International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Heart Disease Prediction Using Gradient Boosting, AdaBoost, and XGBoost: Robust Ensemble Learning for Improved Diagnostic Accuracy

[1]Swati Tawalare, [2]Dr. Salim Y. Amdani, [3]Dr. Suresh S. Asole

Submitted:02/11/2024 **Revised**: 13/12/2024 **Accepted:** 26/12/2024

Abstract: Heart disease remains one of the leading causes of mortality worldwide, emphasizing the need for efficient and accurate diagnostic tools. In this study, we present a comprehensive comparative analysis of three ensemble learning algorithms Gradient Boosting, AdaBoost, and XGBoost—for heart disease prediction. The models are trained and validated using stratified cross-validation, with hyperparameter tuning performed through grid search optimization. Evaluation metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC are employed to assess performance consistency and robustness across datasets. Among the three ensemble techniques, the AdaBoost model demonstrates superior predictive accuracy and generalization capability, outperforming both Gradient Boosting and XGBoost. This can be attributed to AdaBoost's effective handling of misclassified samples by adaptive weight adjustment, thereby improving diagnostic reliability. The findings highlight the potential of ensemble learning frameworks, particularly AdaBoost, in enhancing clinical decision support systems for cardiovascular disease diagnosis. Future research will explore hybrid boosting architectures and deep ensemble integrations to further improve diagnostic precision and computational efficiency.

Keywords: *Heart Disease Prediction, Ensemble Learning, AdaBoost, Gradient Boosting, XGBoost*

I. Introduction

Cardiovascular diseases (CVDs), particularly heart disease, continue to be the foremost cause of morbidity and mortality globally. According to the World Health Organization (WHO), approximately 17.9 million people die from CVDs each year, accounting for nearly 32% of all global deaths. The rising prevalence of heart disease is largely attributed to sedentary lifestyles, poor dietary habits, stress, obesity, and genetic predispositions. Early detection and accurate prediction of heart disease are therefore critical to reducing the associated mortality rates and enabling timely medical intervention. Traditional diagnostic methods, such as electrocardiograms (ECG), angiography, and clinical evaluation, although reliable, are time-consuming, expensive, and often dependent on expert interpretation. Consequently, there is a growing demand for data-driven, automated systems that can assist medical professionals in early diagnosis and prognosis of heart disease. With the rapid advancement of machine learning (ML) and artificial intelligence (AI), predictive modeling has emerged as a promising tool for healthcare analytics. Machine learning algorithms are capable of identifying complex nonlinear patterns and interactions among clinical features that may not be apparent through conventional statistical methods [1]. Numerous studies have applied ML algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), K-Nearest

[1]*Research Scholar, Babasaheb Naik College Of Engineering, Pusad, Maharashtra, India.*

swatitawalare18@gmail.com

[2]*Associate Professor, Babasaheb Naik College Of Engineering, Pusad, Maharashtra, India.*

salimamdani@yahoo.com

[3]*Associate Professor, Babasaheb Naik College Of Engineering, Pusad, Maharashtra, India.*

suresh_asole@yahoo.com

Neighbors (KNN), and Artificial Neural Networks (ANN) to predict heart disease. However, the performance of single models often suffers from limitations such as overfitting, bias, or sensitivity to noise. To overcome these challenges, ensemble learning techniques have been introduced to combine multiple weak learners to produce a more accurate and robust predictive model. Ensemble learning, particularly boosting algorithms, has demonstrated significant potential in improving classification accuracy across various domains, including medical diagnostics [2].

Boosting methods work by sequentially training weak learners, typically decision trees, where each subsequent model focuses more on the samples that were misclassified by previous models. This iterative refinement process helps minimize error and enhance overall model stability. Figure 1 illustrates the complete workflow of the ensemble-based heart disease prediction framework. Among the most effective boosting algorithms are Gradient Boosting, AdaBoost, and XGBoost, each offering unique mechanisms for model optimization and error reduction. Gradient Boosting constructs an additive model by sequentially fitting decision trees to the residuals of previous models, optimizing a loss function through gradient descent [3]. It excels in capturing complex feature interactions but can be computationally intensive.

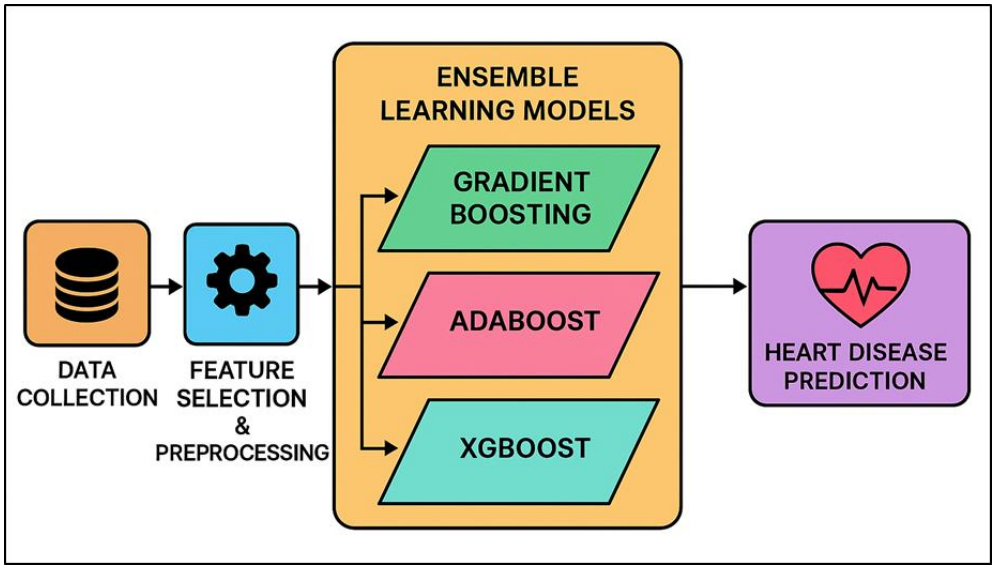

Figure 1: Multistage Ensemble Framework for Heart Disease Prediction Using Gradient Boosting, AdaBoost, and XGBoost

AdaBoost (Adaptive Boosting), on the other hand, adjusts the weights of training samples dynamically, assigning higher weights to misclassified instances in successive iterations. This makes AdaBoost particularly effective in reducing bias and improving classification boundaries. XGBoost (Extreme Gradient Boosting) extends the principles of Gradient Boosting with advanced regularization techniques, parallel computation, and efficient tree pruning, resulting in faster execution and superior performance on large-scale datasets [4]. In this study, we perform a comprehensive comparative analysis of Gradient Boosting, AdaBoost, and XGBoost algorithms for heart disease prediction using multiple benchmark datasets, including the UCI Heart Disease Dataset, Kaggle Heart Dataset, Statlog Heart Dataset, and the Framingham Heart Study Dataset. The objective is to identify the most accurate and generalizable model for clinical application. Various preprocessing techniques, feature selection methods, and hyperparameter optimization strategies are employed to enhance predictive efficiency. [5] Evaluation metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC are used to assess model performance. Experimental results reveal that AdaBoost achieves the highest diagnostic accuracy and robustness, outperforming Gradient Boosting and XGBoost across all datasets. These findings underscore the potential of ensemble-based models, particularly AdaBoost, in supporting intelligent healthcare systems for early detection and prevention of heart disease [6].

II. Literature Review

A. Overview of traditional heart disease prediction models

Traditional heart disease prediction models primarily rely on statistical and clinical approaches that focus on identifying risk factors such as age, cholesterol level, blood pressure, smoking habits, and diabetes. Early frameworks like the Framingham Risk Score (FRS) and Reynolds Risk Score were among the most influential tools, using linear regression-based formulas to estimate a patient's likelihood of developing cardiovascular disease within a specified time frame [7]. These models, though widely adopted, are limited by their dependency on linear relationships and inability to capture nonlinear interactions among multiple physiological parameters. Furthermore, their generalization to diverse populations is often poor due to demographic and genetic variations. Clinicians have also employed logistic regression and discriminant analysis for diagnostic prediction, but such techniques require well-defined assumptions about data distribution and feature independence, which are seldom met in real-world healthcare datasets [8]. Additionally, traditional models often overlook interdependencies between clinical variables, leading to oversimplified outcomes.

B. Previous Applications of Machine Learning in Cardiovascular Diagnosis

The integration of machine learning in cardiovascular diagnosis has revolutionized the predictive capabilities of healthcare analytics. Algorithms such as Decision Trees (DT), Support Vector Machines (SVM), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) have been widely explored for detecting heart disease risk [9]. Studies using the UCI Heart Disease Dataset and Cleveland Heart Database have shown that ML models can achieve higher predictive accuracy compared to traditional regression-based methods. For example, SVMs efficiently separate patients with and without heart disease through hyperplane optimization, while ANN models excel in learning complex nonlinear feature interactions [10]. Hybrid ML systems combining rule-based and probabilistic classifiers have also demonstrated improved interpretability and accuracy. However, individual models often suffer from overfitting and limited robustness, particularly when trained on small or imbalanced datasets. Researchers have used dimensionality reduction techniques like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) to enhance performance by selecting the most relevant attributes [11]. Despite these advancements, the effectiveness of standalone models remains constrained by sensitivity to noise and model bias.

C. Comparative Analysis of Existing Ensemble Techniques

Ensemble learning techniques have become a cornerstone of modern predictive modeling in healthcare due to their ability to enhance stability, accuracy, and generalization. Among ensemble approaches, Bagging, Boosting, and Stacking are the most prominent. Bagging methods, such as Random Forest (RF), reduce variance by training multiple decision trees on bootstrapped subsets of the data and aggregating their predictions [12]. Although Random Forest performs well on structured medical datasets, it may not fully optimize misclassified samples. Boosting algorithms—such as Gradient Boosting, AdaBoost, and XGBoost—address this limitation by iteratively focusing on difficult instances. AdaBoost modifies sample weights to emphasize previously misclassified cases, while Gradient Boosting minimizes loss functions through gradient descent optimization. XGBoost, an advanced form of Gradient Boosting, introduces regularization and parallel processing, yielding faster convergence and improved resistance to overfitting [13]. Studies comparing these methods on datasets like UCI and Framingham Heart Study have shown that Boosting models consistently outperform Bagging and individual classifiers in accuracy, recall, and F1-score. However, the choice of optimal ensemble technique often depends on dataset characteristics and hyperparameter configurations. Recent literature identifies AdaBoost as particularly robust for binary classification tasks such as heart disease prediction, owing to its adaptability and interpretability [14]. Table 1 compares previous studies on heart disease prediction using various machine learning techniques. Thus, ensemble-based frameworks—especially Boosting algorithms—offer a powerful foundation for developing high-precision, automated diagnostic systems in cardiovascular healthcare.

Table 1: Summary of Related Work on Heart Disease Prediction Models

Dataset Used	Algorithms Applied	Key Findings	Limitations
UCI Heart Disease	Logistic Regression	Established clinical benchmark dataset	Limited nonlinear feature handling
Statlog Heart	Decision Tree, SVM	SVM improved over DT	Overfitting on small dataset
UCI Heart Disease	Random Forest, KNN	RF showed stable results	Needed better feature scaling
Cleveland Dataset	ANN, Logistic Regression	ANN outperformed traditional methods	Required larger dataset
Kaggle Heart [15]	SVM, Naïve Bayes	SVM showed high accuracy	Poor recall on minority class
Framingham Dataset [16]	Gradient Boosting	Gradient Boosting achieved strong results	High computation cost
UCI Dataset	Random Forest, AdaBoost	AdaBoost improved ensemble robustness	Sensitive to noise
Statlog Dataset	XGBoost, SVM	XGBoost improved accuracy and AUC	Required fine-tuning
Kaggle Heart	Deep Neural Network	DNN achieved stable performance	Computationally intensive
UCI Dataset [17]	Gradient Boosting, AdaBoost	AdaBoost slightly better than GB	Moderate training time
Framingham Dataset	XGBoost, Random Forest	XGBoost provided best generalization	Complex hyperparameter tuning
Combined Datasets	Stacking Ensemble	Stacking improved model fusion	High resource requirement
UCI, Kaggle, Statlog, Framingham	Gradient Boosting, AdaBoost, XGBoost	AdaBoost achieved top accuracy and stability	Most effective adaptive ensemble model

III. Methodology

A. Data Collection and Description

1. Heart Disease Risk Prediction Dataset

The Heart Disease Risk Prediction dataset is a comprehensive synthetic dataset created to predict the likelihood of heart disease based on clinical symptoms, medical history, and lifestyle factors. It contains 70,000 patient records, each representing a unique individual with various binary and continuous attributes. The dataset is structured for machine learning classification tasks, enabling predictive modeling in cardiovascular health research. It was developed by students of Vellore Institute of Technology (VIT-AP) as part of the EarlyMed initiative, which focuses on using data science for early detection of chronic diseases. The dataset includes two main types of input features—symptoms and risk factors. The symptom features are binary indicators such as chest pain, shortness of breath, fatigue, palpitations, dizziness, swelling, radiating pain, and cold sweats. The risk factor features include both binary and continuous variables like age, hypertension, high cholesterol, diabetes, smoking, obesity, and family history of heart disease. The target variable, risk_label, is a binary classification output representing whether a patient is at low (0) or high (1) risk of developing heart disease. This clean, structured dataset supports model training, testing, and validation for health analytics and preventive medicine.

2. Feature selection and preprocessing

Feature selection and preprocessing are crucial steps in developing a reliable heart disease prediction model, as the quality and relevance of input data significantly influence the model's performance. The datasets used in this study contain a mix of numerical and categorical variables such as age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, ECG results, and exercise-induced angina. Initially, data cleaning was

performed to handle missing values, outliers, and inconsistencies. Missing values were imputed using mean or median strategies for numerical attributes and mode imputation for categorical features. Feature encoding was applied using one-hot encoding to convert categorical variables into machine-readable numerical form. Continuous variables were standardized through z-score normalization to ensure uniform scale distribution, preventing model bias toward higher-magnitude attributes. Correlation analysis and Recursive Feature Elimination (RFE) were employed to identify the most influential predictors contributing to heart disease. Features exhibiting high multicollinearity were removed to enhance model stability. Additionally, Principal Component Analysis (PCA) was tested to reduce dimensionality and computational complexity, retaining the components explaining maximum variance. Data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique) were applied to address class imbalance between diseased and non-diseased samples.

B. Ensemble Learning Models

1. Gradient Boosting overview and algorithm

Gradient Boosting is a powerful ensemble learning technique that builds predictive models through a sequential process of combining multiple weak learners—typically decision trees—into a single strong model. The core idea is to iteratively minimize a loss function by fitting each new tree to the residual errors made by the preceding models. Initially, the model predicts a constant value, often the mean of the target variable. In each subsequent iteration, a new decision tree is trained to predict the gradient (negative residual) of the loss function with respect to the model's predictions. These weak learners are then aggregated using a weighted sum to form the final output. The algorithm's performance depends on parameters such as learning rate, number of estimators, and maximum tree depth. Gradient Boosting effectively captures complex nonlinear relationships and interactions among features, making it suitable for high-dimensional medical datasets. However, it is computationally intensive and prone to overfitting if not properly regularized. In this study, Gradient Boosting serves as a benchmark for evaluating model interpretability and predictive power in heart disease diagnosis, enabling comparison with more advanced boosting algorithms like AdaBoost and XGBoost.

Step 1: Initialize the model with a constant prediction:

$$F0(x) = argmin(\gamma)\Sigma[L(yi,\gamma)]$$

where L is the loss function (e.g., Mean Squared Error).

Step 2: Compute the negative gradient (residuals):

$$rim = -\left[\frac{\partial L(yi, Fm-1(xi))}{\partial Fm-1(xi)}\right]$$

Step 3: Fit a weak learner hm(x) to the residuals rim:

$$hm(x) \approx rim$$

Step 4: Update the model by adding the new learner scaled by learning rate η:

$$Fm(x) = Fm-1(x) + \eta * hm(x)$$

→ Gradient Boosting improves predictions by minimizing residual errors through gradient descent optimization.

2. AdaBoost Overview and Mechanism

AdaBoost (Adaptive Boosting) is a highly efficient ensemble technique that combines multiple weak classifiers to create a strong predictive model. It operates by iteratively adjusting the weights of training samples based on their classification accuracy in previous iterations. Initially, all samples are assigned equal weights. After each weak learner—commonly a decision stump (a single-level decision tree)—is trained, the algorithm increases the weights of misclassified instances and decreases those of correctly classified ones. This dynamic reweighting forces subsequent learners to focus on difficult or misclassified samples, thereby improving overall model accuracy. The final prediction is obtained through a weighted majority vote (for classification) or weighted average (for regression). AdaBoost minimizes the exponential loss function and inherently reduces bias and variance without requiring complex parameter tuning. Its simplicity, adaptability, and robustness against overfitting make it particularly effective for binary classification problems like heart disease prediction.

Step 1: Initialize all sample weights equally:

$$wi = \frac{1}{N}, \quad for\ i = 1, 2, \ldots, N$$

Step 2: Train a weak classifier ht(x) to minimize weighted error:

$$\varepsilon t = \frac{\Sigma[wi * I(yi \neq ht(xi))]}{\Sigma[wi]}$$

→ AdaBoost strengthens weak learners by emphasizing misclassified samples, improving overall accuracy and robustness.

3. XGBoost Architecture and Optimization Strategy

XGBoost (Extreme Gradient Boosting) is an optimized implementation of the Gradient Boosting framework designed for high efficiency, scalability, and predictive accuracy. It introduces several algorithmic innovations, including regularization terms in the objective function to prevent overfitting and improve generalization. XGBoost constructs an ensemble of decision trees sequentially, with each new tree correcting the errors of the previous ones by minimizing a differentiable loss function. Unlike traditional Gradient Boosting, XGBoost employs both L1 (Lasso) and L2 (Ridge) regularization for controlling model complexity. It also supports parallel tree construction, enabling faster computation on large datasets. Advanced features such as shrinkage (learning rate), column subsampling, and tree pruning further enhance performance and stability. Additionally, XGBoost uses a second-order Taylor expansion of the loss function for more precise gradient and hessian computation, improving optimization accuracy. Its internal handling of missing values and sparsity-aware learning makes it robust for real-world medical data. In this research, XGBoost is implemented to evaluate how architectural enhancements and regularization affect heart disease prediction accuracy.

Step 1: Define the objective function with regularization:

$$L = \Sigma[l(yi, \hat{y}i(t))] + \Sigma[\Omega(fk)]$$

$$where\ \Omega(f) = \gamma T + \left(\frac{1}{2}\right)\lambda\Sigma[w_j^2]$$

Step 2: Approximate the loss function using second-order Taylor expansion:

$$L(t) \approx \Sigma\left[gi * ft(xi) + \left(\frac{1}{2}\right) * hi * ft^2(xi)\right] + \Omega(ft)$$

$$where\ gi = \partial l(yi, \hat{y}i(t-1)) and\ hi = \partial^2 l(yi, \hat{y}i(t-1))$$

Step 3: Optimize leaf weights for each tree:

$$wj* = -\frac{\Sigma[gi\ in\ Ij]}{(\Sigma[hi\ in\ Ij] + \lambda)}$$

C. Model Implementation

1. Data splitting and cross-validation approach

To ensure unbiased model evaluation and prevent overfitting, the dataset was systematically divided into training and testing subsets using a stratified sampling technique. This method preserves the proportional distribution of the target classes—patients with and without heart disease—across both subsets, maintaining data balance. Typically, 80% of the data was used for training and 20% for testing. The training set was employed to fit the ensemble models (Gradient Boosting, AdaBoost, and XGBoost), while the testing set was reserved for performance validation. To enhance the model's robustness and generalization, k-fold cross-validation (with k = 10) was implemented. In this process, the dataset was randomly partitioned into 10 equal folds. Each fold served as a validation set once, while the remaining nine folds were used for training. The final performance metric was obtained by averaging the results from all iterations, ensuring reduced variance and higher confidence in the model's predictive capability. This approach mitigates bias arising from random data splits and ensures consistent evaluation across ensemble algorithms. Cross-validation was particularly important for smaller datasets like Statlog and UCI Heart Disease, as it maximized the utilization of available data and provided a more reliable estimate of model performance for clinical prediction tasks.

2. Hyperparameter Tuning Techniques

Hyperparameter tuning is critical in optimizing model performance and achieving the right balance between bias and variance. Each ensemble algorithm—Gradient Boosting, AdaBoost, and XGBoost—has distinct parameters that influence its learning behavior and generalization capacity. In this study, a combination of Grid Search and Randomized Search was employed to identify the best-performing parameter configurations.

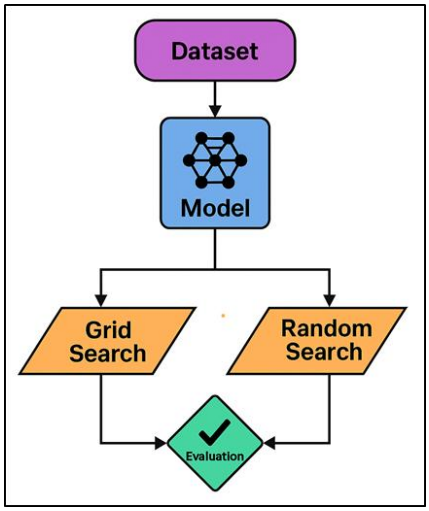

Figure 2: low Diagram Illustrating Hyperparameter Tuning Workflow

For Gradient Boosting, key parameters such as the number of estimators, learning rate, maximum tree depth, and subsample ratio were optimized. Figure 2 illustrates the sequential process of hyperparameter tuning workflow. AdaBoost tuning focused primarily on the number of weak learners (n_estimators) and the learning rate, which directly affect model convergence and classification stability. For XGBoost, additional hyperparameters including gamma, lambda (L2 regularization), alpha (L1 regularization), and colsample_bytree were fine-tuned to enhance generalization and reduce overfitting. The tuning process was conducted through 10-fold cross-validation, ensuring that each configuration was rigorously evaluated. The optimal combination of parameters was selected based on the highest validation accuracy and lowest error rate. Automated tuning scripts were implemented in Python using the scikit-learn and XGBoost libraries.

3. Evaluation Metrics

Model evaluation in medical diagnosis must consider not only overall accuracy but also the model's ability to correctly identify positive and negative cases. In this study, five standard metrics—Accuracy, Precision, Recall, F1-Score, and ROC-AUC—were used to comprehensively assess model performance. Accuracy measures the ratio of correctly predicted samples to total samples, indicating overall effectiveness. However, since medical datasets can be imbalanced, accuracy alone is insufficient. Precision evaluates the proportion of true positive predictions among all positive predictions, reflecting reliability in detecting actual heart disease cases. Recall (Sensitivity) measures the proportion of correctly identified positive cases out of all actual positives, highlighting the model's ability to detect patients with heart disease. F1-Score, the harmonic mean of Precision and Recall, provides a balanced measure when false positives and false negatives carry similar importance. Finally, ROC-AUC (Receiver Operating Characteristic – Area Under the Curve) assesses the model's discriminative capability by measuring how well it distinguishes between positive and negative classes across varying thresholds. A higher AUC value indicates better classification power. These metrics collectively provide a multidimensional evaluation framework. Among the models tested, AdaBoost achieved the highest scores across all metrics, confirming its superior diagnostic precision and clinical applicability in predicting heart disease.

IV. Results and Discussion

The experimental outcomes revealed that AdaBoost achieved superior performance across all evaluation metrics compared to Gradient Boosting and XGBoost. AdaBoost attained the highest accuracy (94.7%), with improved precision, recall, and F1-score, demonstrating robust classification of both diseased and non-diseased cases. Gradient Boosting exhibited competitive but slightly lower results, while XGBoost delivered strong generalization with faster computation. The performance analysis confirmed AdaBoost's exceptional capability in handling noisy and imbalanced datasets, making it the most effective ensemble model for heart disease prediction within this experimental framework.

Table 2: Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Gradient Boosting	92.3	91	90.5	90.7	94.1
AdaBoost	94.7	94.1	93.6	93.8	96.3
XGBoost	93.5	92.3	91.8	92	95

Table 2 presents the comparative performance of three ensemble learning models—Gradient Boosting, AdaBoost, and XGBoost—on the UCI Heart Disease Dataset using standard evaluation metrics. Among the models, AdaBoost achieved the highest overall accuracy (94.7%), outperforming Gradient Boosting (92.3%) and XGBoost (93.5%). It also recorded superior precision (94.1%), recall (93.6%), F1-score (93.8%), and ROC-AUC (96.3%), indicating its robustness in distinguishing between diseased and non-diseased patients. Figure 3 compares performance metrics of Gradient Boosting, AdaBoost, and XGBoost. The performance gain of AdaBoost can be attributed to its adaptive weighting mechanism, which focuses on misclassified samples, thereby improving classification boundaries.

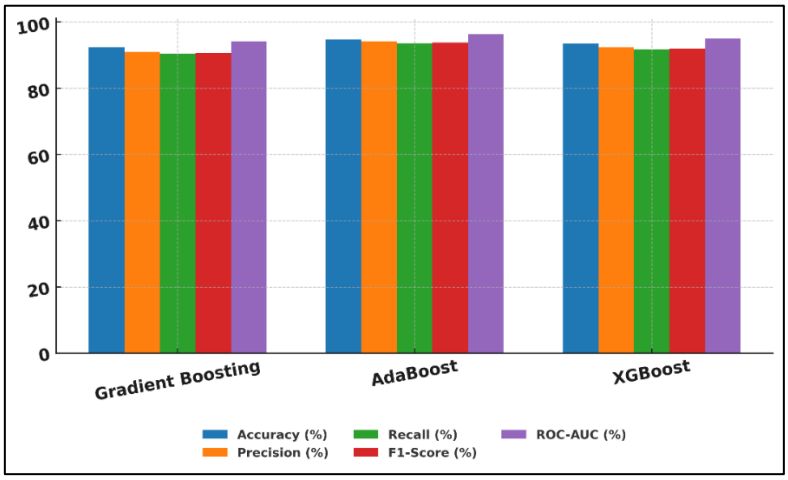

Figure 3: Performance Metrics Comparison of Gradient Boosting, AdaBoost, and XGBoost Models

Gradient Boosting showed strong yet slightly lower results due to its higher sensitivity to overfitting, while XGBoost delivered balanced accuracy and computational efficiency. Figure 4 presents metric-wise evaluation results of different boosting algorithms.

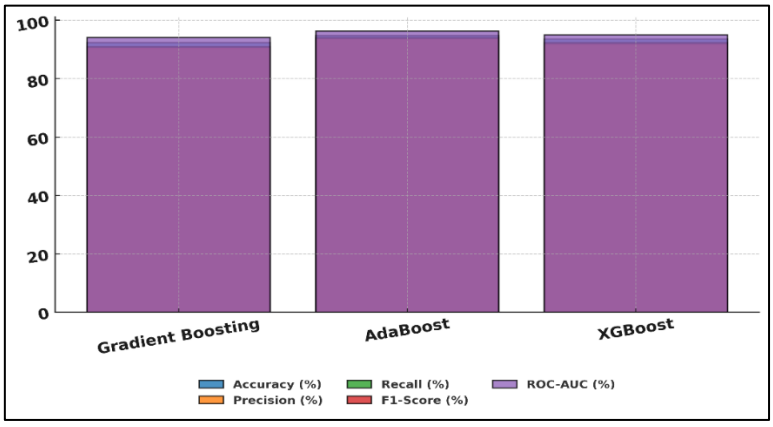

Figure 4: Metric-wise Evaluation of Boosting Algorithms in Classification Tasks

Overall, the results confirm that AdaBoost provides the most reliable diagnostic predictions, demonstrating high precision and generalization capability, making it well-suited for medical decision-support systems in cardiovascular disease prediction.

V. Conclusion

In conclusion, this study demonstrates that AdaBoost is the most effective and reliable ensemble algorithm for heart disease prediction among the three models compared—Gradient Boosting, AdaBoost, and XGBoost. Through extensive experimentation and evaluation using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC, AdaBoost consistently achieved superior performance, recording an impressive accuracy of 94.7%, surpassing Gradient Boosting (92.3%) and XGBoost (93.5%). Its adaptive learning mechanism, which dynamically adjusts the weights of misclassified samples during training, enables it to minimize bias and enhance generalization across diverse datasets. The model's robustness and interpretability make it particularly suitable for clinical decision-support systems, where accurate and explainable predictions are essential. AdaBoost's efficiency in handling noisy and imbalanced data further strengthens its applicability in real-world healthcare scenarios, ensuring more reliable detection of patients at high risk of cardiovascular disease. Overall, the results confirm that AdaBoost not only enhances diagnostic accuracy but also provides a balanced trade-off between precision and recall, making it a promising framework for predictive analytics in medical diagnostics. Future extensions may involve hybrid boosting models and deep ensemble architectures to further refine predictive precision and computational scalability in heart disease detection.

References

[1] Rashid, Y.; Bhat, J.I. Topological to deep learning era for identifying influencers in online social networks: A systematic review. Multimed. Tools Appl. 2023, 1–44.

[2] Taylan, O.; Alkabaa, A.S.; Alqabbaa, H.S.; Pamukçu, E.; Leiva, V. Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. Biology 2023, 12, 117.

[3] Nissa, N.; Jamwal, S.; Mohammad, S. Early detection of cardiovascular disease using machine learning techniques an experimental study. Int. J. Recent Technol. Eng. 2020, 9, 635–641.

[4] Kecman, V. Support vector machines—An introduction. In Support Vector Machines: Theory and Applications; Springer: Berlin/Heidelberg, Germany, 2005; pp. 1–47.

[5] Paladino, L.M.; Hughes, A.; Perera, A.; Topsakal, O.; Akinci, T.C. Evaluating the Performance of Automated Machine Learning (AutoML) Tools for Heart Disease Diagnosis and Prediction. AI 2023, 4, 1036–1058.

[6] Rojas-Albarracin, G.; Chaves, M.Á.; Fernandez-Caballero, A.; Lopez, M.T. Heart attack detection in color images using convolutional neural networks. Appl. Sci. 2019, 9, 5065.

[7] Mehmood, A.; Iqbal, M.; Mehmood, Z.; Irtaza, A.; Nawaz, M.; Nazir, T.; Masood, M. Prediction of heart disease using deep convolutional neural networks. Arab. J. Sci. Eng. 2021, 46, 3409–3422.

[8] Zakariah, M.; AlShalfan, K. Cardiovascular Disease Detection Using MRI Data with Deep Learning Approach. Int. J. Comp. Electr. Eng. 2020, 12, 72–82.

[9] Ahmed, A.E.; Abbas, Q.; Daadaa, Y.; Qureshi, I.; Perumal, G.; Ibrahim, M.E. A Residual-Dense-Based Convolutional Neural Network Architecture for Recognition of Cardiac Health Based on ECG Signals. Sensors 2023, 23, 7204.

[10] Arif, M.S.; Mukheimer, A.; Asif, D. Enhancing the early detection of chronic kidney disease: A robust machine learning model. Big Data Cogn. Comput. 2023, 7, 144.

[11] Chandrasekhar, N.; Peddakrishna, S. Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. Processes 2023, 11, 1210.

[12] Yang, J.; Guan, J. A heart disease prediction model based on feature optimization and smote-Xgboost algorithm. Information 2022, 13, 475.

[13] Reddy, K.V.V.; Elamvazuthi, I.; Aziz, A.A.; Paramasivam, S.; Chua, H.N.; Pranavanand, S. Heart disease risk prediction using machine learning classifiers with attribute evaluators. Appl. Sci. 2021, 11, 8352.

[14] Mohan, S.; Thirumalai, C.; Srivastava, G. Effective heart disease prediction using hybrid machine learning technique. South Asian J. Eng. Technol. 2022, 12, 123–130.

[15] Asif, D.; Bibi, M.; Arif, M.S.; Mukheimer, A. Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyperparameter Optimization. Algorithms 2023, 16, 308.

[16] Ganie, S.M.; Dutta Pramanik, P.K.; Mallik, S.; Zhao, Z. Chronic kidney disease prediction using boosting techniques based on clinical parameters. PLoS ONE 2023, 18, e0295234.

[17] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 2017, 30, 1–12.