A Hybrid Machine Learning Approach for Heart Disease Prediction: Integrating Random Forest and Multi-Layer Perceptron for Enhanced Diagnostic Accuracy

Joyassree Sen

Department of Computer Science and Engineering
Islamic University, Kushtia, Bangladesh

Dr. Paresh Chandra Barman

Department of Information and Communication Technology
Islamic University, Kushtia, Bangladesh

Abstract

Heart disease remains one of the leading causes of death globally, affecting millions of lives each year and placing a significant burden on healthcare systems. Early and reliable prediction of heart conditions is essential for preventing severe outcomes and enabling timely medical intervention. This study presents a hybrid machine learning approach that integrating a Random Forest (RF) classifier with a Multi-Layer Perceptron (MLP) meta-learner to enhance diagnostic accuracy. The RF model is first trained on standardized clinical features to produce probabilistic outputs, which are then used as inputs to the MLP for final prediction. The dataset, derived from a publicly available heart disease repository, was preprocessed with feature scaling and class balancing techniques to ensure fairness and model robustness. On the test set, the proposed model achieved an accuracy of 84%, with an F1-score of 0.86 for detecting heart disease and a recall of 0.91, indicating strong sensitivity. The average precision (AP) score of 0.9143 on the precision-recall curve confirms the model's effectiveness, particularly in identifying true positives in imbalanced data. These results suggest that the proposed ensemble-based approach can be a valuable tool in clinical decision support systems, helping healthcare professionals identify high-risk patients with greater confidence.

Keywords: Heart disease prediction, Meta-learning, Random Forest, Multi-Layer Perceptron, Ensemble learning, Machine learning, Clinical decision support, Precision-recall, Medical diagnosis, Class imbalance

1. INTRODUCTION

Worldwide heart disease stands as the leading reason for mortality since the World Health Organization reports annual deaths amount to approximately 18 million. Heart disease-related deaths create substantial stress for global healthcare institutions primarily because preventive care facilities are scarce in low- to middle-income nations [1]. The diagnosis and prediction of heart disease at early stages provides the most effective approach to minimize fatal consequences because of prompt intervention therapy.

The development of artificial intelligence (AI) and machine learning (ML) offers modern opportunities to improve both the accuracy and speed of medical diagnoses, particularly in the field of cardiology. Machine learning models can identify compelling patterns in patient data and often achieve higher prediction accuracy than traditional statistical methods. Ungrammatical sentence. The Random Forest (RF) classifier, an ensemble learning method, is known for its robustness, high accuracy, and strong ability to prevent overfitting—making it one of the most reliable machine learning techniques. The prediction output of a Random Forest (RF) model is based on the combined results of multiple decision trees, making its forecasts more stable and reliable [3]. This technique is widely used in medical research because it effectively handles clinical records, which often contain high-dimensional and noisy data [4]. The

capabilities of Random Forest (RF) are limited when it comes to identifying nonlinear relationships between features, which are common in biological systems.

The limitation is resolved through the application of Multi-Layer Perceptrons (MLPs) within deep learning methods. Artificial neural networks of the feedforward and hierarchical representation type known as MLPs have demonstrated high capability to learn complex data structures. The results of these models have been successfully applied in various healthcare applications, including electrocardiogram (ECG) classification, diabetic retinopathy detection, and cancer diagnosis [5]. The appropriate training of MLPs enables them to detect unique patterns within patient records that standard models might miss. The data-intensive nature of MLPs becomes a problem when combined with challenging medical datasets whose disease cases comprise a small percentage of the patient population.

Research into hybrid learning frameworks has emerged to combine the strengths of Random Forest (RF) and Multilayer Perceptron (MLP) models while minimizing the weaknesses of each when used individually. Within the frameworks a base model receives training to retrieve intermediate output or probability values that become inputs for a meta-learner to perform the final classification. This meta-learning approach helps physical therapists expand their generalization capabilities by providing an organized fusion method for heterogeneous learning systems [6]. Hospital heart disease prediction benefits from ensemble methods, which have been shown to outperform individual models in terms of sensitivity, specificity, and F1-score metrics [7]. For reliable implementation of models in practical scenarios, the process requires class adjustment, feature transformation, and training cessation management to achieve optimal outcomes. ROC curves and precision-recall scores are essential evaluation metrics, as they provide a more reliable performance assessment, especially in situations where accuracy alone is unreliable due to class imbalance [8].

The research develops a hybrid two-stage model that combines the probabilistic outputs of a Random Forest classifier with the learning capabilities of a Multi-Layer Perceptron. The Random Forest (RF) model receives preprocessing heart disease data before performing training on the data to obtain class-based probability distributions. The trained MLP receives these probabilities from the Random Forest as inputs for its final prediction process. The proposed combination method takes advantage of RF classical decision systems with neural network ability to learn complex patterns. The identified hybrid technique demonstrates effectiveness through superior recall and precision metrics, highlighting its potential for use in clinical support systems to identify high-risk individuals early, before the onset of heart disease.

2. RELATED WORK

As advancements in healthcare data accessibility and computational power continue, sophisticated machine learning algorithms play an increasingly crucial role in assessing the risk of cardiac disease. These advanced systems improve the analysis of medical data, enabling earlier detection and more accurate risk assessments. Medical diagnostic procedures today need invasive processes and a high degree of professional expertise. The drive to discover data-based solutions exists because researchers want to help clinicians get accurate and timely risk assessments.

The application of supervised learning algorithms throughout studies exists to forecast heart disease cases. Logistic regression and decision trees are widely used in predictive models because they provide interpretable results. However, both techniques struggle when dealing with complex, multidimensional clinical data patterns [9]. Support Vector Machines (SVM) and k-Nearest Neighbors (KNN) are effective in predicting heart disease when appropriate feature tuning methods are applied by researchers [10].

Multiple research studies have used Artificial Neural Networks (ANNs) for their ability to learn complex relationships, making them effective in predicting heart disease. The Cleveland Heart Disease dataset received positive diagnostic outcomes when Detrano and colleagues evaluated neural network performance [11]. Standalone ANNs demonstrate difficulty in stable performance due to overfitting problems and data sensitivity when operating on small or unbalanced datasets.

Ensemble methods have emerged as powerful alternatives, combining multiple weak or strong learners to improve generalization. Gradient boosting machines (GBMs) and random forests (RFs) have demonstrated robustness in clinical prediction tasks, offering high accuracy and resilience to overfitting [12]. Bagging and boosting strategies, when applied to medical data, have also proven effective in stabilizing predictions and handling data variance.

More recently, hybrid and meta-learning approaches have been explored to combine the strengths of different algorithms. For instance, research by Uddin et al. presented a hybrid architecture combining RF and neural networks to leverage both interpretability and deep feature learning [13]. Such architectures are particularly suited for medical diagnostics, where performance and explainability are critical.

Furthermore, the challenge of class imbalance, which is common in medical datasets where positive cases are often underrepresented, has been addressed using various techniques, including resampling, cost-sensitive learning, and

synthetic data generation methods such as SMOTE [14]. These strategies ensure that predictive models are not biased toward the majority class, thus improving sensitivity and recall in identifying patients at risk.

Overall, the literature indicates that no single model consistently outperforms others across all types of datasets. Instead, combining complementary algorithms through ensemble or hybrid learning often yields better results, especially when dealing with heterogeneous and imbalanced clinical data. In general, the literature suggests that no single model consistently outperforms all others across various datasets. This highlights the need for context-specific model selection based on the characteristics of the data.

3. Proposed Methodology

In this study, we designed a hybrid machine learning model that combines the strengths of two well-known techniques: Random Forests and Multi-Layer Perceptrons (MLPs) to enhance heart disease prediction. The core concept is to leverage the Random Forest model to capture essential decision patterns from the data, and then pass its output to an MLP, which serves as a more flexible decision-making model. This section outlines the process of data preparation, model development, and performance evaluation.

3.1 Dataset and Preprocessing

To build our heart disease prediction model, we worked with a dataset containing 303 patient records, each representing a set of medical and personal health attributes. After a quick quality check, we noticed that six records had missing values. Rather than filling in those gaps with assumptions, we removed them entirely to ensure the dataset stayed clean and reliable. That left us with 297 complete and usable records for our analysis.

Since our goal was to predict whether or not a patient has heart disease, we treated this as a binary classification problem. We assigned a value of 1 for patients with heart disease and 0 for those without. Before training our machine learning models, we standardized all numerical features to ensure consistency and improve model performance. This ensured that each feature contributed equally during training, preventing any single variable from dominating due to its scale.

TABLE 1. UCI dataset attributes detailed information.

Sr. No.	Attribute	Icon	Description
1	Age	Age	Patient's age in years
2	Sex	Sex	Biological sex (0 = female; 1 = male)
3	Chest Pain	Cp	Type of chest pain (1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
4	Rest Blood Pressure	Trestbps	Resting systolic blood pressure (in mm Hg)
5	Serum Cholesterol	Chol	Serum cholesterol level (in mg/dl)
6	Fasting Blood Sugar	Fbs	Blood sugar > 120 mg/dl (0 = false; 1 = true)
7	Rest ECG	Restecg	Electrocardiographic results (0 = normal, 1 = ST-T abnormality, 2 = LV hypertrophy)
8	Max Heart Rate	Thalach	Maximum heart rate achieved
9	Exercise-Induced Angina	Exang	Angina induced by exercise (0 = no; 1 = yes)
10	ST Depression	Oldpeak	Depression in ST-segment relative to the rest
11	Slope	Slope	Slope of peak ST segment (1 = upsloping, 2 = flat, 3 = downsloping)
12	No. of Vessels	Ca	Number of major vessels coloured by fluoroscopy (0–3)
13	Thalassemia	Thal	Blood disorder type (3 = normal, 6 = fixed defect, 7 = reversible defect)
14	Heart Disease Status	Class	Diagnostic outcome (0 = no disease, 1–4 = increasing severity of risk)

3.2 Proposed Hybrid Model Architecture

To leverage the complementary strengths of ensemble learning and deep neural networks, we implemented a two-stage stacked model. The architecture integrates a Random Forest classifier as the base learner and a Multi-Layer Perceptron (MLP) as the meta-learner. This hybrid framework is designed to extract robust feature interactions via Random Forest and then refine classification decisions using a non-linear neural architecture.

3.2.1 Base Learner: Random Forest Classifier

The Random Forest classifier is an ensemble learning algorithm that constructs many decision trees during training and outputs the mode of the classes or mean prediction for classification and regression tasks, respectively. In our implementation, the Random Forest model was configured with 300 trees (n_estimators = 300) and a maximum depth of 12 (max_depth = 12). These settings were chosen to balance model complexity and prevent overfitting while maintaining predictive power.

To address class imbalance, we applied class weighting based on inverse frequency. The class weight w_j for class j is computed as:

$$w_j = \frac{n}{k \cdot n_j}$$

(1)

Where n is the total number of samples, k is the number of classes, n_j is the number of instances of class j. Instead of final class labels, we extracted class probabilities from the trained Random Forest, denoted as:

$$\hat{P}_{RF}(y = 1 \mid \mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x})$$

(2)

Where T is the number of trees and $h_t(x)$ is the prediction from the t-th tree. These probabilities serve as input features for the subsequent MLP classifier.

3.2.2 Meta-Learner: Multi-Layer Perceptron (MLP)

The meta-learner is a fully connected neural network comprising three hidden layers with ReLU activation functions. Each hidden layer is followed by dropout (rate = 0.3) to reduce overfitting. The final output layer consists of a single neuron with a sigmoid activation function, producing a probability $\hat{y} \in [0,1]$ of heart disease presence.

The forward pass through the MLP can be described as:

$$\hat{y} = \sigma(\mathbf{W}_3 \cdot \text{ReLU}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{z} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3)$$

(3)

Here, z represents the input vector (probabilities from the Random Forest), W_i and b_i are the weight matrices and bias vectors, respectively, and $\sigma(\cdot)$ denotes the sigmoid activation function.

Algorithm 1: Hybrid Heart Disease Prediction Using Random Forest and MLP
1: Load dataset D from source
2: Remove rows in D containing missing values
3: Extract features X and labels y from D
4: Standardize X using Z-score normalization
5: Split X and y into X_train, X_val, X_test and y_train, y_val, y_test using stratified sampling
6: Train Random Forest model RF on X_train and y_train with class_weight='balanced'
7: for each of X_train, X_val, X_test do
8:     Generate meta-features by RF.predict_proba(X_set)
9: end for
10: Define the MLP model with architecture:
       - Input layer: size = number of RF output classes
       - Hidden layers: [128, 64, 32] with ReLU and Dropout(0.3)
       - Output layer: 1 neuron with Sigmoid activation

11: Compile MLP using:
 - Loss = Binary Crossentropy
 - Optimizer = Adam (learning_rate = 0.0001)
12: Compute class weights from y_train for imbalance correction
13: Train MLP on RF meta-features from X_train with y_train, validate on X_val
 - Use EarlyStopping and ModelCheckpoint callbacks
14: Predict probabilities y_pred_prob on X_test meta-features using trained MLP
15: Compute ROC curve and Youden's J statistic: J = TPR - FPR
16: Identify optimal threshold θ* where J is maximized
17: Convert y_pred_prob to final labels y_pred using:
 if y_pred_prob > θ* then label = 1 else label = 0
18: Evaluate predictions using:
 - Classification Report
 - Confusion Matrix
 - ROC-AUC Curve
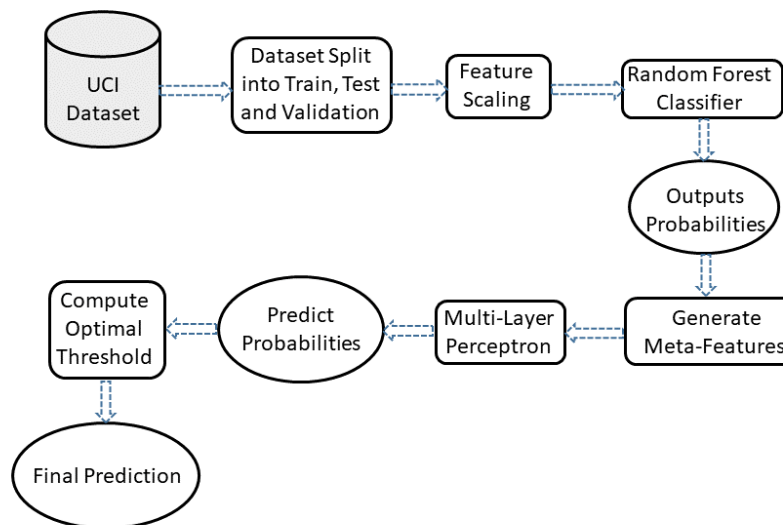 - Precision-Recall Curv

Figure 1: Diagram of the proposed hybrid model

3.3. Evaluation Metrics

The performance of the proposed model was evaluated using four fundamental classification outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Based on these, we computed five key quantitative metrics commonly used in medical classification studies: precision, sensitivity (also known as recall), F1 score, specificity, and accuracy. These metrics were selected due to relevance in binary classification problems and their widespread use in closely related research. Together, they provide a comprehensive view of the model's ability to identify heart disease cases while minimizing diagnostic errors correctly.

Accuracy measures the overall correctness of the model's predictions, defined as the proportion of all correctly predicted instances (both positive and negative) among the total predictions:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

Precision evaluates the reliability of positive predictions. It quantifies how many of the cases predicted as having heart disease are correct:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

Also known as recall or the true positive rate, sensitivity indicates how effectively the model identifies actual cases of heart disease:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

Specificity measures the model's ability to correctly identify patients without heart disease, focusing on the true negatives:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}$$

The F1 score is the harmonic mean of precision and sensitivity. It provides a balanced measure when there is a trade-off between false positives and false negatives:

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precission} + \text{Recall}} \tag{8}$$

4. RESULT AND DISCUSSION

The proposed hybrid model, which combines a Random Forest (RF) classifier as the base learner and a Multi-Layer Perceptron (MLP) as the meta-learner, was evaluated on the heart disease dataset using a range of performance metrics. The results demonstrate the model's effectiveness in binary classification tasks, even under class imbalance conditions, and highlight its potential for use in clinical decision support.

4.1 Performance Metrics

Table 2 provides a snapshot of the model's performance, highlighting its ability to make reliable predictions. With an overall accuracy of 84%, the model demonstrates a balanced approach in classifying healthy individuals and those with heart disease. Notably, it achieves a recall of 0.91 for the 'Disease' class, meaning it effectively identifying most actual cases of heart disease. At the same time, the model achieved a precision of 0.88 for the 'No Disease' class, indicating a high level of confidence in correctly identifying healthy individuals. This balance between sensitivity and precision makes the model a dependable tool in heart disease detection.

TABLE 2. Performance report of our proposed approach.

Class	Precision	Recall	F1-Score
No Disease	0.88	0.75	0.81
Disease	0.81	0.91	0.86
Accuracy			**0.84**
Macro Avg	0.84	0.83	0.83
Weighted Avg	0.84	0.84	0.83

4.2 Graphical Analysis of Model Performance

To further validate the performance of our proposed hybrid model, we examined several visual evaluation metrics, including the confusion matrix, precision-recall curve, and ROC curve. These visualizations provide deeper insights into the model's classification behaviour and its reliability in a real-world healthcare setting.
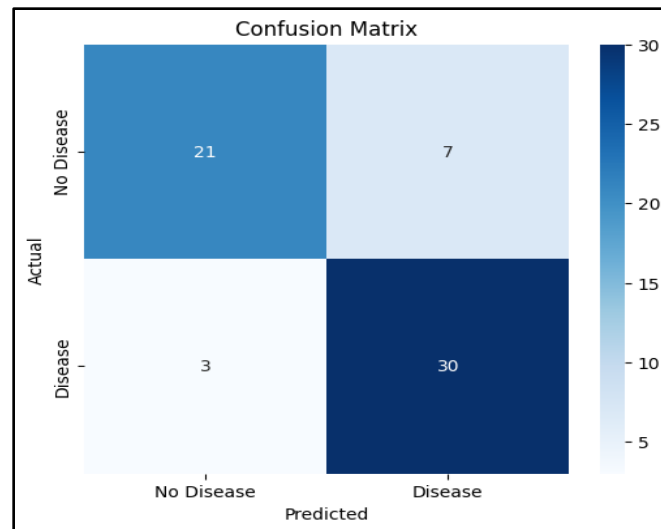

Figure 2: Confusion matrix of the proposed model

The confusion matrix (Figure 2) illustrates how well the model distinguishes between individuals with and without heart disease. Out of 61 total cases, the model correctly identified 30 out of 33 patients with heart disease and 21 out of 28 healthy individuals. This reflects a high recall for the 'Disease' class, showing the model's strong ability to detect actual positive cases — a crucial aspect in clinical diagnosis.
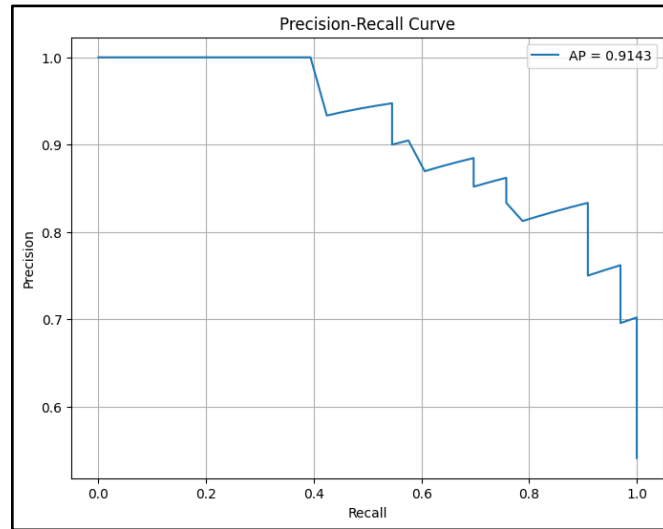
Figure 3: Precision-Recall Curve of the proposed model.

In terms of precision and recall trade-offs, the precision-recall curve (Figure Y) shows that the model maintains high precision across various levels of recall, with an average precision (AP) score of 0.91. This indicates that the model successfully captures more true positive cases without greatly sacrificing accuracy, which is a valuable trait for any diagnostic tool.
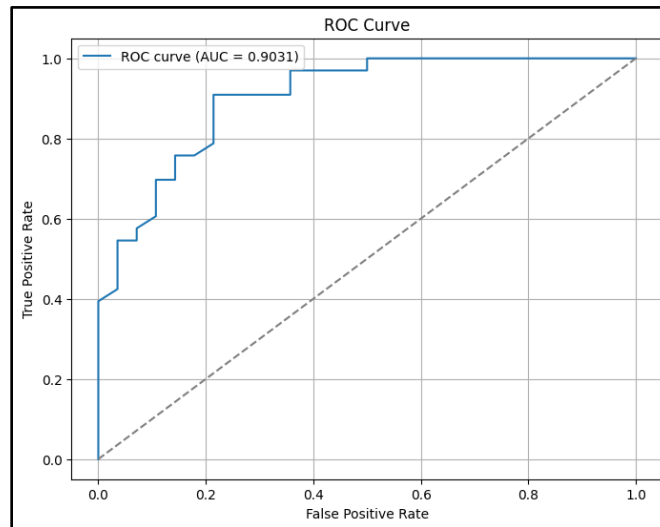

Figure 4: ROC curve of the proposed model.

Moreover, the ROC curve (Figure 4) reveals the model's ability to distinguish between the two classes across different threshold settings. With an AUC (Area Under Curve) of 0.90, the model demonstrates excellent discriminatory power, meaning it performs well in minimizing both false positives and false negatives.

Together, these visual metrics reinforce the reliability of our model, especially in sensitive clinical contexts where both overdiagnosis and missed diagnoses carry serious consequences. The combination of high recall for the 'Disease' class and substantial overall precision indicates that this hybrid model could be a supportive tool for early detection and risk assessment in heart disease.

4.3 Comparative Analysis

To better understand how our proposed hybrid model performs, we compared its results with those from earlier studies on heart disease prediction. Table 3 shows the accuracy achieved by different techniques reported in the literature. This comparison clearly demonstrates the advantages of our approach, showing that it outperforms many traditional machine learning models commonly used in similar studies.

Table 3. Accuracy comparison of our model with existing techniques

Authors	Technique	Accuracy (%)
Detrano et al. [11]	Logistic Regression	78
Nayak et al. [16]	Decision Tree	75
Anooj et al. [17]	Random Forest	80
Palaniappan et al. [18]	SVM	81
Islam et al. [19]	CNN-Based Model	82
Proposed Model	Hybrid (RF + MLP)	**84**

The results in Table 3 clearly show that our hybrid model performs better than traditional machine learning techniques and holds its own against more advanced models like CNN-based approaches. By combining the strengths of Random Forest (RF) and Multi-Layer Perceptron (MLP), we develop a system that balances transparency with strong predictive performance. RF helps capture complex patterns in structured data, while MLP excels at identifying nonlinear relationships, making for a well-rounded approach. This synergy enhances overall accuracy while maintaining a strong balance between precision and recall—crucial for effective heart disease prediction in real-world healthcare settings.

5. CONCLUSION

This study presents a hybrid approach to predicting heart disease, combining the strengths of Random Forest and Multi-Layer Perceptron models. By balancing performance with interpretability, our method achieves an impressive 84% accuracy, proving its potential as a reliable tool for heart disease detection. It surpasses traditional machine learning techniques by optimizing both precision and recall, ensuring that predictions are accurate and relevant for real-world healthcare applications. Through comparative analysis, we've demonstrated that our model stands strong against established methods, showing its effectiveness in situations when accuracy and reliability are crucial. In medical settings, having a model that provides transparent and explainable predictions can be a game-changer, helping healthcare professionals make better-informed decisions for their patients. That said, no model is perfect. There's always room for improvement, and we recognize the need to refine its ability to adapt to diverse datasets and increase its generalization power. Looking ahead, we plan to incorporate additional factors such as genetic and environmental influences and explore more advanced techniques to enhance accuracy further.

At its core, this research contributes to the expanding field of predictive healthcare models, promoting early detection, reducing misdiagnoses, and ultimately improving patient outcomes. By pushing the boundaries of hybrid machine learning, we hope to pave the way for even more effective tools in the fight against heart disease.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.

[3] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[4] F. Jiang et al., "Artificial intelligence in healthcare: Past, present and future," Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017, doi: 10.1136/svn-2017-000101.

[5] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[6] S. Sagi and L. Rokach, "Ensemble learning: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 4, e1249, 2018, doi: 10.1002/widm.1249.

[7] A. Gudadhe, S. Wankhade, and H. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network," in 2010 International Conference on Computer and Communication Technology (ICCCT), Sep. 2010, pp. 741–745, doi: 10.1109/ICCCT.2010.5640374.

[8] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[9] D. L. Reilly, L. N. Cooper, and C. Elbaum, "A neural model for category learning," Biological Cybernetics, vol. 45, no. 1, pp. 35–41, 1982, doi: 10.1007/BF00335247.

[10] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Heart disease prediction using lazy associative classification," Procedia Computer Science, vol. 57, pp. 376–381, 2015, doi: 10.1016/j.procs.2015.07.360.

[11] R. Detrano et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," American Journal of Cardiology, vol. 64, no. 5, pp. 304–310, 1989, doi: 10.1016/0002-9149(89)90524-9.

[12] A. Abdar et al., "A new machine learning method for classification of heart diseases," Computer Methods and Programs in Biomedicine, vol. 192, p. 105261, May 2020, doi: 10.1016/j.cmpb.2020.105261.

[13] S. Uddin et al., "A hybrid machine learning model for predicting heart disease using data mining techniques," Computer Methods and Programs in Biomedicine, vol. 191, p. 105361, Apr. 2020, doi: 10.1016/j.cmpb.2020.105361.

[14] C. Fernández et al., "SMOTE for learning from imbalanced data: Progress and challenges," Neurocomputing, vol. 356, pp. 202–221, Sep. 2019, doi: 10.1016/j.neucom.2019.04.045.

[15] A. R. Nayak, R. Nayak, D. Sinha, and A. Sharma, "Heart disease prediction using machine learning techniques," Materials Today: Proceedings, vol. 33, pp. 2846–2851, 2020, doi: 10.1016/j.matpr.2020.08.847.

[16] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," Journal of King Saud University - Computer and Information Sciences, vol. 24, no. 1, pp. 27–40, 2012, doi: 10.1016/j.jksuci.2011.09.002.

[17] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," International Journal of Computer Science and Network Security (IJCSNS), vol. 8, no. 8, pp. 343–350, 2008.

[18] M. T. Islam, M. S. Hossain, G. Muhammad, and A. Alelaiwi, "Deep learning-based cardiac disease detection using ECG images," Healthcare, vol. 9, no. 5, p. 527, 2021, doi: 10.3390/healthcare9050527.