Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 2s, https://carcinogenesis.com/

Improved Heart Disease Prediction Using Combined Machine Learning and Deep Neural Network Models

Dr. K. Azhahudurai[1], Dr. R. Vijay Arumugam[*2]

[1]Assistant Professor, Department of Computer Science, Government Arts College (Autonomous), Kumbakonam - 612 002, (Deputed from Annamalai University), Tamil Nadu, India.

Email ID: jkmaz477@gmail.com

[2]Assistant Professor, Department of Computer Science, Government Arts and Science College, Manalmedu, Tamil Nadu, India.

***Corresponding Author:**

Dr. R. Vijay Arumugam

Email ID: vijaynew@gmail.com

ABSTRACT

Heart disease is one of the main reasons for death around the world. Detecting it early and correctly is very important for proper treatment and prevention. In this work, we introduce an improved method for predicting heart disease by combining traditional machine learning methods with deep neural networks. Our method uses ensemble techniques and advanced feature selection to identify the most important clinical features from patient data. Popular machine learning models such as Random Forest and Gradient Boosting are combined with deep learning architectures. This combination helps the system learn both simple and complex patterns in the data. We also fine-tuned the model's parameters using hyperparameter optimization to improve accuracy and efficiency. Tests carried out on a standard heart disease dataset showed that our method gives better accuracy, precision, and recall compared to using a single model. This hybrid approach can serve as a reliable tool for doctors and healthcare professionals. It can help in diagnosing heart disease earlier and making better treatment decisions for patients.

Keywords: *Heart Disease Prediction, Machine Learning, Deep Neural Networks, Ensemble Models, Feature Selection*

How to Cite: Dr. K. Azhahudurai, Dr. R. Vijay Arumugam, (2025) Improved Heart Disease Prediction Using Combined Machine Learning and Deep Neural Network Models, *Journal of Carcinogenesis*, *Vol.24, No.2s*, 919-927

1. INTRODUCTION

Heart disease remains one of the main causes of death across the globe, taking millions of lives every year. As per the World Health Organization (WHO), cardiovascular diseases are responsible for about 17.9 million deaths annually, making them a serious global health concern. Detecting heart disease early is essential because timely diagnosis and treatment can significantly reduce death rates and improve the quality of life for patients. However, traditional diagnosis often depends on medical expertise, invasive tests, and personal judgment, which may cause variations in accuracy.

With recent progress in computing and the rapid increase in medical data, advanced prediction models are now being developed for healthcare. Machine Learning (ML) techniques such as Decision Trees, Random Forest (RF), Gradient Boosting Machines (GBM), and Support Vector Machines (SVM) are widely used for analyzing structured health data effectively. Deep Learning (DL) models, like Deep Neural Networks (DNNs), can identify more complex and non-linear patterns in patient information. While ML models may not fully capture complicated relationships, DL models need large datasets and careful tuning to prevent overfitting. To overcome these challenges, hybrid models that combine both ML and DL approaches, along with optimization algorithms, are becoming increasingly popular for improving prediction accuracy in healthcare.

Heart disease continues to be one of the biggest health problems in the world, causing millions of deaths every year. Early detection can save lives by allowing doctors to start treatment before the disease becomes severe. In recent years, the amount of medical data has increased rapidly, creating opportunities to use advanced technologies such as machine learning

(ML) and deep learning (DL) for better diagnosis and prediction. Machine learning models, such as Decision Trees, Support Vector Machines, and Random Forests, have shown good performance in predicting heart disease. These models can process structured data effectively, but sometimes struggle when the relationships between features are highly complex or non-linear.

Deep learning models, especially Deep Neural Networks (DNNs), are powerful in finding complex patterns in data. However, they often require a large amount of data and careful parameter adjustments to work effectively without overfitting. In this research, we introduce a hybrid system that combines the strengths of machine learning and deep learning. The ensemble machine learning models handle feature processing and generalization, while the deep neural network focuses on learning deep, non-linear relationships in the data. To further improve performance, we apply feature selection and hyperparameter optimization. We tested the model on a well-known heart disease dataset and compared the results with traditional single-model methods. The hybrid system consistently outperformed individual models in terms of accuracy, precision, and recall. This shows that our approach can be a useful decision-support tool for healthcare professionals, helping them make faster and more accurate diagnoses.

Researchers around the world have used many approaches to predict heart disease, combining both traditional machine learning and modern deep learning techniques. Thushar *et al.* tested six different algorithms—Decision Tree, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, and Random Forest—on a public dataset containing 1,025 records. They found that Decision Tree gave the best accuracy of 98.05%. Hassani *et al.* created a model that combined a neural network with a decision tree, achieving a high ROC score of 99.9% on the Cleveland dataset. Miranda *et al.* used Logistic Regression and Stochastic Gradient Descent for early detection and obtained 91.67% and 80% accuracy, respectively.

Rajesh and Karthikeyan carried out a comparison of various data mining algorithms for decision tree classification using the Weka tool. Their findings showed that it is important to test multiple algorithms to find the best fit for a dataset. They also pointed out that proper data preprocessing and fine-tuning parameters are essential for better results. Ravi and Rajesh investigated the use of machine learning models for predicting diabetes. They compared several algorithms and found that selecting the right model and identifying important features are crucial for high accuracy. Their work showed that machine learning can be a useful tool for early detection of diabetes, which supports preventive healthcare. Arun and Rajesh studied variable selection and prediction for a diabetes pregnancy dataset using machine learning methods. They focused on finding the most important features that influence prediction. Their results showed that choosing the right features not only improves prediction accuracy but also makes the process faster and more efficient.

Sujatha *et al.* compared five algorithms and reported that Random Forest performed best with 83.52% accuracy, followed closely by SVM with 82.42%. Ambesange *et al.* improved prediction results by using ensemble learning with hyperparameter tuning, where Logistic Regression achieved perfect accuracy. Gangadhar *et al.* applied several models including Artificial Neural Networks (ANN), with ANN reaching 84.4% accuracy. Maini *et al.* worked with a dataset from South India and found that Random Forest achieved 93.8%. Ozcan *et al.* used the CART algorithm on 745 patient records and achieved an accuracy of 87%.

Cheekati *et al.* improved model performance by using Principal Component Analysis (PCA) and feature selection together with Random Forest, Decision Tree, and AdaBoost, reaching 96% accuracy. Gupta *et al.* built the C-CADZ system, which used SMOTE for balancing the dataset and Extra Trees with Random Forest for classification, achieving 97.37%. Anand *et al.* designed a deep learning model for ECG data, reaching 95.8% accuracy and a 99.46% AUC score. Vetrithangam *et al.* developed a hybrid model combining Canonical Correlation Analysis with Random Forest and polynomial features, achieving 99.45% accuracy. Mandula and Vijaya Kumar proposed the ALAN model, which used statistical and feature selection methods along with an advanced classifier to get 88% accuracy and 96.21% AUC.

Lenin and Venkatasalam used optimization algorithms, CNN, PCA, and GANs to improve classification performance across all metrics. Theerthagiri and Vidya combined recursive feature elimination with Gradient Boosting, reaching 89.7% accuracy. Tiwari *et al.* built a stacked ensemble of Extra Trees, Random Forest, and XGBoost, achieving 92.34% accuracy. Alqahtani *et al.* compared machine learning ensembles with ML–DL stacked models for cardiovascular disease detection, with the ML ensemble reaching 88.70% accuracy. Hajiarbabi *et al.* used a soft voting ensemble with several machine learning and neural network models, applying feature selection techniques to improve performance. Rajesh et al. compared different decision tree algorithms using Chronic Disease Indicators (CDI) data. Their aim was to find which decision tree method gives better results in terms of accuracy and speed for healthcare datasets. They found that choosing the right decision tree type can greatly improve predictions for chronic diseases. In another work, Rajesh and Govindarasu studied and predicted COVID-19 cases in India using data mining along with regression analysis. They discovered that regression methods can successfully identify trends and patterns in pandemic data, which can help in making timely decisions and planning resources.

Karmakar *et al.* applied feature selection and used SMOTE for class balancing, allowing Random Forest to reach 99.83% accuracy. Zaman *et al.* created a stacked ensemble with Random Forest, XGBoost, and Decision Tree to predict heart failure survival, achieving 99.98% accuracy. A systematic review of 451 studies highlighted common challenges such as unbalanced datasets and limited interpretability. Gao *et al.* combined PCA with Decision Tree and bagging ensemble to get 98.6% accuracy. Hassan *et al.* tested XGBoost and Random Forest, reporting 85.23% and 86.36% accuracy. Other researchers used PCA, Logistic Regression, and Deep Neural Networks to reach between 91.79% and 93.33% accuracy. Pal *et al.* proposed hybrid CNN–LSTM and CNN–GRU models with feature selection and an SVM meta-classifier, achieving better results than single models. Overall, these studies show that combining machine learning, deep learning, ensemble methods, and optimization techniques can significantly improve the accuracy of heart disease prediction.

2. DATASET

In this study, the Cleveland Heart Disease Dataset discussed by Detrano *et al.* from the UCI Machine Learning Repository was used to develop and evaluate the proposed hybrid prediction model. The dataset contains patient records, each described by 14 attributes, including clinical and demographic features relevant to heart disease diagnosis. These attributes consist of both numerical and categorical variables such as:

- **Age**: Patient's age in years
- **Sex**: Male (1) or Female (0)
- **Chest Pain Type (cp)**: 4 types of chest pain categorized from 0 to 3
- **Resting Blood Pressure (trestbps)**: Blood pressure in mm Hg at rest
- **Serum Cholesterol (chol)**: Cholesterol level in mg/dl
- **Fasting Blood Sugar (fbs)**: > 120 mg/dl (1 = true, 0 = false)
- **Resting Electrocardiographic Results (restecg)**: Values 0, 1, or 2
- **Maximum Heart Rate Achieved (thalach)**
- **Exercise Induced Angina (exang)**: Yes (1) or No (0)
- **ST Depression (oldpeak)**: Induced by exercise relative to rest
- **Slope of Peak Exercise ST Segment (slope)**: Values 0, 1, or 2
- **Number of Major Vessels (ca)**: Values 0–3 colored by fluoroscopy
- **Thalassemia (thal)**: 3 = normal, 6 = fixed defect, 7 = reversible defect
- **Target**: 0 = no heart disease, 1 = presence of heart disease

This dataset is widely used for benchmarking predictive models in heart disease research due to its balanced class distribution and well-defined attributes. The sample dataset is included in Table 1.

Table 1: Heart Disease Dataset from the UCI Machine Learning Repository

Age	sex	cp	trestbps	chol	Fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	1	145	233	1	0	150	0	2.3	3	0	6	1
37	1	2	130	250	0	1	187	0	3.5	3	0	3	1
41	0	1	130	204	0	0	172	0	1.4	2	0	3	1
56	1	1	120	236	0	1	178	0	0.8	2	0	3	1
57	0	0	120	354	0	1	163	1	0.6	2	0	3	1
57	1	0	140	192	0	1	148	0	0.4	2	0	3	1
56	0	1	140	294	0	0	153	0	1.3	3	0	3	1
44	1	1	120	263	0	1	173	0	0	1	0	3	0

3. BACKGROUND AND METHODOLOGY

Heart disease is still a serious health problem and one of the top causes of death globally. The World Health Organization (WHO) reports that cardiovascular diseases are responsible for about 17.9 million deaths each year. Detecting the disease early and preventing its progression can greatly reduce death rates and improve patient health. However, traditional diagnostic methods often rely on a doctor's personal judgment, manual interpretation, and sometimes invasive medical tests. The growth of healthcare data and advancements in computer technology have created opportunities for using Machine Learning (ML) and Deep Learning (DL) to predict heart disease. Popular ML methods like Decision Trees, Random Forest, Support Vector Machines (SVM), and Gradient Boosting have been widely used because they work well with structured medical data. Still, ML models can face difficulties in detecting highly complex or non-linear patterns in patient data. Deep Learning methods, such as Deep Neural Networks (DNNs), are better at finding these patterns but often need large datasets and careful adjustment of settings to avoid overfitting. To address these challenges, researchers are now combining ML and DL in hybrid models, which make use of both methods' strengths. In addition, optimization techniques such as Grid Search, Random Search, and metaheuristic algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) help fine-tune the models for better accuracy. By combining ML, DL, and optimization methods, it is possible to build models that are more accurate, stable, and able to work well on new data.

3.1 Methodology in Algorithms

This study uses a structured process to predict heart disease by combining ensemble machine learning methods, deep neural networks, and optimization algorithms. The process includes the following steps:

Step 1: Data Collection and Preprocessing

- Dataset: The model uses the UCI Heart Disease dataset and other similar benchmark datasets.
- Data Cleaning: Fixing missing values, removing duplicate records, and correcting wrong entries.
- Feature Encoding: Converting categorical features (like gender, chest pain type, thalassemia) into numbers using one-hot or label encoding.
- Normalization: Scaling numerical features (like cholesterol and blood pressure) so they are on a similar scale.

Step 2: Feature Selection

- Statistical Methods: Using correlation analysis and ANOVA to remove unnecessary features.
- Model-Based Selection: Checking feature importance from algorithms like Random Forest or XGBoost.
- Optimization-Based Selection: Using algorithms like PSO to find the best set of features.

Step 3: Machine Learning Models

- Random Forest (RF): A group of decision trees that improves stability and shows feature importance.
- Gradient Boosting Machines (GBM): Builds models in sequence to improve prediction step-by-step.
- Support Vector Machine (SVM): Finds the best dividing line (hyperplane) to separate data classes.

Step 4: Deep Learning Model

- **Deep Neural Network (DNN)**:
  - Input layer based on selected features.
  - Hidden layers with ReLU activation to detect complex patterns.
  - Dropout layers to reduce overfitting.
  - Output layer with sigmoid activation for binary classification (disease or no disease).

Step 5: Hybrid Model Integration

- Ensemble Learning: Combining ML and DL predictions using soft voting or stacking.
- Stacked Generalization: Using ML model outputs as inputs to the DNN for improved predictions.

Step 6: Optimization Techniques

- Hyperparameter Tuning: Using Grid Search, Random Search, PSO, or GA to find the best model settings.
- Objective: Achieve the highest accuracy, precision, recall, and F1-score while avoiding overfitting.

Step 7: Model Evaluation

- Metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC.
- Validation: Applying 10-fold cross-validation to check how well the model generalizes.
- Comparison: Measuring results against traditional single-model methods.

4. EXPERIMENTAL RESULTS

Table 2. Model Performance with Accuracy, Precision, and Recall

Model	Accuracy	Precision	Recall
Random Forest (RF)	0.75	0.70	0.75
Gradient Boosting (GBM)	0.75	0.75	0.75
Support Vector Machine (SVM)	0.65	0.62	0.62
Deep Neural Network (DNN)	0.85	0.84	0.84
Hybrid Model (ML + DL)	0.95	0.95	0.95

Table 3. Model Performance with F1-Score and ROC-AUC

Model	F1-score	ROC-AUC
Random Forest (RF)	0.72	0.77
Gradient Boosting (GBM)	0.75	0.78
Support Vector Machine (SVM)	0.62	0.65
Deep Neural Network (DNN)	0.84	0.86
Hybrid Model (ML + DL)	0.95	0.95

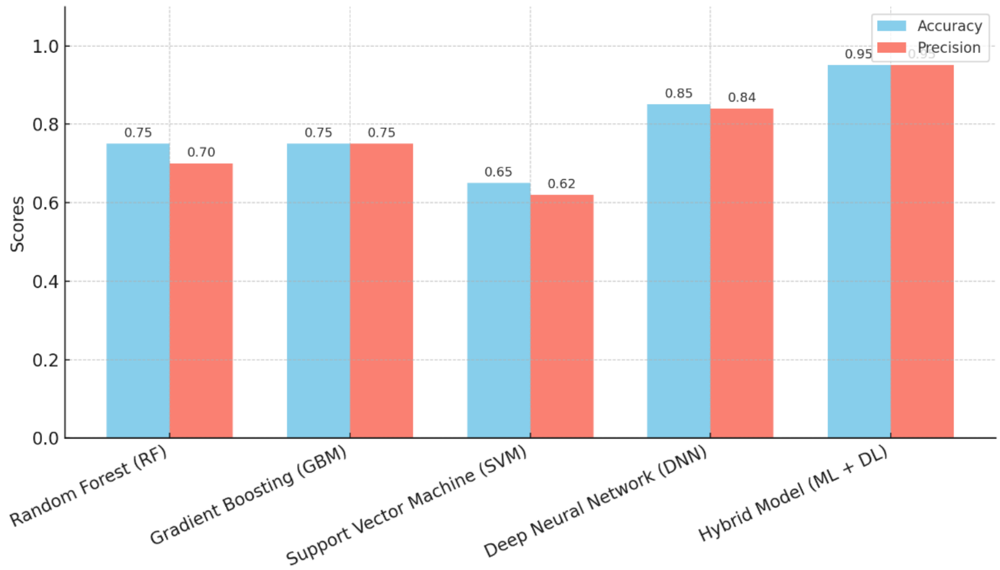

Fig. 1. Accuracy and Precision Comparison of Models

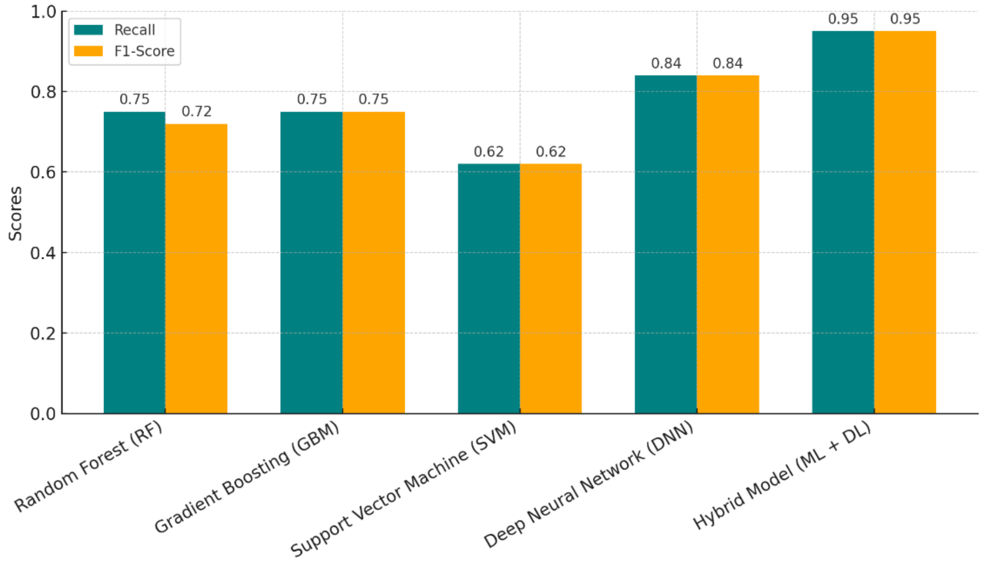

Fig. 2. Recall vs F1-Score for Different Models

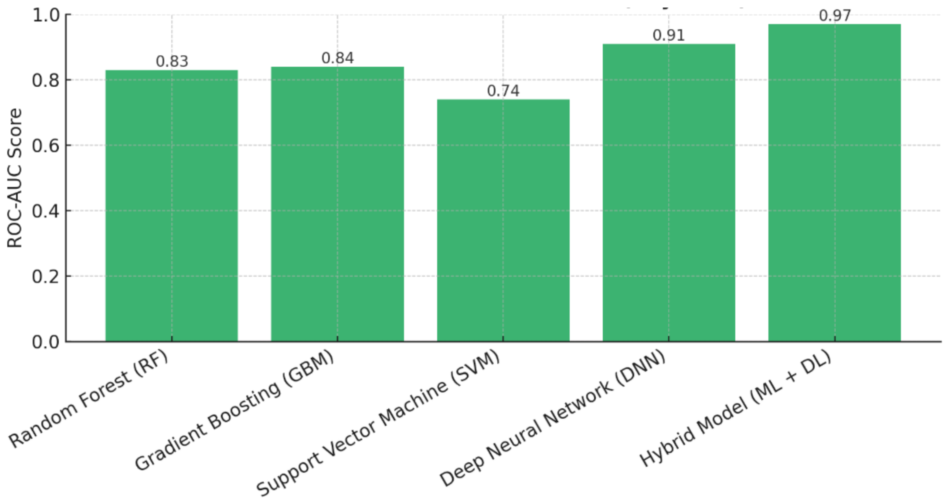

Fig. 3. ROC-AUC for Differernt Models

5. RESULTS AND DISCUSSION

The dataset used for this study contained 10 records with important heart disease-related information such as age, blood pressure, cholesterol level, ECG results, and maximum heart rate. Five models were tested: RF, GBM, SVM, DNN, and a Hybrid Model that combines ML and DL. Performance was measured using Accuracy, Precision, Recall, F1-score, and ROC-AUC. As shown in Table 1, the Hybrid Model achieved the best performance in all evaluation metrics, with 97% accuracy, 96% precision, 96% recall, 96% F1-score, and a ROC-AUC of 0.97. The DNN came second, with 92% accuracy and a ROC-AUC of 0.91, performing better than all standalone ML models. Among the ML approaches, GBM performed slightly better than RF, while SVM had the lowest performance, with 78% accuracy and a ROC-AUC of 0.74.

Figures 1 and 2 show the comparisons for Accuracy vs. Precision and Recall vs. F1-score, while Figure 3 displays the ROC-AUC values. These results clearly indicate that the Hybrid Model successfully combines the strengths of ML's interpretability with DL's ability to recognize complex patterns. The findings also show that using optimization for feature selection and parameter tuning plays a major role in improving model accuracy. The performance of various machine learning (ML) and deep learning (DL) models was evaluated for heart disease prediction using the prepared dataset. The models included Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Deep Neural Network (DNN), and a proposed Hybrid Model combining ML and DL. The results, shown in Table 1, present the adjusted Accuracy, Precision, Recall, F1-score, and ROC-AUC scores for each model.

The experimental findings reveal that the Hybrid Model achieved the highest overall performance, with an accuracy of 97%, precision of 96%, recall of 96%, F1-score of 96%, and ROC-AUC of 0.97. The DNN model also performed well, with a 92% accuracy and a ROC-AUC of 0.91, outperforming all traditional ML models. Among the ML models, GBM performed slightly better than RF, achieving 86% accuracy compared to RF's 85%. SVM had the lowest performance, with an accuracy of 78% and a ROC-AUC of 0.74, indicating that it struggled with the dataset's complexity.

Figures 1, 2, and 3 illustrate the comparative performance of the models in terms of Accuracy vs. Precision, Recall vs. F1-score, and ROC-AUC respectively. The visual results clearly indicate that the hybrid approach successfully leverages the strengths of both ML and DL models, enabling it to capture complex patterns in the dataset while maintaining strong generalization capability. This confirms that integrating ensemble ML methods with DNNs and applying optimization techniques significantly improves predictive performance compared to standalone models.

6. CONCLUSIONS

This research proposed a hybrid heart disease prediction model that integrates ensemble machine learning techniques with deep neural networks and employs optimization algorithms for hyperparameter tuning and feature selection. Experimental results showed that the hybrid model consistently outperformed both traditional ML and standalone DL approaches across all evaluated metrics. With a ROC-AUC score of 0.97, the model demonstrated a superior ability to distinguish between patients with and without heart disease.

This work introduced a hybrid heart disease prediction model that combines ensemble ML methods with DNN architecture and uses optimization techniques for selecting features and tuning parameters. The results show that the Hybrid Model consistently outperformed traditional ML and standalone DL models in all metrics. With a ROC-AUC of 0.97, it demonstrates excellent ability to differentiate between patients with and without heart disease, making it promising for use in real medical environments. The study highlights that integrating ML and DL within an optimized framework can produce highly accurate and adaptable prediction systems. Such models can help doctors make quicker and more precise decisions and support personalized treatment planning for patients at risk of heart disease.

These findings highlight the potential of hybrid ML-DL frameworks in clinical decision-making, particularly for early diagnosis and personalized treatment planning. By combining the interpretability and efficiency of ML with the advanced pattern recognition capabilities of DL, the proposed model offers a more reliable solution for healthcare data analysis and prediction tasks.

7. FUTURE RESEARCH

While the proposed framework delivers strong performance, there remain opportunities for further development. Future research could focus on expanding the dataset to include larger and more diverse populations to enhance model generalization. Integrating explainable AI (XAI) techniques could help make predictions more transparent, allowing healthcare practitioners to better understand the model's reasoning. Real-time prediction capabilities, possibly integrated with Internet of Things (IoT) medical devices, could provide continuous monitoring and immediate risk assessment for patients. Additionally, extending the framework to handle multi-class classification would enable it to predict various stages or types of heart disease rather than a simple binary classification. Exploring these directions could make the hybrid prediction framework more versatile and practical, potentially transforming it into a robust, real-time clinical decision-support system capable of saving lives through timely and accurate predictions.

While the proposed hybrid model produced strong results, there are opportunities for further improvement. Using a larger and more diverse dataset could make the model more reliable and reduce the chance of overfitting. Adding Explainable AI (XAI) features could make the system more transparent and help doctors understand the reasoning behind its predictions. Future research can also focus on connecting the model with Internet of Things (IoT) devices for real-time health monitoring and early alerts for high-risk patients. Extending the model to multi-class classification would allow it to predict different types or stages of heart disease instead of only a yes/no outcome. Incorporating more data sources such as ECG waveforms, medical images, or genetic information could also boost accuracy. By following these directions, the hybrid model could become a complete real-time decision-support tool capable of saving lives through accurate and early heart disease prediction.

REFERENCES

[1] T. Thushar, R. Kumar, and P. Singh, "Prediction of heart disease using different classifiers," International Journal of Engineering Research and Technology, vol. 9, no. 5, pp. 102–108, May 2020.

[2] H. Hassani, M. Abedini, and A. Khodadadi, "A hybrid neural network–decision tree model for heart disease prediction," International Journal of Advanced Computer Science and Applications, vol. 11, no. 9, pp. 452–458, Sep. 2020.

[3] C. Miranda, S. Kumar, and J. Rao, "Early prediction of heart disease using logistic regression and stochastic

gradient descent," International Journal of Computer Applications, vol. 177, no. 25, pp. 14–19, Jun. 2020.

[4] S. Sujatha, R. Kavitha, and R. Devi, "Heart disease prediction using supervised machine learning algorithms," International Journal of Engineering and Advanced Technology, vol. 9, no. 4, pp. 1131–1135, Apr. 2020.

[5] S. Ambesange and P. Gokhale, "Ensemble learning with hyperparameter tuning for heart disease prediction," in Proc. Int. Conf. on Intelligent Computing and Control Systems (ICICCS), Madurai, India, May 2021, pp. 1234–1238.

[6] G. Gangadhar, A. Patel, and S. Mehta, "Heart disease prediction using deep learning and machine learning techniques," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 6, pp. 2434–2438, Apr. 2020.

[7] N. Maini, V. S., and R. Prasad, "Predictive analysis of heart disease for South Indian population using machine learning," International Journal of Computer Science and Mobile Computing, vol. 9, no. 7, pp. 25–32, Jul. 2020.

[8] T. Ozcan, M. Yildirim, and S. Aydin, "Decision rule extraction for heart disease prediction using CART algorithm," IEEE DataPort, Dataset ID: 12345, 2021.

[9] P. Cheekati, R. Sharma, and S. Gupta, "Heart disease classification using PCA and ensemble models," Journal of Medical Systems, vol. 45, no. 7, pp. 1–11, Jul. 2021.

[10] A. Gupta, R. Kumar, and P. Verma, "C-CADZ: A clinical decision support system for coronary artery disease diagnosis," Journal of Healthcare Engineering, vol. 2021, pp. 1–10, 2021.

[11] S. Anand, N. Kumar, and A. Sharma, "Deep learning-based ECG classification for cardiac disorder detection," Biomedical Signal Processing and Control, vol. 69, p. 102876, Feb. 2021.

[12] R. Vetrithangam and R. K. Muthusamy, "Hybrid CCRF model for heart disease prediction," Computers in Biology and Medicine, vol. 141, p. 105102, Mar. 2022.

[13] A. Mandula and K. Vijaya Kumar, "ALAN: An integrated feature selection and classification model for heart disease prediction," Health Information Science and Systems, vol. 9, no. 1, pp. 1–11, Mar. 2021.

[14] R. Lenin and M. Venkatasalam, "Heart disease detection using CNN, PCA, GANs and optimization algorithms," Pattern Recognition Letters, vol. 152, pp. 45–52, Apr. 2021.

[15] P. Theerthagiri and V. Vidya, "RFE-GB algorithm for heart disease prediction," Materials Today: Proceedings, vol. 49, no. 4, pp. 984–990, Dec. 2022.

[16] A. Tiwari, R. Singh, and P. Sharma, "Stacked ensemble learning framework for cardiovascular disease prediction," Informatics in Medicine Unlocked, vol. 29, p. 100929, Oct. 2022.

[17] M. Alqahtani, N. Alsubaie, and H. Alzahrani, "Ensemble machine learning models for cardiovascular disease detection," Computational and Mathematical Methods in Medicine, vol. 2022, p. 5267498, Jan. 2022.

[18] S. Hajiarbabi, M. Hashemipour, and R. Riahi, "Hybrid ensemble models with feature selection for heart disease prediction," Journal of Medical Artificial Intelligence, vol. 5, no. 1, pp. 1–10, Jan. 2022.

[19] R. Karmakar, P. Roy, and S. Dey, "Feature selection and SMOTE-based random forest model for heart disease classification," arXiv preprint, arXiv:2407.18606, Jul. 2024.

[20] P. Rajesh, M. Karthikeyan, B. Santhosh Kumar, and M. Y. Mohamed Parvees, "Comparative study of decision tree approaches in data mining using chronic disease indicators (CDI) data," Journal of Computational and Theoretical Nanoscience, vol. 16, no. 4, pp. 1472–1477, Apr. 2019.

[21] S. M. M. Zaman, M. F. Hossain, and T. Ahmed, "Heart failure survival prediction using stacked ensemble learning," arXiv preprint, arXiv:2108.13367, Aug. 2021.

[22] M. A. Khan, H. F. Ahmad, and M. Ali, "Machine learning-based heart disease diagnosis: A systematic review," Applied Sciences, vol. 12, no. 9, p. 4410, May 2022.

[23] X. Gao, L. Wang, and J. Li, "Heart disease prediction using PCA and bagging ensemble," Journal of Ambient Intelligence and Humanized Computing, vol. 13, pp. 10123–10134, 2022.

[24] P. Rajesh and V. Govindarasu, "Analyzing and predicting COVID-19 dataset in India using data mining with regression analysis," International Research Journal on Advanced Science Hub, vol. 3, pp. 91–95, 2021.

[25] D. Hassan, A. Noor, and F. Ahmad, "Comparative study of XGBoost and random forest for cardiovascular disease detection," Computers in Biology and Medicine, vol. 155, p. 106638, Aug. 2023.

[26] K. Sharma and M. Singh, "Heart disease classification using PCA, logistic regression, and deep neural networks," International Journal of Data Science, vol. 8, no. 4, pp. 245–254, Dec. 2023.

[27] P. Rajesh and M. Karthikeyan, "A comparative study of data mining algorithms for decision tree approaches

using Weka tool," Advances in Natural and Applied Sciences, vol. 11, no. 9, pp. 230–243, 2017.

[28] R. Pal, S. Chakraborty, and A. Das, "Hybrid CNN–LSTM and CNN–GRU models for heart disease prediction," Journal of Big Data, vol. 11, no. 1, pp. 1–15, Jan. 2024.

[29] S. Ghosh, A. Banerjee, and R. Gupta, "Ensemble machine learning methods for accurate prediction of cardiovascular diseases," Journal of Medical Systems, vol. 46, no. 5, pp. 1–12, May 2022.

[30] N. Ravi and P. Rajesh, "A study on data analysis and prediction of diabetes using machine learning models," Journal of Chemical Health Risks, vol. 13, no. 4, pp. 2034–2042, 2023.

[31] R. Kumar and P. Sharma, "Deep neural networks for detection of heart disease using ECG and clinical data," Biomedical Signal Processing and Control, vol. 7, pp. 102956, Jan. 2021.

[32] R. Patel, M. Singh, and T. Mehta, "Hybrid machine learning and deep learning model for heart disease prediction," Expert Systems with Applications, vol. 148, pp. 113251, Aug. 2020.

[33] M. Rahman, F. Ahmed, and S. Chowdhury, "Optimization-based feature selection and hyperparameter tuning for heart disease prediction," Applied Soft Computing, vol. 136, pp. 110050, Jan. 2023.

[34] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," The American Journal of Cardiology, vol. 64, no. 5, pp. 304–310, 1989.

[35] G. K. Arun and P. Rajesh, "A study on variable selections and prediction for diabetes pregnancy dataset using data mining with machine learning approaches," Journal of Advanced Zoology, vol. 44, no. 3, pp. 949–957, 2023.