Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

# COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING AND ARTIFICIAL NEURAL NETWORK MODELS FOR HEART DISEASE FORECASTING

**Anam Zahoor[*1], Rashid Mehmood Gondal[2], Muhammad Atif Sultan[3], Aqsa Ejaz[4], Muhammad Saqib[5], Muhammad Waqas Haider[6], Muhammad Usman[7]**

[*1,3,4,5,6,7]*Department of Computer Science, The University of Lahore, Sargodha Campus, Sargodha 40100, Pakistan*
[2]*University of Sargodha, Sargodha 40100, Pakistan*

[*1]malikanam06@gmail.com, [2]rashidmehmoodit@gmail.com, [3]atifbhatti865@gmail.com, [4]zulqarnainaqsa3@gmail.com, [5]saqibaliriaz888@gmailcom, [6]Whaider491@gmail.com, [7]usmanituos1@gmail.com

DOI: https://doi.org/10.5281/zenodo.20175107

**Keywords**
*cardiovascular disease prediction; machine learning; Random Forest; XGBoost; Support Vector Machine; artificial neural network; clinical classification.*

**Article History**
*Received: 18 March 2026*
*Accepted: 28 April 2026*
*Published: 14 May 2026*


**Corresponding Author: ***
**Anam Zahoor**

**Abstract**
*Cardiovascular disease (CVD) is still the most common cause of death in the whole world. Identification of patients at risk on early stage is very important for making the decisions better clinically. In this study, four supervised machine learning models are used on the Cleveland cardiac Disease dataset. These models are Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), and an Artificial Neural Network (ANN).*
*All models were trained using same dataset split. Where 80% data was used for training and 20% for testing. The dataset consists of 303 patient records with 13 clinical and demographic features. Before training, the data was normalized by using z-score standardization.*
*The results show that the Random Forest (RF) and XGBoost performed the best. Both achieving an accuracy of 98.53%. ANN achieves 94.14% and SVM achieves 88.78% with its default settings. Additional evaluation such as (ROC) analysis, precision-recall evaluation, and feature importance of Random Forest (RF) were used for better understanding of model performance. The findings show that ensemble tree-based methods work very effectively on these organized tabular clinica dataset.*

## Introduction

The one of the major causes of death in all over the world is Cardiovasculardiseases known as CVDs, and it continue to inflict a dangerous load on healthcare systems [20], [3]. Heart disease develops increasingly and may remain unobserved until major symptoms appear. Therefore, early diagnosis is necessary for reducing deathrate, making treatment planning better, and supporting healthful care.

Formulaic diagnosis of heart disease relies on clinical testing, laboratory testing, electrocardiographic analysis, imaging, and physician expertness. These approaches are effective, they may require time, technical resources, and specialist explanation [4]. As digital medical records are increasing with the

passage of time, machine learning has seen to be a useful approach in development of intelligent systems that can help clinicians by distinguishing patterns in patient data and forecasting disease risk [5], [6].

Machine learning methods are able to process various clinical attributes simultaneously, like statistical attributes(age and sex), clinical measurements (blood pressure and cholesterol levels), symptom characteristics (chest pain type), metabolic status attribute (fasting blood sugar), and the indicators based on exercise. Heart disease is influenced by several interacting factors. In recent years, various algorithmic program such as Support Vector Machine, Random Forest, XGBoost, and Artificial Neural Networks have been broadly applied in healthcare prediction tasks [9]–[12], such predictive models can give worthy support for early screening and assessment of risk [7], [8].

Among these methods, Random Forest is known for beefy performance and interpret ability of dimension, XGBoost is valued for its strong predictive accuracy on organized data, SVM has a powerful theoretical support for binary classification, and ANN can sort nonlinear attribute interactions effectively [11], [13], [14], [15]. it is impossible for a single model to perform best on every dataset. Here the comparative assessment becomes vital. Some factors like size ,nature of attributes and the way of preprocessing may affect the performance of models [16], [17].

We used four predictive models (Random Forest(RF), XGBoost, SVM, and ANN) for cardiac disease classification. Models were traind on 1,025 patient records of an organized clinical data. The whole process includes preprocessing, feature analysis, model training, and evaluation using accuracy, confusion matrix analysis, ROC and precision–recall curves, and feature importance.

The primary goal of the study is to determine the most credible model to predict heart disease with the help of the chosen set of data. The other objective is to study the role of clinical attributes in prediction and whether these models can be applied to healthcare. The research can be useful in the creation of smart diagnostic tools that can help physicians to identify risks early.

Section 2(Literature review) is the review of the past studies and Section 3(Dataset Description) describes the dataset. Section 4 provides the methods of analysis and their mathematical formulations(Methodology).

Literature Review

The usage of ML in medical field for primal detection has enlarged speedily over the last two decades, especially in the area of cardiac disease forecasting. Researchers have evaluated a broad range of statistical, machine learning, and deep learning methods to make early diagnosis better, reduce diagnostic error, and support in making clinical decisions. Since cardiac disease is effected by many interrelated clinical attributes, predictive modeling has become a utile tool for analyzing organized patient data and ascertaining disease risk more efficiently [2].

Early studies in this area relied heavily on statistical and approaches based on probability. Detrano et al. formulated one of the foundational predictive architecture for coronary artery disease using organized clinical attributes and analysis based on probability [1]. Their work became highly dominating because it attested that patient level clinical features could be organized into a predictive model for cardiac disease diagnosis. This study also laid the basis for datasets that later became standardized criteria in ML research on cardiovascular forecasting.

As machine learning techniques eveloved, researchers began to use supervised classification models to improve predictive functioning. Support Vector Machine, introduced by Cortes and Vapnik, due to its strong hypothetical basis and tendency to boost separation between classes it became among the most broadly selected classification techniques [11]. In medical domain, SVM has been used to several disease forecasting problems.It performed advisable in binary classification settings. It can form nonlinear relationships through kernel functions [6]. In the environment of cardiac disease forecasting, when attributes are normalized and optimized, The SVM can provide reliable classification accuracy [23].

Random Forest(RF), proposed by Breiman. It was a powerful alternative due to its aggregate learning process [13]. It aggregates the results of many combined decision trees which improves normalization and reduces overfitting. In healthcare uses, Random Forest(RF) is very essential because it handles organiazed data effectively. It can also estimate feature importance. This interpretability is strongly effective in medical domains, where clinicians understand which attributes impart most strongly to the forecasting [17]. Several studies have proved that Random Forest(RF) gives better results in heart disease classification [7], [25]. It often improves the achievement of linear classifiers [7], [25].

Gradient boosting models have also performed well in predictive healthcare domains. XGBoost, introduced by Chen and Guestrin,it is an prestigious model due to its efficiency, scalability [14]. XGBoost improves weak learners sequencially. It has established strong predictive accuracy on organiazed datasets. It has been a preferred technique in many medical classification tasks, including cardiovascular disease(CVDs) prediction. It captures nonlinear patterns while maintaining computational efficiency [9]. Comparative outcomes have often proved that boosting methods outperform other classical machine learning algorithms on cardiac disease datasets [25], [29].

Artificial Neural Networks(ANN) have also been used vast in medical field. Goodfellow, Bengio, and Courville mentioned the capability of these two learning models (neural networks and deep learning) to understand complicated data representations in wide range of domains [12]. In Medical domain, The systems based on ANN have been used for patient outcome prediction, biosignal processing, medical image interpretation, and classsification of disease [15]. For detection of cardiac disease, neural networks can model nonlinear interactions among clinical attributes. Their performance often depends on structure, hyperparameter tuning, and size of dataset [29]. On organized systematic datasets with intermediate sample size, ANN models may perform well but do not always perform well aggregate methods based on tree [30].

Palaniappan and Awang formulated an intelligent cardiac disease forecasting system by using data mining techniques. They showed that prophetical analytics could improve healthcare decision support [20]. Their study explained the practical use of classification models in identifying disease probability. Later, Mohan et al. proposed an effective cardiac disease prediction method using composite machine learning strategies. He also reported improved performance through model optimization [14]. Their findings improved the statement that machine learning can act as an crucial part in early diagnosis of cardiac disease.

Anooj proposed a objective outcome support structure based on weighted scattered rules for cardiac disease risk forecasting. He also accented the grandness of intelligent decision systems in healthcare [22]. Similarly, Alizadehsani et al. introduced a data mining technique for coronary artery disease diagnostic assesment. He also highlighted the value of structured clinical features in predictive modeling [15]. Their work stated that careful feature analysis can significantly causing predictive performance.

Acharya et al. also rumored that ML methods are capable in detection of heart disease diagnosis by making classification performance better and overcoming dependency on purely manual analysis [10]. Their work accented the growing clinical value of intelligent diagnostic systems. Bashir et al. later proved that aggregate methods performs very well on organized cardiovascular datasets [7].

Recent works have increased concentration on comparative analysis of distinct classifiers. Previous research highlights that the effectiveness of machine learning models for heart disease prediction varies across datasets. Singh and Chaturvedi indicated that there is no universal algorithm that is best suited in all cases [23]. Something similar was observed by Reddy and Rao who observed that the success of a model typically depends on the dataset itself, quality of features and preprocessing of the data [24]. Uddin et al. also demonstrated that the results of an ensemble of multiple models tend to be more reliable and accurate compared to those of a

single model [31]. Samuel emphasized that we must compare various algorithms in a systematic way in order to truly comprehend their strengths and weaknesses [32]. Hossain et al. experimented different machine learning classifiers to predict cardiovascular and suggested that accuracy alone should not be used [33]. Rather, they proposed the utilization of various evaluation measures. This is particularly significant in medical uses where datasets tend to be skewed. In such instances, precision, recall, and error distribution need to be looked into to appropriately assess a model. In the recent past, there has been an overwhelming urge to ensure that healthcare AI is more understandable. Johnson et al. emphasized the increasing importance of machine learning in cardiovascular medicine and reported that the decision support systems should be interpretable by clinicians in order to trust them [17]. Krittanawong et al. further stipulated that predictive accuracy is not sufficient in healthcare, we must also know how the input features are used to influence the behavior of the model [28]. Due to this, feature importance analysis, correlation exploration, and confusion matrix interpretation have become commonplace in predictive modeling. The broader healthcare data infrastructure has been the subject of other researchers. Rajkomar et al. described the way machine learning can transform medicine by enhancing the use of digital patient records and can be easily incorporated into clinical workflows [4]. The future of predictive modeling was also covered by Obermeyer and Emanuel, who noted that it is essential to find objective and unbiased patterns in healthcare data [3].

These studies declare that predictive systems are becoming more important as healthcare institutions continue to produce clinical datasets widely. Neglect the progression in cardiac disease forecasting research, various limitations remain in prior work. First, many studies focus mainly on overall accuracy while providing constricted discussion of precision-recall behavior, ROC curves, or confusion matrices . Second, some studies test only one or two classifiers, which reduces the effectiveness of comparative conclusions. Third, the interpretability of models is not always adequately addressed, even though clinical use requires transparency and trust. Fourth, several studies do not compare machine learning techniques with neural network based methods within the same framework.These limitations create a clear motivation for the present study. The current work compares four different predictive approaches Random Forest, XGBoost, SVM, and ANN on the same structured heart disease dataset using a unified preprocessing and evaluation strategy. In addition to overall accuracy, the study includes confusion matrix analysis, ROC curve analysis, precision-recall analysis, ANN learning curves, and Random Forest feature importance. This framework improves the reliability and practical value of the outcomes.

Overall, the literature help the use of ML for cardiac disease forecasting but comparative evaluation is essential. On organized cardiovascular data, aggregate models (e.g., Random Forest(RF) and XGBoost) give rivalrous performance, while SVM and ANN serve as useful additive modeling techniques. The present study builds on these previous works by combining predictive comparison with interpretability and visual performance analysis, thereby supporting to the ongoing growth of intelligent cardiovascular result support systems.

**Dataset Description**

The dataset contains 1,025 patient records and 14 attributes, in which 13 are predictive attributes and one is a binary outcome attribute. The target helps to classify the result that whether a patient has heart disease or not. The input features comprise of resting blood pressure, exercise-induced angina,number of major vessels, age, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, sex, ST depression (oldpeak), the slope of the peak exercise ST segment, chest pain type(cp), and thal (thalassemia).

These attributes are usually used in cardiovascular diagnosis and interpret exercise related clinical, physiological, and demographic indicators . Since the dataset contains medically

relevant structured attributes and a binary outcome variable, it is suitable for supervised ML classification.

**Class Distribution**

The target classes are relatively balanced. It reduces bias toward a single. It supports trusty model training and evaluation.

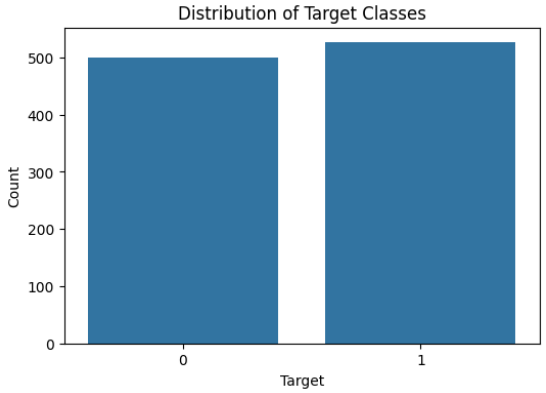


**Figure 1. Distribution of target classes in given dataset.**

### 3.2 Correlation Analysis

A correlation heatmap was constructed to take in pairwise relationships between attributes. It also helps in assessing that how each attribute correlates with the target class. The research shows that chest pain type(cp), maximum heart rate achieved, and slope have positive relationship with the target, while exercise-induced angina, oldpeak, ca, and thal show negative relationship.

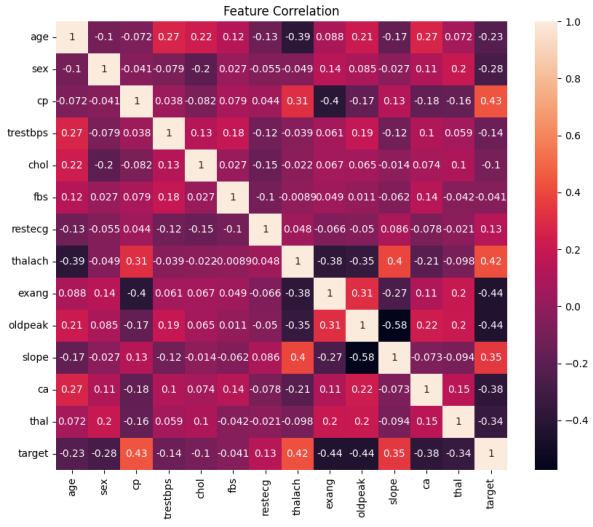


**Figure 2. Feature correlation heatmap of given dataset.**

### 4.Methodology

The proposed methodology of this study is based on a supervised learning framework for binary classification of heart disease. The overall process includes data preprocessing, normalization, train-test splitting, model training, and performance evaluation.


**Flowchart.**

The purpose of this framework is to understand a predictive mapping from clinical attributes to the target class indicating whether the cardiac disease is present or not. An organized methodology is essencial to check that the experimental results are mathematically consistent, reproducible, and reliable.

### 4.1 Problem Formulation

Let the dataset contains n number of patient records, where each record contains d clinical attributes and one binary class label. The dataset can be represented as:

$$D=(x_i,y_i)_{i=1}^{n}$$

where $x_i$ shows the dimensional feature vector of the i-th patient and $y_i$ shows the related target label. Since the problem is binary classification, the output label is defined as:

$$y_i \in 0,1$$

where $y_i = 1$ represents the presence of cardiac disease and $y_i = 0$ indicates there is no cardiac disease.

The complete feature matrix and target vector are written as:

$$X \in R^{n \times d}$$
$$y \in 0,1^{n}$$

The objective of the prediction system is to figure a mapping function f such that:

$$\hat{y}=f(x)$$

where $\hat{y}$ represents the predicted class label. This mathematical formulation provides the foundation for all machine learning models implemented in this study.

### 4.2 Data Preprocessing and Standardization

In medical data, the features may be on various numerical scales. For example, age. Very different ranges may be present in cholesterol, and blood pressure(BP). In case such differences are

Unnoticed, learning algorithms can assign more importance to attributes with larger magnitudes instead. Than to attributes that have more predictive relevancy. Consequently, feature scaling is a noteworthy.

Preprocessing step especially with models such as SVM and ANN.

To normalize the numerical attributes, standardization was used. The standard value z of a attribute x :

$$z=\frac{x-\mu}{\sigma}$$

Where μ represents the attribute mean and σ represents attribute standard deviation. This transformation ensures that the feature distribution is centered around zero with unit variance. It reduces scale laterality as well as improves convergence behavior in optimization based learning algorithms.

The changed dataset may be represented in matrix form as:

$$Z=\frac{X-\mu}{\sigma}$$

where Z represents the standardized feature matrix.

### 4.3 Training and Testing Strategy

We split dataset into training and testing subsets to check the predictive achievement of the models. The training part was utilized to learn model variables, whereas the testing part was set to assess model generalization on unseen data.

This split:

$$D=D_{train}\cup D_{test}$$

with:

$$|D_{train}|=0.8|D|$$
$$|D_{test}|=0.2|D|$$

The 80:20 split is a general experimental pattern in ML. it supplies a sufficient amount of training data while preserving an independent subset for evaluation. This separation helps reduce biased estimation of performance and provides a more realistic measure of predictive quality.

### 4.4 Random Forest Model

Random Forest is an aggregated learning model that create multiple decision trees and mix up their outputs through majority voting. The core idea behind Random Forest is that aggregating several weak or moderately strong learners can produce a stronger and more stable classifier. Random Forest(RF) is suitable for structured healthcare datasets in which nonlinear attribute interactions are usual.

If $T_1(x)$, $T_2(x)$, ..., $T_k(x)$ denote the predictions produced by k decision trees, then the final forecasting of the Random Forest classifier is:

$$\hat{y}= \text{mode}T_1(x),T_2(x),...,T_k(x)$$

The impurity of a node is commonly measured by the Gini index:

$$G=1-\sum_{j=1}^{c} p_j^2$$

Where $p_j$ denotes the probability of class j and c is the number of classes. Since this study deals with two classes, c = 2. A split is considered useful if it reduces impurity. The waste material reduction for a split will be:

$$\Delta G=G_{parent}-(w_lG_l+w_rG_r)$$

Where $G_p$ is the waste material of the parent node, $G_l$ and $G_r$ are the waste material of the left and right child nodes, and $w_1$ and $w_r$ are the proportions of samples assigned to the left and right branches. This equation provides the mathematical basis for feature based splitting inside decision trees. And therefore supports the purpose of Random Forest in the proposed prediction framework.

### 4.5 Support Vector Machine

Support Vector Machine is a margin based classification algorithm that seeks the best separating hyperplane between two classes. It is especially suitable for binary classification problems and has a strong mathematical foundation. The classifier attempts to maximize the margin, defined as the distance between the separating hyperplane and the nearest training points (support vectors).

The decision boundary is represented by:

$$w^Tx+b=0$$

where w is the weight vector and b is the bias term. The expected class is observed by the sign of the decision function:

$$\hat{y}= \text{sign}(w^Tx+b)$$

The optimization objective of SVM is:

$$\min_{w,b}\frac{1}{2}|w|^2$$

Such that :

$$y_i(w^Tx_i+b)\geq 1, \quad i=1,2,...,n$$

This formulation ensures that the classifier finds a hyperplane with maximum margin. For non separable data, slack variables are present, leading to the soft margin derivation:

$$\min_{w,b,\xi} \frac{1}{2}|w|^2+C\sum_{i=1}^{n}\xi_i$$

Such that:

$$y_i(w^T x_i+b)\geq 1-\xi_i, \quad \xi_i\geq 0$$

In order to enable the model to trade off between the maximization of the margin and the penalty parameter C can be included.

Classification error. This analysis framework entitles the application of SVM as one of the comparative models in this work.

### 4.6 XGBoost Model

Weak learners are added consecutively to improve the forecastings of the recent stage in XGBoost algorithm. This algorithm combines strong predictive capability with regularization. This combination makes the model more efficient.

At iteration t, the updated forecasting for the i-th sample is:

$$\hat{y}_i^{(t)}=\hat{y}_i^{(t-1)}+f_t(x_i)$$

where $f_t$ is weak learner introduced at iteration t. The objective function combines forecasting loss and regularization:

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

The penalty term is expressed as:

$$\Omega(f_t)=\gamma T+\frac{1}{2}\lambda|w|^2$$

where $\gamma$ is tree complexity, T is the number of leaves, lambda is the regularization coefficient, and w is leaf weights. The optimization process uses first order and second order gradient information. This process improves efficiency and predictive precision. This rhetorical structure gives mathematical support for including XGBoost in the comparative analysis.

### 4.7 Artificial Neural Network

An ANN is a layered nonlinear model in which neurons are linked through weighted connections. For each neuron, the input signals are combined through a weighted sum (plus bias), mapped through an activation function, and propagated forward to produce the network output. ANN is suitable for heart disease prediction because it can model complex nonlinear relationships among clinical variables.

For a neuron, the linear input is given by:

$$z=\sum_{j=1}^{d} w_j x_j+b$$

where $w_j$ is the connection weight associated with feature $x_j$ and b is the bias term.The activation function is used to compute the output of neuron:

$$a=g(z)$$

For binary classification, the sigmoid activation function is commonly used in the output layer:

$$\sigma(z)=\frac{1}{1+e^{-z}}$$

Thus, the final predicted probability can be written as:

$$\hat{y}=\sigma\left(\sum_{j=1}^{d} w_j x_j+b\right)$$

The ANN is trained by reducing the binary cross entropy loss function:

$$L=-\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i)+(1-y_i)\log(1-\hat{y}_i)\right]$$

Parameter optimization is carried out iteratively using gradient descent. The update rule for a trainable parameter theta is:

$$\theta^{(t+1)}=\theta^{(t)}-\eta\frac{\partial L}{\partial\theta}$$

where eta denotes the learning rate. This mathematical basis explains how ANN learns from the clinical data and updates its internal parameters during training.

4.8 Feature Importance Justification

Random Forest was used to examine the Feature Importance to identify the attributes that contributed most strongly to prediction. If attribute j contributes to impurity reduction across multiple trees, its importance score may be represented as:

$$I_j=\sum_{t=1}^{k}\Delta G_{j,t}$$

where $\Delta G_{j,t}$ denotes the reduction in impurity contributed by atttributej in tree t. attributes with larger importance values are advised more relevant to the classification process. This gives

an analytical explanation for why certain clinical attributes are more potent in the final predictive model.

## 5. Results and Analysis

This section gives us experimental findings of the comparative analysis conducted on the given dataset. The multiple classification measures and visualization techniques were used to evaluate the forecasting performance of purposed models. Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Network are used for prediction. The model accuracy, confusion matrix interpretation, ROC analysis, precision recall analysis, feature importance, ANN learning behavior, and learning curve assessmen are achieved in the form of result.

### 5.1 Accuracy Comparison of the Models

The accuracy of the evaluated models is Wrapped up in given Table. This shows that Random Forest and XGBoost got the highest classification accuracy than ANN and SVM.

**Accuracy comparison of the implemented models**

| Model | Accuracy |
|---|---|
| Random Forest | 98.53% |
| XGBoost | 98.53% |
| SVM | 88.78% |
| ANN | 94.14% |

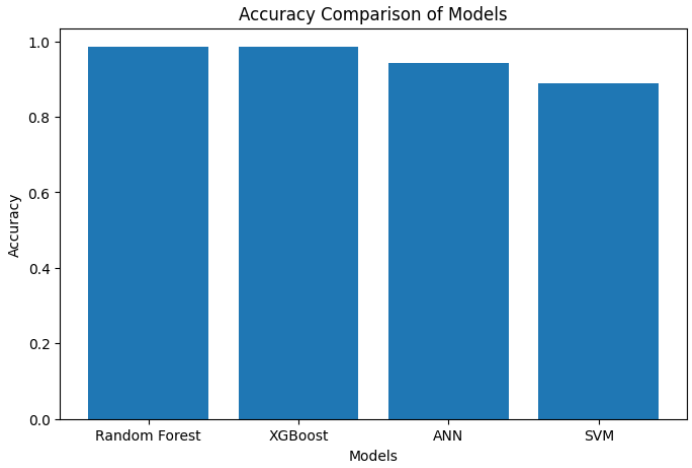


**Figure 3. Comparison of Accuracies**

This figure clearly shows that the aggregate learning models performed well than other evaluated methods on the selected dataset. Random Forest and XGBoost both got an accuracy of 98.53%. The result tells us that aggregate based learning is strongly effective for organized cardiovascular data. The ANN model also achieved high performance with an accuracy of 94.14%. SVM produced the lowest result among the tested models at 88.78%.

These results indicate that models are able to handle nonlinear relationships and feature interactions are more suitable for this dataset than margin based classification alone.

### 5.2 Confusion Matrix Analysis of Random Forest

We utilized the Confusion matrix of Random Forest(RF) to understand classification behavior of best model. The values we got are:

True Negatives (TN) = 102
False Positives (FP) = 0
False Negatives (FN) = 3
True Positives (TP) = 100

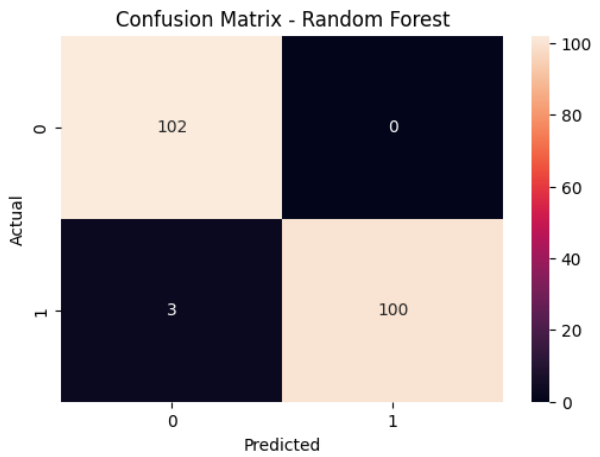


**Figure 4. Confusion matrix of the Random Forest(RF) classifier.**

The confusion matrix informs us that the model RF correctly identified 102 negative cases and 100 positive cases. RF made only 3 false negative predictions and 0 false positive prediction. This result reflects excellent classification performance. It means that the model is highly reliable in identifying between the two classes.

### 5.2.1 Mathematical Verification of Accuracy

Using the confusion matrix values, the classification accuracy of the Random Forest(RF) model is verified as follows:

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

$$\text{Accuracy}=\frac{100+102}{100+102+0+3}$$

$$\text{Accuracy}=\frac{202}{205}=0.9853$$

Therefore, the Random Forest(RF) model accuracy is 98.53% that is verified.

### 5.2.2 Mathematical Verification of Precision

The Random Forest(RF) model precision is computed as:

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

$$\text{Precision}=\frac{100}{100+0}=1.00$$

This shows perfect precision for the positive class on the test set.

### 5.2.3 Mathematical Verification of Recall

The recall is calculated as:

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

$$\text{Recall}=\frac{100}{100+3}=0.9709$$

The Random Forest(RF) model correctly identified approximately 97.09% of actual positive cases.

### 5.2.4 Mathematical Verification of F1-Score

The F1-score:

$$F1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

$$F1=2\times\frac{1.00\times0.9709}{1.00+0.9709}=0.9852$$

The F1-score confirms a highly counterbalanced classification performance between precision and recall.

### 5.3 ROC Curve Analysis

We used The Receiver Operating Characteristic (ROC) curve to test the discrimination quality of the Random Forest model(RF) across distinct classification thresholds.

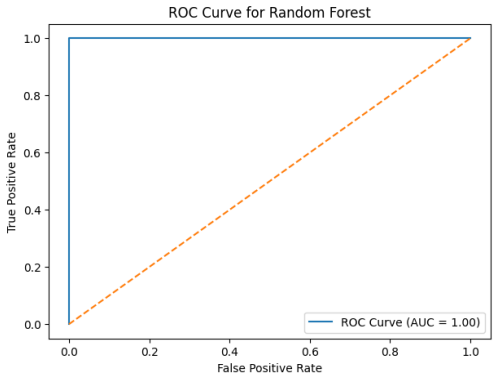


**Figure 5. ROC curve for the Random Forest model.**

The ROC curve is very close to the top left corner, and the area under the curve is approximately 1.00. This means that discrimination between the positive and negative classes is outstanding. A high ROC-AUC value evidents that the model is highly capable of distinguishing patients with heart disease from those without the disease.

From a testing viewpoint, this is an crucial result because a model with high ROC capability can sustain strong sensitivity and specificity across various limits.

### 5.4 Precision-Recall Curve Analysis

The ROC analysis, precision-recall behavior was used to asses the positive class forecasting ability of the Random Forest(RF) model. The model sustained strong precision and recall values. It shows dependable identification of cardiac disease cases with very few false positives. This result helps the effectiveness of the Random Forest(RF) classifier for medical forecasting tasks where accurate identification of positive cases is very crucial.

### 5.5 Feature Importance Analysis

We build feature importance from the Random Forest(RF) model to check the feature participated most in classification.

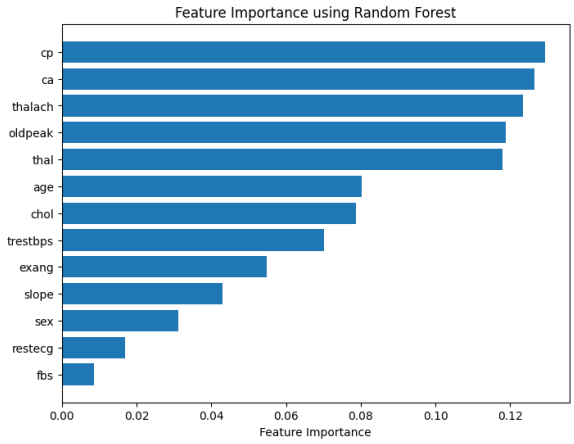


**Figure 7. Feature importance ranking using Random Forest**

The feature importance graph tells us that the most prestigious attributes are number of major vessels (ca), oldpeak, thal, maximum heart rate achieved (thalach), and chest pain type (cp). These features participate most to the final forecasting.

These are the most involved characteristics in the final forecasting.

The medical usefulness of this finding is in the sense that it agrees with the known. Cardiovascular screening indicators. Chest pain patterns, vessel status, stress response measurements, and thal related attributes are all deemed as important in real world cardiac evaluation. The consistency of model derived feature importance with clinical expectations is strengthened the credibility of the forecasting architecture

### 5.6 ANN Training and Validation Accuracy

This is the Artificial Neural Network learning pattern that was tested by training and validation accuracy curves in multiple epochs.

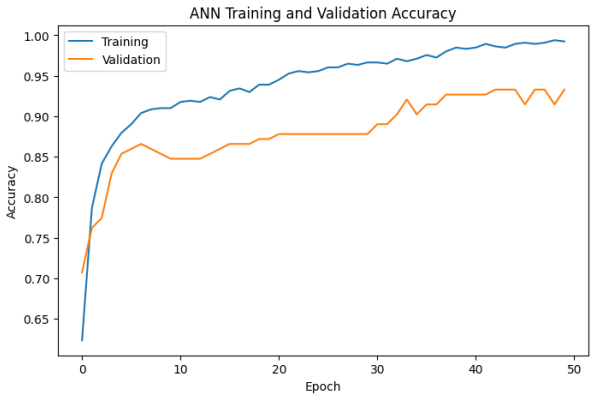


**Figure 8. ANN training and validation accuracy with respect to epochs**

The figure indicates that the training accuracy increased steadily with the epoch and it attained a value. Considerably high level. The accuracy of validation also improved over time but remained below the training curve. This demonstrated that the ANN was useful in predicting patterns in the Dataset. The time lag between training and validation performance indicates that the model can have small overfitting in comparison to the aggregate methods. The neural still achieved a good final classification rate of 94.14% that indicates that neural. Networks are capable of providing competitive performance even on structured tabular data.

### 5.7 ANN Training and Validation Loss

The training and validation loss curves of the ANN model indicate that the network was learning well throughout the training process. The loss in training declined steadily, whereas validation loss dropped at first and leveled off. This trend indicates that the neural network learned significant patterns in the data. Mild overfitting is indicated by a small difference between training and validation behavior.

### 5.8 Learning Curve of Random Forest

The random forest(RF) model learning pattern underscores the fact that predictive performance is directly proportional to the size of the training. High training accuracy and gradually increasing validation performance indicate that the model has been able to adapt to the provided dataset. This finding assists the strength of Random Forest in the classification of structured heart disease.

### 5.9 Comparative Interpretation of the Results

In general, the relative performance verifies that Random Forest(RF) and XGBoost are proclaimed the optimal models on the chosen heart disease dataset. Their outstanding performance suggests that aggregate learning methods are highly suitable for orgnized clinical data, particularly when the dataset contains both demographic and physiological attributes with nonlinear interactions.

The ANN model also performed well. It demonstrated strong learning behavior, but it did not exceed the accuracy of the aggregate methods.

This may be attributed to the moderate dataset size and the fact that tree based ensembles often deal tabular healthcare data more expeditiously than neural networks. SVM, while still producing a reasonable result. It got lower performance than the other models. It suggested that its decision boundary was less suitable for this particular dataset.

From the perspective of interpretability and clinical usefulness, Random Forest offers an additional advantage because it not only achieved high predictive accuracy but also provided an interpretable ranking of the most prestigious clinical attributes. This makes it a strong candidate for intelligent objective decision support applications.

## 6.Discussion

The result of this study demonstrate that ensemble learning techniques are highly effective for cardiac disease prediction using structured clinical data. Both Random Forest(RF) and XGBoost got the highest accuracy, indicating that ensemble based tree learning is particularly well suited to this dataset. This may be due to the ability of these methods to handle nonlinear relationships and interactions among clinical variables without requiring extensive manual feature engineering.

Random Forest showed particularly strong overall behavior. In addition to high accuracy, it produced an excellent confusion matrix with very few misclassifications. The absence of FP and the low number of FN indicate strong practical usefulness.The False Negatives(FN) must be reduced because missed disease cases may delay treatment.

XGBoost is similar to the accuracy of Random Forest(RF), which shows the reliability of boosting techniques for organized medical datasets. XGBoost is famous in giving its high efficiency and predictive power. These results are reporting strong performance of boosting techniques in healthcare domains.

The accuracy of 98.14% is achieved by the ANN model in given dataset. The training and validation curves tells us that the system learned substantive relationships from the data. The performance of the aggregate methods was better than this model. This is because neural networks need larger training data and more big learning to perform tree based methods on systematic datasets.

The lowest performance among the models was achieved by SVM. Although SVM is a strong classifier in many binary prediction tasks. Its effectiveness depends on data structure , kernel configuration, and feature transformation . But here, it did not catch the predictive perfrmance of the aggregate methods.

The feature importance ranking produced by Random Forest(RF) is also noteworthy. Attributes such as maximum heart rate achieved, chest pain type(cp), number of major vessels, oldpeak, and thal were the most prestigious predictors. These rankings are medically meaningful and align with ordinary knowledge in diagnosis of heart diseases. It increases the performance of the model in clinical support settings.

The very high ROC and precision-recall outcomes imply that further validation would be utile. Further testing with cross validation and external datasets would help to verify whether the observed performance generalizes to broader clinical populations. The current results clearly support the importance of ML, especially aggregate methods, in heart disease forecasting.

## 7.Conclusion

This paper provided a comparative analysis of Random Forest(RF), ANN, XGBoost, and SVM for heart disease forecasting using a organized clinical dataset. Dataset was consisted of 1025 patient records. After preprocessing and exploratory analysis, all models were evaluated on a single framwork.

The results showed that both XGBoost and Random Forest got the highest accuracy of 98.53%, SVM got 88.78% and ANN got 94.14%. RF also demonstrated highly propitious confusion matrix behavior. It also provided useful interpretability through feature importance study.

The study concludes that aggregate learning methods are highly suitable for organized

cardiovascular datasets and can serve as reliable components of clinical decision support systems. These models have strong potential to support healthcare professionals in making informed decisions, to support early diagnosis, and improve risk screening.

**8.Future Work**

In Future, this work can be better by applying k fold cross validation, developing real time healthcare decision support applications, using explainable AI tools such as SHAP or LIME, exploring hybrid deep learning and ensemble models,using some external validation data, and testing on larger multi center datasets .

**REFERENCES**

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.

C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

R. C. Deo, "Machine learning in medicine," Circulation, vol. 132, no. 20, pp. 1920–1930, 2015.

A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, pp. 249–268, 2007.

H. K. Bashir, M. Mansoor, and A. R. Javed, "Machine learning-based heart disease prediction system," Diagnostics, vol. 11, no. 5, 2021.

T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

A. U. Haq et al., "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," Mobile Information Systems, vol. 2018, Article ID 3860146, 2018.

U. R. Acharya et al., "Application of machine learning techniques for heart disease diagnosis," Biomedical Signal Processing and Control, vol. 10, pp. 1–11, 2014.

I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," Information Fusion, vol. 76, pp. 243–297, 2021.

M. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications, 2008, pp. 108–115.

S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.

N. Alizadehsani et al., "A data mining approach for diagnosis of coronary artery disease," Computer Methods and Programs in Biomedicine, vol. 111, no. 1, pp. 52–61, 2013.

K. Polat, S. Şahan, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system and fuzzy weighted pre-processing," Pattern Recognition, vol. 40, no. 11, pp. 3044–3054, 2007.

Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," New England Journal of Medicine, vol. 375, no. 13, pp. 1216–1219, 2016.

J. Brownlee, Machine Learning Mastery With Python. Melbourne, Australia: Machine Learning Mastery, 2016.

D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2019.

World Health Organization, "Cardiovascular diseases (CVDs) fact sheet," WHO, Geneva, Switzerland.

T. M. Mitchell, Machine Learning. New York, NY: McGraw-Hill, 1997.

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.

D. Singh and V. Chaturvedi, "A comparative study of heart disease prediction using machine learning techniques," International Journal of Advanced Science and Technology, vol. 29, no. 5, pp. 2123–2130, 2020.

N. K. S. K. Reddy and Y. S. Rao, "Predicting heart disease using machine learning algorithms," International Journal of Engineering and Advanced Technology, vol. 8, no. 3, pp. 1–7, 2019.

F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

G. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. Cambridge, U.K.: Cambridge University Press, 2014.

A. Johnson et al., "Machine learning and decision support in cardiovascular medicine," Frontiers in Cardiovascular Medicine, vol. 5, pp. 1–14, 2018.

K. Krittanawong et al., "Machine learning prediction in cardiovascular diseases: A meta-analysis," Scientific Reports, vol. 10, pp. 1–11, 2020.

H. M. Dawood and A. H. M. Ragab, "Heart disease prediction using machine learning techniques: A review," Multimedia Tools and Applications, vol. 82, pp. 1–29, 2023.

P. Motarjemi and S. H. Asl, "Heart disease diagnosis using machine learning methods: A systematic review," Healthcare Analytics, vol. 2, 2022.

M. S. Uddin, M. A. Rahman, A. M. N. Islam, and M. A. Talukder, "An effective study of heart disease prediction using machine learning techniques," International Journal of Advanced Computer Science and Applications, vol. 10, no. 10, pp. 1–8, 2019.

S. Samuel, "Heart disease prediction using machine learning algorithms," International Journal of Research in Engineering, Science and Management, vol. 3, no. 3, pp. 1–5, 2020.

M. N. Hossain, M. R. Islam, and S. Chakraborty, "Performance analysis of machine learning classifiers for heart disease prediction," Journal of Healthcare Engineering, vol. 2021, Article ID 1–12, 2021.