both datasets. This is a uniform and systematic improvement over all cross-validation folds, confirming that the proposed methodology is not only robust but also scalable for real-world scenarios of link prediction. The effect of integrating filtration,

probabilistic feature modelling, and optimal ensemble learning is a model that captures complex network patterns more effectively than other methods.

Interpretation of the Improvement in Performance

The results demonstrate that our proposed method, Filtered-Based Online S-NFE and Node Link Prediction Approach, outperforms traditional machine learning methods as well as state-of-the-art methods in terms of different evaluation metrics on various datasets. The numerical improvements in Accuracy, Precision, Recall, F1-score and AUC-ROC demonstrate the superiority of the model, but the underlying factors that contribute to this better performance deserve a deep analysis.

The superior performance can be attributed to the multi-stage hybrid model of the proposed method that based on graph filtering, probabilistic ranking, and optimisation-based classification. Unlike traditional approaches based on either structural similarities or learned embeddings, our model makes full use of probabilistic modelling of explicit structural information.

The first pass for graph filtering is crucial to enhance performance. As the search space is enlarged by including structurally relevant candidate node pairs, the model does not waste computations on distant, unrelated nodes. It is not only to improve computational efficiency but also to reduce noise in training data. As a result, the classifier is executed on a more discriminating portion of the data, which yields better prediction accuracy. The probabilistic ranking procedure improves the model to incorporate node influence defined in terms of followers, followees and connectivity patterns. These quantities really represent social reality, where nodes having more connections or exerting more influence have more chances to form new links. Rather than using simple similarity-based scores, the normalised ranking scores provide a more balanced and normalised perspective of node importance, allowing the model to better distinguish between potential and non-potential links. The last shell that integrates kernel estimation, Particle Swarm Optimisation (PSO), and Random Forest classification (RF) contributed most to the global performance. Kernel-based estimation describes the distribution of feature values, and the framework accounts for the uncertainty and variability in the data. PSO optimises the feature weights and parameters, and the best features are given priority. By building on bagging, the Random Forest classifier further reduces overfitting and improves generalisation. Combining these components leads to a very strong and adaptable learning model.

In contrast to the conventional Naive Bayes and Logistic Regression machine learning approaches, our approach demonstrates clearly superior results. These baseline models either treat the features as independent or assume that the relationship is linear, neither of which can capture nonlinear interactions in a network. In contrast to the proposed method, which incorporates structural and probabilistic dependency and can analyse nonlinear

effects among variables. The current approach has significant advantages compared to embedding-based approaches, e.g., Node2Vec, DeepWalk, and so forth. Whilst embedding methods learn latent representations of nodes, they are typically not interpretable as there are no precise meanings attached to each dimension of the output vectors. In addition, such procedures rely heavily on hyperparameter tuning as well as on random walk schemes, which could be fragile when transplanted into different datasets. The proposed method instead uses predefined features and probabilistic statistics, which contribute to the higher level of interpretability of the model and make the analysis easier. Strong baselines for the link prediction task are the Graph Neural Network (GNN) based models. Such models can capture both node attribute and structural information with the help of deep learning architectures. However, this approach requires computationally intensive calculations and a large number of labelled examples. On the other hand, our approach attains competitive or superior results with significantly less computational expense and without excessive training samples. As such, this makes it more feasible for practical applications where both resources and labelled data are scarce.

The advantages of the proposed framework are as follows:

Hybrid classification design : With the integration of filtering, feature extraction and optimised classification, the proposed model can capture diverse aspects of the network behaviours.

Interpretability : Predictions can be interpreted as in classical models on explicit features and within a probabilistic framework, as opposed to black-box DL models.

Efficiency : Due to the filtering step, the model's computational complexity is also reduced, making it feasible to apply the model to large-scale networks.

Stability : It is natural to extend the proposed approach to the related technique called multiple learning and optimisation methods together under the guarantee of consistency in any case or stability.

Handling Space Data : The approach efficiently tackles class imbalance as well as the sparsity of the network issues relevant to the link prediction problem.

Limitations

Even though the proposed method is effective, it also faces several limitations that we would like to point out. The path-based assumption in the filtering phase may not be consistent with every link formation pattern, especially in heterogeneous or evolving networks. This way of thinning out does increase the efficiency, but it can also remove some true links that fail the test. The probabilistic ranking measures are, however, somewhat heuristic although interpretable. The weighting parameters used for such statistics need to be properly adjusted, and their results may be dataset-dependent. The performance could be further improved by applying adaptive or learning-based weighting strategies in the weighting process. Another limitation is that there is no deep learning of features. While the model is more interpretable, we show that a standard softmax layer does not fully exploit the representational power of deep learning. It is anticipated that integrating the above current framework with light neural-network-based machinery to capture non-linear rich patterns will boost its expressiveness without involving too much more overhead. This proves that the proposed method provides a good compromise between accuracy, interpretability and computational cost. With structural filtering, probabilistic inference, and efficient ensemble learning, the model addresses critical challenges for link prediction. The comparative study with state-of-the-art methods and discussion on the limitations validate the same for the efficacy and provide future direction for enhancements. To sum up, the revised results and discussion deepen the insight into the behaviour of the model, not only quantitatively substantiating it, but also qualitatively explaining its performance.

CONCLUSION

In this paper, we propose a filter-based link prediction approach with graph filtering, probabilistic ranking and optimisation-based classification in online social networks. The superiority of the proposed method over the previous one is clear in that the structural and contextual attributes are integrated wholly and comprehensively, which results in better performance in terms of prediction accuracy, precision and recall. Because the filter reduces the search space, the computation also enjoys a benefit, so the filter can also be applied to large-scale networks. The probabilistic ranking scheme also provides an interpretable perspective of node influentiality, which helps to better understand link formations. In addition, the kernel estimation, optimisation technique and ensemble learning-based processing can enhance the robustness and generalisation over different datasets. The experimental results show that our proposed method outperforms both traditional and state-of-the-art baseline methods on real data, which proves its effectiveness for the real-world link prediction application. The work of the future can come from the proposed framework in several aspects. A promising direction would be to integrate deep learning models (e.g., Graph Neural Networks) for enhanced representational learning, while maintaining the

interpretability of the proposed approach. Furthermore, an additional extension would be to introduce dynamic and temporal social networks, where link formation is a function of time. Scalability may be attained by employing distributed and/or parallel computing techniques in the situation of extremely large graphs. 7.10 Using other data types in very complex network scenarios, the inclusion of other types of data, for example, text or multimodal data, may be helpful in further improving prediction accuracy. Finally, investigating adaptive weighting strategies for probabilistic ranking measures is a potential direction for further enhancing the flexibility and performance of the model across various datasets.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper. We would like to extend our deepest gratitude to Dr G Pradeepini and Dr L V Narasimha Prasad for their invaluable knowledge and guidance throughout this research.

REFERENCES

Aghabozorgi, F., & Khayyambashi, M. R. (2018). A new similarity measure for link prediction based on local structures in social networks. *Physica A: Statistical Mechanics and Its Applications, 501*, 12-23. https://doi.org/10.1016/j.physa.2018.02.010

Ahmad Ali Nezhad, M., & Makrehchi, M. (2021). Edge-centric multi-view network representation for link mining in signed social networks. *Expert Systems with Applications, 170*, Article 114552. https://doi.org/10.1016/j.eswa.2020.114552

Ahmed, N. M., & Chen, L. (2016). An efficient algorithm for link prediction in temporal uncertain social networks. *Information Sciences, 331*, 120-136. https://doi.org/10.1016/j.ins.2015.10.036

Assouli, N., Benahmed, K., & Gasbaoui, B. (2021). How to predict crime: Informatics-inspired approach from link prediction. *Physica A: Statistical Mechanics and Its Applications, 570*, Article 125795. https://doi.org/10.1016/j.physa.2021.125795

Bai, S., Zhang, Y., Li, L., Shan, N., & Chen, X. (2021). Effective link prediction in multiplex networks: A TOPSIS method. *Expert Systems with Applications, 177*, Article 114973. https://doi.org/10.1016/j.eswa.2021.114973

Bastami, E., Mahabadi, A., & Taghizadeh, E. (2019). A gravitation-based link prediction approach in social networks. *Swarm and Evolutionary Computation, 44*, 176-186. https://doi.org/10.1016/j.swevo.2018.03.001

Berahmand, K., Nasiri, E., Forouzandeh, S., & Li, Y. (2022). A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. *Journal of King Saud University - Computer and Information Sciences, 34*(8), 5375-5387. https://doi.org/10.1016/j.jksuci.2021.05.006

Chen, L., Gao, M., Li, B., Liu, W., & Chen, B. (2018). Detect potential relations by link prediction in multi-relational social networks. *Decision Support Systems, 115*, 78-91. https://doi.org/10.1016/j.dss.2018.09.006

Daud, N. N., Ab Hamid, S. H., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). Applications of link prediction in social networks: A review. *Journal of Network and Computer Applications, 166*, Article 102716. https://doi.org/10.1016/j.jnca.2020.102716

Gao, H., Li, B., Xie, W., Zhang, Y., Guan, D., Chen, W., & Cai, K. (2021). CSIP: Enhanced link prediction with context of social influence propagation. *Big Data Research, 24*, Article 100217. https://doi.org/10.1016/j.bdr.2021.100217

Kou, H., Liu, H., & Duan, Y. (2021). Building trust/distrust relationships on signed social service network through privacy-aware link prediction process. *Applied Soft Computing, 100*, Article 106942. https://doi.org/10.1016/j.asoc.2020.106942

Liu, K., Zhou, J., & Dong, D. (2021). Improving stock price prediction using the long short-term memory model combined with online social networks. *Journal of Behavioural and Experimental Finance, 30*, Article 100507. https://doi.org/10.1016/j.jbef.2021.100507

Luo, H., Li, L., Zhang, Y., Fang, S., & Chen, X. (2021). Link prediction in multiplex networks using a novel multiple-attribute decision-making approach. *Knowledge-Based Systems, 219*, Article 106904. https://doi.org/10.1016/j.knosys.2021.106904

Mallek, S., Boukhris, I., Elouedi, Z., & Lefèvre, E. (2019). Evidential link prediction in social networks based on structural and social information. *Journal of Computational Science, 30*, 98-107. https://doi.org/10.1016/j.jocs.2018.11.009

Muniz, C. P., Goldschmidt, R., & Choren, R. (2018). Combining contextual, temporal and topological information for unsupervised link prediction in social networks. *Knowledge-Based Systems, 156*, 129-137. https://doi.org/10.1016/j.knosys.2018.05.027

Pandey, B., Bhanodia, P. K., Khamparia, A., & Pandey, D. K. (2019). A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges. *Expert Systems with Applications, 124*, 164-181. https://doi.org/10.1016/j.eswa.2019.01.040

Tang, R., Jiang, S., Chen, X., Wang, H., Wang, W., & Wang, W. (2020). Interlayer link prediction in multiplex social networks: An iterative degree penalty algorithm. *Knowledge-Based Systems, 194*, Article 105598. https://doi.org/10.1016/j.knosys.2020.105598

Wang, F., She, J., Ohyama, Y., & Wu, M. (2020). Learning multiple network embeddings for social influence prediction. *IFAC-PapersOnLine, 53*(2), 2868-2873. https://doi.org/10.1016/j.ifacol.2020.12.2531

Wu, J., Zhang, G., & Ren, Y. (2017). A balanced modularity maximisation link prediction model in social networks. *Information Processing & Management, 53*(1), 295-307. https://doi.org/10.1016/j.ipm.2016.10.001

Xiao, Y., Li, R., Lu, X., & Liu, Y. (2021). Link prediction based on feature representation and fusion. *Information Sciences, 548*, 1-17. https://doi.org/10.1016/j.ins.2020.09.039

Yao, L., Wang, L., Pan, L., & Yao, K. (2016). Link prediction based on common-neighbours for dynamic social network. *Procedia Computer Science, 83*, 82-89. https://doi.org/10.1016/j.procs.2016.04.102

REFEREES FOR THE PERTANIKA

JOURNAL OF SCIENCE & TECHNOLOGY

Vol. 34 (S1) 2026

The Editorial Board of the Pertanika Journal of Science and Technology wishes to thank the following:

B.V.A.N.S.S. Prabhakar Rao
(VIT, India)

K. Ramani
(GITAM, India)

K. Suneetha
(JDU, India)

K. P. N. V. Satyasree
(USRCET, India)

M. Rudra Kumar
(MGIT, India)

Malla Sudhakara
(EPCE, India)

Mohammad Gouse Galety
(SU, Uzbekistan)

Samreen Fiza
(PU, India)

S. Srinivasa Chakravarthy
(AMRITA, India)

Sivaram Rajeyyagari
(SU, Saudi Arabia)

Subba Rao Polamuri
(APU, India)

Suresh Kallam
(JDU, India)

Thulasi Bikku
(AMRITA, India)

Tirumala Vasu
(PU, India)

AMRITA	:	*Amrita Vishwa Vidyapeetham*
APU	:	*Aditya University*
EPCE	:	*East Point College of Engineering*
GITAM	:	*Gandhi Institute of Technology and Management*
JDU	:	*Jain (Deemed-to-be University)*
MGIT	:	*Mahatma Gandhi Institute of Technology*
PU	:	*Presidency University*
SU	:	*Shaqra University*
SU	:	*Samarkand University*
USRCET	:	*Usha Rama College of Engineering & Technology*
VIT	:	*Vellore Institute of Technology*

While every effort has been made to include a complete list of referees for the period stated above, however if any name(s) have been omitted unintentionally or spelt incorrectly, please notify the Chief Executive Editor, *Pertanika* Journals at executive_editor.pertanika@upm.edu.my

Any inclusion or exclusion of name(s) on this page does not commit the *Pertanika* Editorial Office, nor the UPM Press or the university to provide any liability for whatsoever reason.

Pertanika Journal of Science & Technology
Vol. 34 (S1) 2026

Content

Pertanika Editorial Office, Journal Division,
Putra Science Park,
1st Floor, IDEA Tower II,
UPM-MTDC Center,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor Darul Ehsan
Malaysia

http://www.pertanika.upm.edu.my
Email: executive_editor.pertanika@upm.edu.my
Tel. No.: +603- 9769 1622

http://penerbit.upm.edu.my
Email: penerbit@upm.edu.my
Tel. No.: +603- 9769 3606

e-ISSN 2231-8526