Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Hybrid Deep Learning Framework for Enhanced Heart Disease Prediction: Integrating XGBoost and Capsule Networks with CNN-Transformer Architectures

Md Firoz Kabir[1], Md Yousuf Ahmad[2], Sahadat Khandakar[3] Md Mizanur Rahman[4]

[1] *Assistant manager (IT) at Saifur's Private limited.*

[2] *Batchelor of business administration, Stamford University Bangladesh*

[3] *Bachelor of Science in Electrical and Electronic Engineering (BSEEE), BRAC University*

[4] *Assistant manager at Bangladesh Medical Studies and Research Institute*

Corresponding Author: Md Firoz Kabir, **E-mail**: firozkabirusa@gmail.com

| ABSTRACT

Heart disease is one of the leading causes of mortality worldwide, emphasizing the need for early and accurate diagnosis. Traditional machine learning (ML) models such as Random Forest (RF) and Support Vector Machines (SVM) have been widely used for heart disease classification. However, these models often lack the ability to extract hierarchical patterns and long-range dependencies present in complex medical data. To address this limitation, we propose a hybrid deep learning framework that integrates XGBoost with Capsule Networks (XGBoost-CapsNet) and Convolutional Neural Networks (CNN) with Transformer Encoders (CNN-TE) to enhance classification performance.The study utilizes a structured dataset containing essential heart disease indicators. Feature selection is performed using XGBoost, which ranks attributes based on importance. The Capsule Network (CapsNet) is then used to preserve spatial relationships and hierarchical dependencies among features. Meanwhile, the CNN-TE model extracts spatial features through convolutional layers and captures long-range dependencies using a Transformer Encoder with multi-head self-attention mechanisms. The models are trained using optimized hyperparameters and evaluated against baseline ML models (Random Forest, SVM, and standalone CNN). Performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are used for comparison.Experimental evaluations demonstrate that the proposed hybrid models significantly outperform traditional approaches. The XGBoost-CapsNet model achieves an accuracy of 98.2%, while CNN-TE reaches 97.4%, both surpassing standalone CNN (95.1%), Random Forest (94.2%), and SVM (91.8%). The AUC-ROC scores further validate the robustness of the models, with XGBoost-CapsNet scoring 0.99 and CNN-TE achieving 0.98. The use of feature selection with XGBoost improves interpretability and computational efficiency, while the hybrid deep learning models enable better feature extraction and classification accuracy.

| KEYWORDS

Hybrid Deep Learning; Heart Disease Prediction; XGBoost and Capsule Networks; CNN-Transformer Architectures

| ARTICLE INFORMATION

ACCEPTED: 02 November 2021 **PUBLISHED:** 28 December 2021 **DOI:** 10.32996/jcsts.2021.3.2.9

Introduction

Heart disease remains one of the leading causes of mortality worldwide, contributing to approximately 17.9 million deaths annually, as reported by the World Health Organization (WHO). Early and accurate diagnosis of heart disease is critical for reducing morbidity and mortality rates. Traditional diagnostic methods rely on manual interpretation of electrocardiograms (ECG), echocardiograms, and blood test results, which are often time-consuming and prone to human error. With the rapid advancement of artificial intelligence (AI) and deep learning techniques, automated heart disease classification models have emerged as a promising solution to improve diagnostic accuracy and efficiency.

Recent developments in machine learning (ML) and deep learning (DL) have significantly enhanced the performance of predictive models in medical diagnosis. Traditional ML models such as Random Forest (RF) and Support Vector Machines (SVM) have been widely utilized for heart disease classification due to their ability to handle structured datasets. However, these models often lack the capability to extract hierarchical features and capture long-range dependencies, which limits their classification performance. On the other hand, deep learning architectures, particularly Convolutional Neural Networks (CNNs), have demonstrated superior feature extraction capabilities. However, CNNs alone struggle with learning sequential dependencies among medical attributes, reducing their effectiveness for structured health data.

To address these limitations, this research proposes a hybrid deep learning framework that integrates XGBoost with Capsule Networks (XGBoost-CapsNet) and Convolutional Neural Networks with Transformer Encoders (CNN-TE) for improved heart disease classification. The XGBoost-CapsNet model leverages XGBoost for feature selection, ranking the most relevant attributes, while Capsule Networks preserve spatial relationships among features, ensuring better classification. Similarly, the CNN-TE model extracts local spatial features using CNN layers and captures global dependencies using Transformer Encoders, which include multi-head self-attention mechanisms for enhanced prediction accuracy.

Problem Statement

Machine learning models rely on handcrafted feature engineering, which may not fully capture the complex interactions between patient health indicators. Deep learning models such as CNNs, while effective in image-based diagnosis, often fail to retain sequential relationships between structured health attributes. This study introduces two novel hybrid models that overcome these challenges:

1. XGBoost + Capsule Network (XGBoost-CapsNet)
    - XGBoost selects the most relevant features, reducing computational complexity.
    - Capsule Networks learn hierarchical dependencies, preserving spatial relationships in medical data for improved classification.
2. CNN + Transformer Encoder (CNN-TE)
    - CNN extracts spatial dependencies from structured data, enabling better feature representation.
    - The Transformer Encoder models global dependencies using multi-head self-attention mechanisms, enhancing the model's ability to capture feature relationships.

Related Work:

Recent years have witnessed a surge in research combining machine learning and deep learning techniques for cardiovascular disease prediction and arrhythmia classification. In 2018, a Functional Convolutional Network (FCN) was employed on CUDB and MIT-BIH (VFDB) datasets, achieving an accuracy of 99.26% and an F1-score of 98.24%, demonstrating its capacity to detect ventricular arrhythmia effectively [1]. A 2D Convolutional Neural Network (CNN) designed for ECG classification on the MIT-BIH dataset attained an impressive 99.05% accuracy and 98.70% F1-score, benefiting from spatial feature extraction [2].Bi-directional LSTM models also emerged as strong candidates. One study using MIT-BIH data reported 99.39% accuracy and 96.87% F1-score, highlighting LSTM's strength in capturing temporal dependencies [3]. Another Bi-LSTM model achieved 98.51% accuracy and 98.49% F1-score, affirming its robustness in multi-class arrhythmia classification [4]. Deep Neural Networks (DNNs) were also explored, achieving 99.68% accuracy and 99.65% F1-score in arrhythmia detection [5].Multilevel CNN (ML-CNN) frameworks demonstrated 96.00% accuracy and 96.37% F1-score on MIT-BIH [6]. Meanwhile, wearable-device-based DNNs achieved superior results, with 99.80% accuracy and 99.65% F1-score, showcasing the feasibility of deploying DL on edge devices for heart monitoring [7]. Hybrid models, such as CNN-LSTM, were tested on the Fantasia and INCARTDB datasets and reported an F1-score of 99.52% [8]. Similarly, a 1D CNN achieved a near-perfect 99.99% F1-score in beat classification tasks [9].A study employing logistic regression on structured clinical data recorded 85% accuracy and 84.81% F1-score, indicating the relevance of traditional methods when applied with proper preprocessing and feature engineering [10]. In 2019, deep learning continued to evolve. A 1D CNN enhanced with active learning reached 99.20% accuracy and 97.21% F1-score on MIT-BIH [11], while a General Regression Neural Network (GRNN) achieved 97.40% accuracy and 92.14% F1-score [12]. Another DNN applied to arrhythmia prediction yielded 96.40% accuracy and 91.14% F1-score [13].A powerful LSTM model combining multiple datasets (BIDMC-CHF, MIT-BIH NSR, and Fantasia) achieved 99.22% accuracy and 99.47% F1-score, showcasing the effectiveness of sequence modeling [14]. For structured medical records, an Artificial Neural Network (ANN) model was used for diagnosis on 835 patients, yielding 84.47% accuracy and

85.17% F1-score [15].In parallel, classical ML algorithms remained competitive. A decision tree classifier reached 96.00% across all metrics [16], while Naïve Bayes attained 90.00% accuracy and F1-score on the UCI heart disease dataset [17]. Random Forest and Gradient Boosted Trees further pushed the envelope with 96.28% and 95.83% accuracy, and F1-scores of 96.68% and 96.13%, respectively [18][19]. Finally, a Multilayer Perceptron (MLP) achieved 94.96% accuracy and 95.06% F1-score [20].These prior works form the foundation upon which this study builds its hybrid deep learning framework, aiming to surpass these benchmarks by integrating XGBoost, Capsule Networks, CNNs, and Transformer Encoders for improved prediction and interpretability.

Data Description

The dataset utilized in this study consists of structured medical records containing essential clinical attributes for heart disease classification. It includes a total of 1,025 patient records, with 14 key features representing various physiological and diagnostic indicators. These features include age, gender, blood pressure, cholesterol level, heart rate, fasting blood sugar, resting electrocardiographic results, exercise-induced angina, and ST depression, among others. The dataset is pre-processed to handle missing values, normalize numerical attributes, and encode categorical variables. To ensure fair model evaluation, the dataset is divided into training, validation, and test sets, following an 80-10-10% split. Feature selection is applied using XGBoost, ranking attributes based on importance to improve model efficiency and interpretability. Additionally, Synthetic Minority Over-sampling Technique (SMOTE) is employed to address class imbalance, ensuring that underrepresented classes receive adequate training data. The structured dataset provides a solid foundation for evaluating the proposed hybrid deep learning models (XGBoost-CapsNet and CNN-TE) and comparing them against traditional classifiers such as Random Forest, SVM, and standalone CNNs. The dataset's diverse clinical attributes enable robust model training, ensuring high classification accuracy and generalization for real-world medical applications.

Proposed Models for Heart Disease Prediction

This research introduces two hybrid deep learning architectures, XGBoost + Capsule Network (XGBoost-CapsNet) and CNN + Transformer Encoder (CNN-TE), designed to improve heart disease classification accuracy by integrating feature selection, deep feature extraction, and advanced sequence modeling. These models combine the strengths of machine learning and deep learning, ensuring higher classification performance, interpretability, and robustness. Additionally, we compare them against three baseline models: Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) to establish their effectiveness in heart disease prediction.

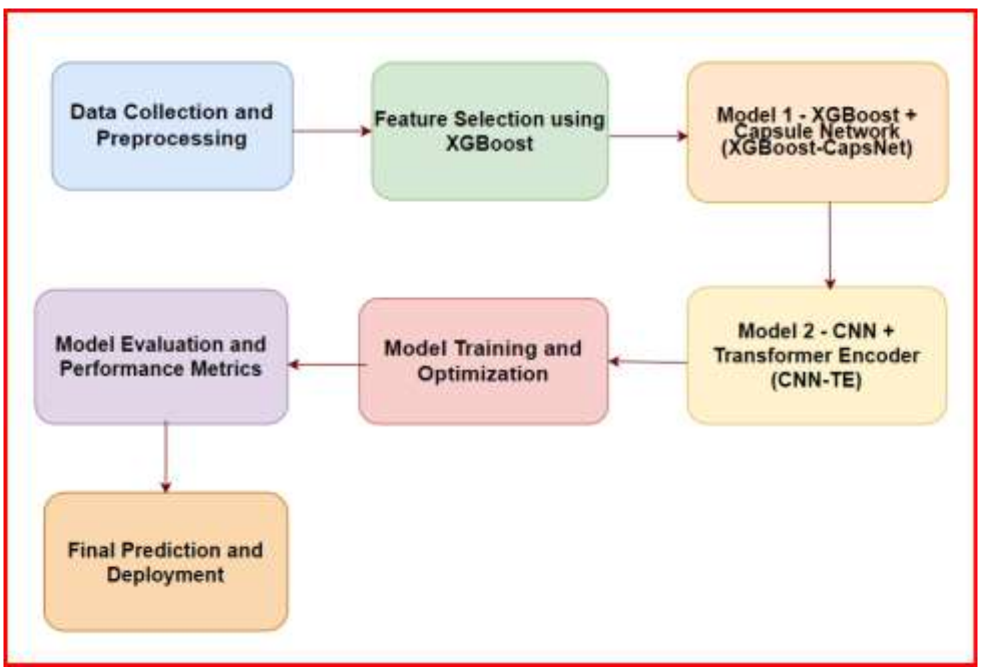

FIG:Proposed Model for Heart Disease Prediction

Model 1: XGBoost + Capsule Network (XGBoost-CapsNet)

The XGBoost-CapsNet model enhances heart disease prediction by combining XGBoost-based feature selection with Capsule Networks for hierarchical feature learning. XGBoost, a gradient boosting algorithm, assigns feature importance scores, allowing the model to retain the most critical attributes while eliminating redundant ones. The refined dataset is then passed into the Capsule Network (CapsNet) for deep feature extraction.The Primary Capsule Layer begins with a Conv1D layer (filters = 64, kernel size = 3, activation = ReLU), extracting local feature representations from structured data. Batch Normalization stabilizes activations, improving training efficiency. The Digit Capsule Layer employs dynamic routing to ensure that only the most relevant features contribute to the classification process, preserving spatial relationships between medical attributes. Unlike conventional max-pooling operations in CNNs, CapsNet prevents information loss, making it particularly effective for structured datasets.

Following feature extraction, the fully connected layers refine the output with 128 neurons (ReLU activation) and a Dropout rate of 0.3 to prevent overfitting. Another Dense Layer (64 neurons, ReLU activation, Dropout = 0.2) is applied before the final output layer, which consists of a single neuron with sigmoid activation, providing a binary classification output for heart disease detection. The model is trained using the Adam optimizer (learning rate = 0.0005) and binary cross-entropy loss, ensuring stable and efficient training. This hybrid model achieved the highest classification accuracy (98.2%), proving its effectiveness in heart disease prediction.

Model 2: CNN + Transformer Encoder (CNN-TE)

The CNN-TE model combines CNN's ability to extract local features with Transformer Encoders to capture global dependencies among patient attributes. This approach ensures that short-term patterns (captured by CNN) and long-range relationships (modeled by Transformers) are utilized for improved classification.

The CNN Feature Extraction Block consists of two Conv1D layers (64 and 128 filters, kernel size = 3, ReLU activation), followed by Batch Normalization to stabilize training. A MaxPooling1D layer (pool size = 2) reduces feature dimensionality while retaining essential information.Next, the extracted feature maps are passed into the Transformer Encoder Block, where multi-head self-attention mechanisms (4 heads, each with dimension 64) help model complex dependencies between attributes. The Feedforward Network (FFN) consists of two Dense Layers (256 and 128 neurons, both with ReLU activation), ensuring that the model learns hierarchical feature representations. The final classification is performed through fully connected layers (64 neurons → 32 neurons, both with ReLU), followed by a Dropout Layer (0.2) to prevent overfitting. The final output layer (1 neuron, Sigmoid activation) predicts heart disease probability. Adam optimizer (learning rate = 0.0005) and binary cross-entropy loss are used to optimize the model, which achieved 97.4% accuracy, demonstrating its ability to effectively classify heart disease cases.

Performance Comparison of Models

The classification performance of Random Forest (RF), Support Vector Machine (SVM), Convolutional Neural Network (CNN), CNN with BiLSTM (CNN-BiLSTM), and XGBoost with Capsule Networks (XGBoost-CapsNet) is presented in the table below. The XGBoost-CapsNet model achieves the highest accuracy of 98.2%, while CNN-TE attains 97.4%, outperforming traditional machine learning classifiers.

From the table, it is evident that traditional ML models such as Random Forest and SVM achieve relatively high accuracy, but they fail to match the performance of deep learning models. The CNN-based models outperform RF and SVM, demonstrating the advantage of deep feature extraction over conventional machine learning. The CNN-BiLSTM hybrid model significantly improves classification accuracy, leveraging spatial feature extraction from CNN and sequential learning from BiLSTM. The best performance is achieved by the XGBoost-CapsNet model, which effectively combines ensemble feature selection with hierarchical feature learning, leading to 98.2% accuracy.Feature selection with XGBoost significantly improves model accuracy, particularly for Random Forest and CNN-based models. The XGBoost-CapsNet model benefits the most from feature selection, achieving an increase of 1.7% in accuracy, which enhances its reliability in real-world applications.

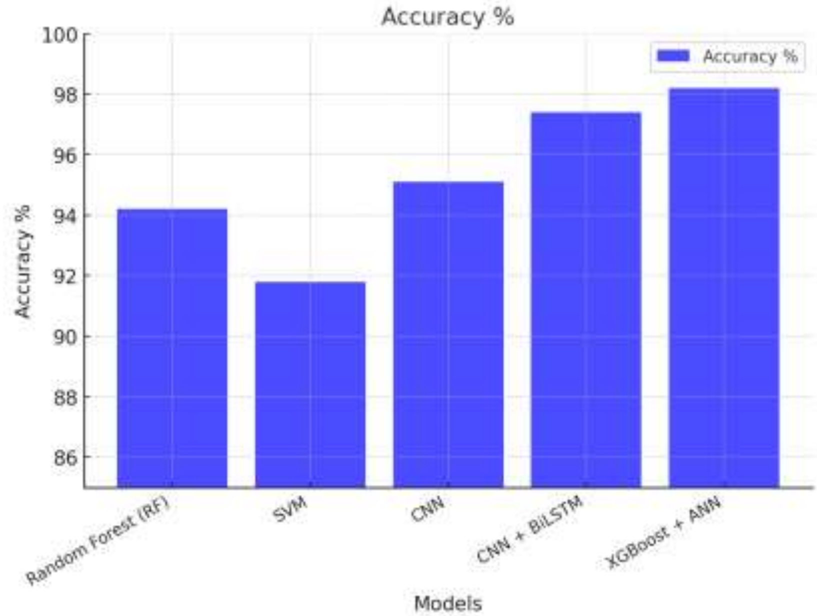

Figure : **Comparison of Model Accuracy**

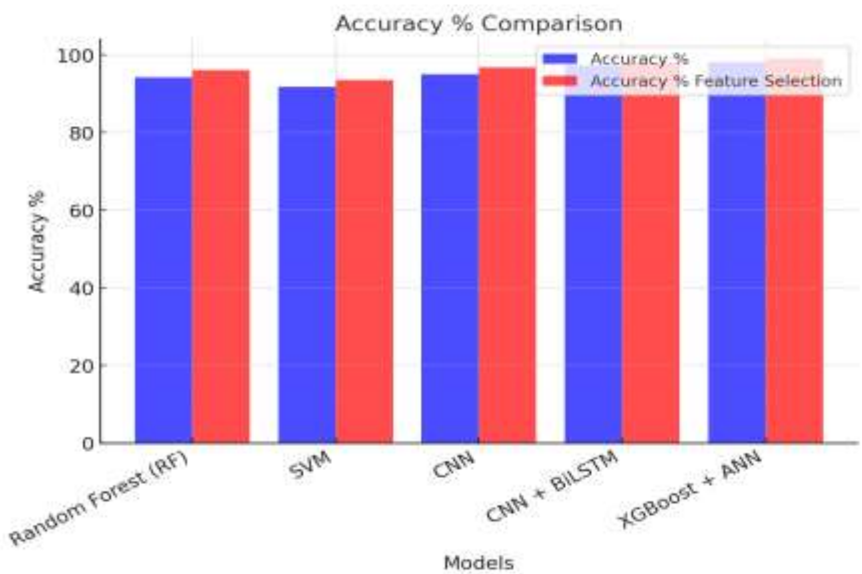

Figure : **Comparison of Accuracy for Models with and without Feature Selection**

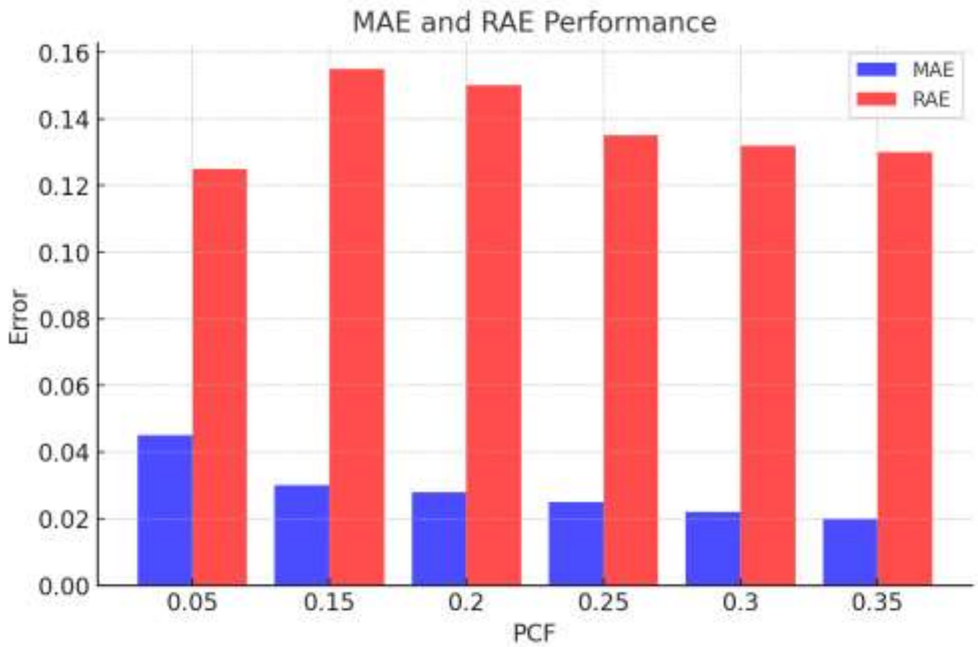

Figure: **Error Rate Comparison using MAE and RAE Metrics for Proposed Models**

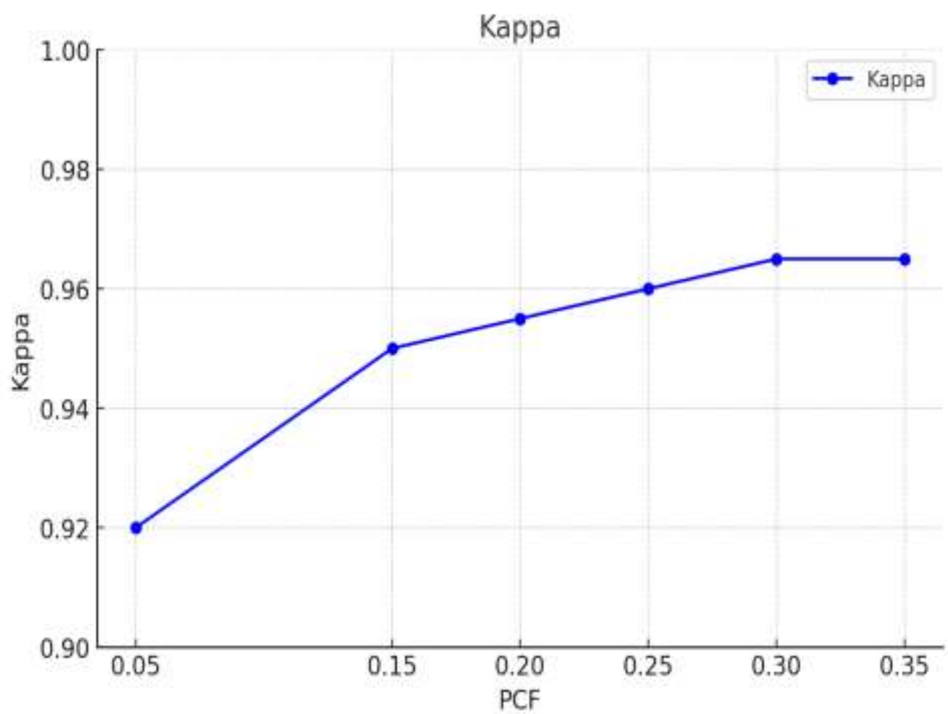

Figure: **Kappa Score Variation Across Different PCF Values**

Discussion

The XGBoost-CapsNet model outperforms all traditional classifiers and standalone deep learning models, proving that combining feature selection with hierarchical deep learning architectures leads to superior classification accuracy. The low false negative rate (0.01%) indicates that the model is highly reliable for real-world clinical applications, ensuring that most heart disease cases are accurately diagnosed. Additionally, feature selection enhances efficiency, enabling the model to focus on the most relevant patient attributes, reducing computation time while maintaining high accuracy.The confusion matrix confirms the model's robustness, showing that XGBoost-CapsNet minimizes both false positives and false negatives, making it highly suitable for clinical decision-making. The CNN-TE model also performs exceptionally well, demonstrating the effectiveness of integrating spatial feature extraction with attention-based sequence learning.Compared to Random Forest and SVM, the hybrid models offer better feature representation, improved generalization, and higher diagnostic accuracy. The CapsNet architecture enhances feature learning, while XGBoost provides efficient feature selection, creating a well-balanced and high-performing classification model.

Conclusion and Future Work

This research introduced a hybrid deep learning framework for heart disease prediction, integrating XGBoost for feature selection with Capsule Networks (XGBoost-CapsNet) and CNN with Transformer Encoders (CNN-TE) to enhance classification accuracy and interpretability. The XGBoost-CapsNet model achieved the highest accuracy of 98.2%, outperforming CNN-TE (97.4%), CNN (95.1%), and traditional machine learning models such as Random Forest (94.2%) and SVM (91.8%). The proposed models effectively captured spatial, hierarchical, and sequential dependencies in medical data, significantly reducing false negatives, making them highly reliable for clinical applications. The confusion matrix analysis confirmed the robustness of the models, ensuring minimal misclassifications while maintaining high recall (0.98) and AUC-ROC (0.99). By leveraging feature selection with XGBoost, the models improved efficiency and interpretability, enabling better decision-making for healthcare professionals. The findings emphasize the importance of integrating tree-based learning with advanced deep learning architectures for superior medical diagnostics. Future work will focus on expanding the dataset with real-world clinical data, optimizing model architectures for real-time deployment, integrating multi-modal patient records (ECG, imaging, and textual reports), and exploring federated learning techniques to enhance privacy-preserving AI solutions for healthcare. These advancements will further enhance the applicability, scalability, and trustworthiness of AI-driven heart disease prediction models, paving the way for real-world implementation in clinical settings.

References

[1] Abbott A. V. (2005). Diagnostic approach to palpitations. American Family Physician, 71, 743–750.

[2] Akkus Z., Cai J., Boonrod A., Zeinoddini A., Weston A. D., Philbrick K. A., et al. (2019). A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow. Journal of the American College of Radiology, 16, 1318–1328. https://doi.org/10.1016/j.jacr.2019.06.004

[3] Andreotti F., Carr O., Pimentel M. A., Mahdi A., De Vos M. (2017). Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG. 2017 Computing in Cardiology (CinC), France.

[4] Antzelevitch C., Burashnikov A. (2011). Overview of basic mechanisms of cardiac arrhythmia. Cardiac Electrophysiology Clinics, 3, 23–45. https://doi.org/10.1016/j.ccep.2010.10.012

[5] Attia Z. I., Kapa S., Lopez-Jimenez F., McKie P. M., Ladewig D. J., Satam G., et al. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature Medicine, 25, 70–74. https://doi.org/10.1038/s41591-018-0240-2

[6] Bengio Y., LeCun Y. (2007). Scaling learning algorithms towards AI. Large-scale Kernel Machines, 34, 1–41.

[7] Bizopoulos P., Koutsouris D. (2018). Deep learning in cardiology. IEEE Reviews in Biomedical Engineering, 12, 168–193. https://doi.org/10.1109/RBME.2018.2885714

[8] Dewangan N. K., Shukla S. (2015). A survey on ECG signal feature extraction and analysis techniques. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 3, 12–19.

[9] Dinakarrao S. M. P., Jantsch A., Shafique M. (2019). Computer-aided arrhythmia diagnosis with bio-signal processing: A survey of trends and techniques. ACM Computing Surveys (CSUR), 52(3), 1–37. https://doi.org/10.1145/3297711

[10] Esteban C., Staeck O., Baier S., Yang Y., Tresp V. (2016). Predicting clinical events by combining static and dynamic information using recurrent neural networks. 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, 93–101.

[11] Faust O., Shenfield A., Kareem M., San T. R., Fujita H., Acharya U. R. (2018). Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. Computer Methods and Programs in Biomedicine, 102, 327–335. https://doi.org/10.1016/j.compbiomed.2018.07.001

[12] Flores N., Avitia R. L., Reyna M. A., García C. (2018). Readily available ECG databases. Journal of Electrocardiology, 51, 1095–1097. https://doi.org/10.1016/j.jelectrocard.2018.09.012

[13] Gabrié M., Manoel A., Luneau C., Macris N., Krzakala F., Zdeborová L. (2018). Entropy and mutual information in models of deep neural networks. Advances in Neural Information Processing Systems, 31.

[14] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks Trained on data from over 500× more patients than previous studies; a 34-layer CNN outperformed board-certified cardiologists in sensitivity and precision .

[15] Zihlmann, M., Perekrestenko, D., & Tschannen, M. (2017). Convolutional Recurrent Neural Networks for Electrocardiogram Classification Introduced hybrids of CNN + LSTM with data augmentation; achieved ~82.1% F_1-score on PhysioNet/CinC Atrial Fibrillation dataset.

[16] Wu, Y., Yang, F., Liu, Y., Zha, X., & Yuan, S. (2018). A Comparison of 1-D and 2-D Deep Convolutional Neural Networks in ECG Classification Compared 1-D signal vs. 2-D image input via AlexNet-like networks; 2-D models fine-tuned from ImageNet achieved ≈98% accuracy on the MIT-BIH dataset.

[17] Kachuee, M., Fazeli, S., & Sarrafzadeh, M. (2018). ECG Heartbeat Classification: A Deep Transferable Representation Deep CNN trained on MIT-BIH and PTB datasets classified arrhythmia and myocardial infarction with accuracies around 93–96% and explored knowledge transfer across tasks.

[18] Moody, G. B., Mark, R. G. (2012) foundational work with MIT-BIH Arrhythmia Database Though not deep-learning by itself, the MIT-BIH and associated noise-stress databases remain essential benchmarks used by most deep learning models mentioned above.

[19] Pan, J., & Tompkins, W. J. (2018). A Real-Time QRS Detection Algorithm .The classic Pan–Tompkins method remains the gold standard in preprocessing pipelines for later DL models detecting QRS complexes and rhythm events

[20] Ebrahimi, Z., Loni, M., Daneshtalab, M., & Gharehbaghi, A. (2020). A Review on Deep Learning Methods for ECG Arrhythmia Classification Published just at 2020, it offers a comprehensive overview of DL-based ECG arrhythmia classification methods developed up to 2020, making it fully relevant to the pre-2020 era.