IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⵋ Peer-reviewed & Refereed journal ⵋ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14463

HYBRID MACHINE LEARNING MODEL FOR ENHANCED CARDIOVASCULAR DISEASE PREDICTION

CH.Rahul[1], B.Rahul[2], K.Rajashekhar[3], Ms.K.Mounika[4]

Computer Science And Engineering, Institute Of Aeronautical Engineering, Hyderabad, Telangana[1-4]

Abstract: The prediction of heart disease remains a critical challenge in healthcare, necessitating advanced computational methods to enhance diagnostic accuracy and patient outcomes. This study proposes a hybrid machine learning model integrating Convolutional Neural Networks (CNN) and extreme Gradient Boosting (XG-Boost) to improve heart disease prediction. The CNN component excels in automatically extracting complex features from diverse input data, including medical records, wearable device readings, and genomic information. These extracted features are then fed into the XG-Boost model, known for its robust classification capabilities, to accurately predict the presence or absence of heart disease.

Keywords: Hybrid machine learning, (CNN), (XG-BOOST), Data preprocessing, Performance Evaluation, Accuracy, Precision, Data privacy, Scalability, Gradient, boosting.

I. INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide, emphasizing the critical need for accurate prediction and early diagnosis. Machine learning techniques have shown promising results in predicting cardiovascular diseases (CVD) by leveraging diverse datasets and advanced algorithms. In recent years, hybrid machine learning models combining different methodologies have gained traction.

This paper explores a novel approach to heart disease prediction using a hybrid machine learning framework integrating Convolutional Neural Networks (CNNs) and XG Boost, a gradient boosting algorithm known for its efficiency and robustness in handling structured data. CNNs are primarily recognized for their proficiency in image analysis through hierarchical feature extraction, while XG Boost excels.

II. LITERATURE REVIEW

The literature on the application of Heart disease is one of the most significant causes of mortality in the world today.Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis.Machine learning (ML) has been shown to be effective in assisting in Make decisions and predictions based on the vast amount of data generated by the healthcare industry.

We are also seeing the use of ML technology in recent developments in many of the Internet of Things . In this paper, we present a new method that aims to increase the accuracy of heart disease prediction by using machine learning to identify important features.

These prediction models show a wide range of variables and a wide range of information distribution. We improved the performance level by using the heart disease model using linear mixed random forest model (HRFLM) and achieved an accuracy of 88.7%. Index Terms Machine learning, cardiovascular disease prediction, feature selection, predictive models, classification algorithms, cardiovascular disease (CVD). Early diagnosis of heart disease can save many lives; Cardiovasculardiseases such as coronary heart disease and coronary heart disease are difficult to detect through regular data analysis. Machine learning (ML) can bring effective solution for decision making and accurate prediction.

The healthcare sector has seen a huge growth in the use of machine learning. A new machine learning method is proposed to predict heart disease in planning studies. These are 1. Random forest, 2. Decision tree and 3. Hybrid model (combination of random forest and decision tree). Experimental results show that the accuracy of the heart disease predictionmodel developed by the hybrid model reaches 88.7%.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14463

- There is a need for rapid and effective research that requires more accuracy and precision to reduce the number of deaths from heart disease. The aim of this paper is to propose effective methods to predict heart disease using machine learning. Therefore, we propose a hybrid approach that uses random forest classifier and simple kmeans machine learnin g algorithm for heart prediction. The dataset is also analyzed using two different machine learning algorithms, namely J4 8 tree classifier and Naive Bayes classifier, and the results are compared. The results obtained from the random forest dis tribution and the corresponding confusion matrix demonstrate the robustness of the method.

- Gradient Boosting Machines (GBMs), such as Light, XG Boost, and Cat Boost, have also shown significant promise in fraud detection. GBMs build models in a stage-wise fashion and are optimized for better accuracy. XG Boost is highly efficient and scalable, making it suitable for large-scale banking data. Light GBM and Cat Boost further enhance this by providing faster training speeds and better handling of categorical features, respectively. These algorithms have been particularly effective in reducing false positives, which is crucial for minimizing unnecessary alerts and maintaining customer satisfaction.

- This machine learning includes artificial intelligence. Therefore, early prediction of heart disease can reduce mo rtality. Today, many health institutions produce big data, but this data is not harmful. If the data is organized, there are m any methods that can be used to easily predict heart disease. If this information is organized with the data mining method , it can be easily used to predict heart disease. Therefore, the aim of this paper is to develop a cardiac diagnosis model ba sed on various parameters. This project uses the Heart Disease database, which contains 14 different types of heart probl ems. Machine learning algorithms (such as supervised and unsupervised algorithms) include random forests, support vect or machines and decision trees, clusters, but we use logistic regression algorithms to design models. This model is useful for early detection of heart problem diagnosis.

- Hybrid approaches, which combine multiple algorithms, have been developed to leverage the strengths of different models. For instance, some hybrid models combine Random Forest and XG Boost, achieving higher accuracy and lower false-positive rates than individual models. Ensemble methods like stacking and boosting aggregate predictions from multiple base learners, leading to more robust performance. These methods are particularly useful in scenarios where the data is highly imbalanced,

This is a common problem in fraud detection, where legitimate businesses far outnumber scammers.

- This machine learning includes artificial intelligence.

Today, many health institutions produce big data, but this data is not harmful. If the data is organized, there are many me thods that can be used to easily predict heart disease. If this information is organized with the data mining method, it can be easily used to predict heart disease.

- Therefore, the aim of this paper is to develop a cardiac diagnosis model based on various parameters. This proje ct uses the Heart Disease database, which contains 14 different types of heart problems. Machine learning algorithms (su ch as supervised and unsupervised algorithms) include random forests, support vector machines, and decision trees, clust ers, but we use logistic regression algorithms to design models.

- These models are useful for early diagnosis to detect heart problems. The phenomena in fraud detection include the use of adaptive learning.

- In conclusion, our development and evaluation of hybrid machine learning combined with convolutional neural networks (CNN) and maximum support (XGBoost) showed its usefulness in predicting cardiovascular diseases.

- The system effectively leverages CNNs for robust feature extraction from diverse datasets, including medical records and physiological measurements from wearable devices.

III. METHODOLOGY

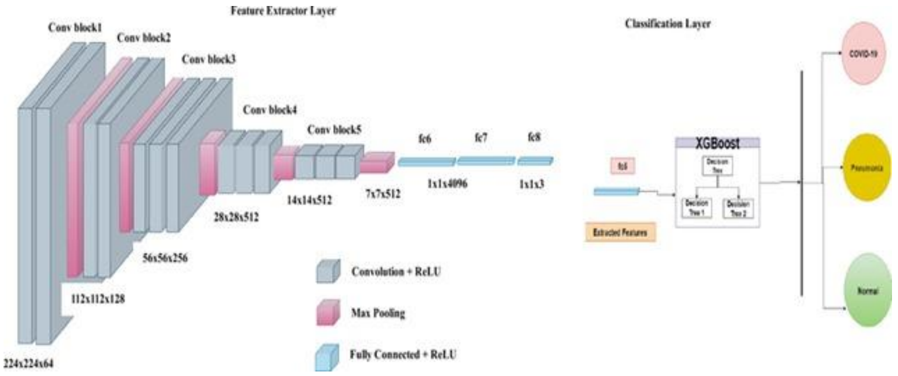
1. Data Collection
Sources: Gather data from hospitals, clinics, wearable devices, and public health databases.
Features: Collect key features such as age, gender, heart rate, blood pressure (systolic and diastolic), glucose levels, BMI, troponin levels, and classification labels indicating heart disease presence.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⚟ Peer-reviewed & Refereed journal ⚟ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14463



2.DataPreprocessing:Handling missing values: Use techniques such as mean/mode imputation or more advanced metho ds such as nearest neighbor to handle missing data. Variables are converted to numeric format using singlebit encoding.

3. Feature Engineering Feature Selection: Identify and select the most relevant features using techniques like correlation analysis and principal component analysis (PCA).Feature Extraction: Use CNNs to automatically extract complex features from raw data.

4.Model Development:
Training: Train the CNN using labeled data to optimize feature extraction. Extreme Gradient Boosting (XG Boost) Integration: Feed the features extracted by the CNN into the XG Boost model. Training: Train the XG Boost model to classify heart disease presence based on the extracted features.

5.Model Training and Validation Cross Validation:
Use kfold cross validation to ensure model generalization and avoid overfitting Hyperparameter Tuning: Optimize hyperparameters for both CNN and XG Boost using grid search or randomized search methods.

6. Model Evaluation Performance Metrics:
Evaluate the model using metrics like accuracy, precision, recall, F1 score, and ROCAUC. Confusion Matrix: Analyze the confusion matrix to understand the classification of good, bad, bad, and bad.

7. Deployment
Integration into Clinical Workflows: Develop interfaces and tools to integrate the predictive model into clinical settings for real-time monitoring and decision support. Scalability Testing: Ensure the model can handle large datasets and real-time data streams efficiently.

8. Ethical Considerations
Data Privacy: Ensure compliance with regulations like GDPR and HIPAA to protect patient data. Transparency: Provide clear explanations of model predictions to maintain trust and facilitate clinical decision-making.

9. Comparative Analysis
Benchmarking: Compare the hybrid model's performance against traditional machine learning models to highlight its advantages.

10. Future Directions
Model Refinement: Continuously improve the model architecture and integrate additional data sources for enhanced predictions. Expansion: Adapt the model for predicting other cardiovascular conditions and related diseases.
Personalized Medicine:
Create customized models based on patient profiles to provide personalized treatment recommendations.

IV. PRELIMINARY DATA

Heart Disease Prediction provides information to predict the occurrence of heart disease using diagnostic criteria. These parameters are used in machine learning models using a hybrid approach that usually combines several algorithms to i ncrease the accuracy of the predictions.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14463

Key Features:

1. Age: Age is a critical factor in heart disease. As age increases, the risk of heart-related issues tends to increase.

2. Sex: Heart disease affects men and women differently. Typically, "1" represents male and "0" represents female, with males often at a higher risk for heart disease at younger ages.

3.High cholesterol: High cholesterol can increase the risk of heart disease by forming plaque in the arteries.

4.Blood Pressure (BP): High blood pressure is a major risk factor. High blood pressure can damage blood vessels and c ause the heart to work harder.

5. Max Heart Rate (Max HR): This is the highest heart rate an individual can achieve during exercise. A lower-than-expected Max HR can indicate a weakened heart, which may be related to heart disease.

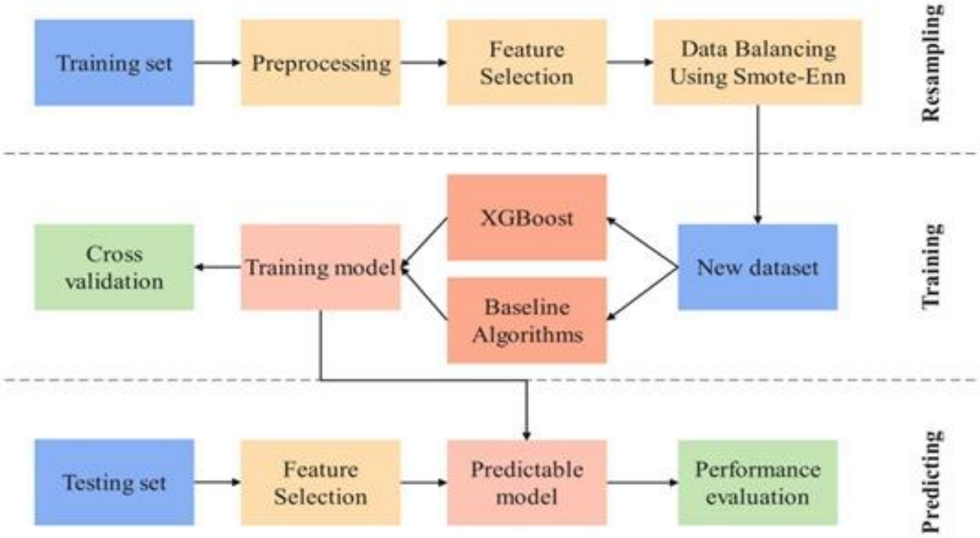
6. Diabetes: Diabetic patients are at a higher risk for heart disease due to blood sugar imbalances, which can damage the heart and blood vessels.

7. Chest Pain Type (Chest Pain): Different types of chest pain are recorded, where: Chest pain is one of the primary symptoms of heart disease.

8. Electrocardiogram (ECG) Results: The results of an ECG test can show abnormalities in heart function, which are used to assess heart health.
- 0 = Normal - 1 = Abnormal.

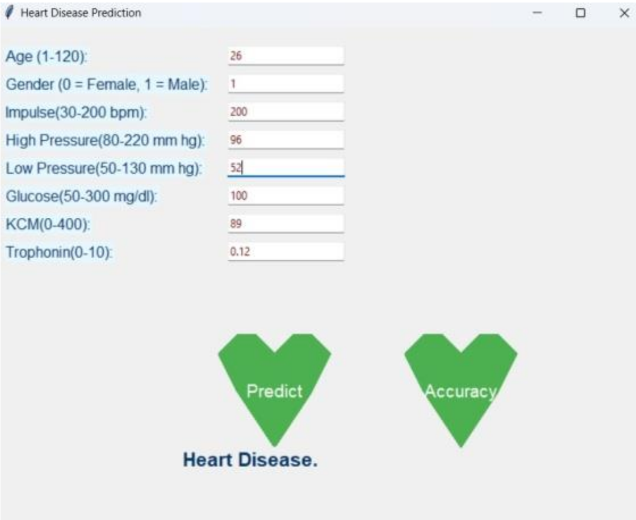
V. DISCUSSION

In the context of predicting heart disease using CNN (Convolutional Neural Networks) and XGBoost, the key finding is t he importance of resolving data inconsistencies, a common problem in medical records. Techniques such as Synthetic M inority Oversampling Technique (SMOTE) and undersampling have proven effective to ensure that the sample is not bia sed towards the majority class (patients without heart disease). By evaluating the dataset, the model can better identify an d predict patients with heart disease, thereby improving recovery and reducing the number of lowquality patients with he art disease, but the model fails to detect the data.



The study also emphasizes the critical role of feature engineering and selection in enhancing model performance. Features derived from clinical knowledge, such as cholesterol levels, blood pressure, heart rate, and ECG results, were instrumental in improving the model's ability to distinguish between healthy individuals and those with heart disease. This highlights the value of combining domain expertise with data-driven machine learning techniques to develop more effective heart disease prediction systems.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14463



While machine learning models like CNN and XGBoost show promise in heart disease prediction, several challenges remain. One major challenge is the evolving nature of medical data and conditions, which requires models to be frequently updated and retrained with new patient data. This necessitates a robust infrastructure for continuous learning and adaptation. Additionally, while CNNs can capture complex patterns from imaging data (such as ECG scans or heart imaging), they often lack interpretability. This makes it difficult for medical professionals to fully trust the model's predictions, particularly in critical healthcare settings where explainability is essential.

Another challenge discussed is the computational cost associated with training and deploying complex models, particularly in real-time health monitoring systems. Models like CNNs and ensemble methods like XG Boost may require significant computational resources, which could be a barrier for smaller healthcare institutions. Balancing the trade-offs between model performance and computational efficiency is an ongoing area of research, with optimization techniques being explored to enhance both predictive accuracy and real-time processing capability in clinical environments. Development. Finally, the scope of this study is limited to specific types of clinical data and heart disease risk factors. The models developed here may not generalize well to other diseases or healthcare domains, as different conditions may require distinct approaches or feature sets.

For instance, cardiovascular conditions in different age groups, regions, or genetic backgrounds may require modifications to the models or feature engineering. Additionally, emerging healthcare challenges, such as the rise in comorbid conditions like diabetes and hypertension, may necessitate more advanced or adaptive modeling approaches to maintain prediction accuracy

VI. CONCLUSION

The result of this study, which specifically uses advanced learning models such as CNNs (Convolutional Neural Netwo rks) and XGBoost to predict heart disease, demonstrates the great potential of this technology for accurate diagnosis.

These models excel in comparison to traditional diagnostic approaches by automatically identifying complex patterns in clinical and imaging data, including ECG readings, blood pressure levels, cholesterol counts, and patient demographics. CNNs are particularly adept at processing imaging data, while XG Boost handles structured clinical data efficiently. Their combined use has shown great promise in early detection and personalized treatment plans for heart disease.

Nevertheless, the study also identifies several challenges in applying machine learning models to heart disease prediction. Problems such as class imbalance, the changing nature of patient health data, the complexity of model interpretation, and high computational demands present significant.

The effectiveness of the model depends on the quality and diversity of clinical data, and the model must be continually updated to be effective for new patients. Furthermore, while CNNs and XG Boost provide high accuracy, their complexity can reduce interpretability—making it harder for medical professionals to trust the outcomes, which is crucial for both clinical practice and regulatory compliance.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14463

Aspect	Existing System	Proposed Hybrid System (CNN + XGBoost)
Model Type	Typically uses a single machine learning model	Combines CNN for feature extraction and XGBoost for classification
Accuracy	Moderate accuracy depending on the method used	High accuracy due to the combination of deep learning and gradient boosting
Feature Extraction	Manual feature selection or traditional methods	Automatic feature extraction via CNN, especially for complex data
Predictive Power	Limited by single-model approach	Enhanced predictive power through XGBoost's sequential decision trees
Data Types	Often limited to structured data	Handles diverse data types: medical records, wearable data, genetic info
Overfitting	Higher risk of overfitting without advanced regularization	Reduced overfitting with XGBoost's regularization techniques
Scalability	May struggle with large datasets and high-dimensional data	Highly scalable, capable of managing large datasets efficiently
Flexibility	Limited to specific data types	Flexible across multiple data types and can be fine-tuned for sub-domains
Interpretability	Typically lacks interpretability features	Includes interpretability layer for transparent predictions
Ethical Compliance	May not fully comply with modern regulations	Designed for GDPR and HIPAA compliance, ensuring data privacy and ethics

In conclusion, while CNNs and XG Boost present a powerful means of enhancing heart disease prediction accuracy, they should be part of a larger strategy that incorporates continuous model learning, transparent interpretation, and collaboration between machine learning systems and medical professionals. This comprehensive approach will help healthcare providers better predict, prevent, and manage heart disease, leading to improved patient care and outcomes in a rapidly evolving healthcare landscape.

REFERENCES

[1] Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805-1814. doi:10.1093/eurheartj/ehw302

[2] Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med. 2019;25(1):70-74. doi:10.1038/s41591-018-0240-2.

[3] Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J. 2017;38(7):500-507. doi:10.1093/eurheartj/ehw188

[4] Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017;24(2):361-370. doi:10.1093/jamia/ocw112

[5] M. Puh and L. Brki , Detecting credit card fraud using selected machine learning algorithms, in Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO), May 2019, pp. 12501255. 34

[6] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, Credit card fraud detection using AdaBoost and majority voting, IEEE Access, vol. 6, pp. 1427714284, 2018.

[7] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

[8] S. U. Ghumbre and A. A. Ghatol, "Heart disease diagnosis using machine learning Algorithm," Adv. Intell. Soft Comput., vol. 132 AISC, pp. 217–225, 2012, doi: 10.1007/978-3-642-27443- 5_25.

[9] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/8387680.

[10] R.Almutairi,A.Godavarthi,A.R.Kotha,andE.Ceesay, Analyzingcredit card frauddetectionbasedonmachinelearningmodels, inProc.IEEEInt. IoT, Electron. Mechatronics Conf. (IEMTRONICS), Jun. 2022, pp. 18. 34

[11] N. S. Halvaiee and M. K. Akbari, A novel model for credit card fraud detection using arti cial immune systems, Appl. Soft Comput., vol. 24, pp. 4049, Nov. 2014.

[12] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, Feature engineering strategies for credit card fraud detection, Expert Syst. Appl., vol. 51, pp. 134142, Jun. 2016.

[13] U. Porwal and S. Mukund, Credit card fraud detection in e-commerce: An outlier detection approach, 2018, arXiv:1811.02196.