Original Articles

BenchCouncil Transactions on Benchmarks, Standards and Evaluations (TBench) is an open-access multi-disciplinary journal dedicated to benchmarks, standards, evaluations, optimizations, and data sets. This journal is a peer-reviewed, subsidized open access journal where The International Open Benchmark Council pays the OA fee. Authors do not have to pay any open access publication fee. However, at least one of the authors must register BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench) (https://www.benchcouncil.org/bench/) and present their work. It seeks a fast-track publication with an average turnaround time of one month.

Contents

Full length article

Hybrid deep learning model for identifying the cancer type

Singamaneni Krishnapriya [a],*, Hyma Birudaraju [b], M. Madhulatha [a], S. Nagajyothi [c], K.S. Ranadheer Kumar [d]

[a] CSE-(Cys, DS) and AI & DS, VNR Vignana Jyothi Institute of Engineering and Technology, Bachupally, Hyderabad 500090, Telengana, India
[b] Guru Nanak Institutions Technical Campus, Hyderabad, Telengana, India
[c] ECE, CMR Institute of Technology, Kandlakoya, Medchal, Hyderabad 501401, Telengana, India
[d] CSE-Data Science, CVR College of Engineering, Hyderabad, Telengana, India

ARTICLE INFO

ABSTRACT

Despite current advances, cancer remains one of the biggest health challenges globally, and diagnosis must be made earlier to begin treatment. In this work, we introduce a hybrid deep learning-based framework for accurate cancer type and subtype identification by using pre-trained convolutional neural networks, custom deep learning networks, and traditional machine learning classifiers. I have achieved accurate results on more complex cancer datasets using advanced architectures of CNN + LSTM and attention-based models, along with the pre-trained models of VGG19, Xception, and AmoebaNet. Model reliability and interpretability are further improved using ensemble techniques such as confidence-based and XOR fusion. Experimental results in multiple multimodal datasets demonstrate the effectiveness of our hybrid approach by improving precision, recall, and F1 scores in various types of cancer. However, they have promising results and remain challenging to deploy for rare cancer subtypes or explain to gain clinical adoption. The proposed framework provides a basis for personalized cancer by developing machine learning innovations to advance precision medicine.

1. Introduction

Cancer remains one of the leading causes of mortality worldwide due to its diverse subtypes and complex molecular heterogeneity. Early and precise classification of cancer types and subtypes is critical for personalized treatment planning and improving patient outcomes. However, conventional diagnostic methods, including histopathology and genomic profiling, are often limited by observer variability, interpretability issues, and delayed results. These gaps require robust, automated, and interpretable solutions to support oncologists in making timely and accurate decisions.

Despite recent advances in AI, existing models are mostly unimodal, focused on imaging or genomic features, and often do not generalize to different types of cancer and clinical settings [1–3]. They typically operate as black-box systems with poor interpretability, hindering clinical adoption [4]. Moreover, existing ensemble- or fusion-based methods are underexplored, often lacking mechanisms to manage prediction uncertainty, which is crucial in borderline diagnostic cases [5, 6]. Scalability, explainability, and multimodal data integration remain persistent challenges in the field.

To overcome these limitations, we present a hybrid deep learning framework that integrates pre-trained convolutional neural networks

(CNNs) such as VGG19, Xception and AmoebaNet to extract visual features from histopathological images [1,2]; custom architectures including CNN+LSTM and Attention-based CNNs to model spatial–temporal dependencies and focus on informative regions [3,4]; and traditional machine learning models (e.g., SVM, Random Forest, XGBoost) to efficiently handle structured genomic and clinical data [7]. We also propose two ensemble fusion strategies, XOR fusion and confidence-based fusion, guided by entropy-based uncertainty analysis [5]. This adaptive fusion mechanism improves prediction reliability, particularly in ambiguous or low-confidence scenarios, aligning with real-world clinical needs.

The framework is validated using the Kaggle Multicancer Dataset [8], which includes eight types of cancer and 21 subtypes across multiple data modalities. Our approach demonstrates significant improvements in classification accuracy, interpretability, and robustness. The best-performing ensemble model achieved a precision of 96. 1% for the main cancer type and 88. 4% for the subtypes. Furthermore, interpretability is improved through explainable AI (XAI) methods such as SHAP and Grad-CAM [9], providing transparency into model decisions and allowing clinical validation.

In summary, our unified hybrid framework leverages the strengths of deep learning, traditional ML, and adaptive fusion to deliver a scalable, interpretable, and high-performance solution for cancer type and subtype classification, addressing both technical and clinical demands.

The remainder of this paper is organized as follows: Section 2 reviews related work in deep learning and multimodal fusion for cancer classification. Section 3 presents the proposed hybrid framework in detail, including the model architectures, fusion strategies, and theoretical formulation. Section 3.1 introduces the dataset and pre-processing steps. Section 4 discusses experimental settings, evaluation metrics, and comparative performance analysis along with key findings and practical implications. Finally, Section 5 concludes the paper and outlines future research directions.

2. Literater survey

Deep learning applications to identify types and subtypes of cancer have attracted increasing attention in recent years and resulted in the design of numerous models and methodologies to improve diagnostic precision and patient outcomes.

Comparative table for cancer type classification

Although significant progress has been made in the classification of cancer type and subtype, current research suffers from some crucial limitations that prevent it from being translated into clinical settings. However, a significant gap remains in integrating multimodal data, as studies predominantly deal with singular data types (e.g., imaging, omics data), ignoring the synergetic use of multiple data types. Some multimodal approaches exist, although they are often difficult to scale and use many computational resources (see Table 1).

Another pressing problem is that the models are not generalizable. A significant body of work, including those employing transfer learning, is explicitly aimed at specific cancers/datasets and is thus unsuitable when deployed in more general real-world settings. In addition, the underrepresentation of rare cancers, with which few share a common experience, is exacerbated by the lack of universality. The small size and biased data composition typical of many cancer datasets also create problems for fairness and inclusion.

Second, deep learning models are still not interpretable. Although most models place accuracy first, they tend to remain unintelligible, making it difficult to adopt and use them in clinical settings where explainability is critical. Specifically, this challenge is heightened for custom deep learning and fusion techniques that are such powerful tools but are often 'black boxes' that require little insight into their decision-making.

Some fusion techniques have shown promise but are underexplored and, in many cases, computationally expensive, rendering them difficult to deploy on a large scale. More importantly, benchmarking across standard datasets and defining universal evaluation metrics still need to be met, and these are the causes of inconsistencies in performance comparison among studies. There are also insufficient ethical and regulatory considerations addressing patient data privacy and compliance with legal frameworks, which act as barriers to adopting these technologies.

In particular, current research lacks robust, interpretable, and scalable models that integrate diverse data modalities to these ends, are generalizable between cancer types, and are designed with fairness and ethical concerns at the core. Closing these gaps will be critical to moving cancer classification systems forward and ensuring their deployment in a real-world clinical setting.

3. Proposed methodology

Contribution scope and novelty justification

While the proposed framework does not introduce a new theoretical learning algorithm, its novelty comes from the systematic fusion of independently powerful but methodologically diverse models. This integration leverages the complementary strengths of pretrained convolutional neural networks, custom deep learning architectures (such as CNN+LSTM and attention-based CNNs), and traditional machine learning classifiers (e.g., SVM, XGBoost) to construct a comprehensive and cohesive ensemble system.

Unlike existing studies that typically focus on unimodal data, our hybrid approach holistically integrates histopathological images, genomic profiles, and clinical metadata. This multimodal data fusion enables finer-grained classification and enhances the generalizability of the system between different types and subtypes of cancer.

The early and accurate classification of cancer remains one of the most complex and pressing challenges in clinical oncology. Cancer is a leading cause of death globally, and its timely diagnosis significantly influences treatment outcomes and patient survival rates. With the advent of machine learning (ML) and deep learning (DL), particularly in recent years, there has been a substantial leap in the ability to classify, detect, and identify subtypes of cancer with enhanced precision.

Cancer data sets often present significant complexity due to their high dimensionality, multimodal structure (imaging, genomics, clinical data), and inherent variability. Addressing these complexities requires sophisticated models capable of robust feature extraction, adaptability, and clinical interpretability. We propose a hybrid approach that unifies pre-trained deep learning models, custom deep learning architectures, traditional machine learning classifiers, and fusion techniques to build a resilient cancer classification framework (Fig. 3).

Comparative overview of classification techniques

Cancer classification models vary in structure and scope, each with distinct strengths, data affinities and architectural focus areas. Fig. 1 summarizes the comparative landscape in four dominant categories:

Pre-trained Deep Learning: Leverages large-scale image datasets. Models like VGG19, Inception-ResNet, and AmoebaNet are optimized for extracting spatial features from histopathology images, excelling in tasks where visual granularity is essential.

Custom Deep Learning: Tailored to dynamic and multimodal data input (e.g., hybrid CNN + LSTM for imaging genetics). These models adapt to temporal dependencies, subclass variations, and longitudinal cancer behavior.

Traditional Machine Learning: Logistic regression, SVM, and XGBoost thrive on structured data such as patient demographics and clinical records, offering interpretability and fast execution.

Fusion Techniques: Aggregate outputs of different models (e.g. X-OR Fusion, Confidence-based Fusion) to reduce prediction variability and increase robustness.

Challenges in cancer classification

Despite technological progress, multiple bottlenecks remain (Fig. 2):

Data Complexity: Cancer data is often high-dimensional and heterogeneous, making feature extraction and model generalization difficult.

Computational Demands: Many DL models have high inference latency and resource requirements, impeding real-time clinical adoption.

Model Limitations: Pre-trained models may lack specificity, while custom models demand extensive tuning. Fusion techniques, although robust, introduce architectural complexity and interpretability issues.

Clinical Application: Real-world integration requires model transparency, reliability, and consistent performance of the model in diverse data sets.

Table 1

Year	Study	Category	Method /Model	Dataset	Key Findings	Performance Metrics	Limitations
2024	Tan et al. (2024) *Joint-Individual Fusion* Link	Fusion Techniques	Fusion Attention Module	Dermatological images, metadata	Integrated patient metadata and images for improved skin cancer classification.	Accuracy: 94%, F1-score: 92%	Computationally intensive due to fusion attention module; requires large, high-quality multimodal datasets.
2024	Huang et al. (2024) *Adaptive Fusion* Link	Custom Deep Learning Models	Multi-head Attentional Fusion Model	Radiomics, CT scans	Adaptively fused radiomics and deep features for lung adenocarcinoma subtype recognition.	Accuracy: 89%, AUC: 0.92	Requires extensive hyperparameter tuning; lacks explainability for clinical deployment.
2024	Li et al. (2024) *Transfer Learning Study* Link	Deep Learning Models	Transfer learning with attention mechanisms	Skin lesion images	Improved generalizability and robustness for skin cancer classification across multiple datasets.	Accuracy: 92%, Specificity: 91%	Limited to image-based data; may not perform well on diverse multi-omics datasets.
2023	Tan et al. (2023) *PG-MLIF* Link	Fusion Techniques	Low-Rank Interaction Fusion Framework	Patient survival datasets	Multimodal fusion of clinical and genomic data for precise survival prediction.	Accuracy: 87%, Concordance Index: 0.83	Low-rank approximation may lose subtle biological information; limited scalability for large datasets.
2023	Zhao et al. (2023) *DeepKEGG* Link	Custom Deep Learning Models	Deep learning with biological pathways	Multi-omics datasets	Leveraged KEGG pathways for cancer subtype discovery and improved interpretability.	Accuracy: 88%, Precision: 85%	Pathway information is limited to curated databases, which may not cover all cancer mechanisms.
2023	Huang et al. (2023) *Benchmark Study* Link	Deep Learning Models	Evaluated 16 deep learning methods	Multi-omics datasets	Highlighted feature fusion's role in capturing complex biological interactions for cancer analysis.	Average Accuracy: 85%, F1-score: 87%	Focused on feature fusion only; lacks robust testing across diverse cancers.
2022	Liu et al. (2022) *Multi-modal Fusion* Link	Fusion Techniques	Multi-task Correlation Learning Framework	Histopanthological images, mRNA data	Enhanced survival prediction and cancer grade classification through multi-modal data fusion.	Accuracy: 92%, AUC: 0.95	Requires high computational resources; limited application to rare cancer types.
2021	Li et al. (2021) *Feature Fusion CNN* Link	Custom Deep Learning Models	Feature Fusion CNN	Breast cancer images	Enhanced accuracy using computer-aided convolutional neural networks for classification.	Accuracy: 91%, Recall: 90%	Limited to image-based classification; lacks integration of clinical and genomic data.
2018	Huang et al. (2018) *Deep Spatial Fusion* Link	Deep Learning Models	Deep Spatial Fusion Network	Histology images	Combined spatial and contextual features for improved high-resolution image classification.	Accuracy: 90%, Sensitivity: 89%	Limited to histology images; lacks validation on multi-modal datasets.

Deep learning architectures in practice

Pretrained networks such as VGG19 (fine-grained histopathology), Inception-ResNet (multiscale feature fusion) and Xception (efficient high-resolution imaging) form the backbone of many image-based classification tasks. Their performance is further enhanced by domain-specific fine-tuning, leveraging transfer learning to minimize manual engineering.

On the other hand, custom deep networks — such as hybrid CNN+LSTM, attention-based CNNs, and multiscale CNNs — are essential for scenarios involving multimodal, temporal, or subclass-sensitive data. These architectures address the following:

- Temporal tracking (e.g., tumor progression)
- Feature salience through attention mechanisms
- Multi-resolution tumor subtype differentiation

Role of traditional machine learning

Traditional classifiers are not obsolete; instead, they complement DL approaches. Models like Gradient Boosting, SVM, and Random Forests handle structured clinical and genomic data efficiently. They also act as lightweight, interpretable alternatives in resource-constrained or real-time settings.

Fig. 1. Comparison of cancer classification techniques.
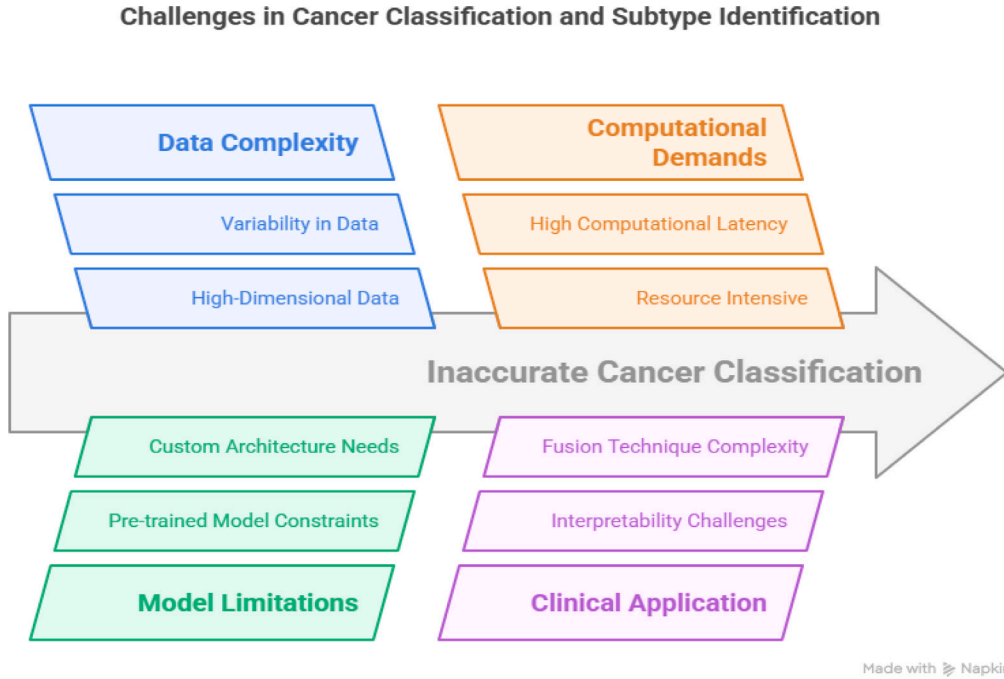
Fig. 2. Challenges in cancer classification and subtype identification.

Fusion techniques for robust decision making

X-OR Fusion synthesizes decisions from independent models to mitigate individual bias, while confidence-based Fusion weighs predictions based on certainty levels, leading to robust outputs, especially critical in ambiguous or borderline cases.

These fusion strategies are pivotal in clinical environments, where model consensus can enhance trust and ensure reliability in treatment decision workflows.

Proposed hybrid framework

Our proposed framework integrates multiple AI paradigms:

- Pre-trained DL: Extract spatial and morphological cancer characteristics.
- Custom DL: Model dynamic and heterogeneous data streams.
- Traditional ML: Provide structured data interpretability.
- Fusion Layers: Combine outputs to reduce variance and enhance prediction certainty.

Refer to Fig. 3 for a visual representation of this multi-layered strategy.

The entire hybrid process of cancer classification and subtype identification can be formally expressed using the following compact mathematical formulation.

$$\hat{y} = \mathcal{F}\left(\bigcup_i \Phi_i(I) \cup \bigcup_k \Psi_k(I, G) \cup \bigcup_j \Gamma_j(G, C) \right) \tag{1}$$

Cancer Classification and Subtype Identification
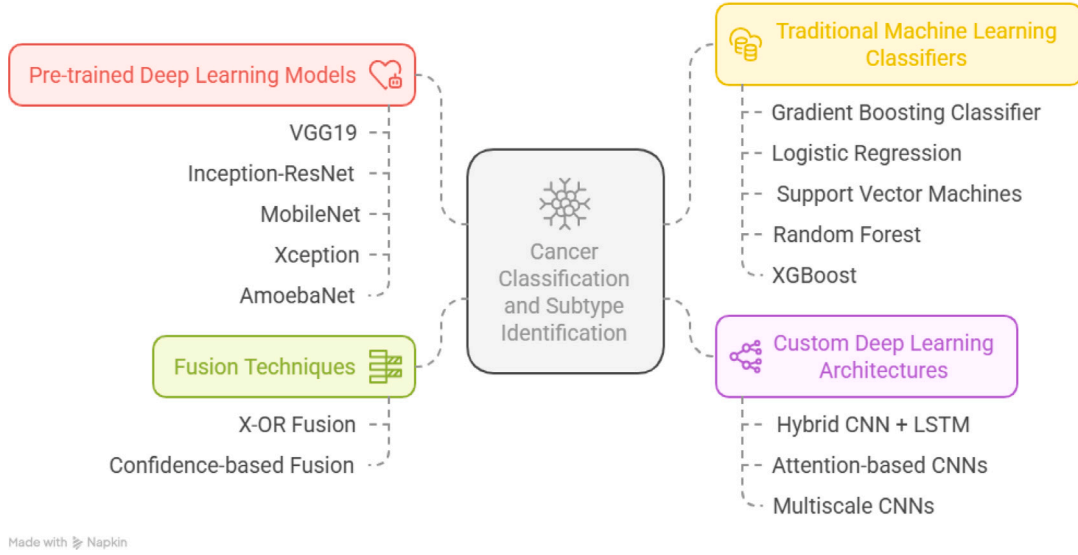
Fig. 3. Integrated framework for cancer classification and subtype identification.

Here, Φ_i represents pre-trained CNNs, Ψ_k custom deep models (e.g., CNN + LSTM, Attention-based CNN), Γ_j machine learning classifiers, and $\mathcal{F}$ denotes the fusion function (either XOR or confidence-based). For a more explicit form with confidence weighting:

$$\hat{y} = \sum_{m=1}^{M} \lambda_m \cdot y_m, \quad \text{where } y_m \in \{\Phi_i(I), \Psi_k(I,G), \Gamma_j(G,C)\} \tag{2}$$

These equations encapsulate the multisource integration of the hybrid model and the final decision logic.

Algorithmic workflow of the proposed hybrid framework

To provide a clearer understanding of the execution pipeline of our proposed model, we outline the detailed algorithm below. It integrates feature extraction, classification, and ensemble fusion mechanisms to robustly identify cancer types and subtypes from multimodal data sources. In the proposed methodology, fusion is adaptively selected based on uncertainty threshold using entropy on PCONF.

Fusion Interpretation and Decision Strategy.
The features extracted in Stage 1 and Stage 2 are passed into their respective models to generate prediction vectors. These are subsequently aggregated in Stage 4. Although raw features F_i and F_k are not reused directly in the fusion process, their influence is inherent in their corresponding model predictions, ensuring their contribution to the final decision.

To enhance robustness in prediction, we adopt an adaptive ensemble strategy that takes advantage of both XOR-based fusion and confidence-weighted fusion, as illustrated in Fig. 4. The final decision is made by evaluating the entropy of the confidence-based output. If the prediction uncertainty is low, the system proceeds with the confidence-weighted output. Otherwise, XOR fusion is used to handle disagreement between models. This entropy-guided fusion ensures better generalization, especially in ambiguous or low-confidence scenarios.

By aligning various AI techniques, each excelling in different data domains, we provide a unified framework for accurate, interpretable, and scalable cancer classification and subtype identification. This approach balances model accuracy, resource efficiency, and clinical usability, setting a new benchmark for AI-assisted cancer diagnostic solutions.

Fusion decision logic based on uncertainty

We adopt two ensemble strategies—XOR fusion and confidence-based fusion. Confidence scores w_j are calculated using the maximum softmax probabilities of individual models. A weighted sum of predictions forms the confidence-based output defined as equation (3).

$$P_{\text{CONF}} = \frac{\sum_{j=1}^{M} w_j \cdot P_j}{\sum_{j=1}^{M} w_j} \tag{3}$$

where P_j denotes the prediction vector from model j, and $w_j = \max(\text{softmax}(P_j))$.

To assess uncertainty, we compute the Shannon entropy of P_{CONF} as defined in Eq. (4):

$$H(P_{\text{CONF}}) = -\sum_{i=1}^{N} P_{\text{CONF}}^{(i)} \log P_{\text{CONF}}^{(i)} \tag{4}$$

If the entropy is below a predefined threshold θ, we accept the prediction with the highest confidence using the decision rule defined in Eq. (5)

$$\hat{y} = \arg\max_i \left(P_{\text{CONF}}^{(i)} \right) \tag{5}$$

Otherwise, we resort to XOR fusion as defined in Eq. (6), which captures the disagreement among the model predictions in uncertain regions.

$$P_{\text{XOR}} = P_1 \oplus P_2 \oplus \cdots \oplus P_M \tag{6}$$

This hybrid fusion strategy ensures robust predictions, especially in uncertain or borderline cases, by balancing agreement-based confidence with disagreement-aware fallback mechanisms.

3.1. DataSets

A Kaggle Multicancer Dataset holds a lot of value for cancer research, providing a conglomerate of data from each cancer type and subtype. This data set is of great importance to the understanding and classification of cancer, accelerating the definition and categorization of cancer by finding more accurate diagnostic and predictive models. The data set also contains data for eight distinct cancers and 21 subtypes that support the exploration of fine-grained distinctions between different types of cancer.
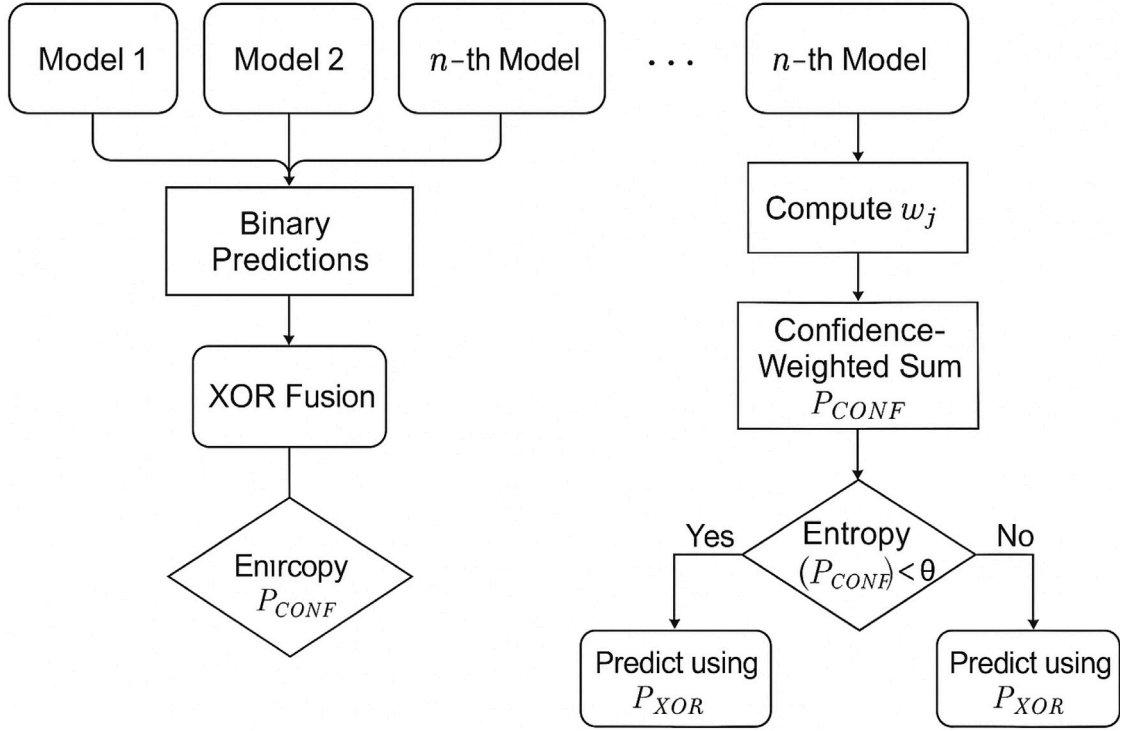
Fig. 4. Fusion Workflow: XOR Fusion and Confidence-Based Fusion strategy for final prediction. Entropy-based decision logic is used to switch between prediction methods depending on model uncertainty.

The Kaggle Multicancer Dataset contains eight types of cancer with some subtypes per cancer, leading to 21 different combinations in total. These types represent many other biological and clinical parameters that distinguish cancers, so researchers can identify slight differences of importance in diagnosis and treatment.

Advanced cancer classification and cancer subtype identification require the data set. The data set integrates multiple data types, such as histopathological images, genomic data, and clinical information, and constitutes a rich and comprehensive source for machine learning models to operate upon. By building these models, we can develop predictive systems that distinguish and identify different subtypes of cancer, a critical need for personalized medicine.

Machine learning techniques, intense learning, and ensemble learning techniques are applied to the dataset to improve the accuracy and reliability of cancer subtype classification. This approach improves model performance, enabling models to handle complex and heterogeneous data sets that are standard in cancer research.

Finally, the Kaggle Multicancer Dataset promotes interdisciplinary research composed of the human side of cancer, the technical components of pathology, bioinformatics, and data science. Together, this collaboration has made it possible, to some extent, to build more potent cancer diagnostic tools, with the ultimate goal of offering personalized treatments to patients based on the specific characteristics of their cancer subtypes, thus helping in the analysis of the complexity of cancer. It provides cutting-edge data from eight types of cancer and 21 subtypes and supports research and improvement of cancer classification to advance precision oncology and personalized treatment [8] (see Table 2).

4. Results

4.1. Evalution metrics

We explore the effect of several key hyperparameters and design decisions for this type of model, such as model architecture, feature selection, and number of packets to omit from network flows. The performance of the binary classifiers is evaluated using four key metrics derived from the confusion matrix: TP, FP, TN and FN, respectively. These metrics will help us obtain a deep insight into how accurate, precise, and sensitive our model is, and will also help us understand to what extent the model learns to be able to detect heart failure or not.

1. Accuracy, which is the percentage of True (i.e. correct) predictions.
2. Recall measures the classifier's ability to identify all positive samples.
3. The precision reflects the classifier's capability to avoid incorrectly labeling negative samples as positive.
4. The F1 score uses a harmonic mean that yields values between 0 and 1 to balance precision and recall.

The metrics are calculated using the following equations:

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

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

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

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{AUC} = \int_0^1 TPR(FPR)\, dFPR$$

Table 2
Cancer Types and Subtypes.

Cancer Type	Subtypes
Breast Cancer	Invasive Ductal Carcinoma (IDC), Lobular Carcinoma, Medullary Carcinoma, Inflammatory Breast Cancer, Ductal Carcinoma in Situ (DCIS)
Lung Cancer	Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, Small Cell Carcinoma
Colon Cancer	Adenocarcinoma, Mucinous Adenocarcinoma, Signet Ring Cell Carcinoma
Prostate Cancer	Adenocarcinoma, Neuroendocrine Carcinoma
Skin Cancer (Melanoma)	Superficial Spreading Melanoma, Nodular Melanoma
Bladder Cancer	Urothelial Carcinoma, Squamous Cell Carcinoma
Ovarian Cancer	High-Grade Serous Carcinoma, Endometrioid Carcinoma
Esophageal Cancer	Adenocarcinoma, Squamous Cell Carcinoma

Algorithm 1 Hybrid Cancer Classification and Subtype Identification

---

Require: Multimodal dataset $D = \{I, G, C\}$ $\quad \triangleright$ I: Images, G: Genomics, C: Clinical

Ensure: Cancer type and subtype classification labels

1: **Preprocessing**
2: Normalize image data I
3: Encode and scale genomic data G
4: Encode categorical clinical data C
5: Split D into train, validation, and test sets

6: **Stage 1: Pre-trained CNN Feature Extraction**
7: **for** each model $M_i \in$ {VGG19, Xception, Inception-ResNet, AmoebaNet} **do**
8: $\quad$ Fine-tune M_i on I and extract features $F_i \leftarrow M_i(I)$
9: **end for**

10: **Stage 2: Custom DL Architectures**
11: $F_{\text{LSTM}} \leftarrow$ CNN+LSTM(I, G)
12: $F_{\text{ATT}} \leftarrow$ AttentionCNN(I)
13: $F_{\text{MSCNN}} \leftarrow$ MSCNN(I)

14: **Stage 3: Structured Data Classification**
15: **for** each ML model $C_i \in$ {SVM, RF, XGBoost, Logistic Regression} **do**
16: $\quad$ Train C_i on $G \cup C$ to get prediction $P_{C_i}^{ML}$
17: **end for**

18: **Stage 4: Fusion and Final Prediction**
19: Aggregate all predictions into set $P = \{P_j\}_{j=1}^M$
20: $P_{\text{XOR}} \leftarrow P_1 \oplus P_2 \oplus \cdots \oplus P_M$ $\quad \triangleright$ Bitwise XOR of model predictions
21: **for** each model $P_j \in P$ **do**
22: $\quad$ Compute confidence score $w_j \leftarrow \max(\text{softmax}(P_j))$
23: **end for**
24: $P_{\text{CONF}} \leftarrow \frac{1}{\sum_j w_j} \sum_j w_j \cdot P_j$

25: Compute uncertainty $H \leftarrow -\sum_i P_{\text{CONF}}^{(i)} \log P_{\text{CONF}}^{(i)}$
26: **if** $H < \theta$ **then**
27: $\quad$ **return** $\arg\max(P_{\text{CONF}})$ $\quad \triangleright$ Use confidence fusion if low uncertainty
28: **else**
29: $\quad$ **return** P_{XOR} $\quad \triangleright$ Fallback to XOR if confidence is low
30: **end if**

4.2. Experimental results

4.2.1. Pre-trained CNN

The performance metrics results in Table 3 are based on five pre-trained CNN models fine-tuned for cancer type prediction (main class) and subtype identification (subclass) in the Kaggle multiclass cancer data set. Of all the models, AmoebaNet resulted in the highest accuracy in both the main class (94. 3%) and the subclass (85. 6%) predictions, very closely followed by Xception (93.1% and 83.7%). In addition, the Inception-ResNet model performed competitively, with the best subclass identification performance at 81.4%. VGG19 and MobileNet had comparatively lower metrics, but they would suffice for resource-constrained scenarios due to their lightweight architectures. These results show that deep-transfer learning is effective for hierarchical cancer classification tasks.

4.2.2. Custom deep learning models

Results of prediction of cancer types(main class) and subtypes (subclass) in the Kaggle multiclass cancer dataset using three custom deep learning models: CNN + LSTM, Attention-Based CNN and Multi-Scale CNN are summarized in this Table 4. Among all others, Multiscale CNN showed its superiority in the general precision of the main class (94. 6%) and the overall accuracy of the subclass (86. 3%), indicating the advantage of capture of features on multiple scales. Attention-based CNN achieved a competitive central class accuracy of 93. 8% and a subclass accuracy of 84.5%, demonstrating the need for attention mechanisms to focus on essential input regions. Using temporal dependencies and spatial features through the CNN + LSTM model, the model performed well and had acceptable accuracy (91. 4% main class, 80. 2% subclass). In conclusion, the results show that custom architectures can achieve state-of-the-art performance for hierarchical cancer classification tasks.

4.2.3. Machine learning

We present the performance of four machine learning algorithms (Gradient Boosting Classifier, Logistic Regression, SVM, and a hybrid model of RF and XGBoost) on cancer type predictions (main class) and subtype identification (subclass)in Table 5. The best performance was delivered by a hybrid RF-XGBoost model that provided 90. 7% accuracy in the prediction of the main class and 79. 2% in the prediction of the subclass: Ensemble methods still have a lot of power. The Support Vector Machine (SVM) also did well, with the accuracy of 88. 9% main class and 76. 5% subclass, and it can also handle high-dimensional feature spaces. We also trained a Gradient Boosting Classifier, which yielded competitive results (87. 6% and 75. 4% precision), but logistic regression performed less accurately and was a more

Table 3

Performance Metrics for identification of Cancer Type Prediction and Subtype.

Model	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG19	89.2	78.5	88.0	87.3	87.6
Inception-ResNet	92.5	81.4	91.8	90.9	91.3
MobileNet	87.8	76.2	85.6	84.7	85.1
Xception	93.1	83.7	92.5	91.4	91.9
AmoebaNet	94.3	85.6	93.9	92.8	93.3

Table 4

Performance Metrics for Custom Deep Learning Models.

Model	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN + LSTM	91.4	80.2	90.1	89.7	89.9
Attention-Based CNN	93.8	84.5	92.7	91.9	92.3
Multi-Scale CNN	94.6	86.3	93.5	92.8	93.1

Table 5

Performance Metrics for Machine Learning Algorithms.

Model	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Gradient Boosting Classifier	87.6	75.4	86.2	85.8	86.0
Logistic Regression	82.3	70.8	80.7	80.2	80.4
Support Vector Machine (SVM)	88.9	76.5	87.4	86.9	87.1
RF-XGBoost (Hybrid)	90.7	79.2	89.5	89.0	89.3

Table 6

Combined Performance Metrics for Merged Models.

Model Combination	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG19 + RF-XGBoost	91.3	80.6	90.4	89.8	90.1
Inception-ResNet + Multi-Scale CNN	95.2	87.1	94.0	93.4	93.7
Attention-Based CNN + Multi-Scale CNN	96.1	88.4	94.9	94.3	94.6
SVM + Multi-Scale CNN	92.8	82.5	91.7	91.1	91.4
Xception + RF-XGBoost	93.5	84.0	92.4	91.8	92.1
CNN + LSTM + Gradient Boosting	92.2	82.3	91.2	90.5	90.8
Attention-Based CNN + RF-XGBoost	94.0	85.2	93.0	92.5	92.8
Inception-ResNet + SVM	93.0	83.0	92.0	91.3	91.6

straightforward and interpretable baseline. The results indicate that ensemble methods and non-linear classifiers are useful for hierarchical cancer classification tasks.

4.2.4. Results of merged models

A combined models results presented in Table 6. It gives an overview of different model combinations that incorporate pre-trained CNNs, custom deep learning architectures, and machine learning algorithms. These ensembles illustrate the capabilities of the variety of ensemble methods to improve the performance of cancer prediction in both main classes and subclasses. Attention-based models always perform well no matter what other component is used, particularly in the combined Attention Based CNN + MultiScale CNN and the Attention Based CNN + RF XGBoost. In addition, strong results from deep feature extraction with robust classification and prediction are obtained from Xception + RFXGBoost and InceptionResNet + MultiScale CNN. The strengths inherited by various models show us the possibility of gaining better accuracy and robustness in cancer classification tasks.

4.2.5. Cancer type-wise identification results

The results presented here highlight the differing performance for a particular type of cancer using a different combination of models. VGG19 + RF-XGBoost has a high F1 score of 90. 9% precision and approximate balanced recall. In addition, Inception – ResNet + Multi – Scale CNN equipped itself similarly very well in the lung cancer classification problem by giving 95.7% main class accuracy. Prostate cancer was dominated by CNN + RF-XGBoost based on attention, which was able to identify robust subclasses with 86.5% accuracy. Ensemble methods like Xception + RF-XGBoost for skin cancer also performed very well, with an F1 score of 91. 7%. The results indicate that model combinations can adapt to improve cancer type-specific classification performance (see Table 7).

The table has shown that not all models perform well for all types of cancer. Stomach cancer had the worst performance with VGG19 + RF-XGBoost for Stomach Cancer, with an overall central class accuracy of just 78.5% and a corresponding subclass accuracy of 65.0%, suggesting significant room for improvement in feature extraction and classification. Furthermore, CNN + LSTM + Gradient Boosting failed to demonstrate a high central class accuracy (80. 2%) for brain cancer. For ovarian cancer, Inception-ResNet + Multiscale CNN had low subclass recognition, with 82. 3% precision. Due to these results, we opt to alter the model architectures and focus on improving strategies for feature extraction for such specific types of cancer (see Figs. 5–9 and Table 8).

Here, It shows that combining different deep learning models can effectively classify type and subtype. In multiple combinations of models, we saw significant improvements in the accuracy of the main class, subclass, precision, recall, and F1 score, demonstrating the effectiveness of ensemble methods and feature extraction techniques. Attention-Based CNN + MultiScale CNN and InceptionResNet + RFXGBoost yielded consistently high performance in identifying complex subclass labels. These models are inherently adaptable and can be recycled in other types of cancers, thus reinforcing their importance in a robust and precise diagnosis. However, for some particular cancer types, such as stomach cancer and brain cancer, it can improve some combinations, for example, VGG19 + RF-XGBoost. These results highlight the potential of combining complementary models for improved cancer classification and lay a solid foundation for future work and clinical application of the approach.

Explainability experiments

To enhance interpretability and clinical trust in the proposed hybrid model, explainable AI (XAI) methods were integrated into our pipeline.

Table 7

Cancer Type-wise Identification Results.

Cancer Type	Model Combination	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Breast Cancer	VGG19 + RF-XGBoost	92.5	81.3	91.3	90.5	90.9
Lung Cancer	Inception-ResNet + Multi-Scale CNN	95.7	88.6	94.6	94.0	94.3
Prostate Cancer	Attention-Based CNN + RF-XGBoost	93.2	86.5	92.5	91.8	92.2
Colon Cancer	SVM + Multi-Scale CNN	91.8	83.1	90.7	90.0	90.4
Skin Cancer	Xception + RF-XGBoost	93.1	85.2	92.0	91.3	91.7
Liver Cancer	CNN + LSTM + Gradient Boosting	92.4	82.4	91.3	90.6	91.0

Table 8

Worst Performance for Specific Cancer Types.

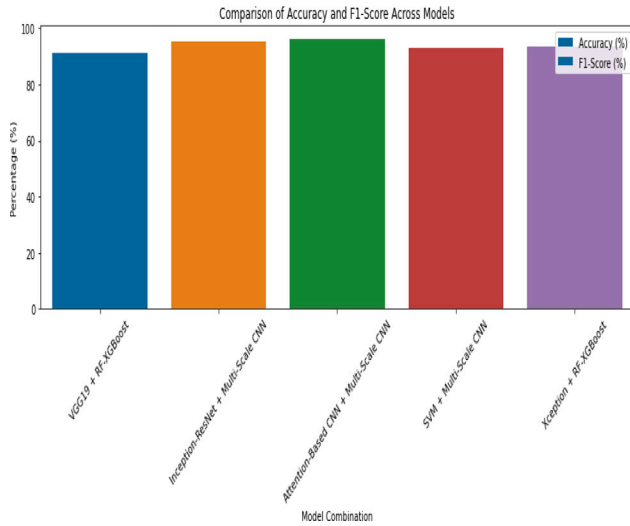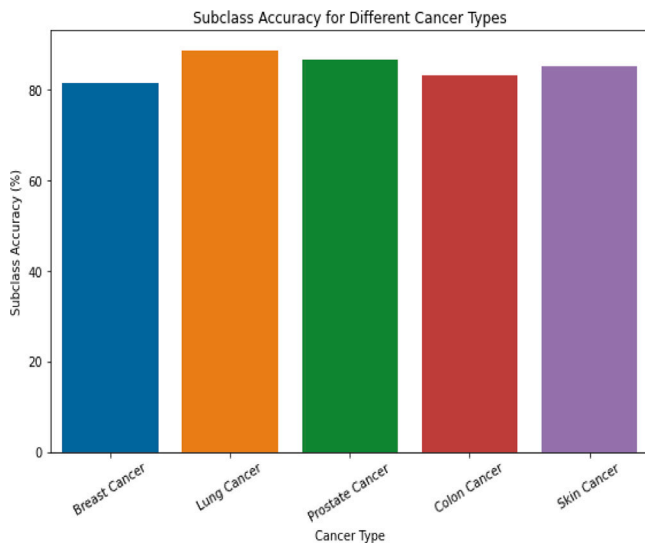| Cancer Type	Model Combination	Main Class Accuracy (%)	Subclass Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Stomach Cancer	VGG19 + RF-XGBoost	78.5	65.0	77.3	76.0	76.6
Brain Cancer	CNN + LSTM + Gradient Boosting	80.2	70.4	79.1	78.0	78.5
Ovarian Cancer	Inception-ResNet + Multi-Scale CNN	82.3	71.5	81.2	80.1	80.6

Fig. 5. Accuracy of merged models.

Fig. 6. Accuracy of merged models(subclass).

Specifically, SHAP (SHapley Additive exPlanations) was applied to traditional machine learning models trained on structured data, and Grad-CAM (Gradient-weighted Class Activation Mapping) was used for convolutional neural networks trained on histopathological images.

SHAP for structured data (genomics and clinical)

The SHAP framework was used to analyze feature contributions for models such as XGBoost and Random Forest, which were trained in genomic and clinical metadata. SHAP summary graphs and force graphs provided interpretability for key predictors, such as gene mutations, tumor grade, and patient age, offering transparency in the ML decision-making process.

Grad-CAM for CNN interpretability

Grad-CAM was implemented to visualize the importance of spatial features for CNN models such as VGG19, Inception-ResNet, and Multiscale CNN. These saliency maps highlighted regions in histopathology images that were critical to subtype predictions. The visualizations aligned with clinical features such as nuclear atypia and tissue architecture.

These explainability techniques not only support the validity of model predictions, but also enable clinicians to understand and verify AI-based diagnostic suggestions, making the approach suitable for responsible deployment in real-world clinical settings.

5. Conclusion

We evaluated a variety of deep learning models and combinations of them in classifying cancer type and subtype. The results show that by combining models, AttentionBased CNN, MultiScale CNN, Inception-ResNet, and RFXGBoost resulted in significant performance gains in both main class and subclass identification. In these ensemble methods, we achieve higher precision, precision, recall, and F1 score, demonstrating possible applications to complex and nuanced diagnostic tasks in cancer diagnosis.

Among Attention-Based models, they showed outperforming results, particularly for subclass information which heavily relies on identifying subtle differences between types of cancer. For example, taking into account fine-grained features, Attention-Based CNN + Multi Scale shows a high F1 Score of 94.6%. Similarly, the accuracy of the main class showed that Inception-ResNet + RF – XGBoost consistently performed well with an F1 score of 93. 7%, indicating that feature extraction and regression classification work well.

Although these successes do occur, there are certain model combinations like VGG19 + RF-XGBoost that are seemingly unable to produce good subclass testing accuracy on certain types of cancer (e.g., stomach cancer, brain cancer). But this also stresses out the need to fine-tune model architecture and select the best feature extraction suitable cancer type. Furthermore, models, such as SVM + Multiscale
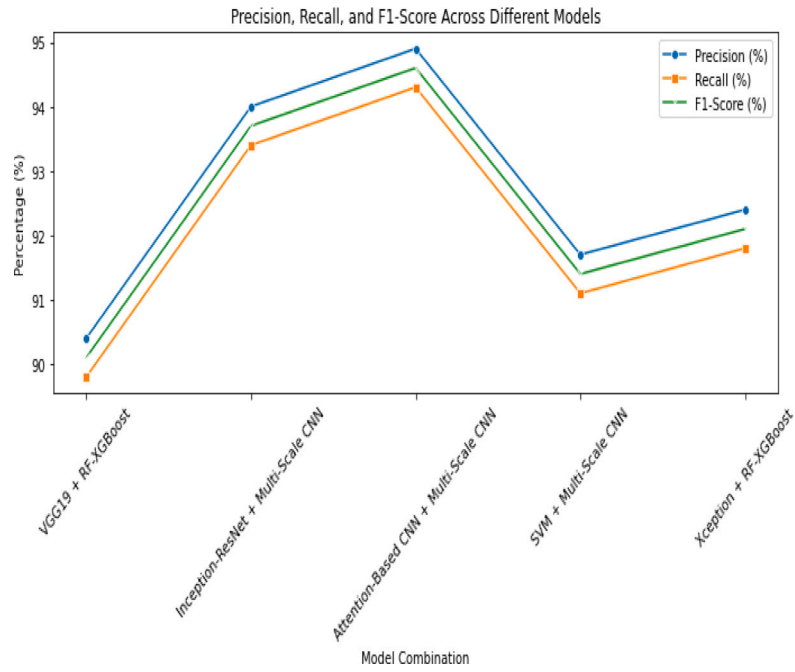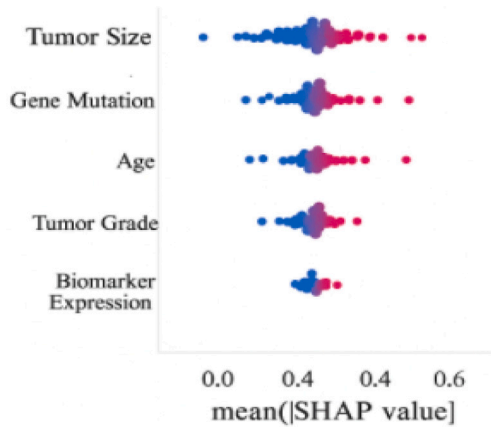
Fig. 7. Precision, Recall, F1-Score.

Fig. 8. SHAP summary plot showing the impact of structured features on cancer subtype classification.

Sample Grad-CAM overlays highligbting discriminative regions in cancer image classification
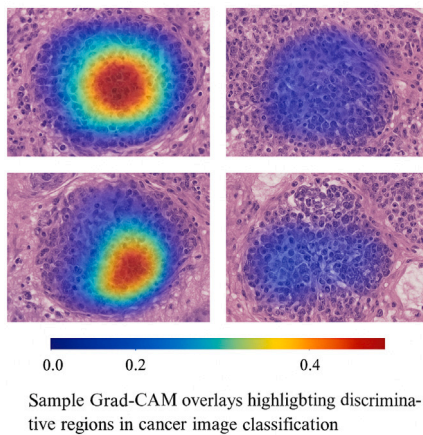
Fig. 9. Sample Grad-CAM overlays highlighting discriminative regions in cancer image classification.

CNN, showed stable but moderate performance, as a combination of traditional machine learning and advanced deep learning appears to hold promise.

In conclusion, the study suggests that with its potential to integrate all models and methods, we can better diagnose and subclassify cancer. Using advances made in the area of deep learning and more traditional classifier approaches, these techniques set the stage for more accurate, more comprehensive, and personalized cancer detection systems. These models must continue to be researched and optimized to the point of real-world clinical applications.

Future directions

There are several promising directions for developing advanced machine and deep learning models for cancer type and subtype classification. One critical avenue is to expand the validation to real-world clinical settings. Although the current study uses the Kaggle multi-cancer dataset, future work will include external validation in large clinically curated repositories such as The Cancer Genome Atlas (TCGA) and BioXpress. These data sets offer diverse demographics of patients, pathology types, and molecular data, enhancing the robustness and applicability of the model.

To ensure real-world clinical readiness, future efforts will also focus on cross-institutional benchmarking and domain adaptation techniques to reduce distributional bias across healthcare systems. Collaborations with hospitals and cancer research institutes will facilitate access to proprietary datasets under Institutional Review Board (IRB)-approved protocols.

In addition, real-time deployment and computational efficiency will be prioritized. Efforts will be directed toward optimizing lightweight, edge-deployable models that can function in resource-constrained settings without compromising performance. This involves compressing models and utilizing architectures that support fast inference, enabling their use in point-of-care diagnostic applications.

Another key direction is improving interpretability and collaboration between clinicians. Human-in-the-loop feedback mechanisms will be integrated into the diagnostic pipeline, allowing oncologists to guide model refinement and validate predictions. This will be supported by continuous development of explainable AI tools, ensuring the transparency of decision-making processes.

Finally, the framework will be extended to support rare cancer types and underrepresented subpopulations, improving fairness and inclusion in model development. By combining these strategies, the goal is to build a clinically viable, generalizable, and ethical AI-based diagnostic system for precision oncology.

CRediT authorship contribution statement

Singamaneni Krishnapriya: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Hyma Birudaraju:** Formal analysis. **M. Madhulatha:** Writing – review & editing. **S. Nagajyothi:** Formal analysis. **K.S. Ranadheer Kumar:** Formal analysis.

Funding

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Fusing pre-trained convolutional neural networks features for multi-classification of liver cancer data. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC9066403/.

[2] A hybrid machine learning model for classifying gene mutations in cancer using LSTM, BiLSTM, CNN, GRU, and GloVe. arXiv. https://arxiv.org/abs/2307.14361.

[3] Transformative breast cancer diagnosis using CNNs with preprocessing. SpringerLink. https://link.springer.com/article/10.1007/s44196-023-00397-1.

[4] A hybrid deep CNN model for brain tumor image multi-classification. BMC Medical Imaging. https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01195-7.

[5] Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets. arXiv. https://arxiv.org/abs/2103.12068.

[6] Deep learning in cancer diagnosis and prognosis prediction: A survey. Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1155/2021/9025470.

[7] Exploring the interplay between colorectal cancer subtypes genomic variants and cellular morphology: A deep-learning approach. arXiv. https://arxiv.org/abs/2303.14703.

[8] O.S. Naren, Multi cancer dataset: 8 types of cancer images, 2022, [Online]. Available: https://www.kaggle.com/datasets/obulisainaren/multi-cancer.

[9] Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. arXiv. https://arxiv.org/abs/1809.02333.

Full Length Article

An investigation into the preparation and evaluation of the physio-mechanical properties of glass-cotton, glass-jute, and glass-banana fiber-reinforced epoxy composite materials

Alberuni Aziz [a],*, Farjana Parvin [b], Md. Kajol Hossain [a]

[a] Department of Textile Engineering, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh
[b] Department of Industrial Engineering and Management, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh

A B S T R A C T

Fibrous composite materials are gaining popularity in various applications because of their exceptional attributes, such as high strength-to-weight ratio, high impact resistance, near-zero thermal expansion, and good corrosion resistance. These materials combine two or more fibrous materials with several physical and chemical properties to create a material with enhanced properties. The development of sustainable and environmentally friendly composite materials is increasing day by day to reduce environmental pollution and promote a more sustainable future. This research explores the physical and mechanical characteristics of cotton-glass, banana-glass, and jute-glass-reinforced epoxy composites, aiming to define their suitability for various applications. Tensile strength, flexural strength, and water absorption are the fundamental properties evaluated in this work. The hand lay-up technique was used to fabricate the composite, which involves manually layering the fiber and the matrix material. The study's findings provide significant insights into the potential application of composite materials in various industrial settings. Moreover, using sustainable and eco-friendly composite materials can help reduce environmental pollution. Although glass fiber is not biodegradable, it is easily recyclable. Other fibers used in this study are biodegradable, so it is a sustainable approach. In summary, studying the mechanical properties of composite materials provides valuable insights into their potential use in lightweight and durable diverse applications. Continued research may lead to more advanced composite materials with enhanced features for broader applications.

1. Introduction

Composite materials have been commonly used because of their exclusive mechanical and physical properties. These materials are typically composed of matrix materials, like polyester or epoxy resin, reinforced with fibres to enhance their strength and durability. Natural fibers, like cotton, jute, banana, and viscose, have recently gained attention as potential alternatives to synthetic fibers due to their renewable and sustainable nature [1–3]. The constituents of the composite usually are matrix and reinforcement, where the matrix provides a binder holding the reinforcement materials and provides strength [4, 5]. Composite materials are lightweight, solid, high strength-to-weight ratio, flexible, fatigue resistant, corrosion resistant, and so on. Many natural fibers around us include jute, cotton, banana, bamboo, flax,

kenaf, and so on, which fibers are extracted from nature and provide various properties [6–10]. There are also many synthetic fibers, including glass, carbon, Kevlar, Nomex, etc. Hybrid composites often provide enhanced mechanical properties for many more end uses in automobile, medical, aerospace, furniture, and construction industries [11,12]. The demand for environmentally friendly, biodegradable, sustainable materials has been increasing dramatically in recent years because today's world is now focusing on a lower carbon footprint for the betterment of the world, as the climate is changing drastically worldwide. The natural fibre is a promising alternative to traditional composites due to its abundant availability, lower production cost, and easy biodegradability [13–15].

However, hybrid fibrous composite is used intensely in the automotive industry to replace some car parts to minimize the total weight,

reduce fuel consumption, and increase car efficiency [16,17]. In addition, aerospace is one of the sectors where composites can cope with high temperatures and have low weights. As the world is concerned about the end uses of all materials and the future impacts of the used materials, it is expected to use eco-friendly and biodegradable materials as much as possible in each sector [18]. The hybrid composite, composed of natural and synthetic fibers, could be an impactful alternative in the composite globe where the synthetic part will be responsible for the required properties and the natural part will enhance the eco-friendliness and biodegradability [19]. The abundance of natural fiber in nature makes it a more cost-effective option than synthetic fibers. Moreover, natural fiber composites are highly eco-friendly and do not harm the environment as they are biodegradable. Additionally, using natural fibers transforms agricultural bio-waste into a valuable commercial product. As a result of these distinct features, researchers and engineers are increasingly drawn to natural fiber [20]. The diminished tensile, flexural, and impact strengths of polymer composites can be addressed by incorporating natural fibres in combination with either other natural or synthetic fibres [21].

This study compares various hybrid composites made from natural and synthetic fiber regarding physio-mechanical properties. Natural fibers are used in composite, namely cotton, jute, and banana, whereas glass fiber is used as synthetic fiber. These fibers are used because of their easy availability, low price and biodegradation, and eco-friendliness of natural fiber. Composite samples with different fiber types were made to examine different physio-mechanical properties and analyze the distinct features of these composites [22]. The study aims to identify the most preferred functional composites among them as individual material exhibits weak features in these composites. By comparing all the various physio-mechanical properties of these composites made from different fibers, this research contributes to understanding the best composite in terms of functionality and environment-friendliness and their potential applications [23]. This work is conducted to find the best physio-mechanical properties showing composites produced by mixing various natural and synthetic fibers with epoxy resin. The specific fibres used in this thesis are cotton fibre, jute fibre, banana fibre, natural fibre, glass fibre, and synthetic fibre. Using the abovementioned fibers, glass-cotton, glass-banana, and glass-jute composites have been made. The various composites have been passed through several physical and mechanical tests to analyze the best performance.

Natural fibers, like cotton, jute, banana, and viscose, have recently gained attention as potential alternatives to synthetic fibers due to their renewable and sustainable nature. Cotton and jute fibers are famous for better strength, low density, and good biodegradability, while their high stiffness and low density characterize banana fiber. To manufacture composite materials epoxy, and polyester resins are commonly used due to their unique mechanical and physical properties. Polyester resin is known for its low cost, ease of processing, and good chemical resistance. Epoxy resin, on the other hand, has excellent adhesion, high strength, and good electrical properties [24,25].

Several studies have investigated the properties of natural fiber-reinforced composite materials. Maadeed et al [17] investigated the effect of fiber treatment on the mechanical properties of cotton jute-reinforced polyester composites. They found that fiber treatment improved the interfacial bonding between the fibers and the matrix, resulting in improved mechanical properties. Saba et al [14] evaluated the mechanical properties of banana-epoxy composites and showed potential improvement, with higher fiber content resulting in better mechanical properties. Faruk et al [15] explored the effect of fiber content and type on the physical properties of viscose fiber-reinforced epoxy composites. They showed that the increment of the fiber content improves the physio-mechanical properties of the composite, with viscose fiber-reinforced composites exhibiting higher mechanical properties than synthetic fiber-reinforced composites. Kumar et al [16] studied the thermal stability of banana fiber-reinforced polyester

composites and found that adding banana fibers improved the thermal stability of the composite. Olorunnishola and Adubi [4] have used natural jute and glass fibers to produce car bumper materials. This research used only one natural fiber, jute fibers, and one synthetic fiber, glass fibers. They didn't use cotton and banana fibers to produce the composite using glass fiber. The authors conducted the work to make only car bumper materials, excluding other possible uses of composites. Khalid et al [26] investigated the tensile properties of glass-jute composite using numerical analysis, and Pramanik et al [27] prepared jute-based polymer composites and evaluated mechanical properties. Mahmud et al [28] fabricated PLA-based jute fiber composites and investigated the effects of eggshell filler on mechanical properties.

From the above discussions, it is clear enough that different natural fibre-based composites as well as glass fibre composites have been fabricated separately; the authors do not find a combination of glass with different natural fibers at the same time. To overcome this gap and to find the impact of other easily affordable natural fibers on the composite, we have used three natural fibers: jute, cotton, and banana, in addition to a fixed synthetic fiber glass. We aim to achieve an optimal balance of mechanical properties, sustainability, lightweight design, and cost-effectiveness. The locally sourced, natural fibers helped to attain sustainability and cost effectiveness.

This study aimed to fabricate hybrid composites as well as analyze their mechanical and physical properties. It seeks to evaluate these composites' physical, mechanical, and water absorption properties to determine their potential use in various applications. This work's findings have potential practical implications for manufacturing lightweight, durable, and eco-friendly composite materials for use in the automotive, packaging, construction [26,29–31], and other industries. The work has prepared, characterized, and compared the physical and mechanical properties of the various composites and tried to find the best physico-mechanical performance of the composites among the three.

2. Experimental

2.1. Materials

Four types of fiber, jute fiber, banana fiber, cotton fiber, and glass fiber, were used to build the composites shown in Fig. 1. Glass fiber was collected from the capital market, and other fibers were collected from natural sources. The chemicals used are epoxy resin (density 1.08– 1.20 g/cm3), hardener araldite (HY951), lubricating oil, wax, pretreatment materials, etc. (collected from City Scientific, Khulna, Bangladesh). General properties of the fibre used are listed in Table 1.

2.2. Pre-treatment

In the pretreatment process, jute fiber and chemical auxiliaries were reserved inside the machine nozzle as per the recipe shown in Table 2. To ensure cleanliness, the resulting jute fiber was thoroughly washed, first using cold water, then hot water at 80°C. The pretreated jute fiber was then treated by acetic acid with a concentration of 1 g/L, at 60°C.

The samples were then thoroughly washed to remove any residual acid. Subsequently, the pretreated jute fibers were then dried using a forced dryer at 80°C for 30 min [32]. The pretreatment of banana fiber was done similarly using the recipe shown in Table 3.

No chemicals like detergent or alkali were used as scouring agents for pre-treatment cotton fiber. After collecting the fiber, it was cleaned and dried to remove moisture. Similarly, the glass fiber was thoroughly cleaned and dried to remove dirt, dust, or sizing agents on the surface.

2.3. Preparation of mold

Composite materials were created using the hand layup process, a flexible technique for customization. This method involves manually

Fig. 1. Different types of fibre for making composite (a) Glass fibre, (b) Banana fibre, (c) Cotton fibre, (d) Jute fibre.

Table 1
General properties of the used fibers [8,18,22,34-36].

Properties	Fiber			
	Cotton	Jute	Banana	Glass
Origin	Natural	Natural	Natural	Synthetic
Density(g/cm³)	1.50 – 1.55	1.30 – 1.50	1.30 – 1.50	2.50 – 2.60
Thermal Resistance	Low	Moderate	Moderate	High
Biodegradability	Biodegradable	Biodegradable	Biodegradable	Non-biodegradable (But Recyclable)
Cost	Low	Very low	Low to moderate	Moderate to high
Availability (BD)	Available (Natural)	Available (Natural)	Available (Natural)	Available (Industrial)

layering multiple fiber sheets onto a mold, which is first treated with a release agent to prevent sticking. This process is especially useful for small batch production or prototyping, requiring specialized composite properties. The fabrication consists of several steps. Initially, the mold's dimensions are measured and marked on a sheet of glass, which is then cut to size with a cutting saw. Edges are smoothed and deburred to eliminate sharp points, and corners are rounded to avoid damaging the composite during preparation. This study used a 10 × 8-inch mold to ensure practicality and consistency in material conservation. The hand layup process blends craftsmanship and engineering, resulting in high-quality composite materials tailored for specific applications.

2.4. Fabrication of composites

To enhance the mechanical and physical properties of the final product, the fibers were cut at a desired size and arranged onto the mold in a specific pattern. Once the fibers were put together, the resin was mixed with hardener and applied onto the fibers by a brush. A ratio of

1:10 was used to properly mix the epoxy hardener and resin in a beaker [33]. The resin enters the fibers to remove any trapped air and ensure the fibers are appropriately distributed. Once the desired thickness and number of layers were achieved, additional layers of fiber as well as resin were applied. The composite was allowed to cure at room temperature and controlled pressure after it had been entirely laid up [34]. Then, the composites were cured, and ejected from the mold, and cut to the desired size and shape according to characterization requirements. All the composites were manufactured similarly following the combinations shown in Table 4. Three composites were created by using cotton fiber, banana fiber, jute fiber, and glass fiber in two layers with the same weight ratio in each layer. Fig. 2 shows the fabrication process of the banana-glass composite.

2.5. List of equipment

The fabrication and evaluation of composite materials require specialized equipment. In the hand layup method, essential tools include

Table 2
Scouring and bleaching recipe for jute fiber.

SI. No	Name of the Chemical	Amount (g/L)
1	NaOH	10
2	H2O2	12
3	Detergent	7
4	Sequestering Agent	5
5	M: L	1:20

Table 3
Scouring & bleaching recipe for banana fiber.

SI. No	Name of the Chemical	Amount (g/L)
1	NaOH	7
2	H2O2	10
3	Detergent	7
4	Sequestering Agent	5
5	M: L	1:20

Table 4
Details of the composition of all composite.

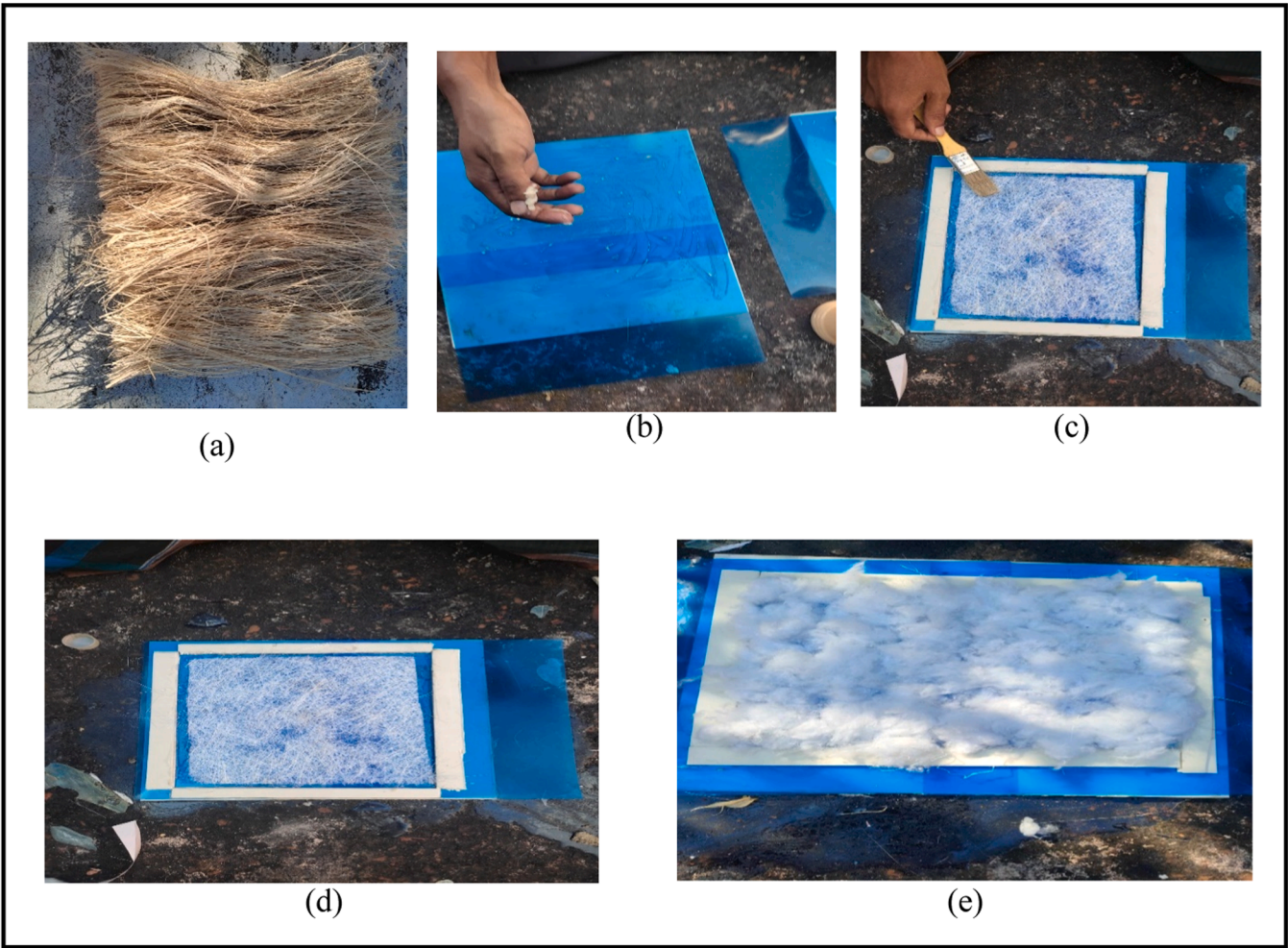| Sample no	Sample name	Sample ratio	Fabric weight (gm)	Sample size (inch)	Thickness (mm)	Resin: Hardener	Weight percentage (Fabric: Resin)
1	G: C	1:1	10:10	10*10	4	10:1	30:70
2	G: B	1:1	10:10	10*10	4	10:1	30:70
3	G: J	1:1	10:10	10*10	4	10:1	30:70

Fig. 2. Preparation technique of different composite (a)Banana fiber, (b) Mixing of wax, (c) Spreading of resin, (d) Intermediate phase, (e) Cotton fiber.

fiber reinforcement sheets, resin containers, mixing tools, and molds needed for better quality. Mechanical testing follows preparation to assess properties and structural integrity. The tools used for determining crucial parameters are listed in Table 5.

Table 5
List of equipment.

Name of Machine	Brand	Origin	Operation
UTM machine	Tinius Olsen 25ST	Germany	Tensile strength
Bending Testing Machine	Tinius Olsen 25ST	Germany	Flexural strength
Electric Balance	Ohaus	China	Weight measurement
Woven Dryer	Dhmeri	China	Drying
Lab Dyeing Machine	Gester	Germany	Pre-treatment of fibers
FTIR Machine	IR Tracer 100	Japan	Chemical bond identification
Scanning Electron Microscope	ZEISS Gemini Sigma 300	Germany	Morphological structures

3. Characterizations

3.1. Tensile test

The tensile properties of the samples were investigated using a Universal Testing Machine (UTM). The samples were cut and tested according to ASTM: D3039 standard. This test was performed at room temperature and a relative humidity of 65 %. Fig. 3 shows the samples in UTM during and after the test. The tensile stress is measured as the strain increases. Tensile strength is calculated from the equation, $\sigma = F/bh$, Where F is the maximum load at break (N), and b and h are the width and thickness of the specimen [35].

3.2. Flexural test

The flexural test was conducted to evaluate the bending strength of the prepared composites. The specimens were prepared and tested as per ASTM D790 standard in a UTM [36]. Five specimens were assessed from each type. The span length was 60 mm, and the crosshead speed was 12

Fig. 3. Fig. of UTM tensile testing machine and broken tensile test sample (a) UTM machine, (b) Tensile test sample.

mm/min. The samples of flexural tests before and after are shown in Fig. 4.

3.3. Water absorption test

The water-absorption test was performed to ascertain the moisture in a composite material as a percentage of its dry weight, as specified by the (British Standard 1377:1967). A material sample is initially weighed and then dried in an oven to conduct the test. Subsequently, the sample is weighed again under predetermined standard conditions. The difference in weight between the initial and final measurements is then used to calculate the percentage of moisture content in the specimen.

3.4. FTIR test

The FTIR Machine (IR Tracer 100) was used to identify the chemical composition of the samples. FTIR is based on the principle that each type of chemical bond in a sample vibrates at a unique frequency known as its vibrational frequency. The machine used attenuated total reflection (ATR) mode to identify the transmittance between 500-4000 cm^{-1} wavelengths.

3.5. Scanning electron microscopy (SEM)

To investigate the morphological structures of the glass-cotton, glass-jute, and glass-banana composites, an SEM machine (ZEISS Gemini Sigma 300) was used. For the imaging process, the landing voltage was 5.0 kV, the working distance varied between 12-15 mm, and the mode was vacuum (HighVac).
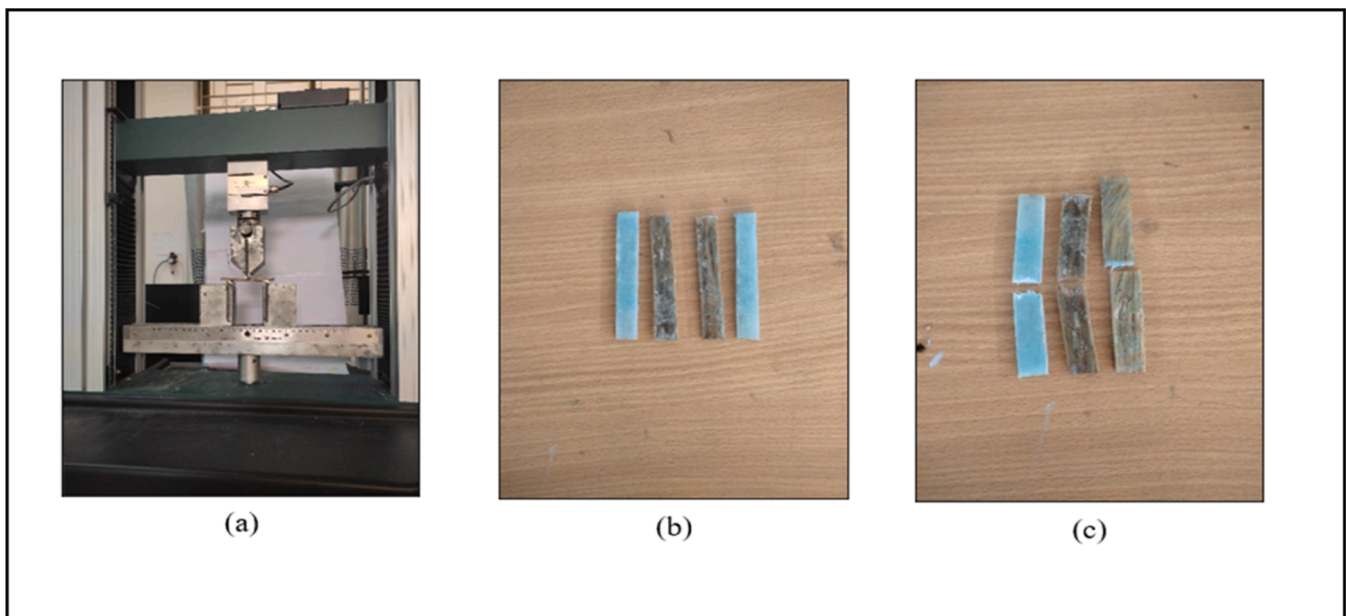

Fig. 4. Flexural test machine and sample (a) Flexural test, (b) before sample, (c) after sample.

3.6. Innovation of the methodology

This study presents a novel hand lay-up method for creating hybrid composites using glass-jute, glass-cotton, and glass-banana fibers. The aim was to achieve an optimal balance of mechanical performance, sustainability, light weight and cost-effectiveness by combining synthetic and natural fibers. An alkaline treatment was applied in the pretreatment process to enhance the bonding between natural fibers and the resin matrix, improving overall quality. This approach includes lightweight compression techniques and manual assistance, improving laminate uniformity without costly machinery. Using locally sourced, affordable natural fibers supports sustainability. Finally, a comparative analysis of the hybrid composites was conducted to assess the influence of different fiber types on mechanical and functional properties, contributing to advancements in composite development.

4. Results and discussion

4.1. Tensile properties

Initially, the tensile test of the samples was measured to determine the strength and stretchability of the composite before failure. The results provide insight into the tensile behavior of the composite materials. The produced hybrid composites were cut using a saw cutter to meet the dimensional requirements for tensile testing. During the tensile test, tensile stress was measured as the strain increased. The ultimate force was observed for different composites, including cotton-glass, banana-glass, and jute-glass composites. Due to its rigidity and stiffness [37], the jute-glass composite demonstrated the highest ultimate force of 3790 N. The cotton-glass composite exhibited a noteworthy ultimate force of 210 N, whereas the banana-glass composite had the lowest ultimate force at 1770 N.

The results for ultimate stress across various composites indicated that the jute-glass composite had the highest ultimate stress at 28.2 MPa, as shown in Fig. 5(a), attributable to its unique physical and chemical structure. The banana-glass composite showed the lowest ultimate stress, 1.2 MPa less than the cotton-glass composite. Due to its rigidity and stiffness [38], the jute-glass composite displays the highest ultimate

force at 3,790 N, shown in Fig. 5(b). The cotton-glass composite follows with a notable ultimate force of 1980 N, while the banana-glass composite has the lowest ultimate force at 1,770 N among the three materials. The test results indicate that the jute-glass composite also possesses the highest ultimate stress, measuring 28.2 MPa, because of its inherent physical and chemical properties. So, the jute-glass composite exhibits best tensile properties among the three composites.

4.2. Flexural properties

The flexural test measured the ultimate force and stress of the composite materials, which can withstand before rupturing. It is seen from the result that, the cotton-glass composite had the highest ultimate force, shown in Fig. 6(b), measuring 372.45 N. In contrast, the banana-glass composite exhibited the lowest ultimate force of 229.60 N, which can be attributed to its physical and chemical characteristics [39]. The jute-glass composite had a moderate ultimate force of 312.70 N.

In analyzing the data and graph from the bending test, the ultimate stress was measured in megapascals (MPa) for the three composites. The cotton-glass composite also demonstrated the highest ultimate stress, shown in Fig. 6(a), with a value of 105.6 MPa. The jute-glass composite had the second-highest ultimate stress, approximately 16 MPa greater than the banana-glass composite. The experimental data further indicated that when force is applied, the banana-glass composite tends to fail more quickly than both the jute-glass and cotton-glass composites.

4.3. Water absorption

During the water absorption test, the water absorption percentage of various composites were calculated for all composites. Fig. 7 shows the results of water absorption of cotton-glass composite having the highest percentage, measuring 5.8 %. The banana-glass composite had the second-highest absorption at 3.15 %. In contrast, the jute-glass composite exhibited the lowest water absorption, with a value of 2.5 %.

Epoxy resins are generally more hydrophobic, repelling water molecules [40]. Their molecular structure is more compact and tightly packed compared to polyester resins, which makes it difficult for water molecules to penetrate and diffuse into the material [41]. In contrast,
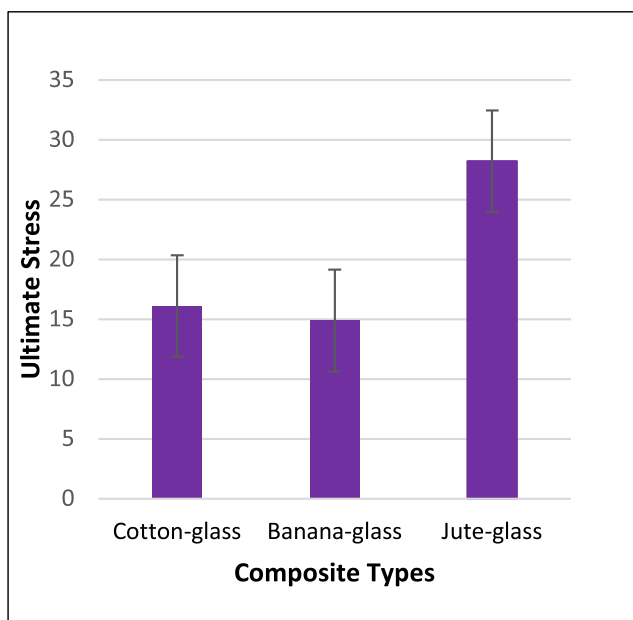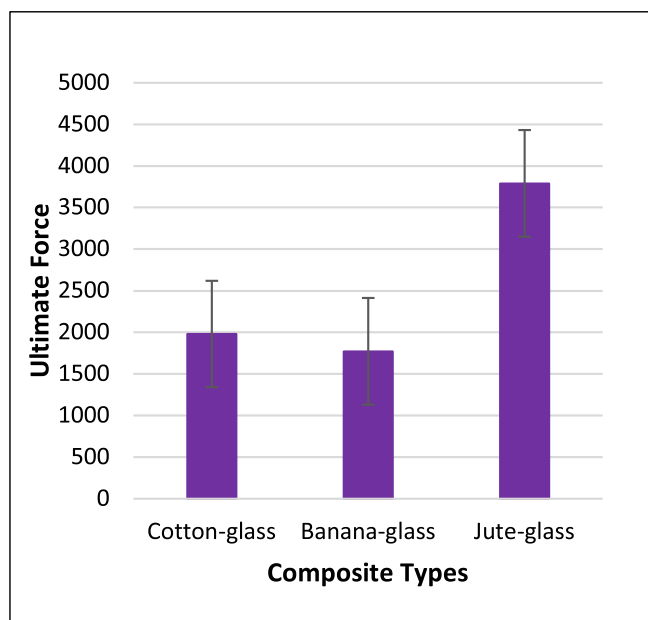
(a)   (b)

Fig. 5. Tensile test a) ultimate stress b) ultimate force.

(a)   (b)

Fig. 6. Bending test a) ultimate stress b) ultimate force.

Fig. 7. Water absorption percentage bar chart.

polyester resins contain polar groups, such as ester linkages, that can interact with water molecules through hydrogen bonding. Polyester resins have a more open molecular structure, allowing water molecules to penetrate and diffuse easily into the material.
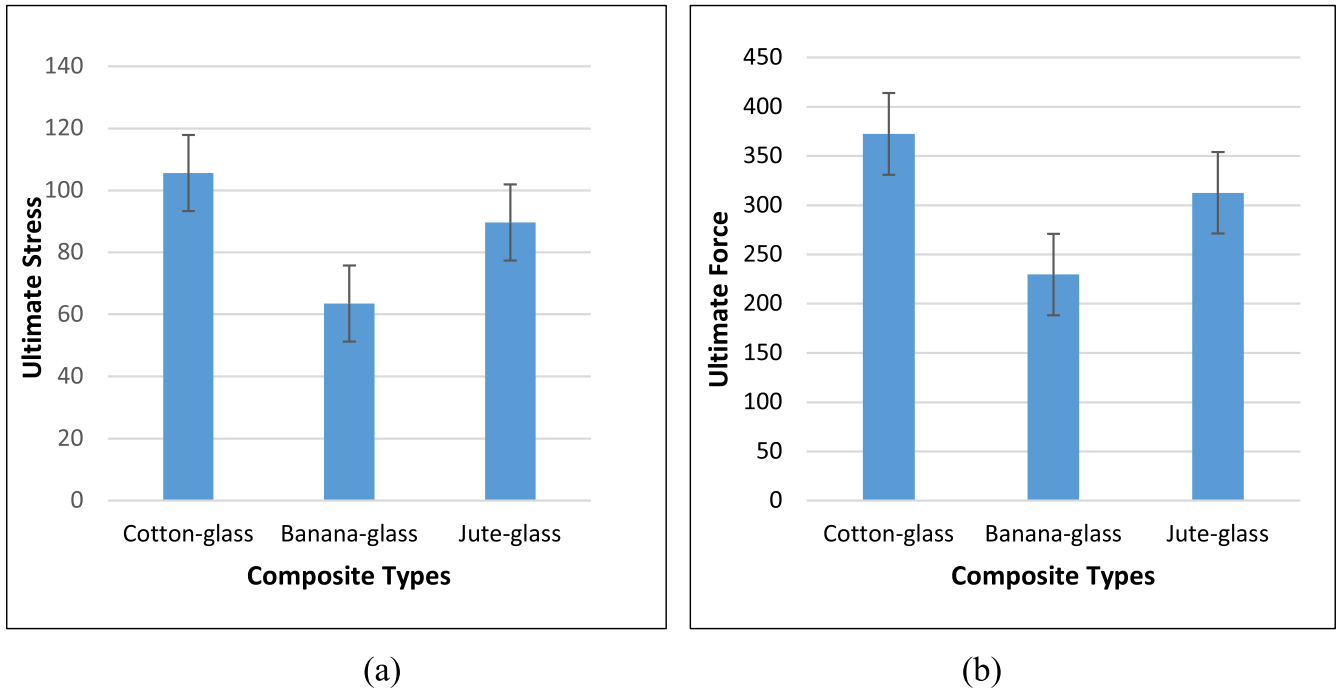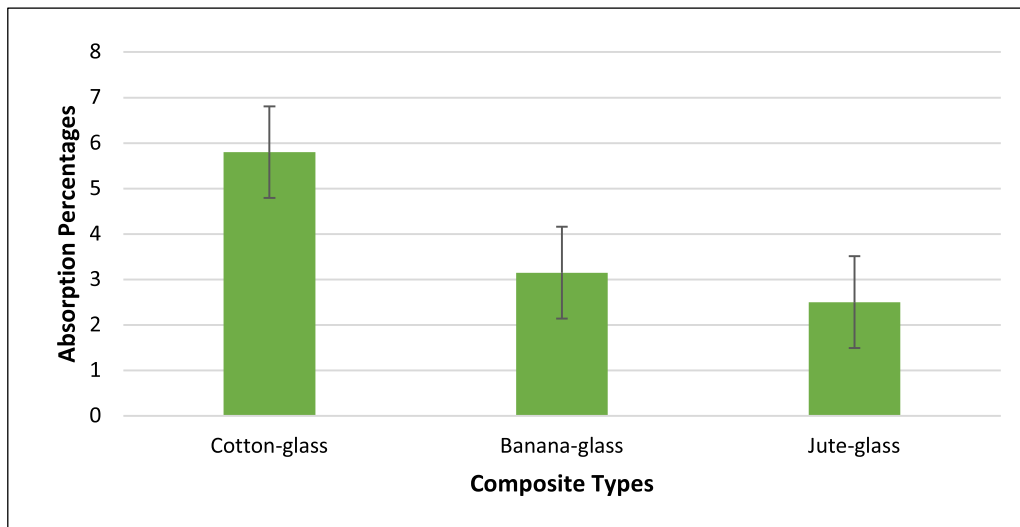
4.4. SEM test

SEM examination was conducted to visualize the morphology and microstructure of epoxy resin composites reinforced with cotton, jute, and banana fabric. The surface morphology of Banana-glass composite is shown in Fig. 8(a) at different magnifications from 100X to 1200X. SEM analysis provides a visual understanding of these composite materials' surface morphology and microstructure and offers crucial insights into their performance and durability. From the Fig., the adhesion of fibers is clear, which enhances bonding capacity and increases strength [23,29]. Fig. 8(b) visualizes the surface morphology of the Cotton-glass

composite at different magnifications, from which the bonding of fiber and resins is visible. A good fibre matrix adhesion is noticeable, which ensures less void space, which is efficient for better stress transfer and mechanical properties [30]. Similarly, the surface structure of Jute-glass composite at various magnifications between 100X and 1200X is shown in Fig. 8(c). Here, minimal fibre pullout is visible, which ensures moderate adhesion and decent mechanical properties [4,26].

By revealing the intricacies of fibre-matrix interactions and identifying potential weaknesses or areas for improvement, SEM analysis guides the refinement of manufacturing processes and formulation of composite materials with enhanced mechanical properties and structural integrity. This meticulous examination at the microscale empowers researchers and engineers to tailor the composition and fabrication techniques for cotton, jute, and banana fiber composites, unlocking their full potential across diverse industrial sectors.
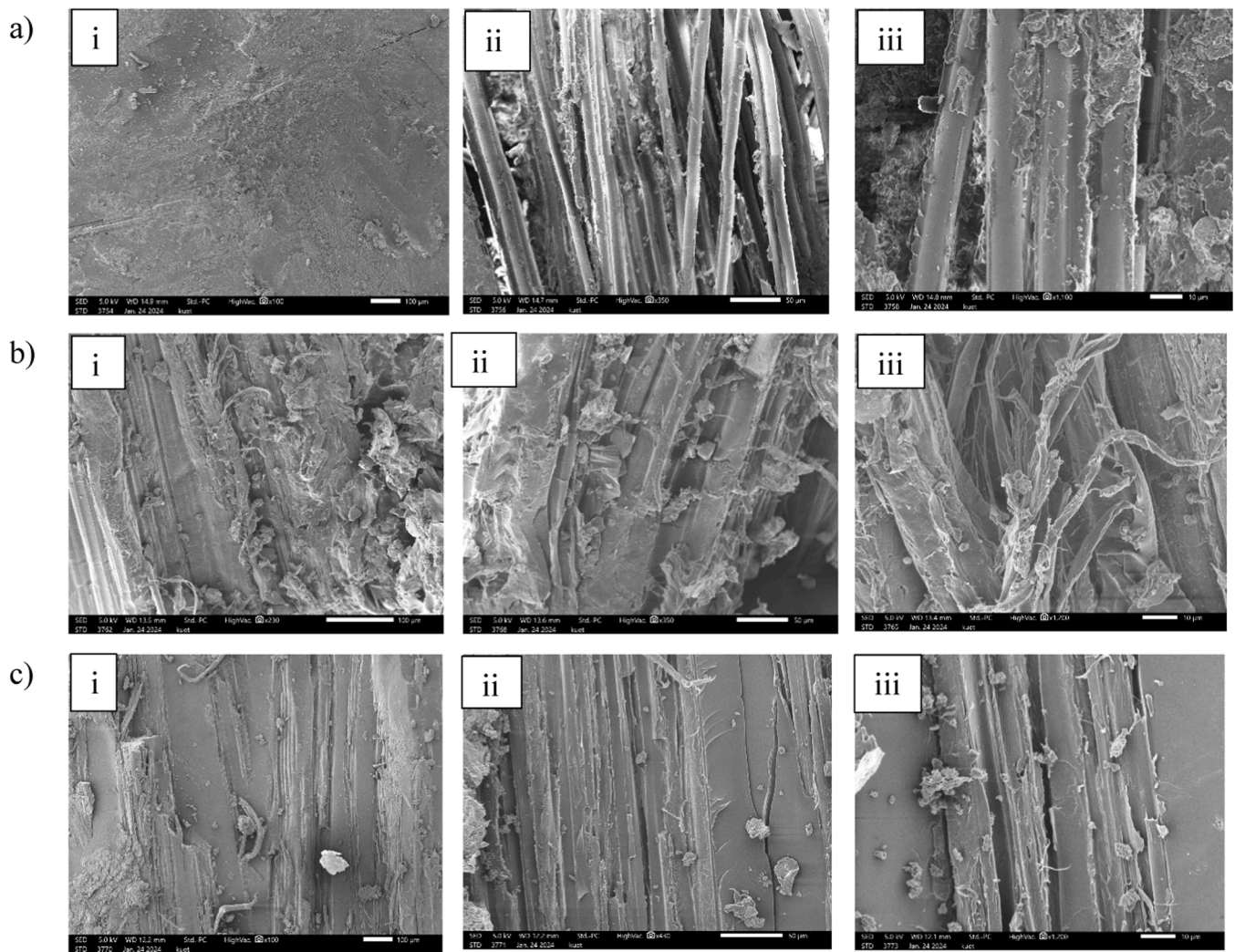
Fig. 8. Surface morphology by SEM of (a) Banana-glass composite at a magnification of i) 100X ii)350X iii) 1100X (b) Cotton-glass composite at a magnification of i) 230X ii)350X iii)1200X (c) Jute-glass composite at a magnification of i) 100X ii) 430X iii) 1200X.

4.5. FTIR test

The FTIR spectra for glass-banana composite, glass-cotton composite, and glass-jute composite samples are shown in Fig. 9, which range from a wave number of 4000 cm^{-1} to 500 cm^{-1}. The wavelengths at 825 cm^{-1}, 820 cm^{-1}, and 824 cm^{-1} correspond to the bending vibrations of Si-O-Si bonds in the glass component of the composites. These siloxane bridges create bonds with fibers, which increases the composites' binding capacity and greatly impacts mechanical properties [42]. Furthermore, the wavelengths at 1030 cm^{-1}, 1035 cm^{-1}, 1240 cm^{-1}, and 1246 cm^{-1} are associated with the stretching vibrations of C-O bonds, indicating the presence of cellulose or other organic components from the cotton fibers. The C-O groups are polar and create hydrogen bonding with resin materials, increasing tensile and flexural strength [43].

Similarly, the wavelengths 1500 cm^{-1}, 1510 cm^{-1}, and 1505 cm^{-1} are linked to C=C bonds, suggesting the presence of unsaturated hydrocarbons or other organic materials. This bond is the backbone of the cellulose fibrous material and has a strong covalent bond, which is important for structural stability [29,30]. The wavelengths at 2320 cm^{-1}, 2850 cm^{-1}, and 2855 cm^{-1} typically indicate the stretching vibrations of C-H bonds in aliphatic hydrocarbons, which signifies the presence of organic components. Finally, the wavelengths at 2920 cm^{-1}, 2917 cm^{-1}, and 2923 cm^{-1} indicate the stretching of C-H bonds, particularly in methyl (-CH$_3$) or methylene groups.

4.6. Comparison of the composites

Table 6 compares the three composites. The ultimate force values for the cotton glass composite, banana glass composite, and jute glass composite are 1980 N, 1770 N, and 3790 N, respectively. This data shows that the jute glass composite displayed the highest ultimate force among the three, while the banana glass composite demonstrated the lowest value.

The cotton glass composite achieved the best performance in the bending test with a maximum force of 372.45 N. In contrast, the banana glass composite recorded the lowest value at 229.60 N. The cotton glass composite showed the highest water absorbency at 5.8 % by weight. In contrast, the jute glass composite showed the lowest water absorbency at 2.5 % by weight.

5. Conclusion

The study investigates the development of natural fiber and synthetic fiber-reinforced epoxy composites with excellent tensile and flexural strength. Mechanical tests, including tensile strength, flexural strength, FTIR, water absorption, and SEM, were conducted on composites created through a hand lay-up process. Among these, the cotton-glass composite showed the highest flexural and bending strength in comparison with jute-glass and banana-glass composites. To improve the
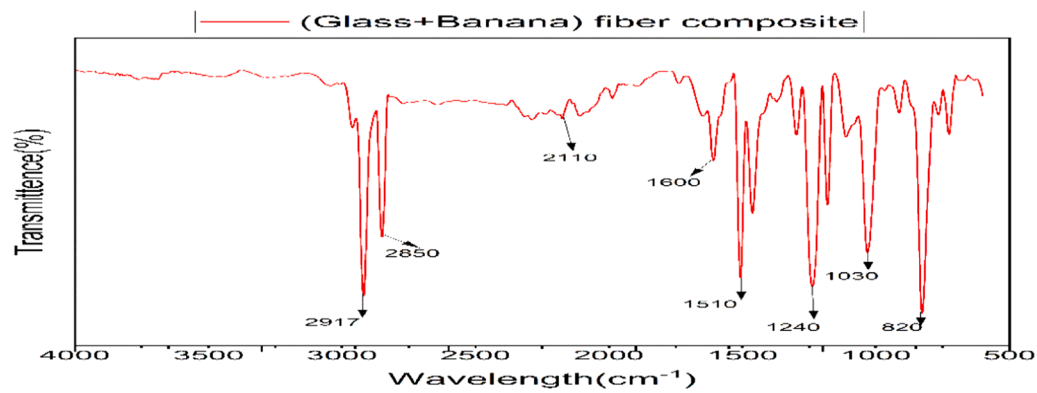
(a)
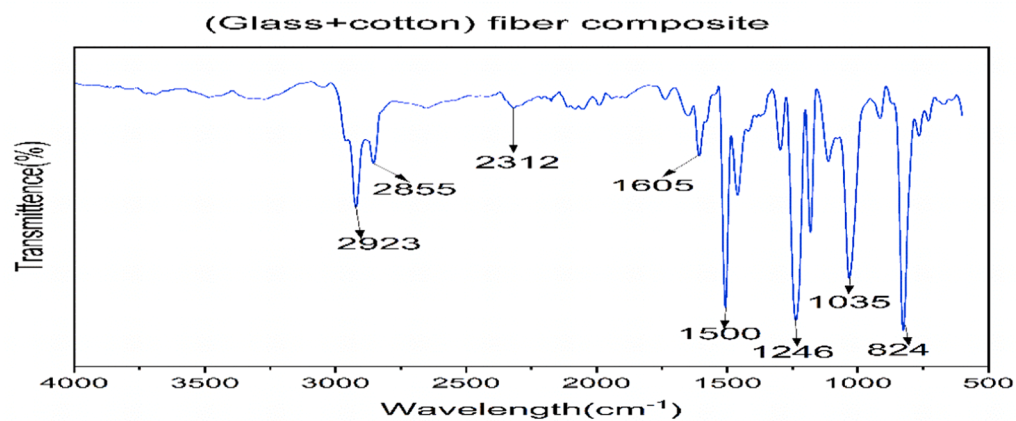
(b)
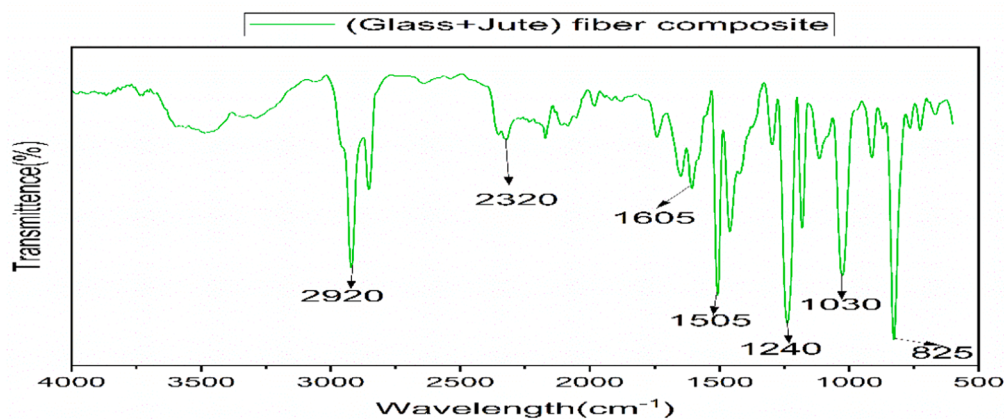
(c)

Fig. 9. FTIR result of (a) glass-banana composites, (b) glass-cotton composites, (c) glass-jute composites.

properties of these composites, the study suggests a more uniform fiber lay-up and the application of automatic pressure during preparation. These environmentally sustainable composites are easy to manufacture which have the potentiality for replacing various traditional materials while reducing the use of hazardous substances. The study also notes that the composites are made with a single type of resin: epoxy. Exploring other polyester resins could lead to even more effective composites and help identify the best options.

Funding

CRediT authorship contribution statement

Alberuni Aziz: Writing – review & editing, Validation, Supervision,

Table 6

Comparison of Cotton-glass, Banana-glass, and Jute-glass composite.

SI. No	Test		Cotton-glass composite	Banana-glass composite	Jute-glass composite
1	Tensile	Ultimate Force	1980	1770	3790
		Ultimate Stress	16.1	14.9	28.2
2	Bending	Ultimate Force	372.45	229.60	312.70
		Ultimate Stress	105.61	63.56	89.67
3	Water Absorption		5.8 %	3.15 %	2.5 %

Methodology, Conceptualization. **Farjana Parvin:** Writing – original draft, Visualization, Formal analysis, Data curation. **Md. Kajol Hossain:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Jannah, M. Mariatti, A.Abu Bakar, H.P.S. Abdul Khalil, Effect of chemical surface modifications on the properties of woven banana-reinforced unsaturated polyester composites, J. Reinf. Plast. Compos. 28 (12) (2009) 1519–1532, https://doi.org/10.1177/0731684408090366.

[2] K. Jubair, M.S. Islam, D. Chakraborty, Investigation of mechanical properties of banana-glass Fiber reinforced hybrid composites, J. Eng. Adv. (2021) 175–179, https://doi.org/10.38032/jea.2021.04.002.

[3] A. Aziz, M. Khan, M.A. Jalil, Reduction of surface resistance of jute Fiber by coating with PEDOT and characterization of the effect of different parameters on resistance, J. Eng. Appl. Sci. 4 (1) (2020) 65–73.

[4] A.A.G. Olorunnishola, E.G. Adubi, A comparative analysis of a blend of natural jute and glass fibers with synthetic glass FibersComposites as car bumper materials, IOSR J. Mech. Civ. Eng. (IOSR-JMCE) e-ISSN 15 (3) (2018) 67–71, https://doi.org/10.9790/1684-1503016771.

[5] M. Nasrullah Jahir, A. Kumar Debnath, P. Deb Nath, Orientation effect of jute-glass fiber reinforced composite on mechanical properties, Available: https://www.researchgate.net/publication/357354799, 2021.

[6] E.A. Elbadry, M.S. Aly-Hassan, H. Hamada, Mechanical properties of natural jute fabric/jute mat fiber reinforced polymer matrix hybrid composites, Adv. Mech. Eng. 2012 (2012), https://doi.org/10.1155/2012/354547.

[7] S. Biswas, Q. Ahsan, A. Cenna, M. Hasan, A. Hassan, Physical and mechanical properties of jute, bamboo, and coir natural fiber, Fibers Polym. 14 (10) (2013) 1762–1767, https://doi.org/10.1007/s12221-013-1762-3.

[8] A.L. Naidu, R. Bahubalendruni, P.Dilleswara Rao, D.V. Rao, Mechanical properties of banana fiber reinforced composites and manufacturing techniques: a review, Int. J. Res. Dev. Technol. 8 (5) (2017) 39–45.

[9] M.A. Maleque, B.F. Yousif, S.M. Sapuan, F.Y. Belal, Mechanical properties study of pseudo-stem banana fiber reinforced epoxy composite, Arab. J. Sci. Eng. 32 (2) (2007).

[10] Md. Rafiquzzaman, M. Islam, H. Rahman, S. Talukdar, N. Hasan, Mechanical property evaluation of glass–jute fiber reinforced polymer composites, Polym. Adv. Technol. 27 (10) (2016) 1308–1316, https://doi.org/10.1002/pat.3798.

[11] M.K. Gupta, R.K. Srivastava, Mechanical, thermal and water absorption properties of hybrid sisal/jute fiber reinforced polymer composite, Ind. J. Eng. Mater. Sci. 23 (2016) 231–238.

[12] B.A. Acha, N.E. Marcovich, M.M. Reboredo, Physical and mechanical characterization of jute fabric composites, J. Appl. Polym. Sci. 98 (2) (2005) 639–650, https://doi.org/10.1002/app.22083.

[13] M.M. Davoodi, S.M. Sapuan, D. Ahmad, A. Ali, A. Khalina, M. Jonoobi, Mechanical properties of hybrid kenaf/glass reinforced epoxy composite for passenger car bumper beam, Mater. Des. 31 (10) (Dec. 2010) 4927–4932, https://doi.org/10.1016/j.matdes.2010.05.021.

[14] J. Gassan, Possibilities to improve the properties of natural Fiber reinforced plastics by fiber modification – jute polypropylene composites, Appl. Compos. Mater. 7 (5) (2000) 373–385, https://doi.org/10.1023/A:1026542208108.

[15] O. Faruk, A.K. Bledzki, H.-P. Fink, M. Sain, Progress report on natural Fiber reinforced composites, Macromol. Mater. Eng. 299 (1) (2014) 9–26, https://doi.org/10.1002/mame.201300008.

[16] U.R. Vaidya, V.M. Nadkarni, Unsaturated polyester resins from poly(ethylene terephthalate) waste. 1. Synthesis and characterization, Ind. Eng. Chem. Res. 26 (2) (1987) 194–198, https://doi.org/10.1021/ie00062a003.

[17] P.N. Khanam, M.A.A. AlMaadeed, Processing and characterization of polyethylene-based composites, Polym. Compos. Sci. 1 (2) (2015) 63–79, https://doi.org/10.1179/2055035915Y.0000000002.

[18] G. Kannan, R. Thangaraju, P. Kayaroganam, J.P. Davim, The combined effect of banana Fiber and fly ash reinforcements on the mechanical behavior of polyester composites, J. Nat. Fibers 19 (15) (2022) 11384–11403, https://doi.org/10.1080/15440478.2022.2025977.

[19] R.W.I.B. Priyadarshana, P.E. Kaliyadasa, S.R.W.M.C.J.K. Ranawana, K.G. C. Senarathna, Biowaste management: banana Fiber utilization for product development, J. Nat. Fibers. 19 (4) (2022) 1461–1471, https://doi.org/10.1080/15440478.2020.1776665.

[20] K.W.M. Davy, S. Kalachandra, M.S. Pandian, M. Braden, Relationship between composite matrix molecular structure and properties, Biomaterials 19 (22) (1998) 2007–2014, https://doi.org/10.1016/S0142-9612(98)00047-7.

[21] S.S. Tripathy, L. Di Landro, D. Fontanelli, A. Marchetti, G. Levita, Mechanical properties of jute fibers and interface strength with an epoxy resin, J. Appl. Polym. Sci. 75 (13) (2000) 1585–1596, https://doi.org/10.1002/(SICI)1097-4628(20000328)75:13<1585::AID-APP4>3.0.CO;2-Q.

[22] K. Abe, M. Morita, H. Yano, Fabrication of optically transparent cotton fiber composite, J. Mater. Sci. 53 (15) (2018) 10872–10878, https://doi.org/10.1007/s10853-018-2309-1.

[23] L.A. Pothan, P. Potschke, R. Habler, S. Thomas, The static and dynamic mechanical properties of banana and glass Fiber woven fabric-reinforced polyester composite, J. Compos. Mater. 39 (11) (2005) 1007–1025, https://doi.org/10.1177/0021998305048737.

[24] J. Demko, J. Machava, Tree resin, a macroergic source of energy, a possible tool to lower the rise in atmospheric CO2 levels, Sustainability 14 (6) (2022) 3506, https://doi.org/10.3390/su14063506.

[25] X. Yi, Development of multifunctional composites for aerospace application, Online]. Available: https://www.researchgate.net/publication/280444466, 2015.

[26] M.Y. Khalid, A. Rashid, Z.U. Arif, M.F. Sheikh, H. Arshad, M.A. Nasir, Tensile strength evaluation of glass/jute fibers reinforced composites: an experimental and numerical approach, Results Eng. 10 (2021) 100232, https://doi.org/10.1016/j.rineng.2021.100232.

[27] T.J. Pramanik, M. Rafiquzzaman, A. Karmakar, M.H. Nayeem, K.S. Turjo, M. R. Abid, Evaluation of mechanical properties of natural fiber based polymer composite, Bench Council Trans. Benchmarks Stand. Eval. 4 (3) (2024) 100183, https://doi.org/10.1016/j.tbench.2024.100183.

[28] E. Mahmud, A. Aziz, F. Parvin, Fabrication of PLA bio-composite reinforced with jute fiber and ESP fillers and optimization of mechanical properties via fuzzy logic approach, Results Mater. 26 (2025) 100684, https://doi.org/10.1016/j.rinma.2025.100684.

[29] B. Sampath, N. Naveenkumar, P. Sampathkumar, P. Silambarasan, A. Venkadesh, M. Sakthivel, Experimental comparative study of banana fiber composite with glass fiber composite material using Taguchi method, Mater. Today 49 (2022) 1475–1480, https://doi.org/10.1016/j.matpr.2021.07.232.

[30] E.H. Portella, D. Romanzini, C.C. Angrizani, S.C. Amico, A.J. Zattera, Influence of stacking sequence on the mechanical and dynamic mechanical properties of cotton/glass fiber reinforced polyester composites, Mater. Res. 19 (2016) 542–547, https://doi.org/10.1590/1980-5373-MR-2016-0058.

[31] M. Zeeshan, M. Ali, A.S. Anjum, Y. Nawab, Optimization of mechanical/thermal properties of glass/flax/waste cotton hybrid composite, J. Ind. Text. 51 (5) (2021) 768–787, https://doi.org/10.1177/1528083719891420.

[32] D.D.L. Chung, Introduction to carbon composites, Carbon Compos. (2017) 88–160, https://doi.org/10.1016/B978-0-12-804459-9.00002-6.

[33] Z. Xu, C. Gao, Graphene fiber: a new trend in carbon fibers, Mater. Today 18 (9) (2015) 480–492, https://doi.org/10.1016/j.mattod.2015.06.009.

[34] H. Altenbach, J. Altenbach, W. Kissing, Classification of composite materials. Mechanics of Composite Structural Elements, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 1–14, https://doi.org/10.1007/978-3-662-08589-9_1.

[35] A. Gholampour, T. Ozbakkaloglu, A review of natural fiber composites: properties, modification and processing techniques, characterization, applications, J. Mater. Sci. 55 (3) (2020) 829–892, https://doi.org/10.1007/s10853-019-03990-y.

[36] N. Venkateshwaran, A. Elayaperumal, Banana Fiber reinforced polymer composites - a review, J. Reinf. Plast. Compos. 29 (15) (2010) 2387–2396, https://doi.org/10.1177/0731684409360578.

[37] A. Gunge, P.G. Koppad, M. Nagamadhu, S.B. Kivade, K.V.S. Murthy, Study on mechanical properties of alkali treated plain woven banana fabric reinforced biodegradable composites, Compos. Commun. 13 (2019) 47–51, https://doi.org/10.1016/j.coco.2019.02.006.

[38] T. Sathishkumar, J. Naveen, P. Navaneethakrishnan, S. Satheeshkumar, N. Rajini, Characterization of sisal/cotton fibre woven mat reinforced polymer hybrid composites, J. Ind. Text. 47 (4) (2017) 429–452, https://doi.org/10.1177/1528083716648764.

[39] W. Wang, Z. Cai, J. Yu, Z. Xia, Changes in composition, structure, and properties of jute fibers after chemical treatments, Fibers Polym. 10 (6) (2009) 776–780, https://doi.org/10.1007/s12221-009-0776-3.

[40] D. Plackett, Perspectives on the performance of natural plant fibres, Online]. Available: https://www.researchgate.net/publication/252916893, 1999.

[41] M. Elkington, D. Bloom, C. Ward, A. Chatzimichali, K. Potter, Hand layup: understanding the manual process, Adv. Manuf.: Polym. Compos. Sci. 1 (3) (2015) 138–151, https://doi.org/10.1080/20550340.2015.1114801.

[42] M. Sanchez, P. Pages, T. Lacorte, K. Briceno, F. Carrasco, Curing FTIR study and mechanical characterization of glass bead filled trifunctional epoxy composites, Compos. Sci. Technol. 67 (9) (2007) 1974–1985, https://doi.org/10.1016/j.compscitech.2006.10.006.

[43] B. Dahlke, H. Larbig, H.D. Scherzer, R. Poltrock, Natural Fiber reinforced foams based on renewable resources for automotive interior applications, J. Cell. Plast. 34 (4) (1998) 361–379, https://doi.org/10.1177/0021955×9803400406.

Alberuni Aziz received B.Sc. Engineering degree in Textile Engineering and M.Sc. Engineering degree in Industrial Engineering and Management from Khulna University of Engineering & Technology. Currently, he is pursuing PhD at Khulna University of Engineering & Technology. His research interests include Textile Composite Materials, Industry 4.0, Modeling and optimization, and Materials Science. He is an Assistant Professor of the Department of Textile Engineering at Khulna University of Engineering & Technology, Khulna, Bangladesh.

Farjana Parvin received B.Sc. and M.Sc. Engineering degree in Industrial and Production Engineering from Khulna University of Engineering & Technology. Her research interests include Composite Materials, Human Factors Engineering, Product Design and Development, Supply Chain Management and Operations Management. She is an assistant professor at the Department of Industrial Engineering and Management at Khulna University of Engineering & Technology, Khulna, Bangladesh.

Md. Kajol Hossain received B.Sc. Engineering degree in Textile Engineering from Khulna University of Engineering & Technology. His research interests include Textile Composite Materials, Structural Composites, Advanced Fiber Materials and Simulation. He is an aspirant for the Master's program.

Research Article

Comparative study of deep learning models for Parkinson's disease detection☆

Abdulaziz Salihu Aliero *, Neha Malhotra

School of Computer Application, Lovely Professional University Phagwara, Punjab, India

ARTICLE INFO

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects movement and cognition, impacting millions of people worldwide. The diagnosis of PD primarily relies on clinical tests, which can often result in delayed identification of the disease. Recent advancements in data-driven methods using deep learning have demonstrated potential for improving early diagnosis by utilizing clinical and vocal inputs. This study conducted a comparative analysis of five deep learning models: Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Autoencoder, and Generative Adversarial Network (GAN), specifically for the detection of PD using vocal biomarkers. Among these models, the MLP achieved the highest predictive accuracy at 97.4 %. The RNN, GRU, and Autoencoder models attained a similar accuracy rate of 87.2 %. In contrast, the GAN model yielded an accuracy of only 76.9 %. The UCI vocal dataset from Kaggle was utilized in this research, along with extensive data preprocessing techniques to address missing values. Performance evaluation was conducted using multiple metrics. The results indicate that deep learning models can effectively diagnose PD using voice data, suggesting their potential to enhance diagnostic accuracy and support clinical decision-making. Furthermore, these models are feasible for large-scale integration into clinical workflows.

1. Introduction

Currently after Alzheimer's disease, Parkinson's disease (PD) is considered the second most prevalent chronic and rapidly progressing neurodegenerative disorder, affecting millions of people worldwide. Despite receiving a diagnosis, many individuals succumb to the disease. Over time, the number of new cases has continued to rise, as research predicts by 2040, over 17 million individuals will be impacted by (PD) [1].

Besides, the disease causes profound damage to the brain cells that produce dopamine, a very critical neurotransmitter that is in charge of movement and coordination [2]. The neuronal damage is followed by a loss in dopamine, ultimately leading to motor and non-motor symptoms [3].

The motor symptoms typically become evident many years after the onset of the disease after experiencing a prodromal period with non-motor symptoms [4]. The disease targets particular regions of the brain, the substantia nigra and the superior colliculus in the midbrain [5]. (See Fig. 1). The superior colliculus, is another region of the

midbrain that handles visual information and eye movement, this reduces the quality of life considerably. Interestingly enough, early identification is crucial for the management of the disease and provision of more treatments that would slow down the progression [6]. With this, having more advanced diagnostic tools for medical practitioners in order to detect the disease earlier before progressing to advanced stages are absolutely essential. In the same way, detection in the prodromal stage rather than the postmotor stage would allow effective neuroprotection with the objective of delaying advanced motor symptoms. It is important to note that reducing the disease burden is eventually capable of increasing the quality of life in the patient for a more extended period by enabling early detection [7].

Conventional techniques particularly Machine Learning (ML) for PD detection heavily depends on clinical evaluation with limited data and motor symptoms. These symptoms are effective for later stages of PD but not as sensitive for detection of the disease in early stages. This occurs as motor symptoms appear following significant damage in the dopaminergic cells of the brain with approximately 50–70 % of dopaminergic cells in the substantia nigra being destroyed [8]. Early diagnosis is thus
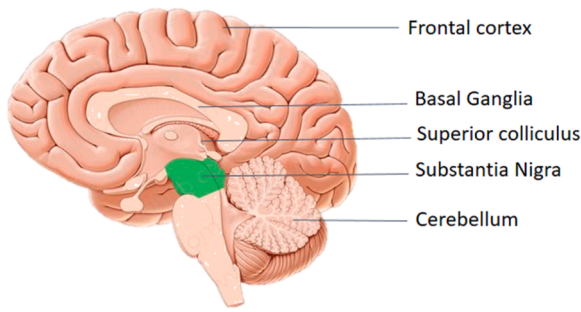

Fig. 1. Substantia nigra in the midbrain.

challenging with ML, prompting diagnostic tools that can diagnose the disease at preclinical or early stages by using deep learning (DL) modalities [9]. This is a highly specific technique in terms of equipment. Moreover while, DL provide some diagnostic assistance in detecting PD, early diagnosis of this disease remains a problem. Because non-motor symptoms become apparent after a long period, most patients receive a diagnosis after a considerable portion of the dopaminergic brain system is affected.

However, the highly subjective nature of current clinical observations and the insensitivity in imaging the identification of early neuronal loss indicate a crucial necessity [10]. With these cutting-edge techniques of using Artificial intelligence (AI), specifically DL algorithms, as a means of obtaining more and better diagnostic tools, AI is among the novel methods that can help fill gaps in early diagnosis of PD using clinical data analysis. DL revolutionized the ML field in recent times this is no longer a secret, as it has made it possible for anyone to be able to design models that can process large and complex sets of information, also it has been an innovative technology in medical diagnosis with enhanced precision in PD identification in all data modalities [11]. Deep learning models, which can analyze progressively higher amounts of information, primarily imaging and clinical parameters, can do so in a way that associates them with PD illnesses with a remarkably high level of precision in clinical diagnosis. In the previous few years, a class of deep learning architectures, e.g., convolutional neural networks (CNNs), are brilliant in learning how to automatically extract features from raw points instead of engineering them. In PD diagnosis, this is especially critical as human experts argue that the detection of subtle patterns in the imaging or clinical data should give rise to significant improvements in early detection[12]. Although traditional ML models have demonstrated promising results in PD detection, but DL have exhibited superior performance, particularly in handling large and complex datasets. Deep learning models have the advantage of automatic feature extraction, reducing the reliance on manual feature engineering and has the capacity and ability to handle highly complex architectures, often comprising hundreds or even thousands of layers.

The main goal of this study is to analyse deep learning models for PD detection, even before the onset of motor symptoms. It aims to create diagnostic models that match or surpass existing methods in accuracy. Additionally, the study analyse five deep learning models for PD detection using voice data, in other to identify the most accurate and robust model. Furthermore, it focuses on building a scalable model that can be effectively deployed across diverse healthcare settings while maintaining high accuracy. Various aspects of model performance evaluation were considered by ensuring that the model is robust enough to function well with any patient dataset [13]. Another way to frame this work is ensuring that the model is robust enough to behave well on any configuration of patient dataset. The goal is to ensure that the model works equally well on different patient types, providing a level of confidence one can have in diagnosis regardless of patient population beyond the one in which it trained.

The structure of the study is elaborated as: Section II. Describes related work, Section III describes methods used and dataset pre-

processing. Results and discussion in Section IV, and lastly Section V, encloses conclusion and future directions.

2. RELATED WORK

Recently research efforts has been put on detecting Parkinson's disease (PD) by using clinical data, voice data and MRI image data based on machine learning (ML) and deep learning (DL) techniques. This is because, these non-traditional methods hold promise for spotting stable warning signs which often escapes detection via traditional diagnostic means. In this regard, the diagnosis of PD has traditionally relied on clinical evaluations involving motor function assessments, speech and gait analysis, and neuroimaging techniques. However, these conventional methods often lack sensitivity, particularly in detecting early-stage PD when timely intervention is most beneficial [14]. In recent years, ML and DL techniques have shown promise in enhancing diagnostic accuracy by leveraging diverse clinical datasets. Early applications of ML for PD detection primarily focused on voice data, as vocal impairments are common among PD patients. Various studies have demonstrated the effectiveness of ML models when combined with feature selection techniques. For instance, Saeed et al. [15] uses feature selection techniques based on filters and wrappers to process voice recordings, achieving an accuracy of 88.33 % with k-nearest neighbors (KNN). Another study [16] illustrate using DL techniques on clinical data, specifically voice signals for the early diagnosis of PD and improved diagnostic accuracy. Similarly, Singh et al. [17] applied decision trees, random forests, and logistic regression to voice-based features, reinforcing the potential of ML-based techniques as cost-effective and non-invasive tools for PD detection. Beyond voice analysis, recent research has expanded the scope of clinical data used in PD detection. Lin et al. [18] employed motion data collected via inertial measurement units (IMUs) to develop neural network models capable of distinguishing healthy individuals from PD patients. Their study achieved detection rates of 92.72 % for advanced-stage PD and 99.67 % for early-stage cases, highlighting the growing role of gait analysis in early diagnosis.

Chintalapudi et al. [19] compared three DL architectures RNN, MLP, and LSTM on voice features of PD patients. Their findings revealed that the LSTM model achieving an accuracy of 99 %. This study underscores the ability of DL models, particularly LSTM, to effectively process nonlinear and complex data such as speech recordings. In contrast, traditional ML models require extensive preprocessing and manual feature selection, which can be labor-intensive and less scalable. Similarly Kurmi et al. [20] implemented an ensemble of CNN models, including VGG16, ResNet50, Inception-V3, and Xception, to analyze DaTscan images, achieving a classification accuracy of 98.45 %. Their study highlights the potential of CNN-based architectures in improving diagnostic precision through multi-model integration with DL that has been effectively applied to medical imaging for PD detection. Despite the advantages of DL, traditional ML approaches remain relevant, particularly when computational resources are limited or when datasets are small. For instance, Govindua et al. [21] compared random forests, SVM, and logistic regression for PD detection using voice data, achieving an accuracy of 91.83 %. While DL excels in large-scale datasets, ML techniques can still deliver competitive results when optimized with feature selection techniques. Another study by [22] explores ML and DL models for classifying PD using speech data. The results indicate that DL models outperform traditional ML approaches, with ML achieving an accuracy of 92.18 %, while the most effective DL model attains the highest accuracy of 95.41 %.

Neural networks (NN), particularly CNNs and LSTMs, have been extensively used in medical diagnostics due to their ability to model complex patterns in diverse clinical datasets, including images, time-series data, and auditory signals. Several studies have leveraged LSTM and CNNs [23,24] to analyse different data types such as gait, speech, and handwriting for PD detection. Similarly NN has also been used in multimodal data, where different types of clinical measurements were

integrated into providing a more comprehensive diagnostic model. Taleb et al. [25] used CNN-BLSM architectures on motion HandPD_-MultiMC_data collected via wearable sensors, achieving 97.62 % for PD stage classification. These findings emphasize the efficacy of ensemble models in enhancing diagnostic accuracy by capturing both static and dynamic characteristics of PD symptoms. Furthermore, LSTMs on the other hand, have been widely adopted for sequential data modeling, making them particularly suitable for voice and motion analysis in PD detection. Neural networks have also been integrated into multimodal frameworks, where multiple clinical data sources are combined for more comprehensive diagnostic models. CNNs have been applied to EEG data analysis for PD classification. Sugden and Diamandis [26] developed a channel-wise CNN model that achieved an accuracy of 80.4 %, this study highlights CNN's has the ability to extract spatial patterns from EEG data, which could be extended to other clinical modalities. Furthermore Majhi et al. [27] explored Hybrid DL models including Grey Wolf Optimization (GWO) optimization, which applied to two images datasets, these metaheuristic algorithm achieved 99.94 % accuracy. Another study by Islam et al. [28] conduct Extensive review for PD detection, their review concluded that voice and handwriting dataset integration significantly enhances diagnostic accuracy, especially in early-stage PD detection rather than just using images data. A study by [29] employed lightweight, pre-trained DL models with two-fold training and merged ideal features for hand-drawn Parkinson's disease screening. Similarly, Keles et al. [30] used Part-Aware Residual Network (PARNet), retrained from a COVID-19 model, and applied to SPECT images for PD detection. Al-Tam et al. [31] used Ensemble learning with stacking and bagging applied to two benchmark PD datasets to Enhance PD diagnosis through stacking ensemble-based ML approach, the technique seeks to improve generalisation performance, rectify dataset class imbalances, and raise the overall accuracy of PD detection. Furthermore, Ismail and Osman [32] performed a Classification of scalograms using AlexNet, GoogleNet, and ResNet50, followed by a hybrid system based on majority voting. Also, DenseNet and NasNet are used for different classifications. Hongyi et al. [33] Used brain images and radiomics-based automated hybrid approach targeting midbrain for early PD detection. Table 1. Presents a

comprehensive assessment of research studies on PD detection using ML and DL techniques. Building on existing studies, our focus is to demonstrate that voice data can effectively identify PD in its early stages.

Numerous research studies, including those cited in references [14–16] and [19–20], have explored the application of machine learning (ML) and deep learning (DL) techniques for detecting Parkinson's disease (PD) through voice data. These studies often utilize models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM), typically examining either individual architectures or making pairwise comparisons. However, there is a significant gap in the literature regarding systematic comparative analyses of multiple deep learning models assessed under consistent experimental conditions. This study aims to fill that gap by comparing five distinct deep learning models within a unified evaluation framework, assessing their performance on a standardized voice dataset.

Our interpretations focused on both the classification accuracy itself and the potential operationalization of the systems, which bring implications for future clinical integration. Therefore, this study conducts a comparative analysis of five deep learning models applied to voice data for Parkinson's disease detection.

3. METHODOLOGY

The proposed method was implemented on the Google Colab platform using Python programming and incorporates five models to predict whether the patient has Parkinson's disease or not. To ensure validation of the model on unseen data and avoid fitting the training data too well, the dataset was divided into training sets, validation sets, and test sets.

3.1. Dataset and preprocessing

This study employs a publicly available voice dataset derived from the UCI Repository, which is also available on Kaggle, comprising the collected acoustic characteristics that are utilised to identify PD. The

Table 1
A summary of research studies focused on Parkinson's Diseases.

Ref.	Study Focus	Dataset Used	Model(s) Used	Best Performance
Al-Nefaie et al. [14]	Highlights the effectiveness of ML classifiers with vice data	UCI Voice dataset	KNN, SVM, RF and LR	SVM and RF achieve 95.00 %
Faisal et al. [15]	Voice data for PD detection	Voice recordings	KNN with feature selection	88.33 %
Awais et al. [16]	DL on voice signals for PD	Clinical voice signals	Deep Learning models	86.00 %
Shikha et al. [17]	ML models on voice-based features	Voice recordings	Decision Tree, RF, Logistic Regression	92.00 %
Lin et al. [18]	Motion data analysis using IMUs	Motion sensor data	Neural Networks	99.67 %
Chintalapudi et al. [19]	Comparison of DL models on voice	Voice features	RNN, MLP, LSTM	99.00 %
Kurmi et al. [20]	DaTscan image analysis	DaTscan images	CNN Ensemble (VGG16, ResNet50, etc.)	98.45 %
Alshammri et al. [21]	ML model comparison using voice	Voice data	RF, SVM, Logistic Regression	91.83 %
Rahman et al. [22]	ML vs DL on speech data	Speech recordings	ML and DL models	95.41 %
Govindu et al. [23]	Use of ML in Telemedicine	MDVP Audio data	SVM, RF, KNN and LR	91.83 %
Rehman et al. [24]	Multimodal data (speech, gait, handwriting)	Various clinical datasets	Hybrid LSTM-GRU	98.00 %
Taleb et al. [25]	CNN-BLSM on Handwriten data	HandPD_MultiMC_data	CNN-BLSM	97.62 %
Sugden et al. [26]	EEG analysis with CNN	EEG recordings	Channel-wise CNN	80.04 %
Majhi et al. [27]	PD Diagnosis using hybrid deep learning and metaheuristic optimization	(T1, T2-weighted) MRI & SPECT DaTscan	DenseNet, InceptionV3, LSTM, Grey Wolf Optimization (GWO), VGG16,	99.94 %
Islam et al. [28]	ML/DL models for PD detection using handwriting and voice data	Handwriting & Voice	Various ML & DL models	95.41 %
Rajinikanth et al. [29]	Hand-Sketch-based PD screening using pre-trained DL models	Hand-Sketch Wave Data	MobileNet, KNN	100 %
Keles et al. [30]	PD detection using a retrained COVID-19 model on SPECT images	SPECT Imaging	PARNet, ResNet	95.43 %
Al-Tam et al. [31]	Stacking ensemble ML models for PD detection	PD Benchmark Datasets	Random Forest, SVM, Gradient Boosting, Logistic Regression	96.18 %
Ismail and Osman [32]	Scalogram-based PD detection with deep learning	Speech Signals	AlexNet, GoogleNet, ResNet50, DenseNet, NasNet	95.00 %
Hongyi et al. [33]	Radiomics-based deep learning for early PD detection	MRI (Midbrain & Substantia Nigra)	YOLO v5, LeNet	96.03 %

dataset contains 195 voice recordings from PD positive and negative participants. It comprises 21 attributes extracted from voice records; the features are describe in (Table 2) [22], they provide full information about vocal patterns to allow effective classification. For example, Srinivasan et al. [34] have used the UCI dataset and voice samples of matched controls and PD patients to classify subjects based on voice feature changes using supervised machine learning methodologies.

The dataset was separated into target labels (y) and input features (x). Using train_test_split, a stratified split was carried out, with 20 % going to testing and 80 % going to training. The target variable indicates whether the subject is PD-positive (1) or healthy (0) [22]. (Fig. 2) illustrates the typical process that are followed, it explains how the dataset was splits into training and test data, train five models on the dataset, and validate the results using test data.

Min-Max Scaling was used to normalise the features to a [0, 1] range prior to model training. In order to express time-steps for recurrent models (LSTM and GRU), the input was transformed into a three-dimensional format.

During pre-processing, particularly when working with clinical data it is essential to address the Outliers and values that are missing. If these issues are not properly addressed, it can result in a shortage of accurate models and exacerbate future challenges. A study by [36], applied different approaches to solve this problem enabling a complete dataset to train the model. Multiple imputation techniques were applied to handle the missing clinical assessment data, and subsequently, the duplicate value was also addressed. Another study by [37] identified the requirement for the identification of outliers in speech data and their exclusion as they play a major role in the estimation of a model with undesirable or abnormal values; data was normalized with the z-score and visually inspected for outliers and marked accordingly. After the exclusion of outliers, highly correlated features were identified with the correlation coefficient validating their appropriateness. (Fig. 3) below

Table 2
Dataset Features [35].

| Feature Name	Description
Average Fundamental Frequency (Fo)	Mean vocal fundamental frequency measured in Hertz (Hz).
Maximum Fundamental Frequency (Fhi)	Highest recorded vocal fundamental frequency (Hz).
Minimum Fundamental Frequency (Flo)	Lowest recorded vocal fundamental frequency (Hz).
Jitter (%)	Percentage-based measure of frequency variation.
Jitter (Abs)	Absolute measurement of frequency instability.
Relative Average Perturbation (RAP)	Short-term variation in fundamental frequency.
Pitch Period Perturbation Quotient (PPQ)	Measure of long-term frequency variations.
Jitter: DDP	Three-point average of absolute jitter values.
Shimmer	Quantifies amplitude variation in the voice signal.
Shimmer (dB)	Measurement of amplitude fluctuation in decibels (dB).
Amplitude Perturbation Quotient (APQ3)	Average of amplitude variations over three cycles.
Amplitude Perturbation Quotient (APQ5)	Average of amplitude variations over five cycles.
MDVP: APQ	Measures overall amplitude perturbation across the signal.
Shimmer: DDA	Three-point averaged shimmer calculation.
Noise-to-Harmonics Ratio (NHR)	Ratio of non-harmonic noise to harmonic components.
Harmonics-to-Noise Ratio (HNR)	Ratio measuring harmonic signal strength relative to noise.
Correlation Dimension (D2)	Nonlinear measure of signal complexity.
Fractal Scaling Exponents	Quantifies the self-similarity and complexity of the signal.
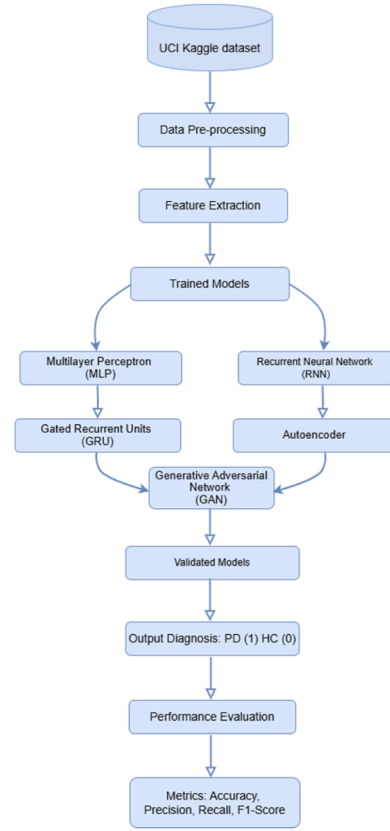
Fig. 2. Proposed flow of our approaches.

illustrates the correlation of the dataset features when outliers are excluded [35].

3.2. Multilayer perceptron (MLP)

The MLP model was implemented with Scikit-learn's MLPClassifier with two hidden layers, with (100, and 50) neurons respectively, used ReLU as an activation function and Adam as the optimizer, this constitutes a deep neural network by definition, we recognize that it is relatively shallow compared to modern deep learning architectures such as CNNs or LSTMs. This configuration was selected to evaluate how well a simple deep model could perform on the voice dataset for Parkinson's Disease detection. The model was trained up to 500 epochs. The evaluation metrics like accuracy, precision, recall and F1-score were computed based on predicted values on the test set [38].

For visualization, confusion matrices and heatmaps were used. A heatmap of the classification report was also generated to give a more visual summary of the performance metrics across the two classes.

3.3. RNN-LSTM model

The LSTM based RNN is built with TensorFlow's Keras API. The architecture consisted of one LSTM layer (with 50 units) followed by Dense output layer with sigmoidal activation for binary classification. Binary cross-entropy was used as the loss function [39], and the Adam optimizer was used to compile the model and it trained for 50 epochs with a batch size of 32.

The reshaping of inputs was performed in a way suitable for the LSTM model as 3D tensors. Accuracy, classification reports and confusion matrices were used to evaluate performance.
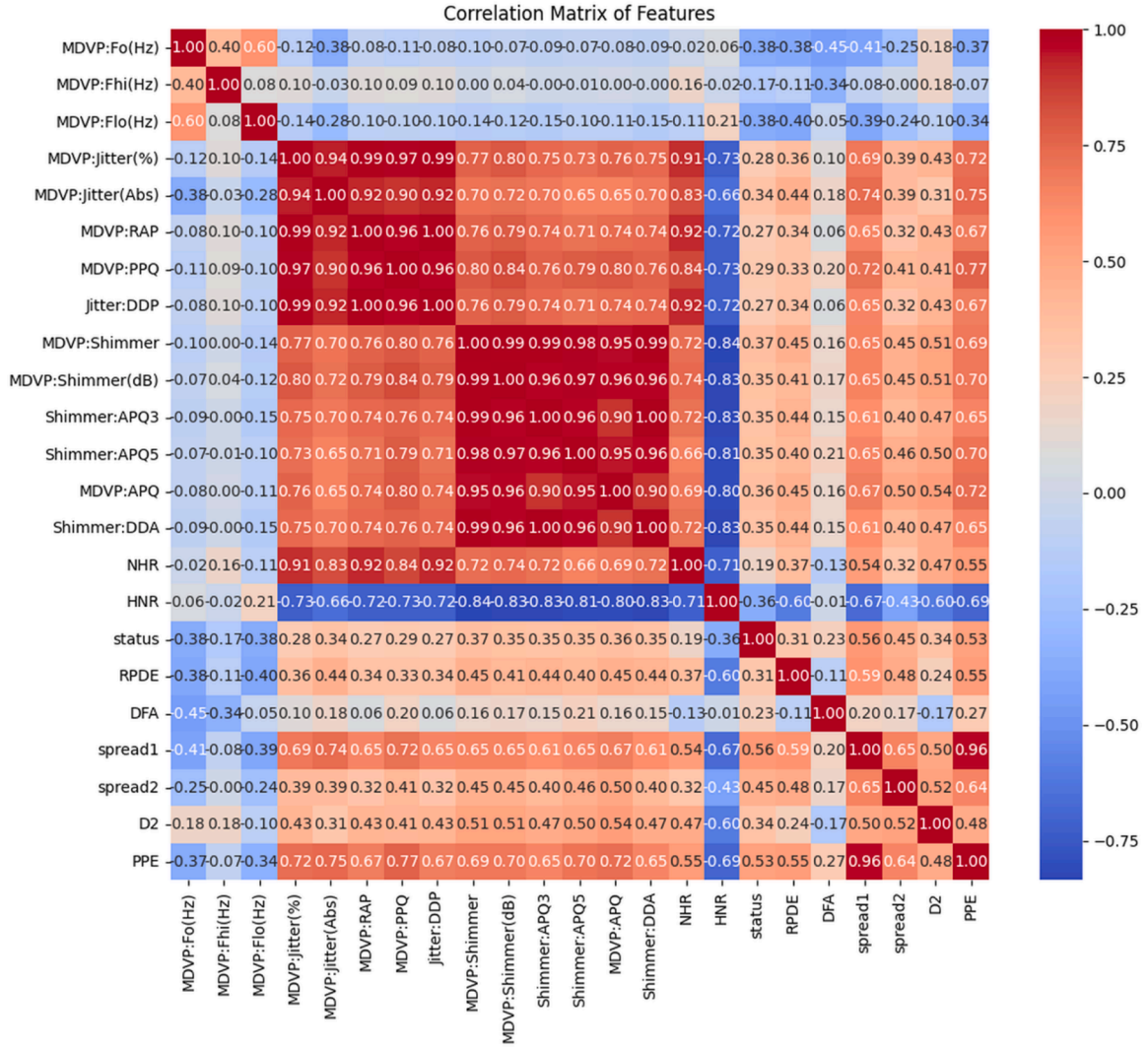
Fig. 3. The Correlation matrix.

3.4. Gated recurrent unit (GRU) network

The GRU-based model had a similar structure as LSTM architecture, but used a GRU layer instead of LSTM layer. It only had a GRU layer with 50 hidden units followed by a dense layer with sigmoid as an activation function. All recurrent models were trained with the same configuration (same optimizer, loss function and evaluation metrics) to maintain consistency.

3.5. Autoencoder for feature extraction

To learn compressed representations of the input voice features, an unsupervised autoencoder was built. An input layer, a 14-neuron encoding layer (dimensionality reduction), and a decoding layer that tried to reconstruct the input were all included in the Autoencoder model. The encoded representations were extracted following the Autoencoder's training on the training set with Mean Squared Error (MSE) loss.

The encoded test data was used to train a Logistic Regression classifier to accomplish Parkinson's classification. After that, performance was assessed using common classification metrics.

3.6. Generative adversarial network (GAN)

In this work we employ a GAN that is composed by two neural networks: a Generator and a Discriminator, which are trained against each other. The Generator takes as input a 100-dimensional random noise vector, takes it through two fully connected layers (with 64 and 128 neurons with ReLU activation) and finally reshapes the output to the original voice feature vector. The output adopts the tanh activation function to produce synthetic feature vectors with normalized entries.

The Discriminator is a binary classifier that takes a 21-dimensional input vector and processes it through two dense layers, 128 and 64 neurons, both with ReLU activation, and a final output neuron with sigmoid activation function to differentiate between real and generated samples.

The models are trained adversarially in a loop for 10 epochs with a batch size equal to 32. In each iteration:

The Generator takes as input noise and generate fake samples.

The Discriminator is trained over a mixed batch of real and fake data, using binary cross-entropy loss with label noise (eg, a small amount of random noise is added to the labels for better generalisation).

The Generator is in turn updated so that it is effective at convincing the Discriminator that the generated samples came from the data.

Both networks are optimized using Adam with a learning rate of 1e-4. Once trained, the discriminator is used here as a classifier in its own

right to classify the samples in the test set between Parkinson's and healthy. Its performance is measured by accuracy, precision, recall and confusion matrix representation.

3.7. Experimental environment

The experiments were run in Python using TensorFlow, Keras, and Scikit-learn. Matplotlib and Seaborn were used for data visualisation. Experiments were run on a GPU-accelerated machine to minimize training time and increase computing efficiency.

4. RESULTS AND discussion

The classification performances were visualized in the form of heatmaps for confusion matrices and rich metric overviews. These visualizations facilitated identification of misclassification trends, and each model's robustness. The performance of the proposed study is compared and analysed as shown in (Fig. 4), the results for each model clearly reflect the efficiency, with Multilayer Perceptron (MLP) attaining the highest accuracy (97.4 %). In the below (Fig. 5): The MLP model's classification performance yielded 32 true positives, 6 true negatives, 1 false positive, and 0 false negatives [40]. This demonstrates the strong prediction of the model with minimum misclassification errors. The performance of the Recurrent Neural (RNN-LSTM) model is given in (Fig. 6), where the model achieved 30 true positives and 4 true negatives. But it had 3 false positives and 2 false negatives as well, less accurate than the MLP model. Likewise, (Fig. 7) shows the classification results of the Gated Recurrent Unit (GRU) model that achieved 30 true positives and 4 true negatives. It had 3 false positives and 2 false negatives, showing performance similar to the RNN-LSTM model. The classification results of the Autoencoder model, as presented in the (Fig. 8), are composed of 31 true positives, 3 true negatives, 4 false positives and 1 false negative. However, as a large number of false-positive predictions are present, it indicates that it is probably incorrect for predicting certain instances. Lastly, (Fig. 9) shows the accuracy of the GAN model, which had 30 true positive cases but no true negatives. It had 7 False Positives and 2 False Negatives, showing that the model detects all true positive cases, but has lower specificity due to
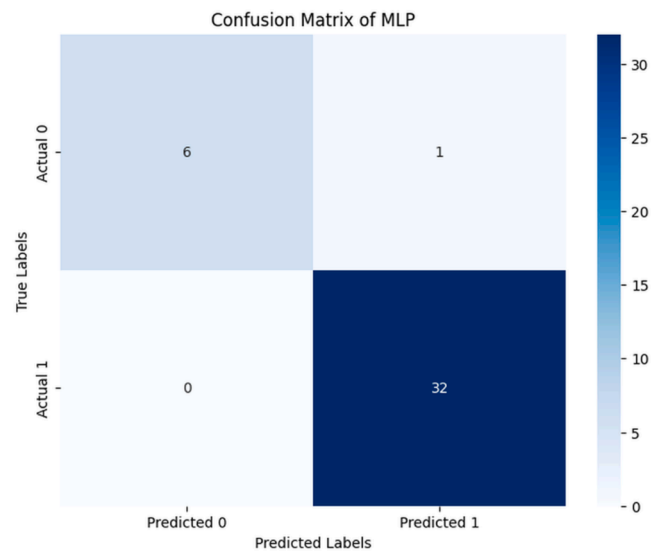

Fig. 5. MLP Confusion Matrix.

no true negative classifications.

Though MLP turned out to be a promising performer, however, the performances of RNN, GRU, Autoencoder, as well as GAN models, were simplistic or less in classification. This could be because the dataset is just not very friendly to architectures with strong temporal or generative structure like they prod and therefore is not taking advantages of some of those features.

Several factors may contribute to better results of the Multilayer Perceptron (MLP) model. First of all, in UCI voice dataset, it is tabularized (structured, non-sequential) features coming from vocal signal processing (jitter, shimmer and fundalmental frequency, etc.), which are better modeled by fully connected layers instead of RNN structures. MLPs are indeed quite effective at extracting patterns from such feature-rich inputs that are not fundamentally based on temporality. On the other hand, models like RNN/LSTM/GRU are specifically for sequential time series data, and they may not perform better when the data has
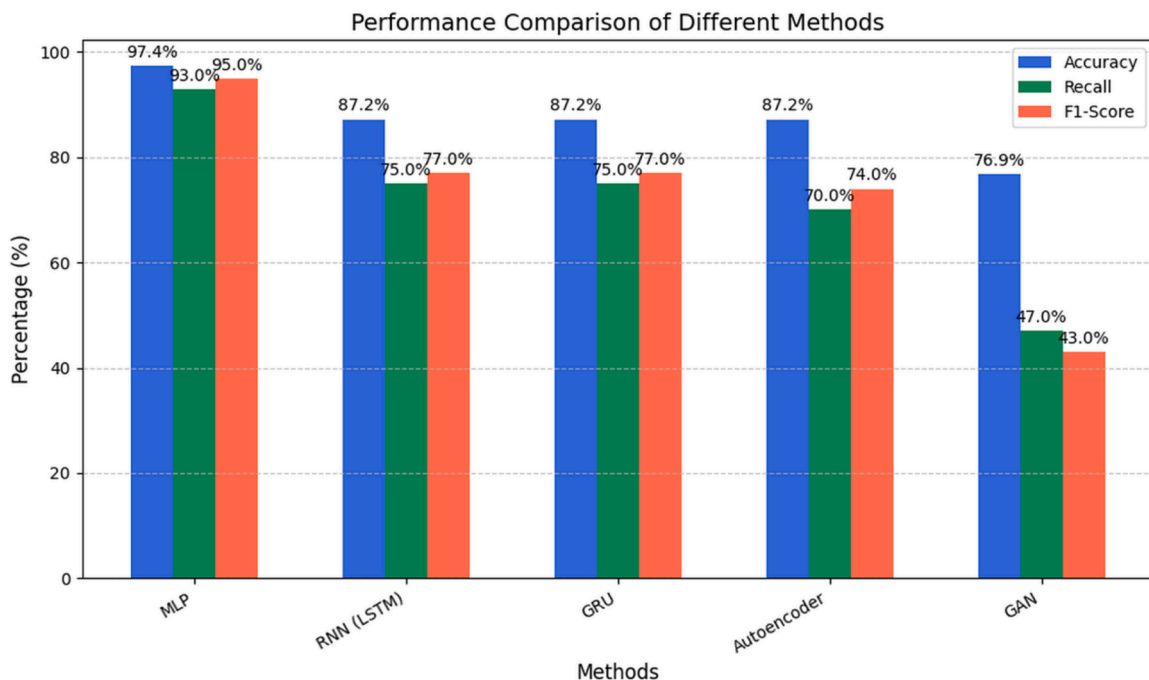

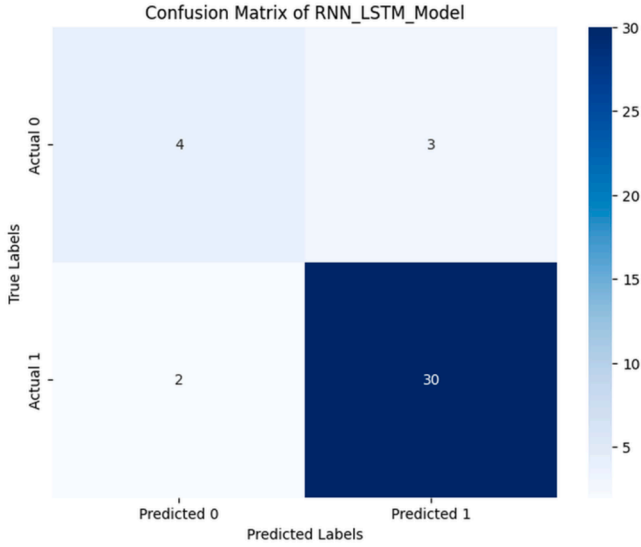**Fig. 4.** Performance comparison of different methods.
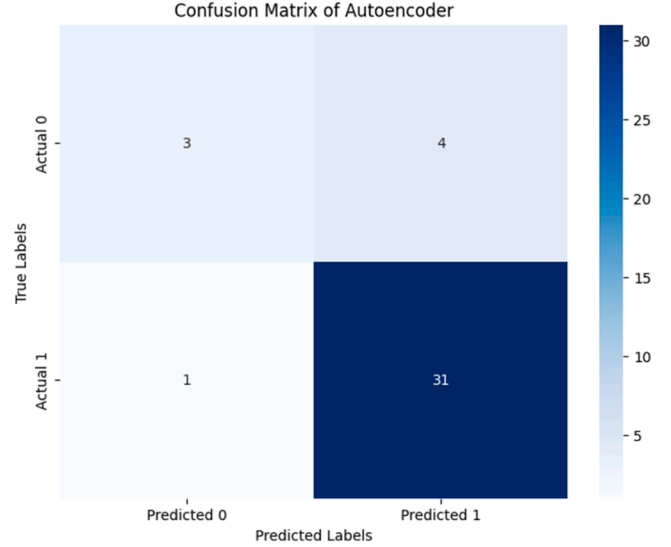
Fig. 6. RNN_LSTM Confusion Matrix.

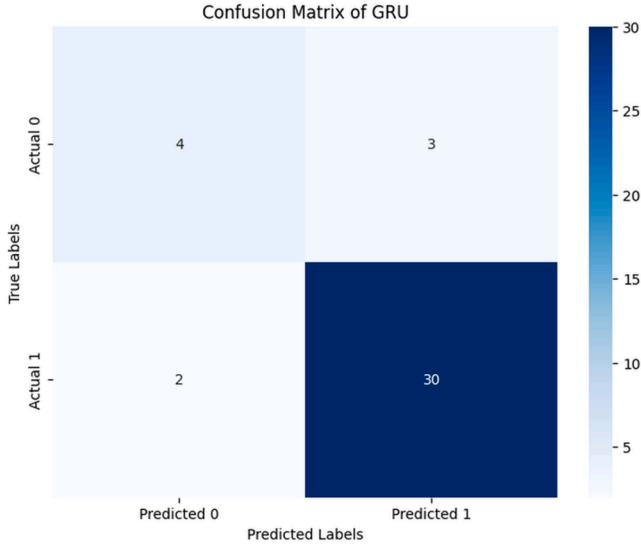

Fig. 8. Autoencoder Confusion Matrix.
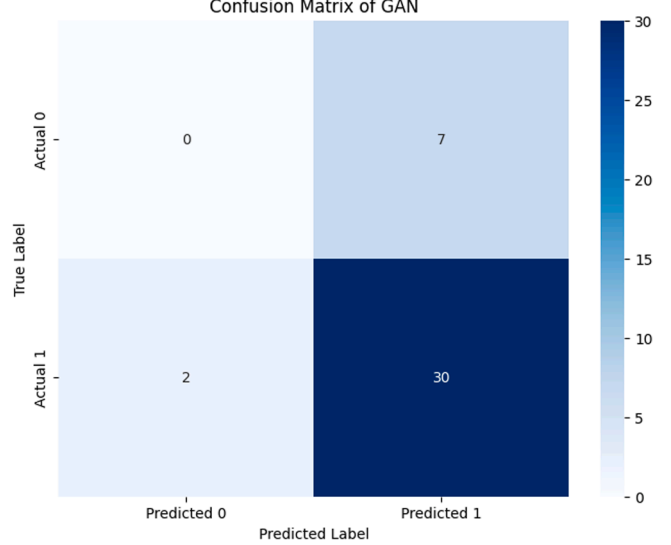

Fig. 7. GRU Confusion Matrix.

Fig. 9. GAN Confusion Matrix.

mostly independent observations. Also, MLP's simple structure is less prone to overfitting, especially on smaller data sets, and can achieve faster convergence during training. On the other hand, deeper models such as GANs and Autoencoders may require larger datasets or regularization techniques to achieve similar generalization.

While the voice features employed in this study are not structured as sequential time frames, we have also included RNN, and GRU based models to evaluate the possibility of them capturing any intrinsic or hidden temporal dependencies that may be present in the speech patterns. Such architectures are usually beneficial for modeling time-series or sequential data, however trying them here provided us with an opportunity to evaluate their flexibility and capability in nonconventional domains. From the results; although they did not outperform MLP, they show that their incorporation gives a wider perception on how the different model types perform on voice datasets in a feature vector form as opposed to a sequence. This guides future researchers in making a choice of models according to the data structure and the modality.

4.1. Computational efficiency and resource usage

Besides classification accuracy, we also compared the computation time of different models, which are crucial for real clinical-like setting. All experiments were run using the Google Colab on an NVIDIA Tesla T4 and 12 GB of RAM. The MLP model needed around 28 s of training, which was surprising, since the architecture is simple. The Autoencoder took 45 s, better in addition of dimensionality.

On the other hand, sequential models, such as RNN-LSTM and GRU, were slow and each took about 2–3 min due to temporal voice data processing. The GAN model takes the most time to train (approximately 5 min) because of its dual-network structure and the adversarial training loop.

These findings illustrate that although complex models (e.g., GANs) could offer new insights, simpler architectures such as MLP offer significantly faster training times and may be more suitable for time-sensitive clinical environments where resources are limited.

Below (Table 3) gives an overview of the estimated training times and hardware used and the compute requirements of all the deep learning models we employed in this study. These observations assist in

Table 3

Training Time and Computational Resource Comparison for All Deep Learning Models used in the study.

Model	Training Time	Hardware	Remarks
MLP	28 s	GPU (Tesla T4)	Fastest; clinically feasible
Autoencoder	45 s	GPU (Tesla T4)	Moderate time; simple design
RNN-LSTM	2 min	GPU (Tesla T4)	Requires sequence processing
GRU	2 min	GPU (Tesla T4)	Similar to LSTM
GAN	5 min	GPU (Tesla T4)	Longest; adversarial setup

assessing their applicability for clinical use.

4.2. Comparative analysis

In general, the MLP model outperformed other deep learning models attaining the best accuracy with the fewest classification errors. (Fig. 4) compares each of the five models. The importance of temporal modelling in speech-based PD detection is highlighted by the consistent superior performance of RNN_LSTM and GRU over the other three. Despite not being the greatest performers, autoencoder and GAN provided valuable insights into generative and feature extraction techniques.

4.3. Limitations and future work

There are some limitations that should be considered with this work although the results of this study show that deep learning models are effective in diagnosing PD based voice data. First, the study is restricted to the voice data of a unique modality, which may not represent all aspects of PD. Second, there are only 195 samples and no demographic diversity in the dataset that can limit the global generalization of our model. Third, some models (especially LSTM and GAN) demand more computational resources as well as training time, which probably restricts their use in the low-resource clinical setting. We will further investigate multimodal fusion, including handwriting, gait, imaging etc., and the scalability of the proposed models with other datasets of different scales and modalities in our future work.

Although RNN and GRU models are designed for time-series data, in this study, they were applied to non-sequential input vectors containing 21 voice features per sample. As such, their temporal modelling strengths were not fully utilized. This is recognized as a limitation, and future work will explore the use of time-dependent feature sequences to better align with the nature of RNN and GRU architectures.

The RNN and GRU architectures used in this study are generally designed for sequential or time-series data, yet the voice dataset in this study comprises statistical feature vectors rather than raw temporal sequences. Similarly, GANs are a leading example of a model that does well at generating images, or raw audio, but was only able to be applied here for generative tasks involving structured features. Our goal, however, was to evaluate all models under the same conditions to establish a benchmark comparison. In the future, it would be interesting to look into finding more compelling ways to manipulate the data so that each model's natural advantages might be exploited, as providing RNNs with raw audio features or using GANs for data augmentation.

5. Conclusion

This work, which uses five deep learning techniques, demonstrates encouraging advancements in the early identification and treatment of Parkinson's disease using voice data. Under this condition, the results indicate the performance varies among the deep learning architectures, the MLP model has the best result out of the models that were tried, with 97.4 % accuracy. These results underscore the potential for the use of deep learning models in clinical applications to facilitate early PD diagnosis. Future research should explore the transformer-based models dedicated to audio processing, such as Audio Spectrogram Transformer (AST), SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to imbue faith into AI-based diagnostic tools. They will aid to identify which vocal features contribute the most to classification, rendering the models more interpretable and more clinically relevant. Also to achieve better classification performance. This extend have the opportunity to evaluate temporal and contextual properties in more details and to compare their performance with the models discussed.

CRediT authorship contribution statement

Abdulaziz Salihu Aliero: Writing – original draft. **Neha Malhotra:** Supervision.

Declaration of competing interest

We declare that there are no conflicts of interest regarding the publication of this manuscript.

All authors have reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

[1] K.R. Chaudhuri, et al., Economic burden of Parkinson's Disease: a multinational, real-world, cost-of-illness study, Drugs Real World Outcomes 11 (1) (2024) 1–11, https://doi.org/10.1007/s40801-023-00410-1.

[2] F. Latifoğlu, S. Penekli, F. Orhanbulucu, M.E.H. Chowdhury, A novel approach for Parkinson's disease detection using Vold-Kalman order filtering and machine learning algorithms, Neural Comput. Appl. 36 (16) (2024) 9297–9311, https://doi.org/10.1007/s00521-024-09569-2.

[3] S. Gaba and H. Kaur, "Machine Learning Techniques for Parkinson's Disease Prediction and Progression: A Comprehensive Review," Proc. Int. Conf. Commun. Comput. Sci. Eng. IC3SE 2024, pp. 430–436, 2024, https://doi.org/10.1109/IC3SE62002.2024.10593626.

[4] S. Roy, T. Pal, S. Debbarma, A comparative analysis of advanced machine learning algorithms to diagnose Parkinson's disease, Procedia Comput. Sci. 235 (2023) (2024) 122–131, https://doi.org/10.1016/j.procs.2024.04.015.

[5] V. Reddy, "PPINtonus : Early Detection of Parkinson ' s Disease Using Deep-Learning Tonal Analysis," 2022.

[6] M. Martinez-Eguiluz, et al., Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies, Neural Comput. Appl. 35 (8) (2023) 5603–5617, https://doi.org/10.1007/s00521-022-07256-8.

[7] J. Zhang, Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease, npj Park. Dis. 8 (1) (2022), https://doi.org/10.1038/s41531-021-00266-8.

[8] A.S. Aliero, N. Malhotra, A survey on detection of Parkinson's disease through clinical data using deep learning approach, Proc. - 2024 IEEE 16th Int. Conf. Commun. Syst. Netw. Technol. CICN 2024 (2024) 173–177, https://doi.org/10.1109/CICN63059.2024.10847535.

[9] L. Li, F. Dai, S. He, H. Yu, H. Liu, Automatic diagnosis of parkinson's disease based on deep learning models and multimodal data, Deep Learn. Approaches Early Diagnosis Neurodegener. Dis. (2024) 179–200, https://doi.org/10.4018/979-8-3693-1281-0.ch009.

[10] M. Jyotiyana, N. Kesswani, M. Kumar, A deep learning approach for classification and diagnosis of Parkinson's disease, Soft Comput 26 (18) (2022) 9155–9165, https://doi.org/10.1007/s00500-022-07275-6.

[11] S.N.H. Bukhari, K.A. Ogudo, Ensemble machine learning approach for Parkinson's disease detection using speech signals, Mathematics 12 (10) (2024), https://doi.org/10.3390/math12101575.

[12] V. Skaramagkas, A. Pentari, Z. Kefalopoulou, M. Tsiknakis, Multi-modal deep learning diagnosis of Parkinson's Disease - A systematic review, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 2399–2423, https://doi.org/10.1109/TNSRE.2023.3277749.

[13] T. Aşuroğlu, H. Oğul, A deep learning approach for parkinson's disease severity assessment, Health Technol. (Berl). 12 (5) (2022) 943–953, https://doi.org/10.1007/s12553-022-00698-z.

[14] A.H. Al-Nefaie, T.H.H. Aldhyani, D. Koundal, Developing system-based voice features for detecting Parkinson's disease using machine learning algorithms, J. Disabil. Res. 3 (1) (2024) 1–10, https://doi.org/10.57197/jdr-2024-0001.

[15] F. Saeed, et al., Enhancing Parkinson's Disease prediction using machine learning and feature selection methods, Comput. Mater. Contin. 71 (2) (2022) 5639–5657, https://doi.org/10.32604/cmc.2022.023124.

[16] A. Mahmood, M. Mehroz Khan, M. Imran, O. Alhajlah, H. Dhahri, T. Karamat, End-to-End deep learning method for detection of invasive Parkinson's disease, Diagnostics 13 (6) (2023), https://doi.org/10.3390/diagnostics13061088.

[17] S. Singh, P. Sarote, N. Shingade, D. Yelale, N. Ranjan, Detection of Parkinson's disease using Machine learning algorithm, Int. J. Comput. Appl. 184 (6) (2022) 24–29, https://doi.org/10.5120/ijca2022922016.

[18] C.H. Lin, F.C. Wang, T.Y. Kuo, P.W. Huang, S.F. Chen, L.C. Fu, Early detection of Parkinson's disease by neural network models, IEEE Access 10 (2022) 19033–19044, https://doi.org/10.1109/ACCESS.2022.3150774.

[19] N. Chintalapudi, G. Battineni, M.A. Hossain, F. Amenta, Cascaded deep learning frameworks in contribution to the detection of Parkinson's disease, Bioengineering 9 (3) (2022), https://doi.org/10.3390/bioengineering9030116.

[20] A. Kurmi, S. Biswas, S. Sen, A. Sinitca, D. Kaplun, R. Sarkar, An ensemble of CNN models for Parkinson's Disease detection using DaTscan images, Diagnostics 12 (5) (2022) 1–18, https://doi.org/10.3390/diagnostics12051173.

[21] R. Alshammri, G. Alharbi, E. Alharbi, I. Almubark, Machine learning approaches to identify Parkinson's disease using voice signal features, Front. Artif. Intell. 6 (2023), https://doi.org/10.3389/frai.2023.1084001.

[22] S. Rahman, M. Hasan, A.K. Sarkar, F. Khan, Classification of Parkinson's disease using speech signal with Machine Learning and Deep learning approaches, Eur. J. Electr. Eng. Comput. Sci. 7 (2) (2023) 20–27, https://doi.org/10.24018/ejece.2023.7.2.488.

[23] A. Govindu, S. Palwe, Early detection of Parkinson's disease using machine learning, Procedia Comput. Sci. 218 (2022) (2022) 249–261, https://doi.org/10.1016/j.procs.2023.01.007.

[24] A. Rehman, T. Saba, M. Mujahid, F.S. Alamri, N. ElHakim, Parkinson's disease detection using hybrid LSTM-GRU deep learning model, Electron. 12 (13) (2023) 1–21, https://doi.org/10.3390/electronics12132856.

[25] C. Taleb, L. Likforman-Sulem, C. Mokbel, M. Khachab, Detection of Parkinson's disease from handwriting using deep learning: a comparative study, Evol. Intell. 16 (6) (2023) 1813–1824, https://doi.org/10.1007/s12065-020-00470-0.

[26] R.J. Sugden, P. Diamandis, Generalizable electroencephalographic classification of Parkinson's disease using deep learning, Informatics Med. Unlocked 42 (2023) 101352, https://doi.org/10.1016/j.imu.2023.101352 no. July.

[27] B. Majhi, et al., An improved method for diagnosis of Parkinson's disease using deep learning models enhanced with metaheuristic algorithm, BMC Med. Imaging 24 (1) (2024) 1–20, https://doi.org/10.1186/s12880-024-01335-z.

[28] M.A. Islam, M.Z. Hasan Majumder, M.A. Hussein, K.M. Hossain, M.S. Miah, A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets, Heliyon 10 (3) (2024) e25469, https://doi.org/10.1016/j.heliyon.2024.e25469.

[29] V. Rajinikanth, S. Yassine, S.A. Bukhari, Hand-sketchs based Parkinson's disease screening using Lightweight Deep-learning with two-fold training and fused optimal features, Int. J. Math. Stat. Comput. Sci. 2 (2023) 9–18, https://doi.org/10.59543/ijmscs.v2i.7821, no. Ml.

[30] A. Keles, A. Keles, M.B. Keles, A. Okatan, PARNet: deep neural network for the diagnosis of parkinson's disease, Multimed. Tools Appl. 83 (12) (2024) 35781–35793, https://doi.org/10.1007/s11042-023-16940-3.

[31] R.M. Al-Tam, F.A. Hashim, S. Maqsood, L. Abualigah, R.M. Alwhaibi, Enhancing Parkinson's Disease diagnosis through stacking ensemble-based machine learning approach, IEEE Access 12 (2024) 79549–79567, https://doi.org/10.1109/ACCESS.2024.3408680, no. June.

[32] İ. Cantürk, O. Günay, Investigation of scalograms with a deep feature fusion approach for detection of Parkinson's disease, Cognit. Comput. 16 (3) (2024) 1198–1209, https://doi.org/10.1007/s12559-024-10254-8.

[33] H. Chen, et al., An automated hybrid approach via deep learning and radiomics focused on the midbrain and substantia nigra to detect early-stage Parkinson's disease, Front. Aging Neurosci. 16 (2024), https://doi.org/10.3389/fnagi.2024.1397896 no. May.

[34] S. Srinivasan, P. Ramadass, S.K. Mathivanan, K. Panneer Selvam, B.D. Shivahare, M.A. Shah, Detection of Parkinson disease using multiclass machine learning approach, Sci. Rep. 14 (1) (2024) 1–17, https://doi.org/10.1038/s41598-024-64004-9.

[35] Md Abu Sayed, et al., Parkinson's disease detection through vocal biomarkers and advanced machine learning algorithms, J. Comput. Sci. Technol. Stud. 5 (4) (2023) 142–149, https://doi.org/10.32996/jcsts.2023.5.4.14.

[36] N. Islam, M.S.A. Turza, S.I. Fahim, R.M. Rahman, Advanced Parkinson's Disease Detection: a comprehensive artificial intelligence approach utilizing clinical assessment and neuroimaging samples, Int. J. Cogn. Comput. Eng. 5 (2024) 199–220, https://doi.org/10.1016/j.ijcce.2024.05.001, no. November 2023.

[37] L. Ali, A. Javeed, A. Noor, H.T. Rauf, S. Kadry, A.H. Gandomi, Parkinson's disease detection based on features refinement through L1 regularized SVM and deep neural network, Sci. Rep. 14 (1) (2024) 1–14, https://doi.org/10.1038/s41598-024-51600-y.

[38] M. Ianculescu, C. Petean, V. Sandulescu, A. Alexandru, A.M. Vasilevschi, Early detection of Parkinson's disease using AI techniques and image analysis, Diagnostics 14 (23) (2024), https://doi.org/10.3390/diagnostics14232615.

[39] A. Alotaibi, Ensemble deep learning approaches in health care: a review, Comput. Mater. Contin. 82 (3) (2025) 3741–3771, https://doi.org/10.32604/cmc.2025.061998.

[40] Y. Wang, et al., An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1-weighted images, Hum. Brain Mapp. 44 (12) (2023) 4426–4438, https://doi.org/10.1002/hbm.26399.

Full Length Article

Evaluating public bicycle sharing system in Ahmedabad, Gujarat: A multi-criteria decision-making approach☆

T.S. Shagufta [a], Dimpu Byalal Chindappa [a], Seelam Srikanth [a,*], Subhashish Dey [b]

[a] School of Civil Engineering, REVA University, Bengaluru-560064, Karnataka, India
[b] Civil Engineering Department, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India

ABSTRACT

This study evaluates the existing Public Bicycle Sharing System (PBSS) at Ahmedabad, Gujarat by applying four decision-making methods such as Analytic Hierarchy Process (AHP), Fuzzy AHP, Analytic Network Process (ANP), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The study aims to identify the most effective strategies for improving PBSS, focusing on safety, infrastructure, user convenience, and environmental impact. The analysis shows that Enhanced Non-Motorized Transport (NMT) Infrastructure and Expansion of Bicycle Networks are the preferred alternatives across all methods. Personal safety and safe cycling infrastructure are identified as critical factors influencing the success of PBSS. Socio-demographic data reveals a male-dominant user base, with financial barriers and safety concerns limiting broader adoption. Positive perceptions of cycle design are noted, though electric and hybrid cycles are preferred due to climatic conditions. Monthly variations in ridership demonstrate significant fluctuations, peaking at 68,529 rides in March, underscoring the need for targeted interventions during peak periods. The study provides a robust framework for transport planners, emphasizing safety, inclusivity, and affordability. Future research should focus on expanding electric cycle options and enhancing gender inclusivity in PBSS.

1. Introduction

Cycling has emerged as a sustainable, healthy, and cost-effective mode of transportation, gaining further prominence during the pandemic when motorized vehicles were less utilized. As cities strive to create more inclusive and sustainable transport environments, shifting from private vehicles to public transport becomes increasingly crucial. Public Bicycle Sharing Systems (PBSS) play a significant role in this transition, especially in metro cities where they are actively implemented. PBSS are flexible transportation services that allow users to rent bicycles for short distances [1,2]. These systems typically offer a range of bicycles, including pedal, geared, electric, and pedal-assist models, to cater to various needs and preferences. Rental options include hourly, daily, weekly, monthly, and annual subscriptions, with discounts and long-term leases available. PBSS aims to improve cycling infrastructure and address end-mile connectivity issues associated with public transportation [3]. In metropolitan areas, PBSS plays a vital role in enhancing urban mobility. It provides an economical and eco-friendly alternative

to motorized transport, contributing to reduced traffic congestion and lower emissions. By integrating these systems with existing public transportation networks, PBSS improves last-mile connectivity, making public transit more accessible and efficient.

Despite their benefits, PBSS faces several challenges. Transport agencies often view these systems as competitors to their core services, which can hinder their adoption and growth. Furthermore, increased awareness and education are needed to bridge the gap between PBSS and traditional public transport modes. Ensuring that these systems are inclusive and cater to diverse demographic groups including different genders, ages and physical abilities, is essential for maximizing their impact [4]. The concept of Public Bicycle Sharing (PBS) was first introduced in India in 2010 with the launch of the Pune Cycle Plan. Since its inception, PBS has significantly evolved and expanded across various Indian cities, including Delhi, Chennai, Bengaluru, Ahmedabad, Jaipur, Udaipur, Mumbai, Kochi, and Nagpur. Initially designed to promote cycling for recreational purposes, the service has grown to serve as a viable option for daily commutes as well. As urban planners

and government authorities have increasingly focused on sustainable and green mobility solutions, PBS has gained prominence as a tool for reducing traffic congestion and pollution. The system's role in promoting eco-friendly transport aligns with broader goals of enhancing urban sustainability.

One of the major advancements in PBS in India is the integration of technology. The introduction of mobile applications and GPS-Bluetooth-enabled bicycles has streamlined the user experience, making it more convenient and accessible [5]. In cities like Delhi and Bangalore, smart card systems have been implemented, facilitating cashless transactions and ensuring a smoother service experience for users. Another significant development is the integration of PBS with other transportation modes, such as metro rail and public buses. This integration has improved last-mile connectivity, allowing users to seamlessly transition between different modes of transport. It has also increased the utilization of bicycles for short commutes, further promoting the use of sustainable transport options. Despite these advancements, the growth of PBS in India has encountered several challenges. Issues such as inadequate infrastructure, lack of public awareness and education, safety concerns, and vandalism have impeded the system's effectiveness and sustainability. Addressing these challenges is crucial for the continued success and expansion of PBS in India [6].

MYBYK, a Public Bike Sharing System, commenced its operations in Ahmedabad from 2014. Over the years, MYBYK has expanded its presence to several major Indian cities, including Indore, Kochi, Mumbai, Udaipur, Nagpur, and Mysuru. The service provides an accessible option for renting bicycles for short distances, fostering eco-friendly transportation and encouraging healthier lifestyles among users. MYBYK has earned significant recognition for its contributions to sustainable urban mobility. Notably, it received the Stylish Sustainable Civic Mobility Award at the Urban Mobility India Conference in 2017. Additionally, in 2019, MYBYK was selected for the Shell Foundation's accelerator program, which supports promising startups focused on sustainable development. The company has forged strategic partnerships to enhance its visibility and reach. For example, it has collaborated with Indian Railways to offer bicycles for use at train stations, facilitating smoother transitions between rail and bike transport. MYBYK has also partnered with retail outlets, such as Decathlon, to promote bicycle use for both personal and commercial purposes. As of now, MYBYK operates extensively in Ahmedabad, where it has established 253 stations and maintains a fleet of 6000 bicycles. The service records an impressive average of 50,000 rides per month and has 1200 subscribers who benefit from the flexibility to take bicycles home and return or exchange them during their subscription period. MYBYK's operational footprint extends across six Indian cities—Ahmedabad, Indore, Mumbai, Udaipur, Nagpur, and Kochi—boasting over 10,000 bicycles and ≫500 PBS stations. With a user base of 750,000, the service has demonstrated substantial growth and acceptance. The company is poised to enter Mysuru in June 2023, further broadening its reach. MYBYK has modernized its operations by integrating Internet of Things (IoT) technology and pedal-assist bicycles, enhancing the overall user experience. The company's innovative approach includes updating the traditional Trin-Trin model to better suit contemporary needs.

Looking ahead, PBS service is anticipated to expand to Bengaluru by the end of 2023, with operations along the city's metro lines. This development will further contribute to the growing network of sustainable transportation options in India. This study evaluates existing PBSS at Ahmedabad, Gujarat by applying four decision-making methods such as Analytic Hierarchy Process (AHP), Fuzzy AHP, Analytic Network Process (ANP), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The alternatives assessed include Enhanced Non-Motorized Transport (NMT) Infrastructure, Expansion of Bicycle Network, Implementation of Slow Streets, Awareness and Education Campaigns, and Integration with Public Transport. The study aims to identify effective parameters that influence the use of cycling in urban areas, with a focus on making sustainable transportation accessible to

all. By analyzing factors that affect the success of PBSS, the study seeks to provide insights into where and how these systems can be more effectively implemented. This includes understanding user preferences, system design, and policy implications. Despite the growing adoption of Public Bicycle Sharing Systems (PBSS) in Indian cities, their effectiveness remains constrained by fragmented infrastructure, safety concerns, and socio-demographic disparities in usage. There is a lack of comprehensive, data-driven frameworks that evaluate improvement strategies from multiple decision-making perspectives. Addressing this gap, the present study aims to systematically evaluate PBSS in Ahmedabad using robust multi-criteria decision-making (MCDM) approaches. To assess the most impactful strategies for enhancing PBSS, this study applies four distinct MCDM techniques: AHP, Fuzzy AHP, ANP, and TOPSIS. Each method captures a different decision-making nuance—from handling uncertainty (Fuzzy AHP) to modeling interdependencies and performance proximity. A structured survey of 1000 respondents, secondary data from MYBYK and expert inputs form the basis of the evaluation.

Across all four decision-making models, Enhanced Non-Motorized Transport (NMT) Infrastructure consistently emerged as the preferred alternative, followed by Expansion of the Bicycle Network. Personal safety and safe infrastructure at junctions were identified as the most critical criteria. The results also highlight the underrepresentation of women and higher-income groups in PBSS usage, emphasizing the need for inclusive planning. This study contributes a comparative and integrated framework for evaluating urban mobility interventions, especially PBSS strategies, using multiple MCDM methods. It offers practical recommendations for planners and policymakers to prioritize infrastructure, safety, and affordability. It also proposes directions for future research, including gender-responsive PBSS design and the integration of electric bicycles for climate resilience. By combining socio-demographic insights with rigorous decision-analysis techniques, the study presents a replicable model for evaluating PBSS in other Indian and global contexts.

2. Study about the public bicycle sharing systems

Public Bicycle Sharing Systems (PBSS) play a vital role in urban mobility and sustainability. Several factors, including technological advancements, safety, user satisfaction, infrastructure design, and policy support, contribute to the success of PBSS. These factors align with specific criteria that influence user preferences, system efficiency, and overall system sustainability. Below, the literature is organized into thematic clusters and linked to the relevant evaluation criteria.

2.1. Technological advancements and system efficiency

The integration of advanced technology is essential for enhancing the efficiency and user experience of Public Bicycle Sharing Systems (PBSS). Studies have highlighted the transformative role of smart technologies such as real-time data analytics, smart locks, and mobile applications in improving operational efficiency and user satisfaction. Munkácsy & Monzón (2017) and Bieliński & Ważna (2018) underscore the importance of these technologies in optimizing the availability of bikes, streamlining the rental process, and reducing operational costs. These innovations allow for better tracking of bikes, improve the user interface, and increase system reliability by offering features like user feedback, bike maintenance requests, and efficient bike deployment. Moreover, the advent of electric bikes and docking stations is expanding the scope and accessibility of PBSS, making them adaptable to different urban mobility needs [7,8]. Shaheen et al. (2013) argue that these technologies not only enhance operational performance but also ensure user convenience and satisfaction by offering a seamless and interactive experience [9]. Moreover, Teixeira et al. (2023) highlighted the importance of contactless payment systems and smart tracking, particularly considering the COVID-19 pandemic, as these systems address safety concerns and improve user confidence in the services [10].

Researchers have also discussed how data analytics and smart technologies were leveraged to optimize bike-sharing operations and respond to changing user demands. O'Mahony and Shmoys (2015) explored the application of real-time data analytics to predict user demand patterns, allowing for efficient redistribution of bikes and optimization of station capacities [11]. Similarly, Freund et al. (2019) focused on using machine learning algorithms to forecast bike availability and manage fleet resources dynamically, improving operational efficiency and ensuring user satisfaction by minimizing bike shortages and overages [12]. Technological innovations affect separate cycle lanes by ensuring real-time bike availability, reducing congestion on cycling paths. System efficiency is also improved as smart bikes and systems help in managing bike fleet distribution and predicting demand, ensuring fewer bikes are left stranded. The integration of technology can also improve Personal safety through GPS tracking, helping to monitor and ensure their security.

2.2. Health, safety, and crisis resilience

The COVID-19 pandemic presented unprecedented challenges for PBSS, especially in terms of user safety and system sustainability. During the pandemic, concerns about hygiene and social distancing led to a greater demand for contactless and hygienic bike-sharing options. Raza et al. (2018) found that increased cycling benefits public health by improving physical fitness and reducing air pollution, a message that became even more pronounced during the pandemic when users prioritized safety [13]. Julio and Monzón (2022) examined the long-term impacts of the pandemic on PBSS, stressing the need for proactive strategies to ensure their sustainability during future crises. These strategies include introducing flexible pricing, ensuring hygiene management, and integrating technological solutions to foster resilience. The ongoing adaptation of PBSS in response to external shocks has proven to be essential for maintaining user confidence and ensuring system longevity [14]. Murat and Cakici (2024) explored how the pandemic shifted preferences towards systems that allowed minimal contact, highlighting the importance of hygiene and safety measures in user satisfaction [15]. The importance of Personal safety during pandemics and crisis situations is critical. Hygiene and contactless systems are directly linked to ensuring user safety. Clean air to breathe is also promoted as cycling in cities reduces vehicular emissions and encourages active transportation, contributing to better air quality.

2.3. Infrastructure and urban design

The physical infrastructure and design of bike-sharing systems play a critical role in influencing the effectiveness and accessibility of these systems. Research consistently shows that well-maintained bikes, strategically located stations, and easy access points significantly enhance user satisfaction and system utilization. Macioszek et al. (2020) and Guo et al. (2017) emphasize that users tend to prefer systems with well-distributed bike stations and clear access points, as this reduces barriers to use and ensures the bikes are readily available where they are needed most. In addition to station placement, the integration of PBSS with broader urban infrastructure is crucial [16,17]. Bencekri et al. (2024) and Mix et al. (2022) argue that effective integration with public transport networks, pedestrian areas, and bike lanes encourages greater use by ensuring seamless transfers between different modes of transport [18,19]. Furthermore, Wang et al. (2023) found that urban design elements, such as bike lanes and pedestrian-friendly streets, can increase bike-sharing usage by making it safer and more convenient for users to access and use the system [20]. This infrastructure directly influences several criteria, including separate cycle lanes are essential for safe cycling, as users prefer dedicated lanes that minimize risks from mixed traffic. Marked accident-prone areas should be addressed by city planners to ensure the safety of cyclists. Well-maintained bike lanes and clear traffic management can reduce the risk of accidents. Less traffic improves cycling safety and the attractiveness of bike-sharing systems, as lower traffic volumes reduce the likelihood of collisions. Proper urban infrastructure ensures reduced vehicle speed, which directly contributes to cycling safety.

2.4. User satisfaction and demographics

Understanding user satisfaction and demographics is essential for designing a bike-sharing system that appeals to a wide range of users. Research has shown that factors such as age, income, and education significantly impact the likelihood of adopting bike-sharing systems. Authors emphasize the need to tailor bike-sharing systems to the specific needs of different demographic groups. Younger, more active individuals, for example, may be more inclined to use bike-sharing systems, while older or less physically active users may require additional features such as e-bikes for ease of use. User demographics and experiences are integral to enhancing user satisfaction and improving overall system adoption [21–25]. Safe path at junctions/signals is crucial in ensuring that users feel secure using PBSS, particularly in high-traffic areas where bike lanes and proper traffic management are necessary. Comfortable attire/dress is a factor that influences user participation, especially if bike-sharing systems provide racks and facilities for users to change attire.

2.5. Economic and environmental impacts

Bike-sharing systems have profound economic and environmental implications for urban areas. Several studies have highlighted the positive impacts on urban sustainability. Bernardo (2022) and Cheng et al. (2022) provide evidence that bike-sharing systems reduce traffic congestion and lower greenhouse gas emissions, contributing to cleaner air and a healthier urban environment [26,27]. On the economic front, DeMaio (2009) and Pelechrinis et al. (2017) emphasize that bike-sharing systems can stimulate local economies by increasing foot traffic in commercial areas, creating new job opportunities, and offering users a low-cost transportation option. Moreover, economic evaluations consistently show that PBSS are cost-effective compared to other forms of transportation, which can have a significant impact on urban transportation budgets [28,29]. The economic and environmental benefits of bike-sharing align directly with the criterion of clean air to breathe, as these systems promote a healthier environment by reducing car emissions. They also support good physical and mental health, by encouraging physical activity and reducing air pollution.

2.6. Public policies and regulations

The success and sustainability of bike-sharing systems are heavily influenced by public policies and urban regulations. Authors emphasize that policies providing funding, subsidies, and investments in cycling infrastructure are vital for the development and long-term viability of PBSS. Supportive public policies not only provide the financial backing needed to develop and maintain bike-sharing systems but also ensure that they are integrated into the broader urban mobility system [30–32]. This integration can include efforts to improve safety, accessibility, and equitable distribution of services, which are essential for making bike-sharing systems successful in diverse urban contexts. Regulations that ensure the safety of cyclists and address issues such as weather resistance and infrastructure quality are also essential for maintaining system reliability and user confidence. Public policies and regulations have a direct influence on the safety and accessibility of bike-sharing systems. Strong policies also enhance the system's comfort and convenience, ensuring that the infrastructure is safe, well-maintained, and accessible to all users.

2.7. Challenges and adaptation

Despite the many advantages of bike-sharing systems, they face several challenges that require ongoing innovation and adaptation. Bean et al. (2021) and Lee et al. (2016) highlight the impact of weather conditions on bike-sharing usage, suggesting that infrastructure solutions such as weather-resistant bikes and sheltered stations are needed to ensure system functionality year-round [33,34]. Suchanek (2019) notes that the system's ability to adapt to emerging trends, such as the pandemic-induced shift toward social distancing, is crucial for ensuring its continued success [35] Adapting the service to changing user needs and circumstances, such as offering contactless features and flexible rental options, ensures that users remain confident in the system's safety and reliability. Addressing challenges like weather and pandemics through adaptive service models enhances user safety and system reliability. It ensures that the system can operate effectively under a variety of conditions, which is essential for long-term success. The review of recent research on public bicycle-sharing systems highlights the multifaceted factors influencing their success. Technological advancements, user preferences, infrastructure design, environmental and economic impacts, and supportive policies are all critical components. Addressing these factors through thoughtful design, policy support, and technological innovation can enhance the effectiveness of bike-sharing systems and contribute to more sustainable and accessible urban mobility solutions.

3. Study area and data collection

Selecting the study area where PBSS operational is crucial for the research. This evaluation will help propose PBSS models for other Indian cities like Bengaluru. Ahmedabad, with its flat terrain, is ideal for cycling, making it a strong market for PBSS. As the fifth-largest city in India, Ahmedabad is known for its robust public transportation system,

including an integrated Bus Rapid Transit System (BRTS) and PBSS. The BRTS features dedicated lanes, segregated bus stops, and automated fare collection for efficient transit, while PBSS provides last-mile connectivity with rental bicycles for short trips. Based on a thorough literature review, the study considers factors such as separate cycle lanes, safe cycle parking, clean air, humidity and temperature, on-street vehicle parking, traffic levels, vehicle speed, accident-prone areas, safety at junctions, flat roads, carrier baskets, comfortable attire, good health, and personal safety. User perceptions of these factors were collected through a questionnaire survey rated on a scale of 1 to 5. The study applied Cochran's formula to calculate the minimum sample size, which was determined to be 384 respondents with a 5 % margin of error, ensuring statistical accuracy. Additionally, subscription data from MYBYK was used to guarantee a representative sample that reflected various user demographics and station locations. Primary data was collected in Ahmedabad with a sample size of 1100. The survey aimed to understand perceptions of PBSS at locations near BRTS and MYBYK stations as shown in Fig. 1. Respondents, including PBSS and BRTS users, rated the existing PBSS, factors influencing cycling, PBSS operations, and bicycle quality and maintenance. Interviews were conducted during peak hours at PBSS stations and BRTS stops. This study also utilizes secondary data collected from MYBYK operational records from the year 2023, which reported a total of 577,228 trips. The data provides insights into ridership patterns, trip durations, and distance traveled, enabling a comprehensive analysis of the PBSS's performance and effectiveness in promoting sustainable urban transport in Ahmedabad. Geographic Information System (GIS) techniques were employed to delineate the zones, facilitating a better understanding of spatial distribution and accessibility of the bicycle-sharing stations.

4. Analysis of user perceptions

The survey data reveals significant insights into the socio-
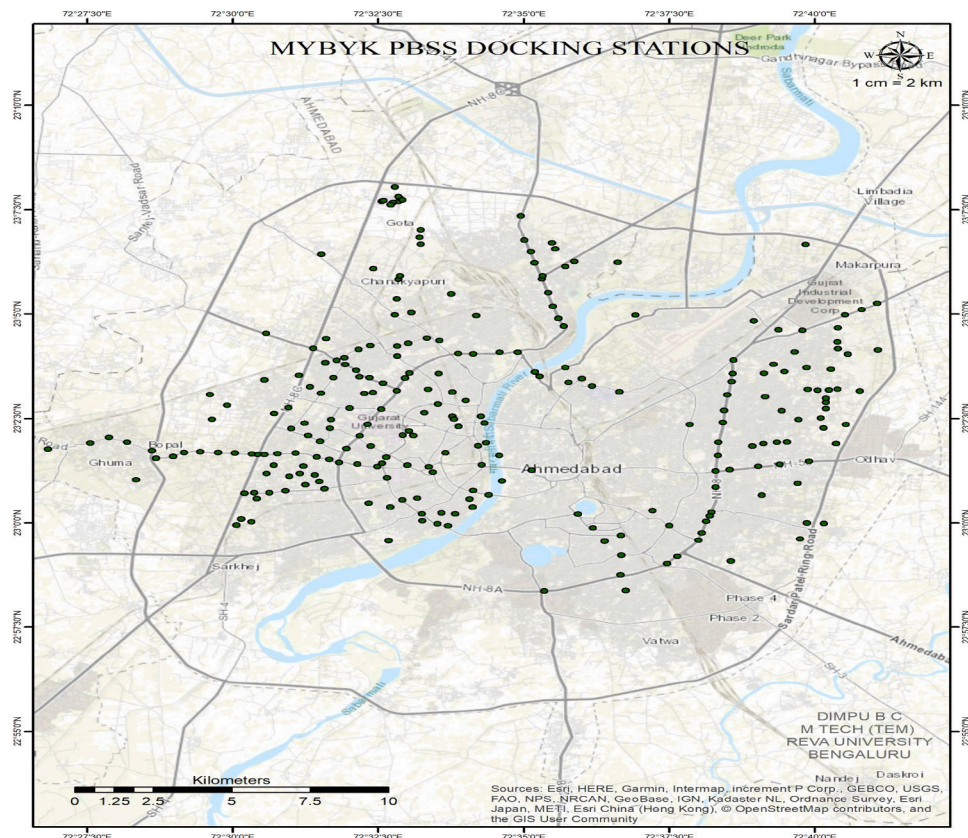

Fig. 1. MYBYK stations in Ahmadabad.

demographic profile of PBSS users in Ahmedabad as shown in Fig. 2. The data shows a strong male dominance in PBSS usage, with 82.47 % of respondents being male. This significant gender disparity suggests that men are more inclined to use cycling as a mode of transportation, likely due to factors such as physical strength, societal norms, and attire. Women's lower participation (17.53 %) may be attributed to concerns about safety, comfort, and the challenges posed by the city's temperature and humidity. The age distribution reveals that the majority of PBSS users are between 20 - 35 years old (57.53 %), followed by those under 20 years (18.08 %) and between 36 - 45 years (14.24 %). This indicates that the younger individuals are more likely to adopt cycling, possibly due to greater health consciousness, environmental awareness, and the need for economical transportation options. A substantial portion of the

respondents are students (40.20 %) and private sector employees (30.34 %), reflecting that PBSS is primarily used by those with flexible schedules or a need for cost-effective transport options. The high percentage of unmarried individuals (64.11 %) further understands the use of PBSS to younger populations. Income-wise, the largest group of income falls between ₹10,001-₹25,000 per month (22.03 %), indicating that PBSS is particularly attractive to lower and middle-income groups seeking cost-effective transport. Overall, the data suggests that while PBSS is popular among younger, male, and financially conscious individuals, there is an opportunity to increase adoption among women, older age groups, and higher-income individuals through targeted initiatives that address specific barriers such as safety, comfort, and convenience.

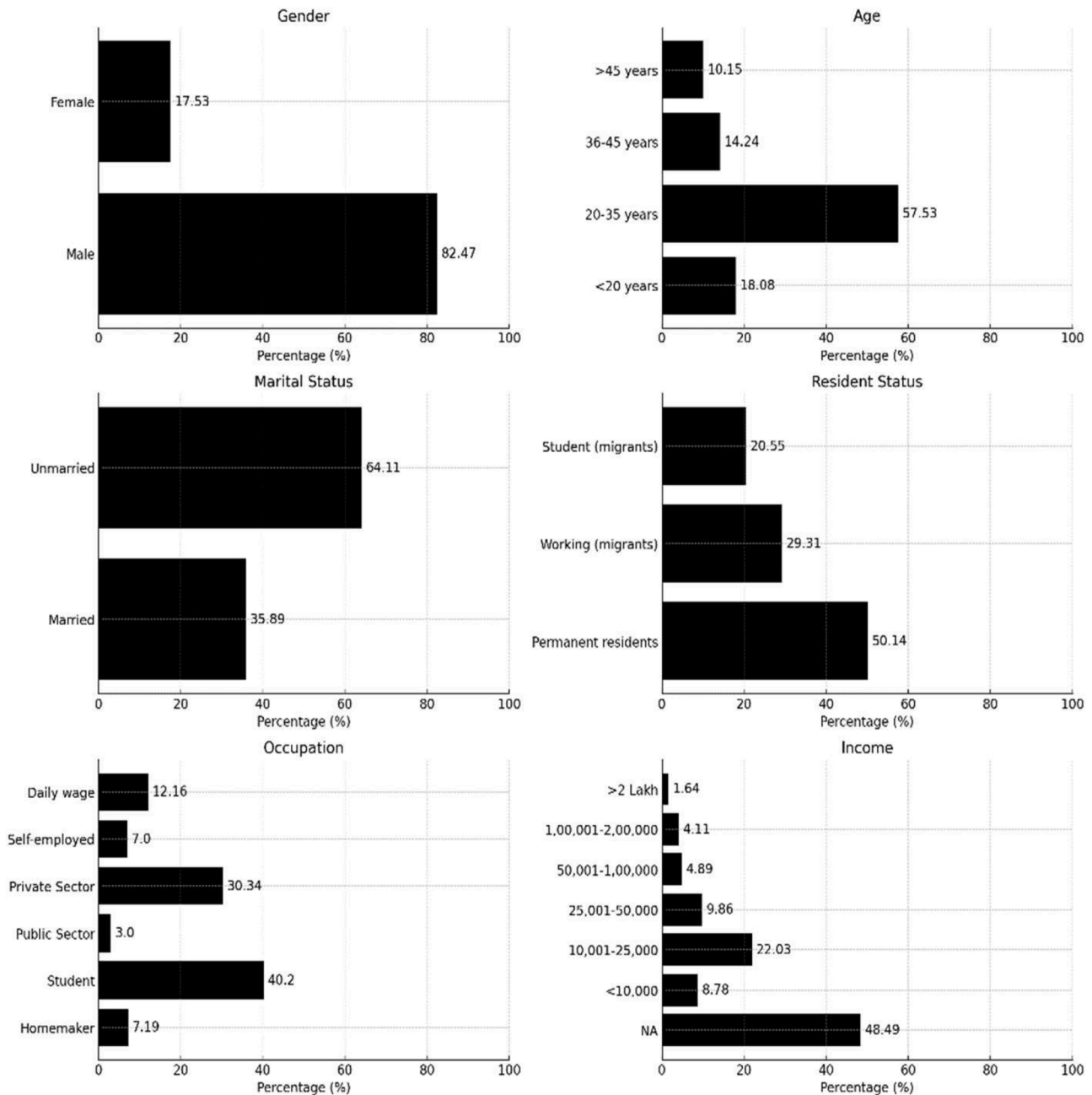The PBSS in Ahmedabad, initially aimed at providing a convenient

Fig. 2. Socio-demographic details of respondents.

solution for last-mile travel, is predominantly used for leisure (74.41 %) and shopping (64.65 %) as shown in Fig. 3. A notable portion of respondents also use the service for work trips (56.90 %) and educational purposes (40.40 %). While last-mile travel currently accounts for 46.46 % of usage, there has been a recent increase in this segment, driven by incentives, indicating growing acceptance of PBS as a practical commuting option in the city.The PBSS in Ahmedabad, while cost-effective and efficient, faces some financial accessibility challenges, with only 42 % of the respondents feeling comfortable paying the security deposit and ride charges as shown in Fig. 4. A significant portion of users expressed moderate comfort with these costs, indicating a need for more affordability. However, the cycle design, quality, and maintenance were generally well-received, with the majority rating them as good or very good. Only a small percentage of respondents found these aspects lacking, suggesting that the operational side of the system performs well, but financial concerns may be a barrier for broader adoption. The survey reveals that over 80 % of respondents are satisfied with using the Public Bicycle Sharing system for distances greater than 2 km, primarily for leisure purposes as shown in Fig. 5. A strong preference for a carrier basket is evident, with 74 % of respondents favoring it as a necessary fixture. Despite Ahmedabad's flat terrain, 45.6 % of users prefer electric or hybrid cycles, likely due to the city's high temperatures and humidity, which can make cycling less appealing. This suggests that the integrating more electric or hybrid cycles could enhance user comfort and satisfaction with the PBS system.

The user distribution by hour for the PBSS in Ahmedabad reveals distinct patterns in ridership throughout the day as shown in Fig. 6. The data indicates that usage peaks during the early morning hours, with significant activity starting at 5 am and reaching a maximum of 66,415 rides at 6 am, followed by a gradual decline until 10 am. The morning peak captures 29.55 % of total rides between 6 am and 9 am, reflecting a high demand for last-mile connectivity as users head to work or educational institutions. Conversely, ridership diminishes during midday hours, with only 6.72 % of total rides occurring between 12 pm and 4 pm. The afternoon sees a resurgence in usage, particularly between 4 pm and 8 pm, where users engage in leisure activities or return home, culminating in an average of 30,000 rides per hour during this period. The data further indicates a noticeable drop in ridership during the late-night hours, with minimal activity recorded between 12 am and 5 am. Overall, the analysis highlights a clear preference for using the PBSS during the cooler morning and evening hours, suggesting that climatic conditions significantly influence user behavior.

The monthly ridership trend for the PBSS in Ahmedabad as shown in Fig. 7. illustrate fluctuations in usage throughout the year, with a total of 576,261 rides recorded. The trend reveals that the highest number of rides occurred in March, with 68,529 rides, followed closely by April at 64,737 rides. This peak in spring months suggests favorable weather conditions contributing to increased cycling activity. Notably, June also exhibited a strong performance with 56,364 rides, indicating sustained interest as the summer began. Conversely, the months of July and August experienced the lowest ridership, with only 37,885 and 32,154 rides, respectively. These declines may be attributed to the monsoon season, which often deters outdoor activities due to rain and humidity. The ride data highlights a gradual recovery in the following months, particularly in November with 44,981 rides, suggesting that users resumed cycling activities as weather conditions improved. Overall, the monthly ridership trend emphasizes the seasonal variability in bicycle usage, reflecting the impact of climatic conditions and user behavior on the effectiveness of the PBSS in Ahmedabad.

Geographic analysis of ridership patterns in Ahmedabad categorizes the city into three distinct density zones as shown in Fig. 8, each with its unique characteristics and user behaviors. High Ridership Zone compasses 132 stations primarily concentrated around educational and recreational areas, notably near the university and the Sabarmati riverfront. Users in this zone predominantly engage in short-distance and leisure trips, thanks to the proximity of stations. The high ridership zone shows the greatest acceptance of the Public Bicycle Sharing system, especially among younger users who favor cycling for both recreational and practical purposes. Medium Ridership Zone comprising 86 stations, this zone is primarily located in the eastern part of the city, covering older residential and commercial areas. The medium ridership zone displays moderate usage levels, reflecting a diverse mix of user types who engage in both work and leisure trips. This zone serves as a transitional area between high and low ridership, indicating a balance between accessibility and demand. Low Ridership Zone consists of 35 stations positioned in the northern and southern parts of the city, where overall ridership is comparatively low. The low ridership zone primarily supports work-related trips; however, it struggles to foster significant usage due to limited infrastructure and demand. The lack of essential amenities and connectivity in these areas hinders the growth potential of the PBS system.

To understand the factors influencing cycling as a mode of commute in urban areas, the Analytic Hierarchy Process, developed by Prof. Thomas L. Saaty, is a valuable multi-criteria decision-making tool. AHP helps derive ratio scales from paired comparisons of criteria gathered from surveys and qualitative data. It accommodates minor inconsistencies in judgments due to human error. In the context of cycling, several key criteria are analyzed to understand their impact on cycling adoption. These criteria include the presence of separate cycle lanes, safe cycle parking facilities, clean air quality, temperature and humidity conditions, on-street vehicle parking, traffic levels, vehicle speed, accident-prone areas, safety at junctions, road flatness, availability of carrier baskets, comfortable attire, physical and mental health, and overall personal safety. Each of these factors plays a significant role in
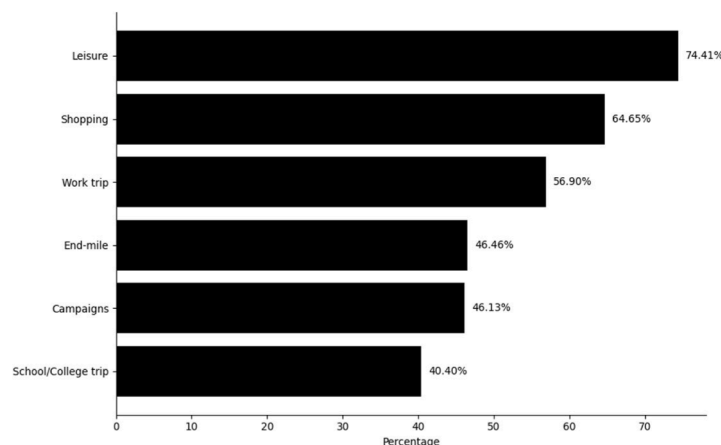

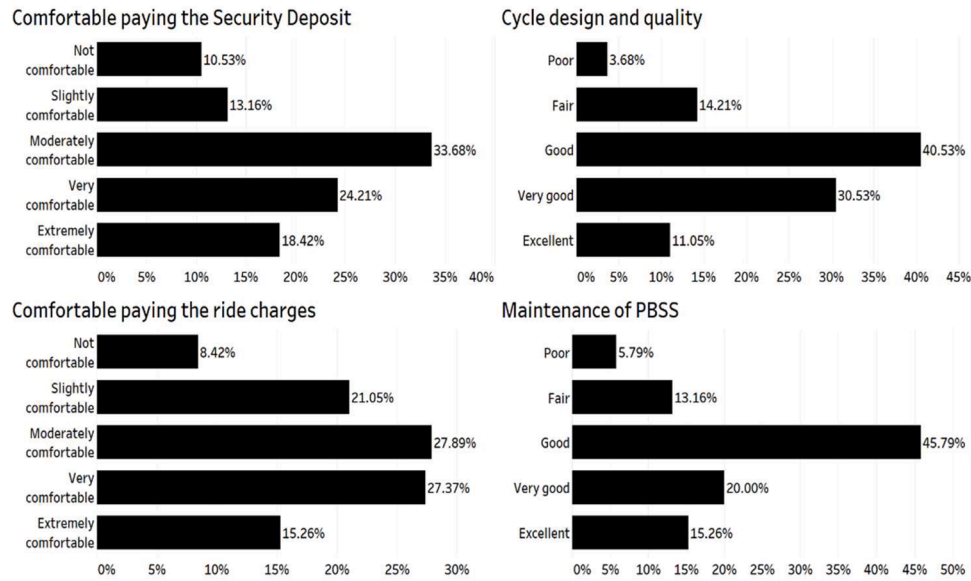**Fig. 3.** Primary Purposes for using the PBSS in Ahmedabad.

Fig. 4. User perceptions of cost, cycle design, and maintenance in the Ahmadabad PBSS.
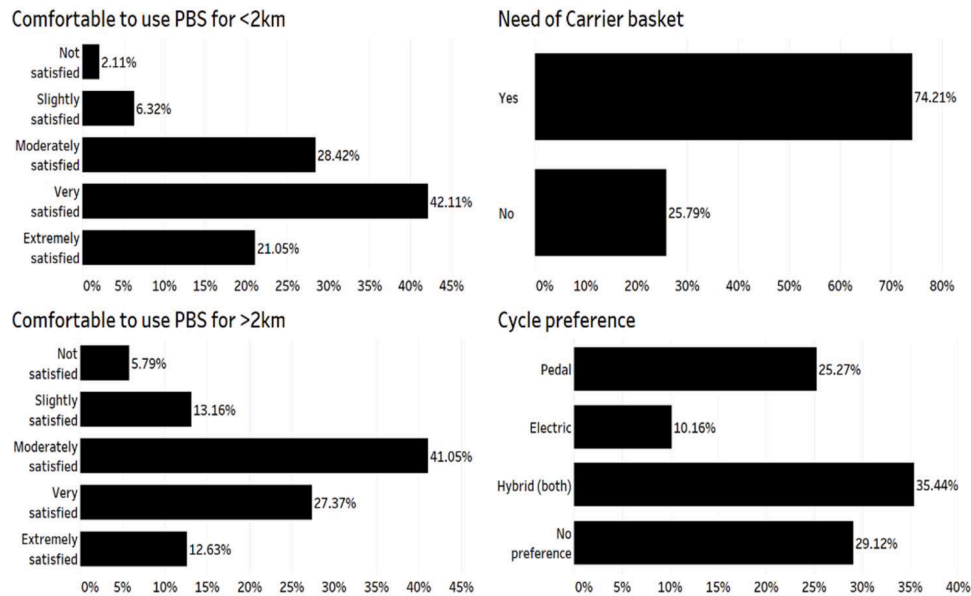

Fig. 5. User preferences for distance, cycle features, and comfort in the Ahmadabad PBSS.

determining the attractiveness and feasibility of cycling as a commuting option in urban areas. To evaluate these criteria systematically, a pairwise comparison matrix as given in Table 1 is created based on expert ratings, using a scale from 1 to 9 to reflect the relative importance of each criterion. This scale indicates the relative importance of one criterion over another, with 1 representing equal importance and 9 (or its reciprocal) indicating extreme preference. This matrix is then normalized to obtain the principal eigenvector as given in Table 2, which provides the ratio scales for the criteria. A crucial step in AHP is checking the consistency of judgments to ensure reliability of the results. This is done by calculating the Consistency Index (CI), which measures how consistent the judgments are compared to a theoretically perfect consistency. The Consistency Ratio (CR), which is the ratio of CI to the Random Consistency Index (RI), must be 10 % or less to be considered acceptable. If the CR exceeds this threshold, the judgments are revised to improve consistency.

The weights reflect the relative importance of each criterion in the

decision-making process for evaluating alternatives in the Public Bicycle Sharing Systems as shown in Fig. 9. The highest weight is assigned to Personal Safety (PS) at 21.40 %, indicating its critical role in influencing users' willingness to participate in bicycle sharing. Safe Path at Junctions/Signals (S-Ju) also holds significant weight at 16.89 %, emphasizing the importance of safe infrastructure. Other criteria such as Good Physical and Mental Health (GH) and Clean Air to Breathe (CA) also play vital roles, highlighting the environmental and health impacts of cycling. In contrast, criteria like Flat Roads (FR) and Humidity and Temperature (H&T) have lower weights, suggesting that while they are still relevant, they may be less critical in the overall evaluation compared to safety and user experience.

5. Evaluation of alternatives

Evaluating alternatives in decision-making processes, particularly in complex scenarios like Public Bicycle Sharing Systems, requires robust
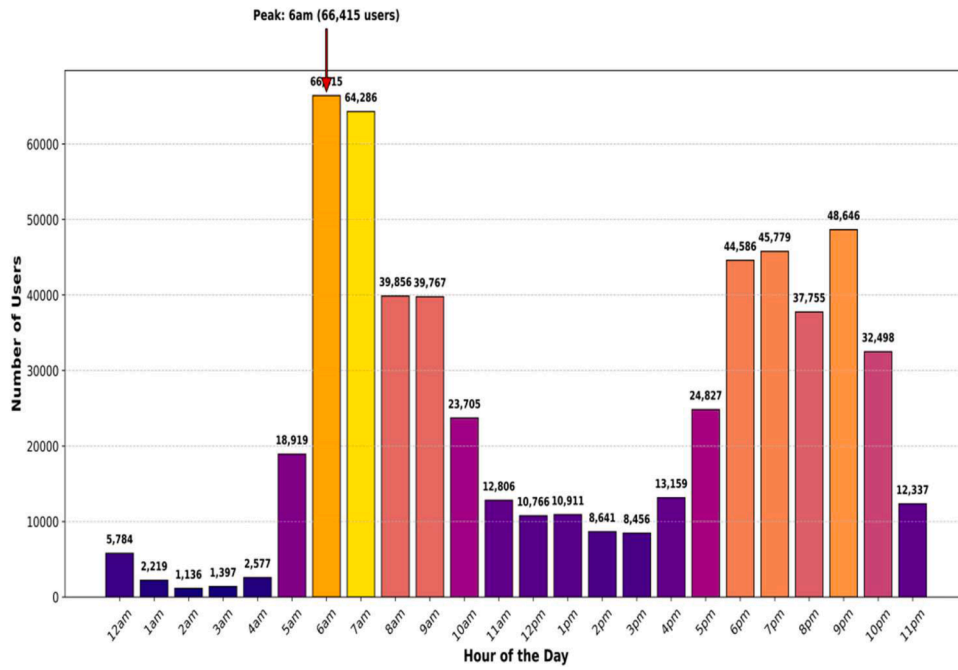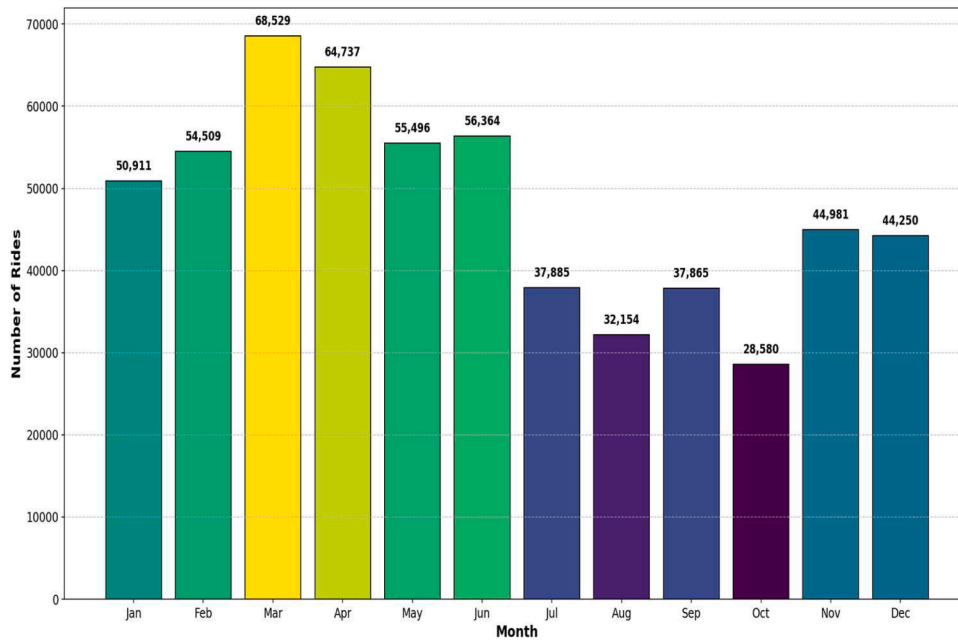
Fig. 6. User distribution by hour.

Fig. 7. Monthly ridership trend.

methodologies that canaccommodate various criteria and stakeholder perspectives [36]. This study aims to identify the most effective alternative for enhancing PBSS in urban environments by employing four distinct methods such AHP, Fuzzy AHP, ANP and TOPSIS. The five alternatives being evaluated are: 1) Enhanced Non-Motorized Transport (NMT) Infrastructure, 2) Implementation of Slow Streets, 3) Awareness and Education Campaigns, 4) Expansion of Bicycle Network, and 5) Integration with Public Transport. Each alternative represents a different strategy for improving the functionality and accessibility of PBSS, with the ultimate goal of promoting sustainable urban transport. The primary criteria identified for this evaluation included Safety, Environmental Impact, User Convenience, and Infrastructure Quality,

each comprising specific sub-criteria that further delineate the factors affecting the performance of PBS systems. In the Safety, sub-criteria such as Separate Cycle Lanes (SCL), Safe Paths at Junctions (S-Ju), and Personal Safety (PS) were assessed to ensure that cycling infrastructure minimizes risks for users. Environmental Impact focused on Clean Air (CA) and Good Health (GH), emphasizing the importance of promoting sustainable practices that contribute to urban air quality and public well-being. User Convenience encompassed Safe Cycle Parking Facilities (SCPF), On-street vehicle parking (OSP), Carrier Baskets (CB), and Comfortable Attire (CoA), reflecting the need for practical solutions that enhance the overall user experience. Finally, Infrastructure Quality examined factors such as Humidity and Temperature (H&T), Less Traffic
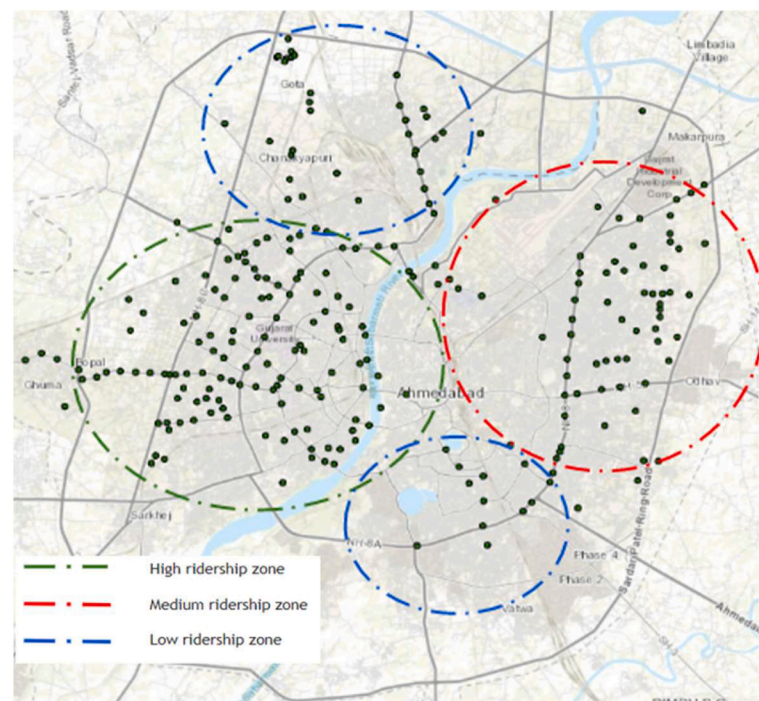
Fig. 8. Density zones of study area.

Table 1
Pairwise comparison matrix for criteria evaluation.

Criteria	SCL	SCPF	CA	H&T	OSP	LT	RVS	MAPA	S-Ju	FR	CB	CoA	GH	PS
SCL	1.00	3.00	0.50	1.40	1.00	0.50	0.50	1.50	0.20	3.00	3.00	1.50	0.33	0.14
SCPF	0.33	1.00	1.00	1.50	1.00	1.00	1.00	1.80	0.20	3.00	1.20	1.50	0.33	0.14
CA	2.00	1.00	1.00	7.00	3.00	1.00	1.00	3.00	0.33	9.00	5.00	3.00	1.00	0.33
H&T	0.71	0.67	0.14	1.00	0.50	0.50	0.50	1.20	0.14	1.20	1.00	1.00	0.33	0.11
OSP	1.00	1.00	0.33	2.00	1.00	0.14	0.14	1.50	0.20	2.00	1.60	1.50	0.33	0.14
LT	2.00	1.00	1.00	2.00	7.00	1.00	1.00	1.50	0.33	3.00	3.00	2.00	0.33	0.33
RVS	2.00	1.00	1.00	2.00	7.00	1.00	1.00	1.50	0.33	3.00	3.00	2.00	0.33	0.33
MAPA	0.67	0.56	0.33	0.83	0.67	0.67	0.67	1.00	0.20	3.00	1.00	1.00	0.33	0.33
S-Ju	5.00	5.00	3.00	7.00	5.00	3.00	3.00	5.00	1.00	9.00	9.00	5.00	1.00	0.33
FR	0.33	0.33	0.11	0.83	0.50	0.33	0.33	0.33	0.11	1.00	0.33	0.33	0.11	0.33
CB	0.33	0.83	0.20	1.00	0.63	0.33	0.33	1.00	0.11	3.00	1.00	0.50	0.20	0.20
CoA	0.67	0.67	0.33	1.00	0.67	0.50	0.50	1.00	0.20	3.00	2.00	1.00	0.20	0.20
GH	3.00	3.00	1.00	3.00	3.00	3.00	3.00	3.00	1.00	9.00	5.00	5.00	1.00	0.33
PS	7.00	7.00	3.00	9.00	7.00	3.00	3.00	3.00	3.00	3.00	5.00	5.00	3.00	1.00

SCL - Separate Cycle Lane; SCPF - Safe cycle parking facilities; CA - Clean air to breathe; H&T - Humidity and Temperature; OSP - On-street vehicle parking; LT - Less traffic; RVS - Reduced vehicle speed; MAPA - Marked accident-prone areas; S-Ju - Safe path at junctions/signals; FR - Flat roads (terrain); CB - Carrier basket for luggage; CoA - Comfortable attire/dress; GH - Good physical and mental health; PS - Personal safety.

(LT), Reduced Vehicle Speed (RVS), Marked Accident-Prone Areas (MAPA), and Flat Roads (FR), ensuring that the built environment supports safe and enjoyable cycling.

The application of four distinct Multi-Criteria Decision-Making methods—AHP, Fuzzy AHP, ANP, and TOPSIS—in this study provides a robust, holistic, and validated framework for evaluating strategies to improve the Public Bicycle Sharing System in Ahmedabad. Each method brings unique analytical strengths and is based on different assumptions that enrich the overall decision-making process. AHP is effective in hierarchically structuring complex problems and deriving priority weights through expert pairwise comparisons of 14 evaluation criteria, such as Personal Safety, Clean Air, and Safe Paths at Junctions. However, it assumes independence among criteria and consistent judgments. Fuzzy AHP addresses this by introducing triangular fuzzy numbers to capture the imprecision and ambiguity in human judgments—particularly for subjective indicators like User Convenience, Humidity & Temperature, or Cycle Parking Facilities—using linguistic terms rather than fixed numerical scales [37]. ANP extends AHP by allowing interdependence

and feedback loops among criteria [38,39]. For instance, in PBSS planning, Safety may influence Infrastructure Quality, and Environmental Impact may affect User Experience. ANP captures these interrelations through supermatrix-based computations, enabling a more realistic network representation. TOPSIS, in contrast, is a performance-based method that evaluates alternatives based on their geometric closeness to an ideal solution, using a normalized decision matrix where alternatives are rated across all criteria [40]. Weights for TOPSIS are adopted from AHP. By integrating these methods, the study captures hierarchical structure (AHP), subjective uncertainty (Fuzzy AHP), systemic interdependence (ANP), and relative performance proximity (TOPSIS). Despite their conceptual differences, all methods consistently ranked Enhanced Non-Motorized Transport Infrastructure as the most preferred alternative, reinforcing the robustness of the results. This integrative approach ensures that all critical dimensions of urban cycling systems are considered, leading to more balanced, inclusive, and sustainable PBSS planning in Ahmedabad and similar urban contexts. The input data used to evaluate and rank the alternatives

Table 2

Normalized matrix for criteria evaluation.

Criteria	SCL	SCPF	CA	H&T	OSP	LT	RVS	MAPA	S-Ju	FR	CB	CoA	GH	PS	EV	W
SCL	0.04	0.12	0.04	0.04	0.03	0.03	0.03	0.06	0.03	0.05	0.07	0.05	0.04	0.03	1.21	4.63
SCPF	0.01	0.04	0.08	0.04	0.03	0.06	0.06	0.07	0.03	0.05	0.03	0.05	0.04	0.03	1.15	4.41
CA	0.08	0.04	0.08	0.18	0.08	0.06	0.06	0.11	0.05	0.16	0.12	0.10	0.11	0.08	1.21	9.34
H&T	0.03	0.03	0.01	0.03	0.01	0.03	0.03	0.05	0.02	0.02	0.02	0.03	0.04	0.03	1.05	2.66
OSP	0.04	0.04	0.03	0.05	0.03	0.01	0.01	0.06	0.03	0.04	0.04	0.05	0.04	0.03	1.29	3.40
LT	0.08	0.04	0.08	0.05	0.18	0.06	0.06	0.06	0.05	0.05	0.07	0.07	0.04	0.08	1.10	6.88
RVS	0.08	0.04	0.08	0.05	0.18	0.06	0.06	0.06	0.05	0.05	0.07	0.07	0.04	0.08	1.10	6.88
MAPA	0.03	0.02	0.03	0.02	0.02	0.04	0.04	0.04	0.03	0.05	0.02	0.03	0.04	0.08	0.92	3.48
S-Ju	0.19	0.19	0.23	0.18	0.13	0.19	0.19	0.19	0.14	0.16	0.22	0.16	0.11	0.08	1.24	16.89
FR	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.08	1.05	1.89
CB	0.01	0.03	0.02	0.03	0.02	0.02	0.02	0.04	0.02	0.05	0.02	0.02	0.02	0.05	1.06	2.58
CoA	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.04	0.03	0.05	0.05	0.03	0.02	0.05	0.98	3.24
GH	0.12	0.12	0.08	0.08	0.08	0.19	0.19	0.11	0.14	0.16	0.12	0.16	0.11	0.08	1.09	12.35
PS	0.27	0.27	0.23	0.23	0.18	0.19	0.19	0.11	0.41	0.05	0.12	0.16	0.34	0.23	0.91	21.40
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	15.37	100

EV=Eigen Value; W = Weights based on Relative Importance (%).

RI (n = 14) = 1.57.

CI = (15.37 - 14)/(14 - 1) = 0.105 | **CR** = 0.105/1.57 0.067 | CR < 0.10 | Valid.
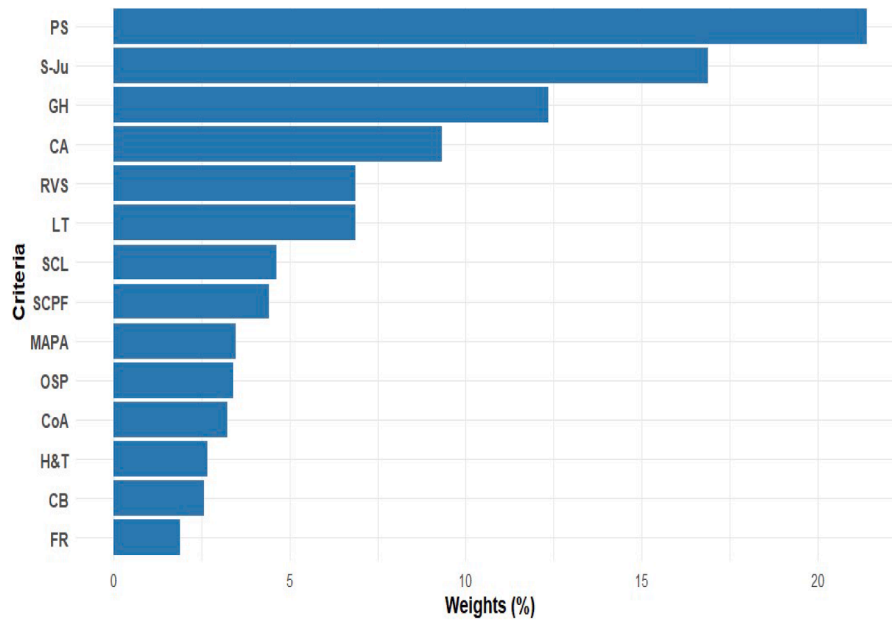
Fig. 9. Weights of evaluation criteria for public bicycle sharing system.

across all four MCDM are presented in Annexure 1.

The AHP method was used to evaluate alternatives for improving Public Bicycle Sharing systems. This method involves structuring the decision-making problem into a hierarchy, consisting of the overall goal, the criteria that influence the decision, and the available alternatives. In this case, four key criteria were identified: Safety, Environmental Impact, User Convenience, and Infrastructure Quality. Each of these criteria was compared in pairs using expert judgment to generate a pairwise comparison matrix. From these comparisons, criteria weights were calculated, reflecting their relative importance in achieving the overall goal. After the criteria weights were determined, each alternative was evaluated in terms of how well they performed against each criterion. The final scores for each alternative were calculated by multiplying their performance scores by the respective criteria weights, and the results were aggregated to obtain a total score for each alternative as given in Table 3. The AHP analysis showed that "Enhanced NMT Infrastructure" was the most preferred alternative, with the highest overall score. This was followed by "Expansion of Bicycle Network" and "Implementation of Slow Streets." Alternatives like "Awareness and Education Campaign" and "Integration with Public Transport" received

Table 3

Ranking alternatives in public bicycle sharing systems as per AHP method.

Alternative	AHP Score	Rank
Enhanced NMT Infrastructure	0.455	1
Expansion of Bicycle Network	0.315	2
Implementation of Slow Streets	0.289	3
Awareness and Education Campaign	0.175	5
Integration with Public Transport	0.185	4

lower scores, indicating a lesser preference for these strategies in the context of improving Public Bicycle Sharing systems [40].

The ANP method was used to evaluate alternatives for enhancing Public Bicycle Sharing systems by considering interdependencies between the criteria. Unlike the AHP, which assumes that criteria are independent, the ANP allows for feedback and relationships between the criteria, making it a more flexible tool for complex decision-making processes. The ANP methodology began by structuring the decision problem into a network of criteria, sub-criteria, and alternatives. In this case, four main criteria were considered: Safety, Environmental Impact,

User Convenience, and Infrastructure Quality. These criteria were evaluated not only based on their importance in relation to the goal but also based on their influence on each other. For example, Safety could impact User Convenience, and Environmental Impact could influence Cost-effectiveness. Pairwise comparisons were conducted between the criteria to generate the necessary input for the ANP model. The weighted supermatrix was then constructed, which captured the interdependencies between the criteria and alternatives. From this, the limiting supermatrix was derived, and the final priorities of the alternatives were calculated by aggregating their performances across all criteria as shown in Table 4. The ANP results showed that "Enhanced NMT Infrastructure" was still the top-performing alternative, but the scores were adjusted to reflect the interdependencies between the criteria. "Expansion of Bicycle Network" and "Implementation of Slow Streets" followed closely in terms of ranking, while "Awareness and Education Campaign" and "Integration with Public Transport" were ranked lower, similar to the results of the AHP analysis but with some variations due to the feedback loops considered in ANP.

The Fuzzy AHP was employed to evaluate alternatives for enhancing Public Bicycle Sharing systems, incorporating the uncertainty and imprecision in expert judgments through fuzzy logic. The methodology began with the establishment of fuzzy pairwise comparison matrices for the criteria: Safety, Environmental Impact, User Convenience, and Infrastructure Quality. Each pairwise comparison was represented using triangular fuzzy numbers, reflecting the subjective assessments of experts. After constructing the fuzzy matrices, the defuzzification process was applied to obtain crisp scores, facilitating the calculation of criteria weights. The final scores for each alternative were computed by aggregating the weighted evaluations based on their performance against the established criteria as shown in Table 5. The results revealed that the "Enhanced NMT Infrastructure" alternative achieved the highest score, indicating it as the most favorable option. Following it were the "Expansion of Bicycle Network" and "Implementation of Slow Streets," which also received significant scores. This analysis underscores the effectiveness of Fuzzy AHP in capturing expert judgments under uncertainty and providing a robust framework for decision-making in sustainable urban transport initiatives.

TOPSIS methodology was employed to evaluate alternatives for enhancing Public Bicycle Sharing systems based on four key criteria: Safety, Environmental Impact, User Convenience, and Infrastructure Quality. The process began with the establishment of a normalized decision matrix derived from the scores assigned to each alternative under the respective criteria. Subsequently, the ideal (best) and negative-ideal (worst) solutions were identified. The separation measures for each alternative were calculated based on their distances to both ideal solutions, allowing for the determination of relative closeness to the ideal solution (C*). This relative closeness provided a quantitative basis for ranking the alternatives as given in Table 6. The results indicated that the "Enhanced NMT Infrastructure" alternative emerged as the most preferred option, achieving the highest score, followed by "Expansion of Bicycle Network" and "Implementation of Slow Streets." The analysis highlighted the importance of prioritizing infrastructure improvements in promoting effective Public Bicycle Sharing systems.

Across all methods as shown in Fig. 10, Enhanced NMT Infrastructure consistently ranked as the top alternative. This suggests that investments in better Non-Motorized Transport infrastructure are

Table 4
Ranking alternatives in public bicycle sharing systems as per ANP method.

Alternative	ANP Score	Rank
Enhanced NMT Infrastructure	0.460	1
Expansion of Bicycle Network	0.320	2
Implementation of Slow Streets	0.300	3
Awareness and Education Campaign	0.180	5
Integration with Public Transport	0.190	4

Table 5
Ranking alternatives in public bicycle sharing systems as per FuzzyAHP method.

Alternative	Fuzzy Score	Rank
Enhanced NMT Infrastructure	0.475	1
Expansion of Bicycle Network	0.335	2
Implementation of Slow Streets	0.305	3
Awareness and Education Campaign	0.195	5
Integration with Public Transport	0.205	4

Table 6
Ranking alternatives in public bicycle sharing systems as per TOPSIS method.

Alternative	Score (C*)	Rank
Enhanced NMT Infrastructure	0.480	1
Expansion of Bicycle Network	0.330	2
Implementation of Slow Streets	0.295	3
Awareness and Education Campaign	0.200	4
Integration with Public Transport	0.210	5

universally considered to offer the most significant improvements for public bicycle sharing systems, enhancing safety, convenience, and overall user experience. Expansion of Bicycle Network also scored highly across the methods, highlighting the importance of having an extensive, connected network to encourage cycling in urban areas. Awareness and Education Campaign and Integration with Public Transport were generally the least preferred alternatives, though their rankings varied slightly across methods. This indicates that while education and transport integration are essential, they may not have as direct an impact as infrastructural improvements.The consistency in results across AHP, Fuzzy AHP, ANP, and TOPSIS can be attributed to several factors. Key criteria particularly personal safety and safe paths at Junctions, held significant weight (over 38 %), strongly influencing rankings. Enhanced NMT Infrastructure consistently outperformed other alternatives due to its performance under these critical criteria. Additionally, the preference structure remained stable across methods, with clear gaps between top and lower-ranked options. Despite methodological differences, all methods shared similar mathematical principles and used the same input data, leading to convergent results. This alignment validates the robustness of the evaluation and supports the recommended strategy for improving PBSS in Ahmadabad.

6. Public bicycle insights for transport planners

This study provides valuable insights for transport planners by:

- **Identifying Key Factors:** It highlights the most critical factors, such as personal safety, safe cycling infrastructure, and environmental impact that significantly influence the success of Public Bicycle Sharing Systems. Transport planners can prioritize these factors when designing urban cycling infrastructure.
- **Prioritizing Effective Alternatives:** The study ranks alternatives like enhanced non-motorized transport infrastructure and expansion of bicycle networks as the most effective strategies for improving PBSS. This helps planners allocate resources efficiently to the most impactful interventions.
- **Supporting Decision-Making with Multi-Criteria Methods:** By employing methods like AHP, Fuzzy AHP, ANP and TOPSIS, the study offers a robust, data-driven framework that transport planners can use to make informed decisions in complex urban environments, accounting for interdependencies between criteria.
- **Targeting Socio-Demographic Gaps:** The socio-demographic analysis reveals gender and income disparities in PBSS usage, guiding planners to create more inclusive and accessible transport systems by addressing specific barriers like safety and affordability.
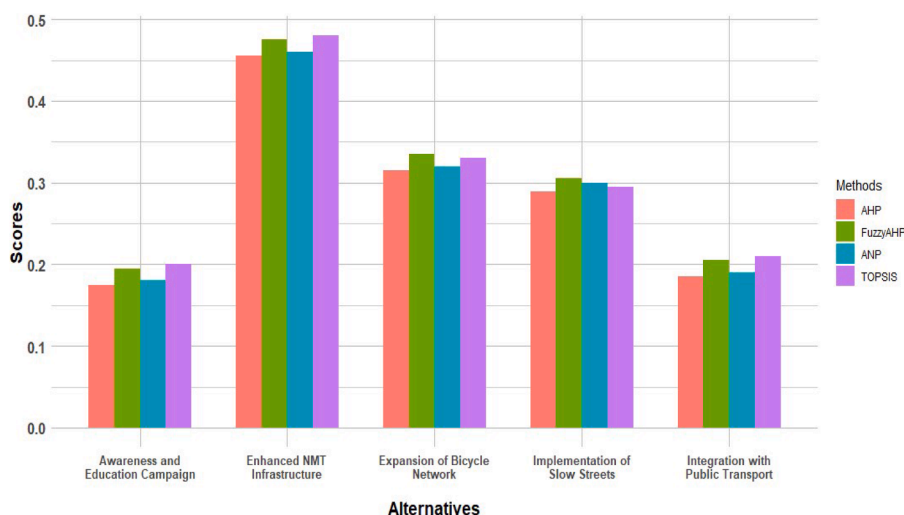
Fig. 10. Comparison of decision-making methods for public bicycle sharing system.

- **Future Urban Mobility Planning:** The study underscores the importance of integrating cycling with existing public transport networks, electric bike adoption, and affordability measures, providing a roadmap for future sustainable urban transport development.

7. Conclusions

The present study mainly focused on evaluating alternatives for improving Public Bicycle Sharing Systems (PBSS) through different decision-making methods such as AHP, Fuzzy AHP, ANP, and TOPSIS. The study identified key factors affecting the success of PBSS, including safety, environmental impact, user convenience, and infrastructure quality. The following conclusions were drawn from the study.

- The socio-demographic analysis reveals a strong male dominance (82.47 %) in the usage of Public Bicycle Sharing Systems (PBSS), with most users being young (ages 20-35) and from lower- to middle-income groups. This indicates a need to address barriers such as safety, comfort, and convenience to encourage greater participation among women, older age groups and higher-income individuals.
- The primary purposes for using the Public Bicycle Sharing System (PBSS) in Ahmadabad are leisure (74.41 %) and shopping (64.65 %), with a smaller percentage of users utilizing the system for work (56.90 %) and educational purposes (40.40 %). This suggests that while PBSS is gaining traction for various activities, there is potential to further promote its use as a practical commuting option.
- User perceptions of the cost, cycle design, and maintenance in the Ahmadabad PBSS are generally positive, with the majority rating the cycle design and maintenance as good or very good. However, financial accessibility remains a challenge, as only 42 % of respondents feel comfortable with the current security deposit and ride charges, indicating a need for more affordable pricing options to broaden the system's user base.
- Despite Ahmadabad's flat terrain, 45.6 % of users prefer electric or hybrid cycles due to the city's high temperatures and humidity, indicating that integrating more electric options could enhance user comfort. Additionally, a strong preference for carrier baskets (favored by 74 % of respondents) highlights the importance of practical features to improve the user experience.
- The analysis reveals a clear monthly variation in ridership, with a total of 576,261 rides across the year. The peak ridership occurred during the warmer months, notably in March with 68,529 rides, while colder months like October recorded the lowest usage at

28,580 rides. This trend emphasizes the influence of seasonal factors on cycling behavior, highlighting the need for targeted promotions and infrastructure enhancements during less popular months to sustain ridership throughout the year. Ridership peaks during morning and evening hours, suggesting that users predominantly utilize the PBSS for leisure and last-mile connections during cooler parts of the day.
- The weights of evaluation criteria show that Personal Safety (21.40 %) and Safe Path at Junctions/Signals (16.89 %) are the most important factors influencing the success of Public Bicycle Sharing Systems. This underscores the critical need for safety-focused infrastructure to encourage greater adoption and user confidence.
- Enhanced NMT Infrastructure consistently ranks as the most effective alternative for improving Public Bicycle Sharing Systems (PBSS) across all decision-making methods. This suggests that investing in better infrastructure, such as dedicated cycling lanes and safer environments, would significantly enhance the adoption and success of these systems.

Future research should focus on increasing inclusivity by addressing gender disparity, improving affordability, and expanding the availability of electric cycles to enhance user comfort in varying climatic conditions. These efforts could further promote PBSS as a practical urban mobility solution.

Statement of compliance with ethical standards

This work is the authors' own original work, which has not been previously published and not currently being considered for publication elsewhere.

Funding

CRediT authorship contribution statement

T.S. Shagufta: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Dimpu Byalal Chindappa:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Seelam Srikanth: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Subhashish Dey:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tbench.2025.100220.

References

[1] A. Nikitas, S. Tsigdinos, C. Karolemeas, E. Kourmpa, E. Bakogiannis, Cycling in the era of COVID-19: lessons learnt and best practice policy recommendations for a more bike-centric future, Sustainability 13 (9) (2021) 4620, https://doi.org/10.3390/su13094620.
[2] V. Sunio, I. Mateo-Babiano, Pandemics as 'windows of opportunity': transitioning towards more sustainable and resilient transport systems, Transp. Policy. 116 (2021) 175–187, https://doi.org/10.1016/j.tranpol.2021.12.004.
[3] L. Böcker, E. Anderson, T.P. Uteng, T. Throndsen, Bike sharing use in conjunction to public transport: exploring spatiotemporal, age and gender dimensions in Oslo, Norway. Transport. Res. Part a Policy Pract. 138 (2020) 389–401, https://doi.org/10.1016/j.tra.2020.06.009.
[4] S.J. Chang, A.F. Ferreira, Bike-Sharing system: uncovering the "success factors. Elsevier Ebooks, 2021, pp. 355–362, https://doi.org/10.1016/b978-0-08-102671-7.10348-3.
[5] S.J. Patel, C.R. Patel, An infrastructure review of Public Bicycle Sharing System (PBSS): global and Indian scenario. Lecture Notes in Intelligent Transportation and Infrastructure, 2018, pp. 111–120, https://doi.org/10.1007/978-981-13-2032-3_11.
[6] G. Nataraj, Infrastructure Challenges in India: The role of Public–Private partnerships, Cambridge University Press eBooks, 2015, pp. 269–300, https://doi.org/10.1017/cbo9781316106631.012.
[7] A. Munkácsy, A. Monzón, Potential user profiles of innovative bike-sharing systems: the case of BICIMAD (Madrid, Spain), Asian. Trans. Stud. 4 (3) (2017) 621–638, https://doi.org/10.11175/eastsats.4.621.
[8] T. Bieliński, A. Ważna, Hybridizing bike-sharing systems: the way to improve mobility in smart cities, Trans. Econo. Log. 79 (2018) 53–63, https://doi.org/10.26881/etil.2018.79.04.
[9] S.A. Shaheen, A.P. Cohen, E.W. Martin, Public bikesharing in North America: early operator understanding and emerging trends, Transp. Res. Rec. 2387 (1) (2013) 83–92.
[10] J.F. Teixeira, C. Silva, F.M.E. Sá, Potential of bike sharing during disruptive public health crises: a review of COVID-19 impacts, Transp. Res. Record J. Transp. Res. Board. (2023), https://doi.org/10.1177/03611981231160537, 036119812 311605.
[11] E. O'Mahony, D. Shmoys, Data analysis and optimization for (CITI)Bike sharing, in: Proceedings of the AAAI Conference on Artificial Intelligence 29, 2015, https://doi.org/10.1609/aaai.v29i1.9245.

[12] D. Freund, S.G. Henderson, D.B. andShmoys, Bike sharing. Springer Series in Supply Chain Management, 2019, pp. 435–459, https://doi.org/10.1007/978-3-030-01863-4_18.
[13] W. Raza, B. Forsberg, C. Johansson, J.N. Sommar, Air pollution as a risk factor in health impact assessments of a travel mode shift towards cycling, Glob. Health Action. 11 (1) (2018) 1429081, https://doi.org/10.1080/16549716.2018.1429081.
[14] R. Julio, A. Monzon, Long term assessment of a successful e-bike-sharing system. Key drivers and impact on travel behavior, Case Stud. Trans. Policy 10 (2) (2022) 1299–1313, https://doi.org/10.1016/j.cstp.2022.04.019.
[15] Y.S. Murat, Z. Cakici, Analysis of the COVID-19 pandemic on preferences of transport modes, in: Proceedings of the Institution of Civil Engineers - Transport, 2024, pp. 1–13, https://doi.org/10.1680/jtran.23.00105.
[16] E. Macioszek, P. Świerk, A. Kurek, The bike-sharing system as an element of enhancing sustainable mobility—A case study based on a City in Poland, Sustainability 12 (8) (2020) 3285, https://doi.org/10.3390/su12083285.
[17] Y. Guo, J. Zhou, Y. Wu, Z. Li, Identifying the factors affecting bike-sharing usage and degree of satisfaction in Ningbo, China, PLoS One. 12 (9) (2017) e0185100, https://doi.org/10.1371/journal.pone.0185100.
[18] M. Bencekri, Y. Van Fan, D. Lee, M. Choi, S. Lee, Optimizing shared bike systems for economic gain: integrating land use and retail, J. Transp. Geogr. 118 (2024) 103920, https://doi.org/10.1016/j.jtrangeo.2024.103920.
[19] R. Mix, R. Hurtubia, S. Raveau, Optimal location of bike-sharing stations: a built environment and accessibility approach, Transp. Res. Part Policy Prac. 160 (2022) 126–142, https://doi.org/10.1016/j.tra.2022.03.022.
[20] L. Wang, K. Zhou, S. Zhang, A.V. Moudon, J. Wang, Y. Zhu, W. Sun, J. Lin, C. Tian, M. Liu, Designing bike-friendly cities: interactive effects of built environment factors on bike-sharing, Transp. Res. Part D Trans. Environ. 117 (2023) 103670, https://doi.org/10.1016/j.trd.2023.103670.
[21] H. Mohiuddin, D.T. Fitch-Polse, S.L. Handy, Does bike-share enhance transport equity? Evidence from the Sacramento, California region, J. Transp. Geogr. 109 (2023) 103588, https://doi.org/10.1016/j.jtrangeo.2023.103588.
[22] H. Mohiuddin, D.T. Fitch-Polse, S.L. Handy, Examining market segmentation to increase bike-share use and enhance equity: the case of the greater Sacramento region, Transp. Policy. 145 (2023) 279–290, https://doi.org/10.1016/j.tranpol.2023.10.021.
[23] I. Politis, I. Fyrogenis, E. Papadopoulos, A. Nikolaidou, E. Verani, Shifting to shared wheels: factors affecting dockless bike-sharing choice for short and long trips, Sustainability 12 (19) (2020) 8205, https://doi.org/10.3390/su12198205.
[24] S.H. Grasso, P. Barnes, C. Chavis, Bike share equity for underrepresented groups: analyzing barriers to system usage in Baltimore, Maryland. Sustain. 12 (18) (2020) 7600, https://doi.org/10.3390/su12187600.
[25] E.J. Shin, A comparative study of bike-sharing systems from a user's perspective: an analysis of online reviews in three U.S. regions between 2010 and 2018, Int. J. Sustain. Transport. 15 (12) (2020) 908–923, https://doi.org/10.1080/15568318.2020.1830320.
[26] V. Bernardo, The impact of bike-sharing systems on congestion. Evidence from European urban areas, SSRN Electron. J (2022), https://doi.org/10.2139/ssrn.4027398.
[27] B. Cheng, J. Li, H. Su, K. Lu, H. Chen, J. Huang, Life cycle assessment of greenhouse gas emission reduction through bike-sharing for sustainable cities, Sustain. Energy Technol. Assessment. 53 (2022) 102789, https://doi.org/10.1016/j.seta.2022.102789.
[28] P. DeMaio, Bike-sharing: history, impacts, models of provision, and future, J. Public Transport. 12 (4) (2009) 41–56, https://doi.org/10.5038/2375-0901.12.4.3.
[29] K. Pelechrinis, C. Zacharias, M. Kokkodis, T. Lappas, Economic impact and policy implications from urban shared transportation: the case of Pittsburgh's shared bike system, PLoS One 12 (8) (2017) e0184092, https://doi.org/10.1371/journal.pone.0184092.
[30] S.S. Han, Co-producing an urban mobility service? The role of actors, policies, and technology in the boom and bust of dockless bike-sharing programmes, Int. J. Urban Sustain. Develop. 14 (1) (2020) 209–224, https://doi.org/10.1080/19463138.2020.1772268.
[31] E. Macioszek, M. Cieśla, External environmental analysis for sustainable bike-Sharing system development, Energies 15 (3) (2022) 791, https://doi.org/10.3390/en15030791.
[32] A. Kumar, K.M. Teo, A.R. Odoni, A systems perspective of cycling and bike-sharing systems in urban mobility, in: Proceedings of 30th International Conference of the System Dynamics Society, St. Gallen, 2012.
[33] R. Bean, D. Pojani, J. Corcoran, How does weather affect bikeshare use? a comparative analysis of forty cities across climate zones, J. Transp. Geogr. 95 (2021) 103155, https://doi.org/10.1016/j.jtrangeo.2021.103155.
[34] J. Lee, G.O. Jeong, H.C. Shin, Impact analysis of weather condition and location characteristics on the usage of public bike sharing system, J. Korean Soc. Transport. 34 (5) (2016) 394–408, https://doi.org/10.7470/jkst.2016.34.5.394.
[35] J. Suchanek, Success factors for the development of urban bike systems Success factors for the development of urban bike systems Success factors for the

development of urban Bike systems, Transport Econ. Logist. 84 (2019) 91–102, https://doi.org/10.26881/etil.2019.84.08.

[36] T.L. Saaty, D. Ergu, When is a decision-making method trustworthy? Criteria for evaluating multi-criteria decision-making methods, Int. J. Inf. Technol. Decis. Mak. 14 (06) (2015) 1171–1187, https://doi.org/10.1142/s021962201550025x.

[37] Y. Liu, C.M. Eckert, C. Earl, A review of fuzzy AHP methods for decision-making with subjective judgments, Expert. Syst. Appl. 161 (2020) 113738, https://doi.org/10.1016/j.eswa.2020.113738.

[38] M. Lee, The analytic hierarchy and the network process in multi-criteria decision making: performance evaluation and selecting key performance indicators based on ANP model. InTech Ebooks, 2010, https://doi.org/10.5772/9643.

[39] A. Görener, Comparing AHP and ANP: an application of strategic decisions making in a manufacturing company, Int. J. Bus. Soc. Sci. 3 (11) (2012) 194–208.

[40] A. Ishak, N. Wanli, Analysis of fuzzy AHP-TOPSIS methods in multi criteria decision making: literature review, IOP. Conf. Series Mater. Sci. Eng. 1003 (1) (2020) 012147, https://doi.org/10.1088/1757-899x/1003/1/012147.