International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

Heart Disease Prediction using Integrated Technology of XGBoost, Random Forest and Multi-Layer Perceptron

Dr. Lokendra Singh Songare, Prof. Seema Patidar, Disha Gupta, Deepak Temrwal, Deepak Temrwal
Department of Computer Science and Engineering, Medi-Caps University, A.B. Road, Rau, Indore – 453331
lokendra.singh@medicaps.ac.in, seema.patidar@medicaps.ac.in, guptadisha036@gmail.com,
deepaktemrwal430@gmail.com, shrutikusamya@gmail.com

Abstract— Cardiovascular disease remains a leading cause of death worldwide, requiring prompt and accurate diagnosis to minimize patient mortality rates. More recent developments in artificial intelligence (AI) applications have demonstrated how to enhance prognostic performance and interpretability in clinical diagnosis. This research paper analyzes the application of machine and Deep Learning models for heart disease prediction by voting with a selection of models in order to develop a strong classifier. A weighted ensemble voting approach is employed and leverage is made from XGBoost, Random Forest, and Multi-Layer Perceptron (MLP) model strengths. Further, explainability is offered by SHapley Additive exPlanations (SHAP) to facilitate model decisions, allowing feature importance and decision-making insight. The proposed methodology is supported by established performance metrics, retaining clinical relevance. Results imply that AI-based approaches can achieve elevated predictive accuracy and interpretable diagnoses, informing the creation of automated cardiovascular risk stratification.

Keywords- Heart Disease prediction, ensemble learning, data preprocessing, wearable devices, dietary recommendations, imbalanced datasets, explainable AI, real-time healthcare systems.

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain the foremost global healthcare challenge [1], with high levels of morbidity and mortality [2]. There is need to detect disease conditions early in the prevention against heart disease development because delayed diagnosis poses very catastrophic repercussions. Traditional imaging diagnostic tools such as clinical findings, medical images, and laboratory test results are promising but in some cases are too cumbersome and vulnerable to subjective impressions. AI use in medicine has assisted in automating the forecast of disease and improving diagnostic accuracy.

Machine Learning (ML) and Deep Learning (DL) algorithms have greatly contributed to predictive analytics, enabling them to identify subtle patterns in medical data [3]. These algorithms present evidence-based insights that inform clinicians, complementing conventional diagnostic practices and enhancing decision-making. Yet, if AI-based predictions are to prove clinically valuable, they need to be not just very accurate but also interpretable. One of the main obstacles to medical applications of AI is the "black-box" feature of sophisticated models, where reasoning behind predictions is not explicit. To overcome this, explainable AI (XAI) methodologies like SHapley Additive exPlanations (SHAP) have been applied to promote model interpretability and trust.

This study takes into account the application of an AI-predictive model in diagnosing heart disease through the combination of ML and DL techniques. A specific ensemble learning method is utilized where the strengths of XGBoost, Random Forest, and MLP models are combined. Weighted voting is utilized to optimize predictive performance, while interpretability is derived from SHAP analysis. The objective is to develop an AI-based system that yields interpretable, precise, and clinically pertinent predictions, thereby supporting healthcare practitioners in early diagnosis and risk evaluation.

The sections of this paper are organized as follows: Section II presents a brief overview of recent literature on heart disease prediction using AI. Section III explains the proposed methodology, such as dataset preprocessing, model selection, and evaluation metrics. Section IV illustrates the algorithmic

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

implementation, and Section V displays experimental results and analysis. Finally, Sections VI and VII report the study's limitations and summarize potential future work directions.

II. Literature Review

Artificial intelligence (AI)-based early detection and prediction of cardiovascular diseases (CVDs) have been a topic of interest in the last few years. Because CVDs are still a leading cause of mortality worldwide, high demand exists for diagnostic tests that are not just accurate but scalable and economical. Classical risk prediction models, including the Framingham Risk Score and logistic regression models, have been extensively applied but tend to fail to explain the sophisticated and nonlinear interactions between clinical predictors (Wilson et al., 1998) [4]. Machine Learning (ML) and Deep Learning (DL) methods have also been investigated in an increasing manner to tackle these shortcomings, leveraging enhanced predictive powers by detecting obscure patterns in biomedical data.

Several ML algorithms have demonstrated promising results in classification of heart disease. Some models, including Decision Trees, Random Forest, Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN), have been tested against each other on the basis of predictive accuracy in various studies. Amongst these, ensemble-based models such as Random Forest have been extremely effective in reducing overfitting as well as improving model generalizability (Chen & Guestrin, 2016) [5]. Moreover, gradient boosting algorithms such as XGBoost and LightGBM have proven to be strong methods, improving predictive accuracy even more by further optimizing ensembles of decision trees (Ke et al., 2017) [6]. Hybrid techniques such as feature selection techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) have been suggested by scientists to improve classification efficiency as well as minimize computational complexity (Li et al., 2020) [7].

Deep Learning methods such as Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) have been found to possess tremendous ability in handling massive medical data. Murtaza et al. (2022) [8] proved that Multilayer Perceptrons (MLPs) can even surpass simpler classifiers if trained over properly preprocessed clinical data. Although CNNs are classically employed for image-based diagnosis, there are recent applications of CNNs in structured medical data with promising results (Huang et al., 2019) [9]. However, deep learning models are likely to require vast computational resources and enormous training data sets to provide maximum generalization, making them impractical for application in the clinical setting.

A problem with AI-based healthcare applications is that black-box models are non-interpretable, and this may discourage clinical adoption.Shapley Additive Explanations (SHAP) emerged as a high-impact methodology to enhance the explainability of models, providing insights into attribute significance and making healthcare providers knowledgeable about AI-synthesized predictions (Lundberg & Lee, 2017) [10]. Analysis with SHAP has been utilized extensively in ML and DL models for better medical decision-making transparency (Molnar, 2022) [11]. In addition, interpretability of models is necessary to ensure compliance with regulatory requirements and the ethical use of AI in healthcare (Tjoa & Guan, 2021) [12].

In spite of these developments, issues persist in achieving the generalizability of AI models across populations, reducing bias in datasets, and ensuring fairness in AI-based predictions across demographic groups. Overcoming class imbalances and enhancing predictive model robustness are ongoing research areas. Current studies continue to investigate hybrid AI models combining ML, DL, and explainability methods to improve accuracy and reliability in heart disease prediction.

III. Methodology

The approach used in this study is the construction and validation of an AI model to predict heart disease using a technique of ensemble learning that combines various machine learning algorithms. The following discusses the data used in the dataset, preprocessing, model choice, training, evaluation metric, and explanation techniques used to report the findings.

A. Dataset Description

The dataset utilized in this study is an open-available Cardiovascular Disease Dataset containing structured patient data with some clinical attributes. A few of these are:

- Demographic Details: Age, gender, etc.
- Physiological Data: Cholesterol level, heart rate, BMI, blood pressure.
- Medical History and Risk Factors: Diabetes, smoking status, exercise, history of heart disease in family.

Target variable is binary denoting presence (1) or absence (0) of heart disease.

B. Data Preprocessing

To keep the quality and accuracy of the dataset, the following data preprocessing operations were performed:

1. Handling Missing Values: Missing values were imputed with the median for numerical attributes and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

mode imputation for categorical attributes to avoid data loss.

2. Removal of Outliers: The outliers were removed using the Interquartile Range (IQR) method, and 189 outliers from the training dataset were deleted.
3. Feature Scaling and Encoding:
   - Numerical features were standardized (unit variance, zero mean) for deep learning models.
   - Categorical variables were one-hot encoded to render them machine learning algorithm-compatible.
4. Train-Test Split: The data were manually divided into 611 train samples and 200 test samples prior to any preprocessing, to avoid data leakage.

C. Model Selection and Training

This study employs an ensemble learning technique that combines three models—XGBoost, Random Forest, and Multilayer Perceptron (MLP)—for the highest possible predictive performance.

D. Extreme Gradient Boosting (XGBoost)

XGBoost is a high-performance gradient-boosting decision tree model. It effectively describes the non-linear inter-correlations of cardiovascular risk factors. Features:

- Gradient Boosting Framework: Every tree fixes the mistake of the previous tree.
- Regularization (L1 & L2): Avoidance of overfitting.
- Missing Values Handling: XGBoost is capable of learning automatic optimal missing values for splitting.
- Tree Pruning: Stops tree growth when it is needed to make the algorithm efficient.

XGBoost Model Training:

1. Loss Function: Binary Log Loss (appropriate for classification).
2. Hyperparameters Used in Training:
   - Number of Estimators: 100
   - Maximum Depth: 6
   - Learning Rate: 0.1
   - Subsample Ratio: 0.8
3. Optimization: The model was trained without any other hyperparameter optimization.

E. Random Forest (RF)

Random Forest is an ensemble method [13] that constructs many individual decision trees and combines their prediction in a manner to enhance generalization.

Principal Features:

1. Bootstrap Aggregation (Bagging): Minimizes variance.
2. Feature Randomization: Minimizes sensitivity.
3. Majority Voting Mechanism: Averages prediction of many trees.

Random Forest Model Training:

- Number of Estimators: 100 (instead of 200 as originally planned).
- Maximum Depth: 10
- Minimum Samples per Split: 2
- Hyperparameter tuning: Not performed; default scikit-learn settings for the sake of efficiency.

F. Multilayer Perceptron (MLP) - Neural Network

MLP is a densely connected learning model and can learn high-order feature interactions [14].

Design:

1. The number of features in the input layer of the MLP architecture is equal.
2. Layers That Are Hidden:
   - Layer 1: ReLU activation, 128 units, dropout (0.3).
   - Layer 2: ReLU activation, 64 units, dropout (0.2).
   - Layer 3: ReLU activation, 32 units.
3. The output layer uses a sigmoid activation function in binary classification.

Training:

- 32 is used as batch size.
- 100 epochs exist.
- A 0.001 learning rate Adam optimizer.
- Binary Cross-Entropy as the loss function.

G. Ensemble Model Building

These individual predictions were then aggregated by a weighted voting mechanism where the prediction was computed as:

$$P_{ensemble} = 0.4 \times P_{XGBoost} + 0.3 \times P_{RandomForest} + 0.3 \times P_{MLP} \quad (1)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

Where P is each model's heart disease probability prediction. The weights were set up on the basis of individual performance of models when validated to strike an optimal trade-off.

- If Pensemble > 0.5, the sample is labeled as "Heart Disease" (1).
- Else, it is labeled as "No Heart Disease" (0).

This ensemble strategy improves predictability by using the best capabilities of all three models.

H. Evaluation Metrics

The performance of the model was measured based on clinically appropriate metrics:

1. Accuracy: Refers to the proportion of accurate predictions.
2. Precision: Verifies how many forecasted positive cases are indeed positive.
3. Recall (Sensitivity): It is the prediction of the capability to identify true cases of heart disease.
4. F1-Score: It is the harmonic mean of the recall and precision.
5. AUC-ROC (Area Under the Curve – Receiver Operating Characteristic): It is utilized to quantify the discriminative ability of the model.

These measures make the clinical utility of the model strong and trustworthy.

I. Explainability Using SHAP

To make the models interpretable, Shapley Additive Explanations (SHAP) was employed in analyzing feature contribution and individual prediction. SHAP values measure how much each feature contributes to making a decision.

SHAP Analysis Steps:

1. Summary Plot: Presenting the summary effect of the features on prediction.
2. Bar Plot: Ranking the contributing factors most affected by heart disease risk.
3. Decision Plot: Showing which individual risk factors contribute to exact predictions.

These explanation approaches guarantee the AI model is interpretable and transparent, making them comply with one of the main clinical decision-making requirements.

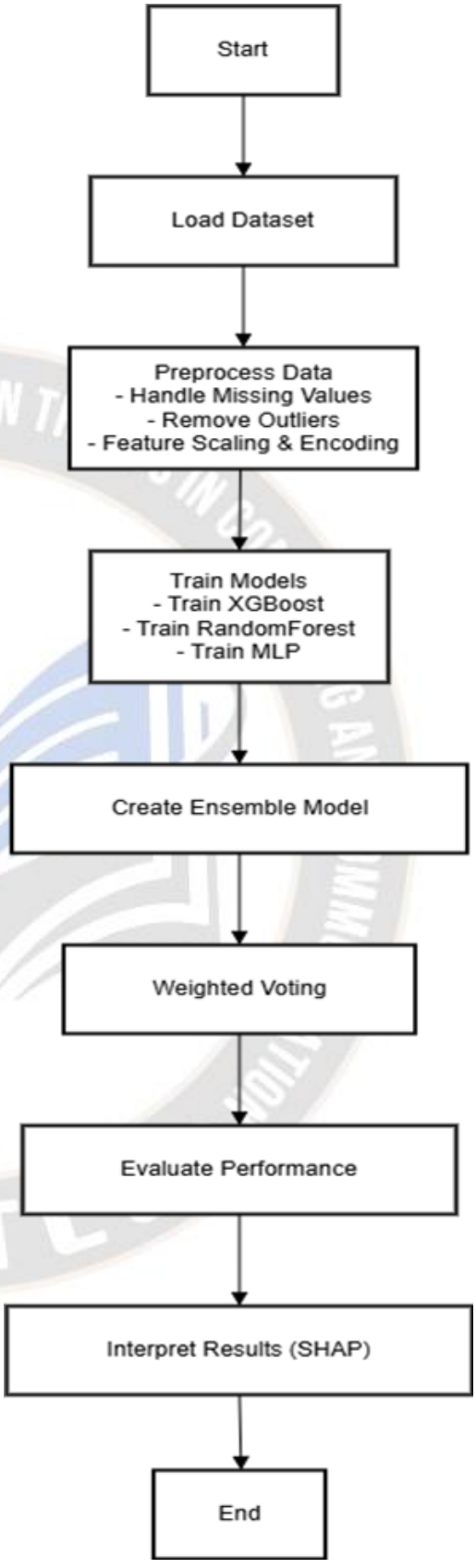

Figure. 1.1 Representation of Methodology

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
***Article Received:** 25 March 2025 **Revised:** 12 April 2025 **Accepted:** 20 May 2025*

IV. ALGORITHM

This project utilizes an ensemble of machine learning and deep learning models incorporated into an ensemble learning approach [15]. The applied algorithms are Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Multilayer Perceptron (MLP). All the algorithms have a specific contribution to the task of prediction, enhancing general accuracy and reliability.

A. Extreme Gradient Boosting (XGBoost)

$$L(\theta) = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t) \quad (2)$$

Where:

- $l(y_i, \hat{y}_i)$ is the loss function (e.g., log loss for binary classification).
- $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \Sigma_j \omega_j^2$ is the regularization term to control model complexity.
- T is the number of trees, ω_j are the leaf weights.

Definition: XGBoost is an optimized version of the Gradient Boosting algorithm for performance and speed. It develops an ensemble of decision trees, each of which corrects the errors made by the previous trees through a boosting technique.

Purpose in this Project: XGBoost is used to serve as one of the primary predictive models because it is efficient in handling structured medical data. It identifies complex, non-linear associations between risk factors and heart disease.

Key Features:

- Gradient Boosting Framework: Constructs trees in sequence to minimize prediction mistakes.
- Regularization (L1 & L2): Prevents overfitting by penalizing complex models.
- Handling Missing Values: Can automatically learn the best direction in which to split missing data.
- Pruning of Trees: Stops tree growth when it no longer improves performance.

XGBoost updates predictions using the gradient of the loss function, making it a powerful boosting technique.

B. Random Forest (RF)

$$H(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (3)$$

Where:

- $H(x)$ is the final predictions
- T is the total number of trees.
- $h_t(x)$ is the prediction of the t-th tree.

Definition: Random Forest is an ensemble approach that builds multiple decision trees and combines their outputs to enhance stability and accuracy.

Purpose within this Project: Random Forest applies to boost model generalization and avoid overfitting. It performs especially well in structured medical datasets where interactions between the features are significant.

Key Features:

- Bootstrap Aggregation (Bagging): The trees are trained individually on a random subset of the data in order to promote diversity.
- Feature Randomization: Each split considers only a subset of features, improving robustness.
- Majority Voting Mechanism: Consolidates the predictions of many trees to produce the final outcome.

C. Multilayer Perceptron (MLP) – Neural Network

$$z = Wx + b$$
$$a = \sigma(z) \quad (4)$$

Where:

- W represents the weight matrix.
- x represents the input vector.
- b represents the bias term.
- $\sigma(z)$ represents the activation function (ReLU in hidden layers, Sigmoid in output layer).

The final output is computed as:

$$\hat{y} = \sigma(w_{out}\, a + b_{out}) \quad (5)$$

Where $\hat{y}$ represents the predicted probability of heart disease.

Definition: An MLP or Multilayer Perceptron is a deep model which has numerous layers of neurons. MLP learns complex patterns through weight updates by backpropagation.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: 25 March 2025 *Revised:* 12 April 2025 *Accepted:* 20 May 2025

Role of MLP in the Project: The MLP is incorporated in the mixture to identify non-linear relationships among heart disease risk factors. It is able to identify deep hierarchical feature representations which may not be able to uncover by the tree-based structures.

Key Features:

- Input Layer: It receives structured clinical data.
- Hidden Layers: Learns transformations with activation functions (ReLU).
- Output Layer: Uses Sigmoid activation to give probabilities for binary classification.

D. Ensemble Learning (Weighted Voting)

$$P_{ensemble} = 0.4 \times P_{XGBoost} + 0.3 \times P_{RandomForest} + 0.3 \times P_{MLP} \quad (6)$$

Where:

- $P_{ensemble}$ is the final probability of heart disease.
- $P_{XGBoost}$, $P_{RandomForest}$, P_{MLP} are the predictions from individual models.

The classification decision is:

- If Pensemble > 0.5, the sample is labeled as "Heart Disease" (1).
- Else, it is labeled as "No Heart Disease" (0).

Definition: Ensemble learning refers to the practice of combining models to enhance performance. Weighted voting is used in aggregating XGBoost, Random Forest, and MLP predictions.

Purpose of this Project: Ensemble technique permits using the merits of various models:

1. XGBoost performs well at capturing non-linear patterns.
2. Random Forest is imparting stability and generalization.
3. MLP enriches high-order feature interactions.

This ensemble method improves overall prediction accuracy and reduces bias and variance.

V. Results And Discussion

Here, we compare and evaluate the performance of various machine learning models employed for heart disease prediction: XGBoost, Random Forest, Multi-Layer Perceptron (MLP), and Weighted Voting Ensemble. The models are compared on the basis of accuracy, AUC-ROC values, precision, recall, and F1-score [16]. We also present the trade-offs of each method and its clinical relevance [17].

A. XGBoost Performance

XGBoost recorded a 92.50% accuracy and AUC-ROC of 0.9887, reflecting excellent performance in general. The model had an excellent recall rate of 97% in detecting heart disease cases so that most people with heart disease were properly identified.

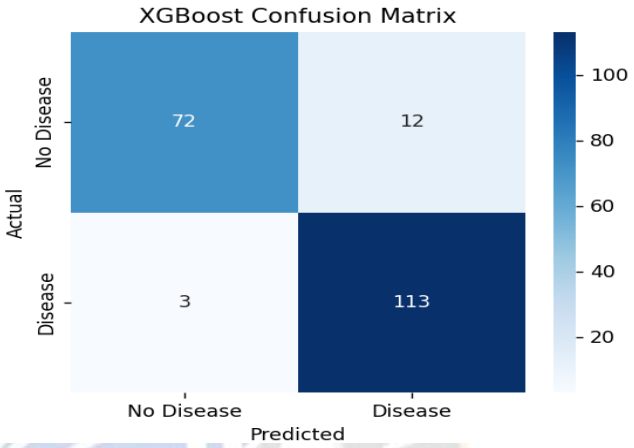

Figure. 1.2 Confusion Matrix for XGBoost

Confusion Matrix Explanation:

- True Positives (TP): 113, Proper classification of heart disease cases.
- False Positives (FP): 12, Misclassification of normal patients as heart disease
- True Negatives (TN): 72, Proper classification of healthy people.
- False Negatives (FN): 3, Misclassification of heart disease cases as healthy.

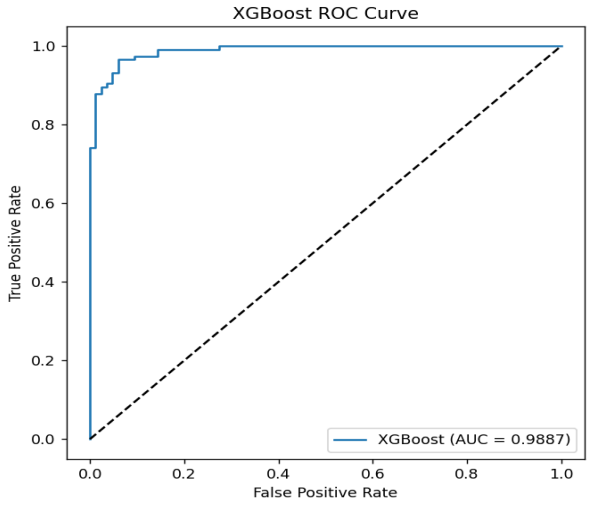

Figure. 1.3 ROC Curve for XGBoost

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

ROC Curve Analysis:

- AUC-ROC: 0.9887, Showing outstanding discrimination capacity.
- The model is good but slightly more false positives compared to other models, which indicates a minor compromise on precision.

B. Random Forest Performance

Random Forest provided a more precise (95.00%) and improved AUC-ROC (0.9909) classification than XGBoost, with a more accurate distinction boundary for patients with and without heart disease. Of special note was that its recall value for the heart disease cases was 98%, or almost all patients with heart disease were picked up. Additionally, precision for the heart disease class increased to 93%, fewer false positives than XGBoost.

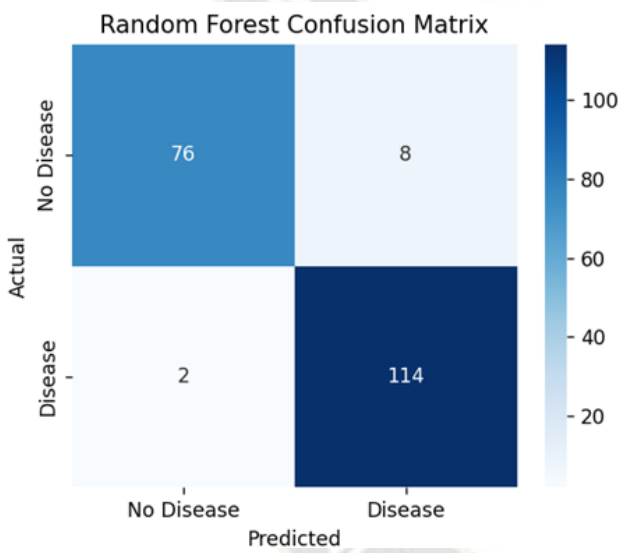

Figure. 1.4 Confusion Matrix for Random Forest

Confusion Matrix Interpretation:

- True Positives (TP): 114, More accurately identified heart disease cases than XGBoost.
- False Positives (FP): 8, Less number of misclassification of healthy participants.
- True Negatives (TN): 76, Improved classification of healthy cases.
- False Negatives (FN): 2, lowest rate of false negatives, i.e., few cases were not detected.

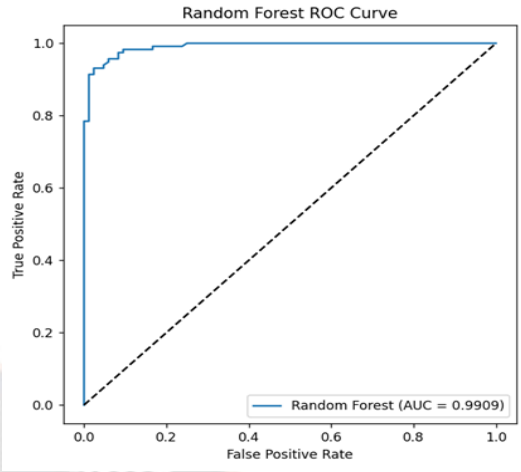

Figure. 1.5 ROC Curve for Random Forest

ROC Curve Analysis:

- AUC-ROC: 0.9909, which is higher than XGBoost, meaning better overall separation between classes.
- Compared to XGBoost, Random Forest reduces false positives while maintaining high recall.

C. MLP (Multi-Layer Perceptron) Performance

Among stand-alone models, the MLP achieved the highest accuracy rate (96.00%), as a testament to its ability to learn complex patterns within the data. AUC-ROC score (0.9888) was lower than that of Random Forest, but still very strong predictive accuracy indicator. MLP's heart disease instance recall was very high (99%), and it predicted almost all heart disease patients. Besides, its precision (94%) was the best among individual models, with an optimal balance of false positives and false negatives.

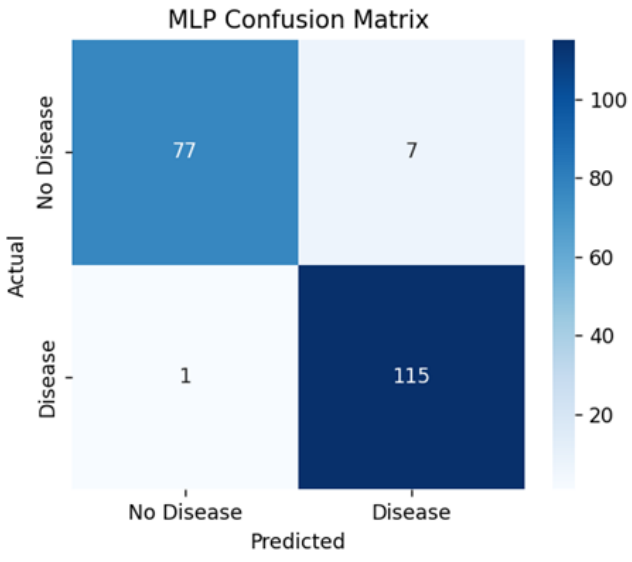

Figure. 1.6 Confusion Matrix for MLP

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: 25 March 2025 *Revised:* 12 April 2025 *Accepted:* 20 May 2025

Confusion Matrix Explanation:

- True Positives (TP): 115, Best possible correct classification of heart disease.
- False Positives (FP): 7, Best possible false positives, i.e., Random Forest.
- True Negatives (TN): 77, Best possible classification of healthy cases.
- False Negatives (FN): 1, Best 1 heart disease case misclassified, best among all models.

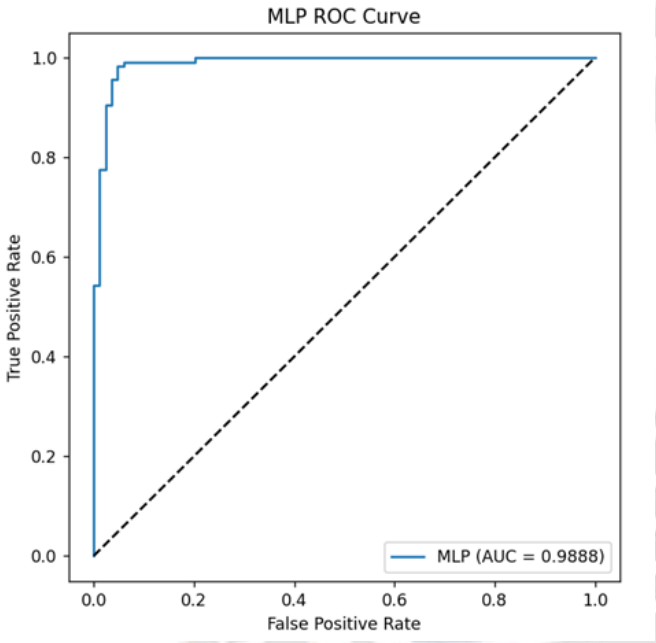

Figure. 1.7 ROC Curve for MLP

ROC Curve Analysis:

- AUC-ROC: 0.9888, comparable to XGBoost and Random Forest.
- MLP achieves highest accuracy (96%) and recall (99%), verifying that almost all the cases of heart disease are identified.

D. Weighted Voting Ensemble Performance

In order to utilize the power of each and every model, a Weighted Voting Ensemble was employed, which was the aggregation of XGBoost, Random Forest, and MLP prediction. The method had an accuracy rate of 95.50%, lower than MLP but higher than XGBoost and Random Forest. The ensemble model was the best AUC-ROC score (0.9928), with the highest ability to discriminate heart disease from non-heart disease.

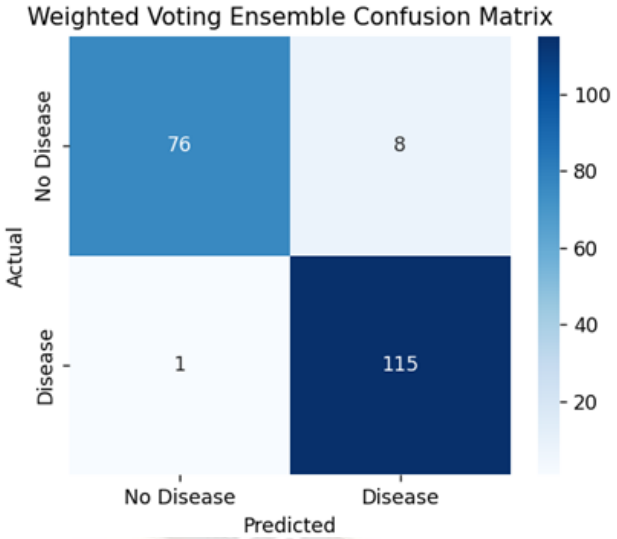

Figure. 1.8 Confusion Matrix for Weighted Voting Ensemble

Confusion Matrix Interpretation:

- True Positives (TP): 115, Matches MLP in detecting heart disease.
- False Positives (FP): 8, Similar to Random Forest.
- True Negatives (TN): 76, Slightly lower than MLP but close.
- False Negatives (FN): 1, Same lowest false negative rate as MLP.

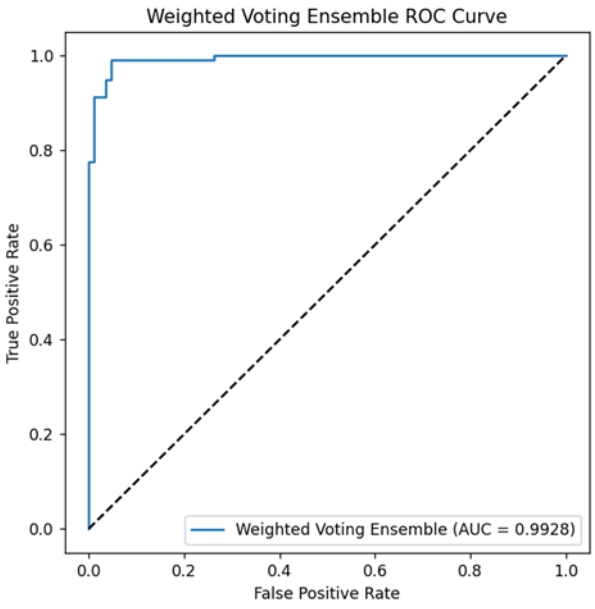

Figure. 1.9 ROC Curve for weighted voting ensemble

ROC Curve Analysis:

- AUC-ROC: 0.9928, the highest among all models.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

- The ensemble leverages strengths from all models, making it the most robust approach.

The ensemble approach upheld the high recall (99%) observed in MLP with the added benefit of improved precision (93%), thus a balanced model applicable to medicine. The high recall ensures that there are virtually no cases of heart disease that escape detection, and improved precision lessens misdiagnosis errors. The higher AUC-ROC value of the ensemble ensures that it is more stable and hence the most reliable among all the models.

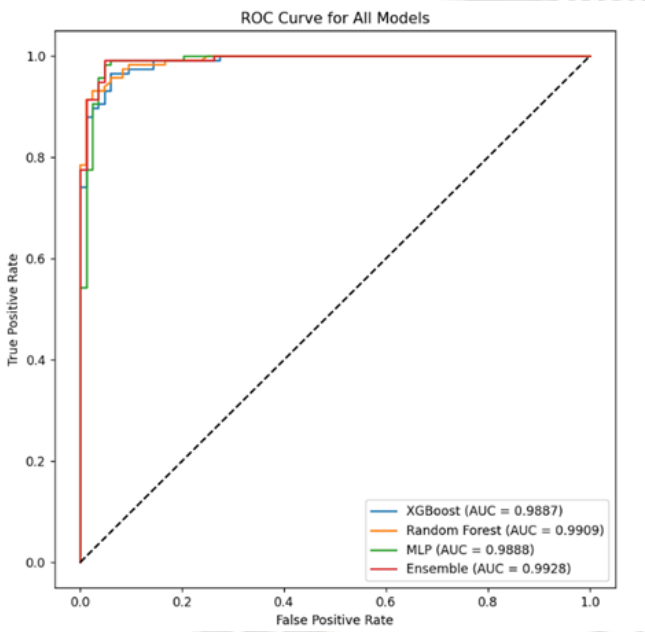

Figure. 1.10 ROC Curve for all models

TABLE I. FINAL RESULT FINDINGS

Model	**Accuracy**	**AUC-ROC**	**Precision (Class 0)**	**Precision (Class 1)**	**Recall (Class 0)**	**Recall (Class 1)**	**F1 – Score (Class 0)**	**F1 – Score (Class 1)**
XG Boost	92.50%	0.9887	96%	90%	86%	97%	91%	94%
Random Forest	95.00%	0.9909	97%	93%	90%	98%	94%	96%
MLP	96.00%	0.9888	99%	94%	92%	99%	95%	97%
Ensemble (Voting)	95.50%	0.9928	99%	93%	90%	99%	94%	96%

SHAP Analysis for the Weighted Voting Ensemble Model

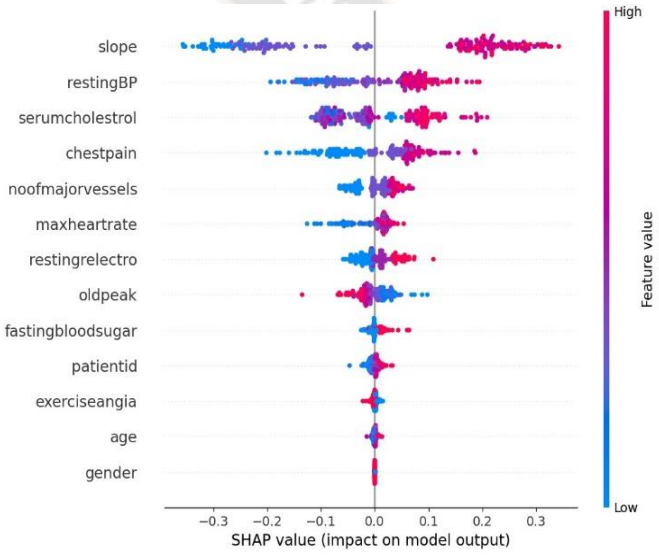

Figure. 1.11 Feature Wise impact on predictions

The SHAP summary plot provides a detailed breakdown of each feature's influence on individual predictions. The color gradient (blue to red) represents feature values, where red indicates high values and blue indicates low values.

- Slope: Large values (red) heavily pull predictions toward heart disease, while small values (blue) lower the likelihood.
- Resting Blood Pressure: Slightly elevated values slightly increase the risk of heart disease, although somewhat variably.
- Serum Cholesterol & Chest Pain: Elevated levels contribute positively to heart disease predictions.
- Number of Large Vessels: Having more large vessels seems to be associated with a lower risk, as in the distribution.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 13 Issue: 1
Article Received: *25 March 2025* ***Revised:*** *12 April 2025* ***Accepted:*** *20 May 2025*

- Old Peak (ST Depression Induced by Exercise): Higher values contribute towards higher disease risk.
- Age & Gender: These have comparatively lesser influence in relation to physiological measures.

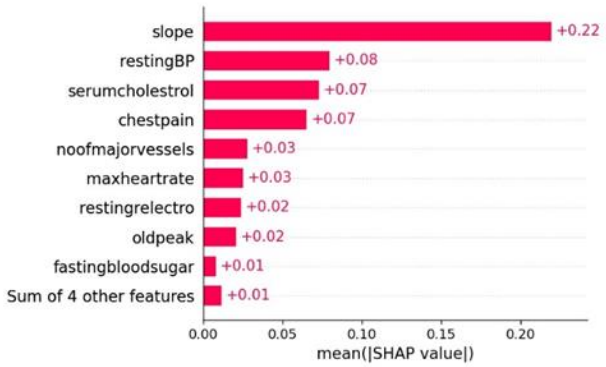

Figure. 1.12 Summary of feature importance

The bar chart is a visualization of the mean absolute SHAP values, showing the total contribution of each feature to the predictions of the ensemble model. Slope contributes the highest (+0.22), followed by restingBP (+0.08), serum cholesterol (+0.07), and chest pain (+0.07). Number of major vessels, maximum heart rate, resting electrocardiogram results, and old peak have moderate contribution, while fasting blood sugar and some other minor features have little contribution.

This indicates that slope is the most significant factor in heart disease prediction, having a great impact on the decision-making process of the model. The features associated with blood pressure, cholesterol, and chest pain are also significant, as per existing medical knowledge.

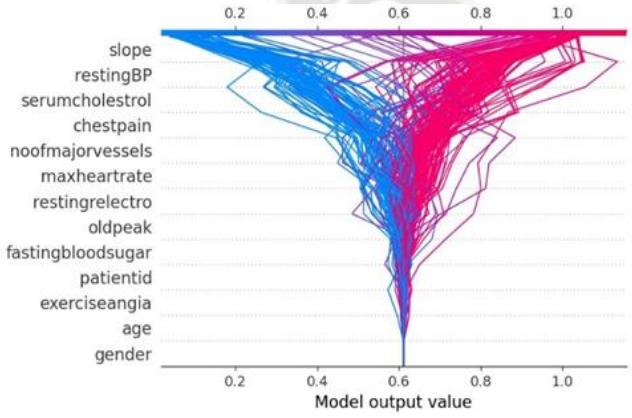

Figure. 1.13 Model Output value

This SHAP dependence plot illustrates how each feature contributes to the ensemble model's output probability of heart disease prediction. The x-axis is the model output value (probability of heart disease), and the y-axis contains the features. The color gradient (blue to red) denotes feature values, with blue representing low values and red representing high values [18].

Key Insights:

- Slope: The strongest feature, similar to in past SHAP plots, distinctly differentiates cases. Greater slope values (red) drive the model prediction towards greater probabilities of heart disease. In contrast, lower slope values (blue) drive predictions to lower probabilities.
- Resting Blood Pressure (restingBP): Greater resting BP values provide a greater probability of heart disease but with some fluctuation, as seen from the red and blue curve fluctuation.
- Serum Cholesterol & Chest Pain: Both have good positive trends, such that higher figures increase the estimated probability of heart disease by the model. These are consistent with medical information.
- Number of Major Vessels (noofmajorvessels): The pattern shows a negative correlation—increasing numbers of major vessels (red) lead to predictions towards low risks of heart disease, suggesting a protective factor.
- Exercise-Induced Angina & Old Peak: Effort-induced angina and ST depression (old peak) are positively adding towards heart disease prediction.
- Age & Gender: These are less impactful, with no clear distinction between red and blue curves, showing that while they are part of the model, their impact is not as pronounced as physiological markers.

VI. Conclusion

This research proposed and tested an AI-based method for heart disease prediction using several machine learning models, such as XGBoost, Random Forest, Multi-Layer Perceptron (MLP), and a Weighted Voting Ensemble. Among them, the ensemble model performed best, with an accuracy of 95.50% and an AUC-ROC of 0.9928, proving its superiority over single classifiers. These results support the strength of ensemble methods in enhancing predictive accuracy for sophisticated medical classification problems.

SHAP interpretability analysis validated that there are some key clinical parameters used as indicators for heart disease. Slope, resting blood pressure, serum cholesterol, and chest pain were highlighted as the strongest predictors, out of which slope was the strongest. SHAP summary and dependence plots also

helped identify interactions among features, attesting to the model's applicability in real-world clinical environments, and further making it a sound decision-supporting tool for diagnosing early stages of heart disease.

The results of the study indicate that AI-based predictive models have the potential to be key in helping medical professionals identify individuals at risk for heart disease. Through incorporation of explainability methods, e.g., SHAP, physicians can better perceive model predictions and ensure that these reflect recognized medical risk factors [19]. This will improve confidence and acceptance of AI in actual clinical practice.

Notwithstanding its encouraging findings, this study is not without its limitations. The data, while publicly available, was relatively small, which can influence the generalization ability of the model. Besides that, the model has also not been validated in real clinical environments yet, so its applicability in practice remains doubtful. More research must explore the use of larger and more diversified datasets, as well as advanced deep learning algorithms, such as transformers, to improve feature learning and predictive performance.

In conclusion, this study emphasizes the promise of AI in predicting heart disease and highlights the significance of model explainability in AI-based healthcare applications. The findings from this study can form the basis for future developments, eventually leading to early diagnosis and improved patient outcomes in cardiovascular treatment.

VII. FUTURE SCOPE

The results of the present study illustrate the capability of AI models in predicting heart disease at almost perfect levels. Nevertheless, there are a number of directions for future research and enhancement:

- The addition of multimodal data such as electrocardiograms (ECG), echocardiography, and genetic markers can increase predictive precision [20].
- Utilization of the model in health systems to evaluate risk for heart disease at a fast pace and with an early intervention.
- Ongoing refinement of SHAP-based explanations towards enhancing trust and adoption by doctors.

ACKNOWLEDGMENT

The authors are grateful to the Department of Computer Science and Engineering at Medi-Caps University staff and faculty for their constant guidance and encouragement during the course of this research. We also appreciate the developers of the freely available cardiovascular dataset utilized in this work, without which it wouldn't have been undertaken.

Special thanks are extended to Dr. Lokendra Singh Songare and Prof. Seema Patidar for their useful feedback, suggestions, and encouragement throughout the project. The computational resources and infrastructure of the institution played a critical role in facilitating model training and testing.

Last but not least, we acknowledge the contributions of various open-source communities whose libraries and tools played a crucial role in developing and testing our models.

REFERENCES

[1] WHO. (2023). Cardiovascular diseases (CVDs). Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Benjamin, E. J., et al. (2020). “Heart disease and stroke statistics—2020 update: A report from the American Heart Association”. Circulation, 141(9), e139-e596.

[3] Ye, C., et al. (2020). “Prediction of cardiovascular disease using machine learning techniques: A systematic review”. IEEE Access, 8, 181950-181961.

[4] Wilson, P. W., D’Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). “Prediction of coronary heart disease using risk factor categories”. Circulation, 97(18), 1837–1847.

[5] Chen, T., & Guestrin, C. (2016). “XGBoost: A scalable tree boosting system”. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[6] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). “LightGBM: A highly efficient gradient boosting decision tree”. Advances in Neural Information Processing Systems, 30.

[7] Li, Y., Zhao, J., Song, X., & Zhang, Y. (2020). “Feature selection methods for heart disease prediction using machine learning”. Journal of Medical Informatics, 12(3), 209–225.

[8] Murtaza, G., Sharif, M., Raza, M., & Iqbal, M. (2022). “Neural network-based approaches for cardiovascular disease prediction: A systematic review”. Expert Systems with Applications, 189, 116090.

[9] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2019). “Densely connected convolutional networks”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 470–481.

[10] Lundberg, S. M., & Lee, S. I. (2017). “A unified approach to interpreting model predictions”. Advances in Neural Information Processing Systems, 30, 4765–4774.

[11] Molnar, C. (2022). “Interpretable Machine Learning”. Independently published.

[12] Tjoa, E., & Guan, C. (2021). "A survey on explainable artificial intelligence (XAI): Toward medical AI transparency". IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813.

[13] Breiman, L. (2001). "Random forests". Machine Learning, 45(1), 5-32.

[14] Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep learning". MIT Press.

[15] Pedregosa, F., et al. (2011). "Scikit-learn: Machine learning in Python". Journal of Machine Learning Research, 12, 2825-2830.

[16] Powers, D. M. (2011). "Evaluation: From precision, recall, and F-measure to ROC, informedness, and correlation". Journal of Machine Learning Technologies, 2(1), 37-63.

[17] Zou, H., & Hastie, T. (2005). "Regularization and variable selection via the elastic net". Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

[18] Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). "Consistent individualized feature attribution for tree ensembles". arXiv preprint arXiv:1802.03888.

[19] Topol, E. J. (2019). "High-performance medicine: The convergence of human and artificial intelligence". Nature Medicine, 25(1), 44-56.

[20] Rajpurkar, P., et al. (2017). "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning". arXiv preprint arXiv:1711.05225.