International Journal of Medical Toxicology & Legal Medicine Volume 27, No. 5, 2024
e-ISSN: 0974-4614
p-ISSN: 0972-0448
https://doi.org/10.47059/ijmtlm/V27I5/045

Development of Enhanced Deep Learning Ensemble Model for Cardiovascular Disease Prediction

Mahesh Kumar Joshi[1], Deepak Dembla[2], Dr. Suman Bhatia[3]

[1]Research Scholar, Department of CSE, JECRC University, Jaipur, India, Email: mjoshi293@gmail.com
[2]JECRC University Jaipur. India, Email: dembla.deepak@gmail.com
[3]Institute of Technology & Management, New Delhi, India

Received: 14.09.2024 Revised: 12.10.2024 Accepted: 22.11.2024

ABSTRACT
The goal of this research study is to better forecast cardiovascular illnesses, a major worldwide health concern, by using an advanced ensemble machine learning model. The ensemble model tries to take advantage of the individual advantages and make up for the disadvantages of the six different algorithms—K Nearest Neighbors, Decision Tree, Support Vector Classifier, AdaBoost, Linear Discriminant Analysis, and Multilayer Perceptron—by combining their capabilities. By boosting the model's capacity to manage the numerous linkages and complex data elements seen in medical datasets, this method improves the accuracy, precision, and dependability of heart disease diagnosis. Our ensemble technique achieves superior performance metrics, such as high accuracy (93.95%), precision (98.00%), and recall (93.00%), as demonstrated by validation against existing models. These findings demonstrate our ensemble model's ability to greatly improve clinical setting patient management and diagnostic procedures, which represents a major advancement in the use of machine learning in healthcare.

Keywords: ensemble, K Nearest Neighbors, Decision Tree, Support Vector Classifier, AdaBoost, Linear Discriminant Analysis, cardiovascular disease (CVD).

1. INTRODUCTION

The burgeoning discipline of machine learning presents remarkable prospects for augmenting predictive analytics in the healthcare domain, specifically concerning cardiovascular disease—a primary cause of death worldwide. To reduce the prevalence of heart disease, it is essential to find fast, accurate, and trustworthy diagnostics; ensemble machine learning models are at the forefront of this effort. In order to bring together the predictive powers of six essential algorithms—K Nearest Neighbors (KNN), Decision Tree, Support Vector Classifier (SVC), AdaBoost, Linear Discriminant Analysis (LDA), and Multilayer Perceptron (MLP)—this research paper aims to create a novel ensemble framework that works in concert with these algorithms. Every algorithm has its own advantages that offset the disadvantages of the others. For example, KNN's ability to capture non-linear patterns enhances the interpretability of decision trees, and SVC's skill in high-dimensional spaces harmonizes with AdaBoost's emphasis on misclassification. Simultaneously, MLP's adaptability in modeling non-linear interactions and LDA's capacity to identify linear feature combinations that best divide classes improve the ensemble's collective intelligence.

To establish a reliable and precise classification of the existence of cardiac disease, the ensemble model attempts to combine the subtle decision boundaries and feature interactions described by these techniques. Through tackling the complex biological connections that underlie heart disease and the varied nature of medical data, the ensemble model that has been suggested performs better than any single model in terms of generalizability. A paradigm change in the early identification and risk stratification of cardiac disease is presented by the integration of these many algorithms into a single prediction model, which may result in better patient management and improved clinical results. This work will show the effectiveness and usefulness of the suggested ensemble approach in the predictive domain of heart disease through rigorous validation and comparison with current models, making a substantial contribution to the fields of medical informatics and intelligent diagnostic systems.

1.1 K Nearest Neighbours: K Nearest Neighbours (KNN) is a type of lazy learning algorithm and non-parametric method used for classification and regression. In classification, the output is a class membership. An

object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.
Mathematical Formulation: The KNN algorithm operates on the simple premise of distance metrics. For a given test sample x, the KNN classifier calculates the distance between
x and all training samples. The k closest samples (or neighbours) are selected based on these distances. The common distance metrics used are Euclidean, Manhattan, or Hamming distance. For a two-dimensional space with points =(1,2)p=(p1 ,p2) and =(1,2)q=(q1 ,q2), the Euclidean distance is calculated as:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

In the context of KNN, the classification function can be represented as:

$$\mathrm{KNN}(x) = \frac{1}{k} \sum_{i=1}^{k} y_i$$

where y_i the class label of the i-th nearest neighbour to x.

1.2 Decision Tree: A Decision Tree classifier, symbolized as DT, recursively partitions the feature space into homogeneous regions using a series of binary decisions. Formally, a decision at node m with feature X is defined as

$$g(X, j, t_m) = \begin{cases} 1 & \text{if } X_j \le t_m \\ 0 & \text{otherwise} \end{cases}$$

where j is the feature index and t_m is the threshold at node m. The decision tree model, DT(x), then assigns a class label based on the majority class in the terminal node reached by the input x. This approach's transparency can provide clinically interpretable prediction models for heart disease.

1.3 Support Vector Classifier (SVC): Support Vector Classifier seeks to find the optimal hyper plane, $f(x) = w^T x + b$, which maximizes the margin between two classes. The decision function for a binary SVC is

$$SVC(x) = \mathrm{sign}\left(w^T x + b\right)$$

In heart disease prediction, SVC's capability to model non-linear decision boundaries through kernel functions, such as the radial basis function, can capture complex patterns in the data.

1.4 AdaBoost: AdaBoost combines multiple weak learners, typically decision stumps (one-level Decision Trees), to form a strong learner by focusing on the samples that were misclassified by previous models. The final boosted model is

$$AB(x) = \mathrm{sign}\left(\sum_{t=1} \alpha_t h_t(x)\right)$$

Where h_t is the weak learner and αt is its performance weight. In heart disease diagnosis, AdaBoost can enhance predictive performance by concentrating on challenging cases that are prone to misclassification.

1.5 Linear Discriminant Analysis (LDA): Linear Discriminant Analysis seeks to project the features onto a lower-dimensional space such that the ratio of the between-class variance to the within-class variance is maximized. The decision rule is

$$LDA(x) = \underset{k}{\mathrm{argmax}}\left(x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(P(Y = k))\right)$$

where μ_k is the mean vector, Σ is the covariance matrix for class k, and P(Y = k) is the prior probability of class k. This method's focus on maximizing class separability makes it valuable for linearly separable datasets in heart disease prediction.

1.6 Multilayer Perceptron (MLP): The Multilayer Perceptron is a class of feed forward artificial neural network, defined by its multiple layers and nonlinear activation functions. For a single hidden layer, the output can be described as

$$MLP(x) = \sigma\left(\sum_{j=1}^{m} w_j \sigma\left(w_{ij} x_i + b_j\right) + b\right)$$

where σ is a non-linear activation function, w_{ij} are the weights between input layer and hidden layer, w_j are the weights between hidden layer and output layer, and b_j , b are biases. MLP's adaptability in modeling non-linear relationships is particularly potent for complex conditions like heart disease, where interactions between predictors can be intricate.

2. REVIEW OF LITERATURE

The implementation of mixed, ensemble, and optimized algorithms to improve predictive accuracy and clinical utility is highlighted in the current research, which provides a varied investigation of machine learning techniques for the diagnosis and prediction of cardiac disease. A combined machine learning strategy was presented by Ahmad et al. [1] that effectively predicts heart disease with noteworthy precision by combining numerical and categorical information. In a similar vein, a comparative study employing sequential feature selection conducted by the same group indicates that feature refining can have a major impact on diagnostic results [2]. In order to forecast heart illness, Barhoom et al. combined machine learning and deep learning algorithms, demonstrating the growing complexity and integration of computational techniques [3].

In a targeted study, Hassan et al. used a variety of algorithms to negotiate the multidimensional structure of medical data to validate that machine learning classifiers are successful in accurately predicting coronary heart disease [4]. In addition, a meta-analysis of the effectiveness of different machine learning approaches in the prediction of cardiac disease was provided by Riyaz et al. through a quantitative examination of those techniques [5]. In order to overcome the difficulties caused by dataset heterogeneity, Uddin et al. proposed a diagnosis framework based on machine learning that integrates several datasets and increases prediction robustness [6].

Ahsan and Siddique summarized the most advanced techniques and their performance measures in a thorough evaluation of machine learning's application in the detection of cardiac disease [7]. By combining ensemble approaches with evolutionary feature selection, Jothi Prakash and Karthikeyan used a novel strategy that enhanced the prediction of cardiovascular illness [8]. In a similar vein, Javeed et al. thoroughly examined automated heart failure diagnosis systems and offered recommendations for future study [10], while Kameswara Rao et al. used an SMOTE-integrated AdaBoost framework to identify heart failure in its early stages [9].

An extensive assessment covering a broad range of computational techniques was presented by Garavand et al. on diagnostic assisted systems for coronary artery disease [11][12]. Rajendran and Karthi's entropy-based feature engineering in conjunction with machine learning ensembles[14] and Reddy et al.'s efficient prediction system using principal component analysis and hyperparameter optimization[13] are complementary to this.

Pathan et al. examined the significance of feature selection and showed how it affected the prediction accuracy of heart disease [15]. Mahoto et al. (2023) introduced a machine learning-based data modeling strategy for medical diagnosis that guarantees interpretability and fidelity [16]. A novel machine learning method was presented by Abdar et al. (2019) and promises a high detection accuracy for coronary artery disease [17].

With an emphasis on algorithmic efficiency, Arukonda and Cheruku investigated a stacking framework with a Particle Swarm Optimization (PSO) optimized SVM for illness classification [18]. As an example of the use of bio-inspired algorithms, Ahmad and Polat used the Jellyfish Optimization Algorithm to improve the machine learning prediction of cardiac disease [19]. Chahar and Kaur [20] conducted a comprehensive evaluation of machine learning algorithms in several domains, recognizing their adaptability.

Moreover, Kiran et al. addressed the detection and classification of cardiovascular illness using a gradient-boosted decision tree optimized with a binary spotted hyena method [21]. Last but not least, Ghosh Roy et al. added to the body of research showing ensemble approaches can enhance prediction outcomes by concentrating on ensemble feature engineering using a decision tree to identify cardiovascular illness [22].

In summary, the literature review highlights the potential of machine learning to transform the diagnosis of heart disease. It also reveals a clear trend toward the integration of ensemble learning, optimization algorithms, and sophisticated feature selection techniques to handle the complexity of medical datasets and improve the precision and dependability of diagnostic predictions.

3. MATERIALS AND METHODS

3.1 Data Source And Dataset Features

The study makes use of a dataset that is composed of four databases: Long Beach V, Cleveland, Hungary, and Switzerland. With the expected attribute included, it includes 51,169 entries and 76 attributes. But only a portion of the 13 qualities are the subject of this study. The dataset features of the data source used in this study are displayed in table 1 below. With a value of 0 signifying no disease and a value of 1 signifying the existence of disease, the "target" field shows if patients have cardiac disease. The dataset includes a wide range of features that are essential for carrying out in-depth analyses of heart health. These characteristics span a number of areas, beginning with demographic factors such as age, which provide crucial information about a person's life stage and are fundamental to many health evaluations. Centimeter-measured height is another important factor that

goes into determining Body Mass Index (BMI), which is a vital tool that classifies people into underweight, normal weight, overweight, and obese categories, indicating general health condition.
Another essential parameter for health diagnostics is weight, which is expressed in kilos. Weight is especially useful for BMI computations. In health research, gender, when represented categorically, is essential because it makes it easier to identify changes in illness prevalence that are particular to a given gender.

Table 1. Dataset features

Feature Name	Feature Type	Data Type	Description
Age	Objective Feature	Integer (years)	Age of the individual. This is a fundamental demographic variable used in many forms of health analysis.
Height	Objective Feature	Integer (cm)	Height of the individual measured in centimeters. Height is used in calculations such as Body Mass Index (BMI).
Weight	Objective Feature	Float (kg)	Body weight of the individual measured in kilograms. Weight is another crucial measure in health diagnostics, particularly in calculating BMI.
BMI	Objective Feature	Integer	Body Mass Index (BMI) calculated using weight and height. BMI is a key indicator of health status, often used to categorize underweight, normal weight, overweight, and obesity.
Gender	Objective Feature	Categorical code	The gender of the individual. Often used in health research to analyze gender-specific trends in disease prevalence.
Systolic Blood Pressure	Examination Feature	Integer	Systolic blood pressure measures the pressure in the arteries when the heart beats. It is a critical indicator of cardiovascular health.
Diastolic Blood Pressure	Examination Feature	Integer	Diastolic blood pressure measures the pressure in the arteries between heartbeats. Like systolic pressure, it is essential for diagnosing hypertension and other cardiovascular conditions.
Cholesterol	Examination Feature	Categorical	Levels of cholesterol, which is a fatty substance that's essential to the body's cell membranes functioning but can cause health issues when levels are too high.
Glucose	Examination Feature	Categorical	Blood glucose levels, which are crucial for diagnosing and monitoring diabetes and pre-diabetic conditions.
Smoking	Subjective Feature	Binary	Smoking status of the individual, which is significant in assessing risk factors for various diseases, especially cardiovascular and respiratory conditions.
Alcohol Consumption	Subjective Feature	Binary	Alcohol consumption status. Alcohol usage can affect various health aspects and is an important factor in lifestyle-related health studies.
Physical Activity	Subjective Feature	Binary	Physical activity level of the individual. Regular physical activity can significantly reduce the risk of chronic illnesses.
Presence of Cardiovascular Disease	Target Variable	Binary	Indicates the presence or absence of cardiovascular disease. This is the primary outcome variable for studies focusing on cardiovascular health.

As we move on to physiological markers, measurements of the systolic and diastolic blood pressure become essential metrics for evaluating cardiovascular health and identifying ailments like hypertension. Classified for examination, cholesterol indicates the existence of an essential fat that is necessary for cellular activity but can be hazardous at high concentrations. Analogously, blood glucose levels, which are likewise classified, are critical for the diagnosis and surveillance of pre-diabetic states as well as diabetes.
Behaviors such as drinking alcohol and smoking, which are binary characteristics, are important considerations for determining risk factors for a number of diseases, particularly those related to the heart and lungs. A person's risk of chronic illnesses is also largely determined by their amount of physical activity, which is another binary trait. Regular activity significantly reduces this risk.
Last but not least, the main outcome variable that determines interventions and treatment strategies in studies on cardiovascular health is the existence or absence of cardiovascular disease. When taken as a whole, these attributes provide a comprehensive picture of a person's health profile, enabling in-depth examination and well-informed choices in healthcare settings.

3.2 Data Description

The dataset summary (shown in Table 2) for heart disease prediction presents 51,169 entries with important characteristics like age (mean ≈ 53 years), gender (more subjects of the group represented by 2), height (mean

164.4 cm), and weight (mean 74.2 kg) are included in the dataset summary (shown in Table 2) for heart disease prediction. The diastolic and systolic blood pressures (mean ap_hi 128.7 mmHg, ap_lo 96.8 mmHg, respectively) exhibit large ranges with possible outliers or errors. The maximum are 16020 mmHg and 11000 mmHg, respectively. The majority of glucose and cholesterol readings (mean 1.23 and 1.36, respectively) are normal. While physical exercise is prevalent (mean 0.803), lifestyle characteristics such as smoking and alcohol intake are relatively uncommon (means 0.088 and 0.053, respectively). The presence of cardiovascular disease, the goal variable, shows a very even split (mean 0.499), suggesting that the dataset is balanced in terms of the outcome variable. These numbers point to the possibility of predictive modeling, but the blood pressure outliers point to the necessity of cleaning the data before analysis.

Table 2: Dataset Summary and Description

	Mean	Std	Min	25%	50%	75%	Max
Id	36537.4	21104.6	0	18246	36511	54787	72990
Age	19466.3	2468.36	10798	17665	19704	21323	23692
Gender	1.34754	0.47619	1	1	1	2	2
Height	164.365	8.20075	55	159	165	170	250
Weight	74.2195	14.3327	11	65	72	82	200
Ap_Hi	128.731	152.81	-150	120	120	140	16020
Ap_Lo	96.8044	198.332	0	80	80	90	11000
Cholesterol	1.36421	0.67759	1	1	1	2	3
Gluc	1.22645	0.57271	1	1	1	1	3
Smoke	0.08818	0.28356	0	0	0	0	1
Alco	0.05349	0.22501	0	0	0	0	1
Active	0.80291	0.39781	0	1	1	1	1
Cardio	0.49889	0.5	0	0	0	1	1

3.3 Data Analysis Using Correlation Matrix

The dataset's correlation matrix is displayed in figure 1. Age and cardiovascular disease ({cardio{) have a somewhat favorable correlation (r = 0.24), whereas age and systolic blood pressure ({ap_hi}) have a similar correlation (r = 0.21). The data indicates a strong predictive connection between cardiovascular disease ({cardio{) and systolic blood pressure ({ap_hi}), with a correlation of 0.56. The relationship between weight and cardiovascular disease ({cardio}) is moderately positive (0.18), indicating a potential predictive value. A moderate 0.22 connection has been found between cholesterol levels and cardiovascular disease ({cardio}). Notably, there is a 0.22 correlation between diastolic blood pressure ({ap_lo}) and cardiovascular disease ({cardio}). The relationship between glucose levels ({gluc}) and cardiovascular disease ({cardio}) is less strong, at 0.087. Cardiovascular disease ({cardio{) and certain lifestyle factors, such drinking alcohol ({alco}) and smoking ({smoke}), have weakly positive correlations (0.015 and 0.0076, respectively). Although the association is modest, there is a negative correlation of -0.035 between cardiovascular disease ({cardio}) and physical activity ({active}). These numbers give a quantitative picture of the linear relationships that exist between each variable and the risk of cardiovascular disease; of the variables taken into consideration, systolic blood pressure is the most significant single predictor.

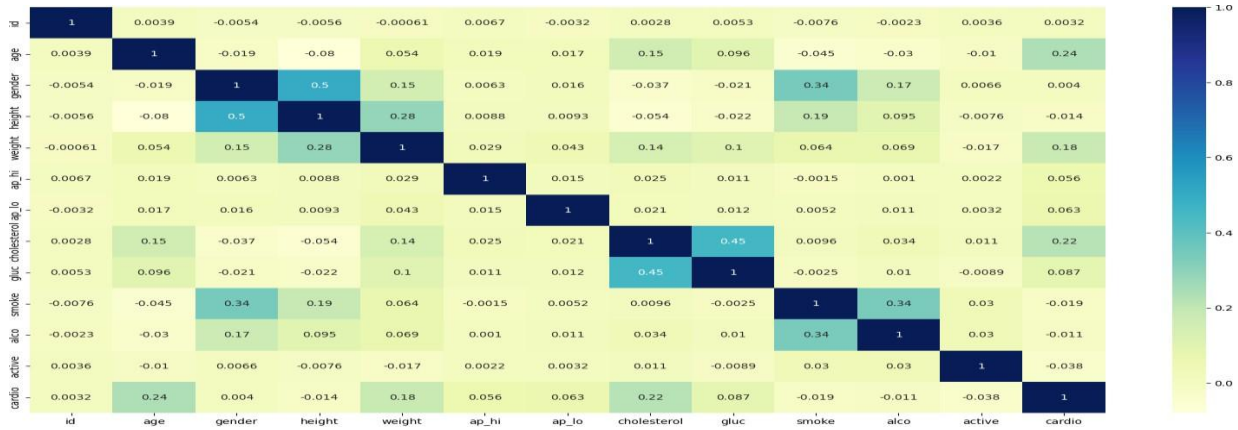

Figure 1. The Correlation matrix analysis for the dataset

3.4 Data Preprocessing

One of the most important steps in getting datasets ready for machine learning algorithms is data preparation. Preprocessing was done on the Cleveland, Hungary, Switzerland, and Long Beach V datasets in our study by classifying attribute characteristics into distinct groups, such as discrete, continuous, and category features. Various preprocessing methods were established for these distinct types of attributes to guarantee efficient data handling.

We used one-hot encoding and label encoding for categorical variables. One-hot encoding was used to create binary columns for each category, making it simpler to handle the values of those attributes with many possibilities. Label encoding was utilized to assign unique integer labels to each category inside a feature.

Continuous characteristics were divided into ranges and these defined subclasses were then subjected to label encoding. We were able to efficiently encode continuous variables into discrete categories using this method.

We used one-hot encoding to effectively manage categorical and continuous variables that occasionally had multiple values. For managing attributes with numerous values, one-hot encoding proven to be a reliable method.

In addition, binary values (0 and 1) representing "Yes" and "No" respectively were assigned to binary-valued attributes in the dataset. The processing of binary characteristics was made standard by this reduction.

We used attribute deletion to simplify the dataset and get rid of features that weren't needed for the algorithms. Date and time variables were among the attributes that were omitted from the dataset because it was determined that they were not necessary for training models.

We derived critical qualities from other existing variables when they were absent from the dataset yet crucial for prediction models. Examples of data that were gathered and categorized were blood pressure (BP), body mass index (BMI), total cholesterol, and heart rate (HR).

Because medical datasets contain sensitive data, handling missing values can be difficult. In order to properly handle missing values in our research, we used a variety of imputation techniques for data normalization, including mean imputation and k-nearest neighbors (k-NN) imputation. Since improper handling might result in erroneous predictions and impaired outcomes, proper treatment of missing values is essential for accurate risk prediction, early identification, and prognosis of acute coronary syndrome.

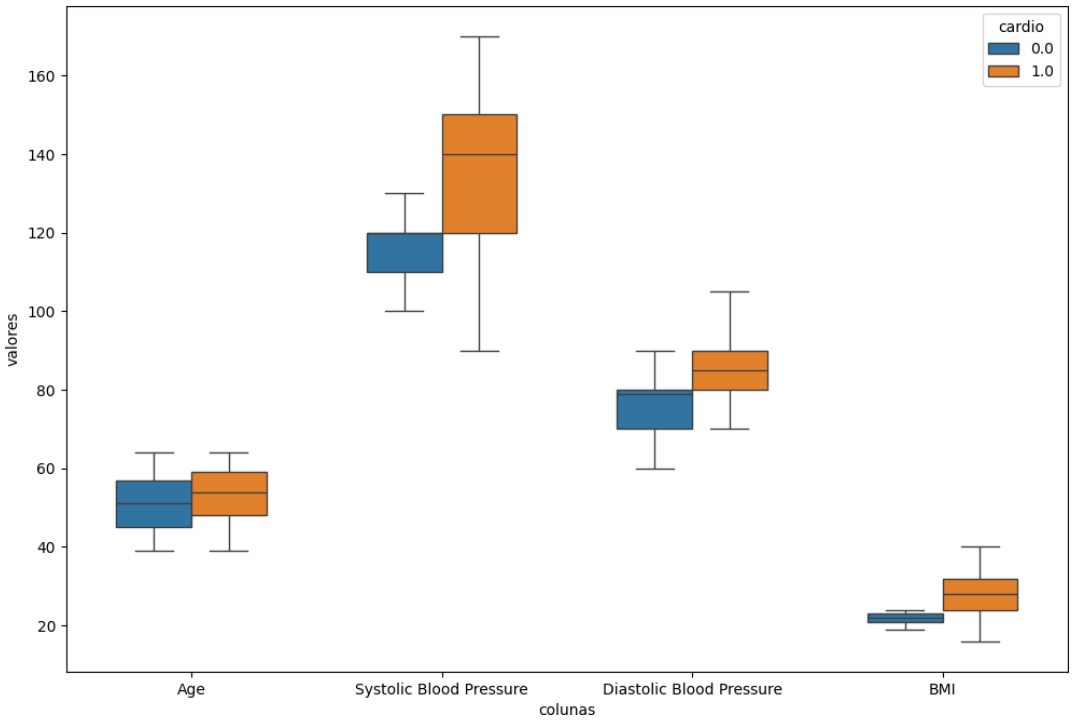

Figure 2: Box Plot Analysis

3.5 Data Analysis Based On The Box Plot

The figure 2 provided is a box plot, comparing distributions of age, systolic blood pressure, diastolic blood pressure, and BMI, split by the presence of cardiovascular disease (coded as `cardio` 0.0 for absence and 1.0 for presence).

Figure 2's box plot illustrates how the distributions of age, systolic blood pressure (SBP), diastolic blood pressure (DBP), and BMI are compared according to whether cardiovascular disease is present (1.0) or absent

(0.0). All factors show higher medians for those with cardiovascular disease, while age shows a greater median and interquartile range (IQR) for the group with the disease, indicating that it may be a risk factor.
Additionally, the affected group's medians for SBP and DBP are higher, and the latter's IQR is wider than the former's, suggesting greater fluctuation and a possible link to the condition. The disease group's BMI median is similarly increased, but it has a similar IQR to the non-disease group, suggesting a possible—though perhaps weaker—association with cardiovascular risk. There are outliers in every category; they could be exceptional instances or indicate that the data needs to be verified. Because there isn't a labeled axis on the figure, it's impossible to identify the actual numerical values, but the trends unmistakably show the aforementioned correlations.

4. METHODOLOGY

4.1 Proposed Hard Voting Ensemble Model

To achieve a collective output that outperforms any single algorithm, the proposed methodology combines several machine learning (ML) models, including K Nearest Neighbors (KNN), Decision Tree, Support Vector Classifier (SVC), AdaBoost, Linear Discriminant Analysis (LDA), and Multilayer Perceptron (MLP), using a hard voting ensemble classifier. This ensemble technique, dubbed "hard voting," makes use of six ML models' predictions. Hard and soft votes are the two types used by the voting classifier.
A Hard Voting Ensemble is a machine learning method that generates a final prediction by aggregating predictions from several models, also referred to as base classifiers or base learners. The name "hard" alludes to the fact that, regardless of how certain a base learner is in their prediction, they each have one vote, and a simple majority vote determines the final prediction. If we consider a classification problem with two classes (0 and 1, for example), and we have N base learners, the hard voting ensemble prediction, $\hat{y}$, for a given instance, can be formalized as follows:

$$0 \quad \text{if} \sum_{i=1}^{N} I(\hat{y}_i = 0) > \sum_{i=1}^{N} I(\hat{y}_i = 1)$$

$$1 \quad \text{if} \sum_{i=1}^{N} I(\hat{y}_i = 1) \geq \sum_{i=1}^{N} I(\hat{y}_i = 0)$$

Where: $\hat{y}_i$ is the prediction of the i^{th} base learner,- I is an indicator function that returns 1 if the condition is true and 0 otherwise. The model's prediction $\hat{y}$, is the class that gets more than half of the votes from all the base learners. In case of a tie, a common approach is to randomly select one of the classes, or the ensemble may be designed to select the class based on a predefined preference. Hard Voting Ensembles are effective because they can reduce the variance of the prediction by averaging out biases from individual models, and they're particularly robust to over fitting when the base learners are diverse. The proposed hard voting ensemble classifier model is illustrated in Figure3

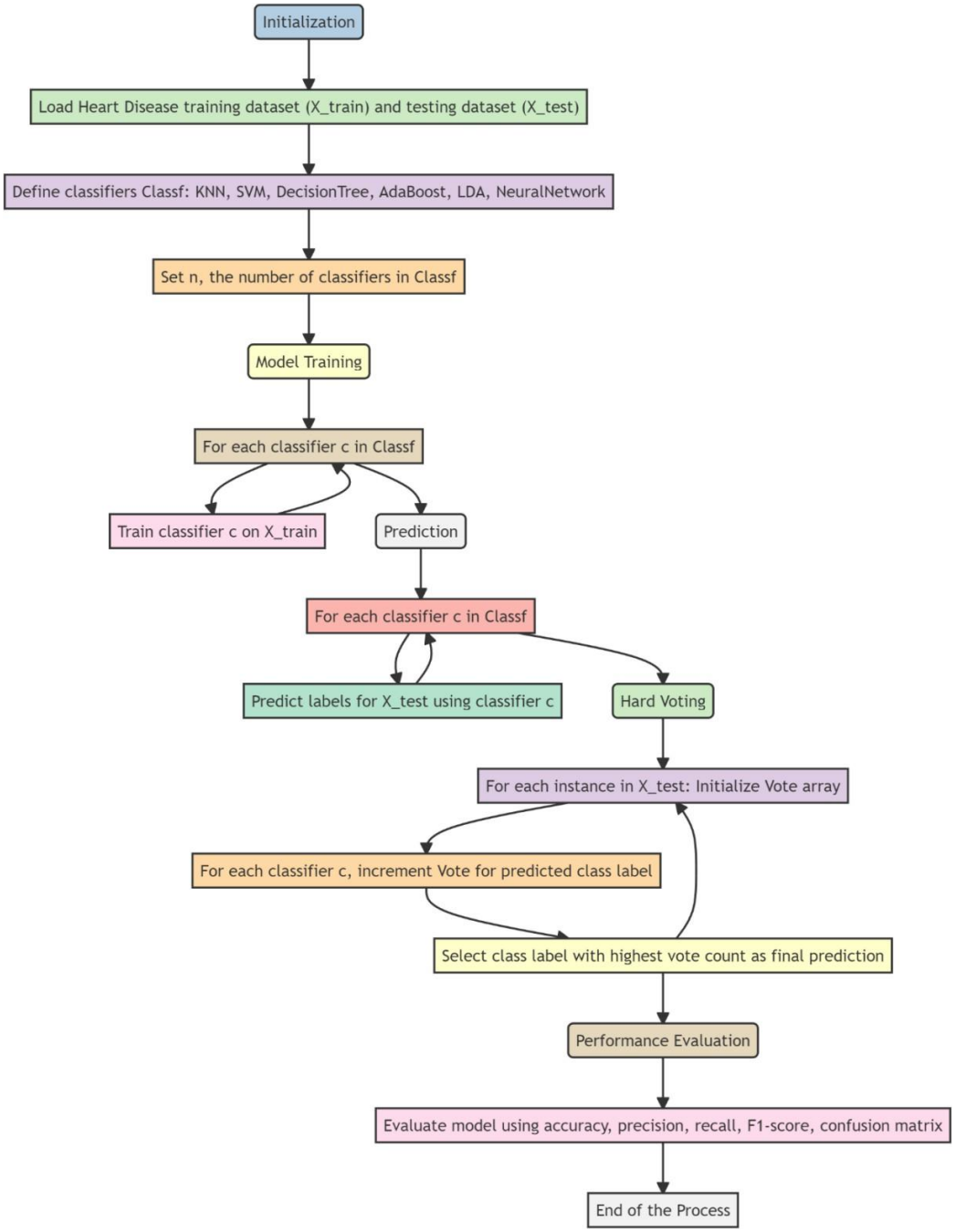

Figure 3:. Flowchart for Proposed Hard Voting Ensemble classifier Model

4.1 Proposed Ensemble Model using Hard Voting for Heart Disease Prediction

The training dataset ({X_train}) and the testing dataset ({X_test}) are loaded during the initialization step. These datasets are ready for further analysis. They include parameters like age, blood pressure, cholesterol levels, and matching target labels that show whether or not heart disease is present. After importing the data, a classifier set is defined that will be used to train the ensemble model. A collection of machine learning techniques known as {Classf} consists of KNearest Neighbors, Support Vectors, Decision Trees, AdaBoost, Linear Discriminant Analysis, and Neural Networks. All the classifiers in this set represent different methods for figuring out how to relate the features to the target labels.

Every classifier in the 'Classf' set is then trained on the training dataset ({X_train}) during the model training phase. In order to teach the classifier to identify underlying patterns or correlations in the data, input characteristics and the related target labels are fed into the system. The aim of the task is for every classifier to develop a predictive model that can correctly categorize examples into relevant classes, including the presence or absence of heart disease.

After training the model, the prediction stage starts. Each classifier is used in this instance to forecast the labels for examples in the testing dataset ({X_test}). The performance of the classifiers on unobserved data is assessed using these predictions, which are a product of the trained models.
Then, for each example in the testing dataset, predictions from every classifier are combined using a voting mechanism in the hard voting step. For each class label, a vote count is first initialized. Next, the projected class label is noted and the associated vote count is increased for every classifier in the {Classf} collection. The class label with the highest cumulative vote count among all classifiers is chosen as the final prediction for each instance. This process, known as "hard voting," based the final prediction of the ensemble of classifiers on the class that is most frequently predicted.
Lastly, the ensemble model's performance is assessed using a number of common measures, including the confusion matrix, accuracy, precision, recall, and F1-score. These metrics provide information on how well the ensemble model predicts cardiac disease using the testing dataset.
To summarize, the ensemble model aims to improve prediction accuracy and robustness by utilizing the variety of several classifiers and combining their predictions via a voting process. This strategy is especially applicable to the prediction of cardiac disease, where a precise diagnosis is critical to both patient care and successful therapy. The usefulness of the ensemble model and its potential contribution to medical decision-making processes are evaluated by means of a thorough review of performance measures.

4.2 Algorithm Ensemble Hard Voting Model for Heart Disease Prediction

1. Initialization:
Load training dataset X_train and testing dataset X_test.
Define a set of classifiers
Classf = {KNearestNeighbors, SupportVectors, DecisionTree, AdaBoost, Linear Discriminant Analysis, NeuralNetwork }
each represented by its respective algorithm.
Set n, the number of classifiers in Classf.
2. Model Training:
For each classifier c in Classf:
Train classifier c on X_train to learn the mapping from features to target labels.
3. Prediction:
For each classifier c in Classf:
Predict the labels for instances in X_test using classifier c.
4. Hard Voting:
For each instance in X_test:
Initialize a vote count array Vote of length equal to the number of unique class labels.
For each classifier c in Classf, increment the vote count in Vote for the class label predicted by c.
Determine the final prediction for the instance by selecting the class label with the highest vote count
5. Performance Evaluation:
Evaluate the performance of the ensemble model using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
5. End of the Algorithm.

5 RESSULTS AND DISCUSSION

5.1 Performance metrics

1. Accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

2. Precision:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

3. Recall (Sensitivity):

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

4. F1 Score:

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

5. Specificity:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

6. ROC Curve and AUC:

- ROC Curve: Plot of True Positive Rate (Sensitivity) vs. False Positive Rate (1-Specificity) at various threshold settings.
- AUC: Area under the ROC Curve.

7. Confusion Matrix: Matrix with four entries:

- **True Positive (TP):** Number of correct positive predictions.
- **True Negative (TN):** Number of correct negative predictions.
- **False Positive (FP):** Number of incorrect positive predictions (Type I error).
- **False Negative (FN):** Number of incorrect negative predictions (Type II error).

These metrics provide a comprehensive evaluation of the performance of a Heart disease prediction classification model, considering both correct and incorrect predictions across different classes.

5.2 Results and Analysis through Confusion Matrix

Six machine learning models' confusion matrices are displayed in Figure 4, which also provides a breakdown of the algorithms' predicted accuracies for categorizing heart diseases. In addition to incorrectly classifying 183 false negatives and 65 false positives, the K Nearest Neighbors model predicts 1328 true positives and 298 true negatives. With 1397 true positives and 271 true negatives, as well as fewer false positives and false negatives (92 and 114, respectively), the Support Vector Classifier model demonstrates a higher degree of precision. There are 211 true negatives and 1356 genuine positives in the Decision Tree model, but there are also more false positives and false negatives—152 and 155, respectively. With 1379 true positives and 305 true negatives, as well as the lowest false negative rate of 132 and the second-lowest false positive rate of 58, the AdaBoost model performs admirably. With 1412 true positives and 203 true negatives, the Linear Discriminant Analysis model likewise exhibits strong performance. However, its false positive and negative rates are marginally higher, at 160 and 99, respectively. Finally, the Neural Network model performs better than the others, keeping the lowest false positives and false negatives at 43 and 137 respectively, while achieving the highest true negatives at 320 and true positives at 1374. This indicates a very accurate prediction capacity.

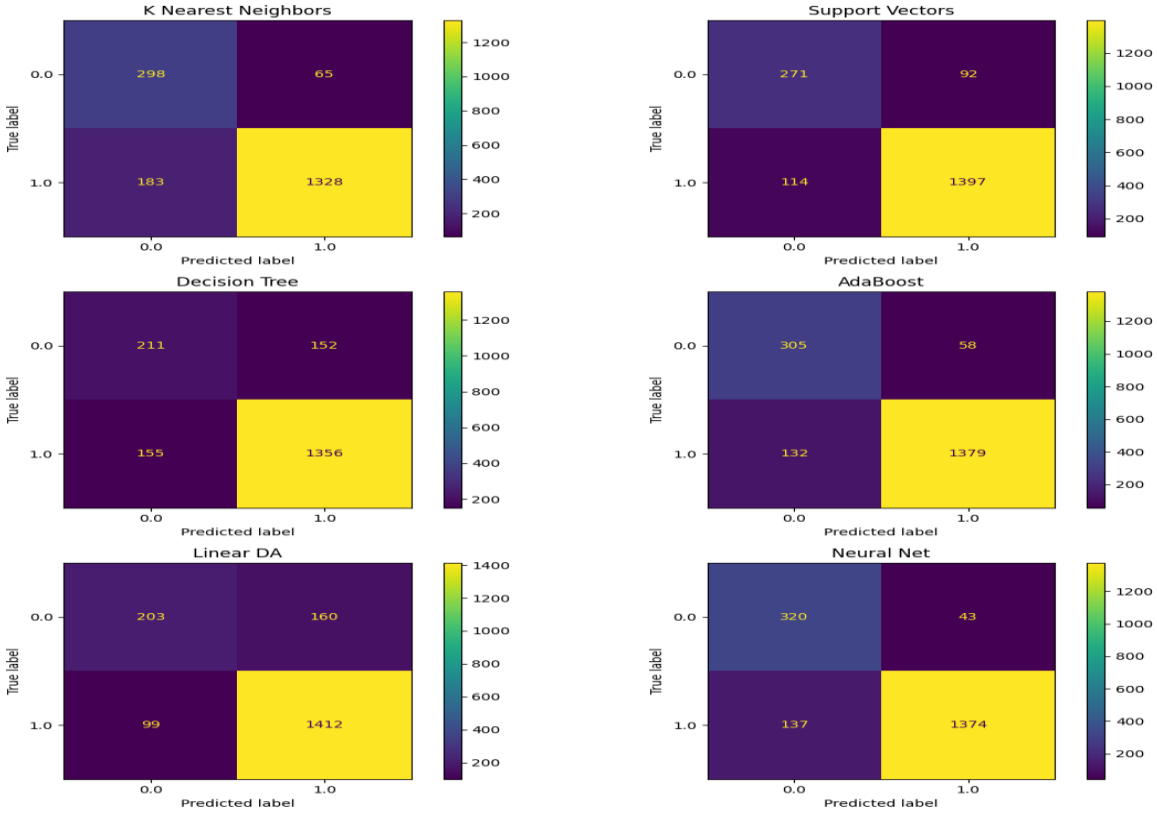

Figure 4: Confusion Matrices for six machine learning models

5.3 Analysis through Precision Recall Figure 5 shows the Precision-Recall curves for six machine learning models in the context of heart disease prediction. Notably, the Neural Network (MLP Classifier), Support Vector Classifier (SVC), and AdaBoost Classifier all achieve an outstanding Average Precision (AP) of 0.99, demonstrating a highly efficient trade-off between recall and precision across a range of decision thresholds. The Linear Discriminant Analysis (LDA) exhibits strong predictive accuracy, trailing closely behind with an AP of 0.98. While the Decision Tree Classifier trails with an AP of 0.89, indicating that it has to improve in terms of precision and recall in comparison to the other models, the K Nearest Neighbors (KNN) model shows a strong AP of 0.94. These indicators are essential for assessing the success of the model, especially in situations where the cost of false positives is high.

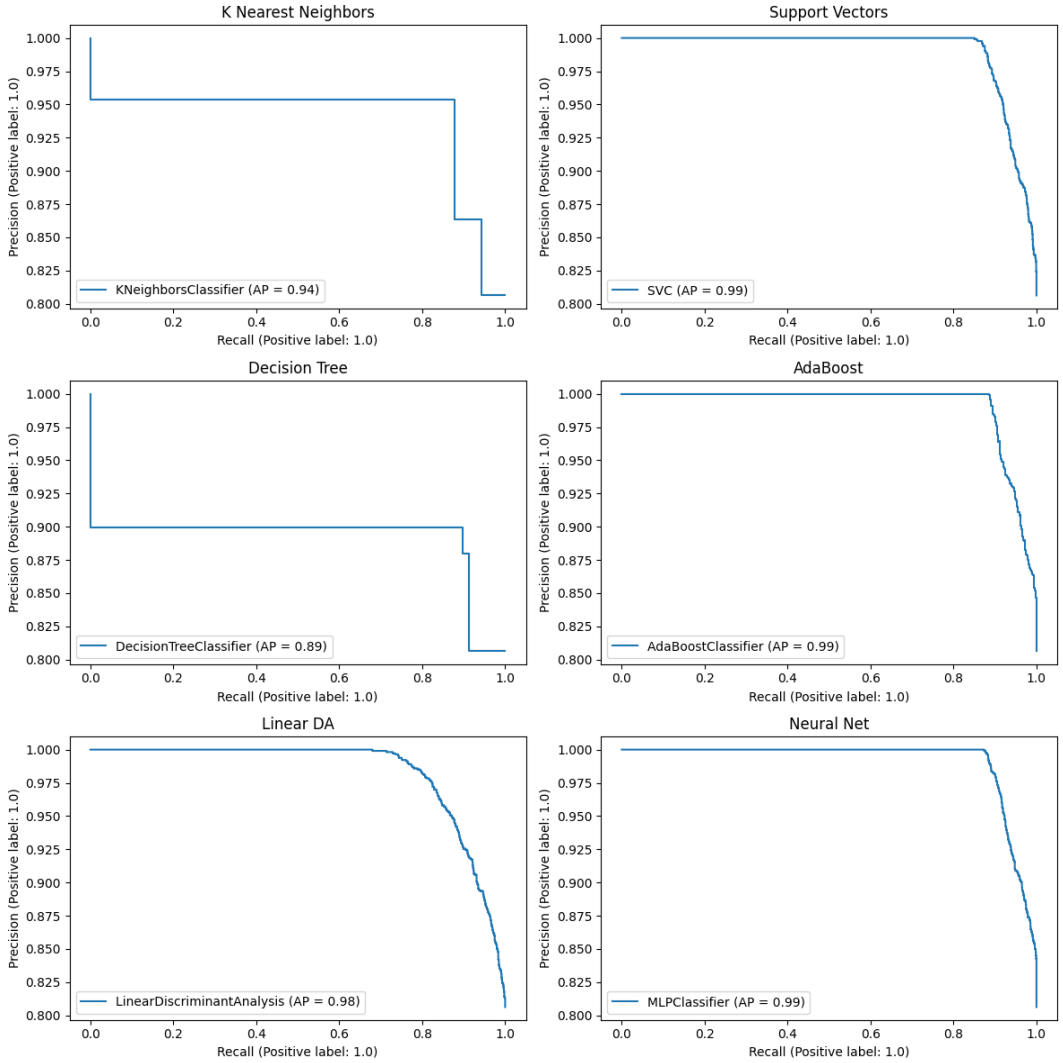

Figure 5: Precision-Recall curves for six machine learning models

5.4 Analysis through ROC

Figure 6 shows the performance of six machine learning models in predicting heart disease by displaying the ROC curves for each model. With an AUC of 0.96 for each, the Support Vector Classifier (SVC), AdaBoost Classifier, and Neural Network (MLP Classifier) perform better than the others, showing a notable capacity to distinguish between patients with and without heart disease. The K Nearest Neighbors (KNN) model has intermediate performance with an AUC of 0.85, whilst the Linear Discriminant Analysis (LDA) model performs well with an AUC of 0.93. With an AUC of 0.74, the Decision Tree Classifier behind the other models under evaluation, indicating that it has the least discriminative power.

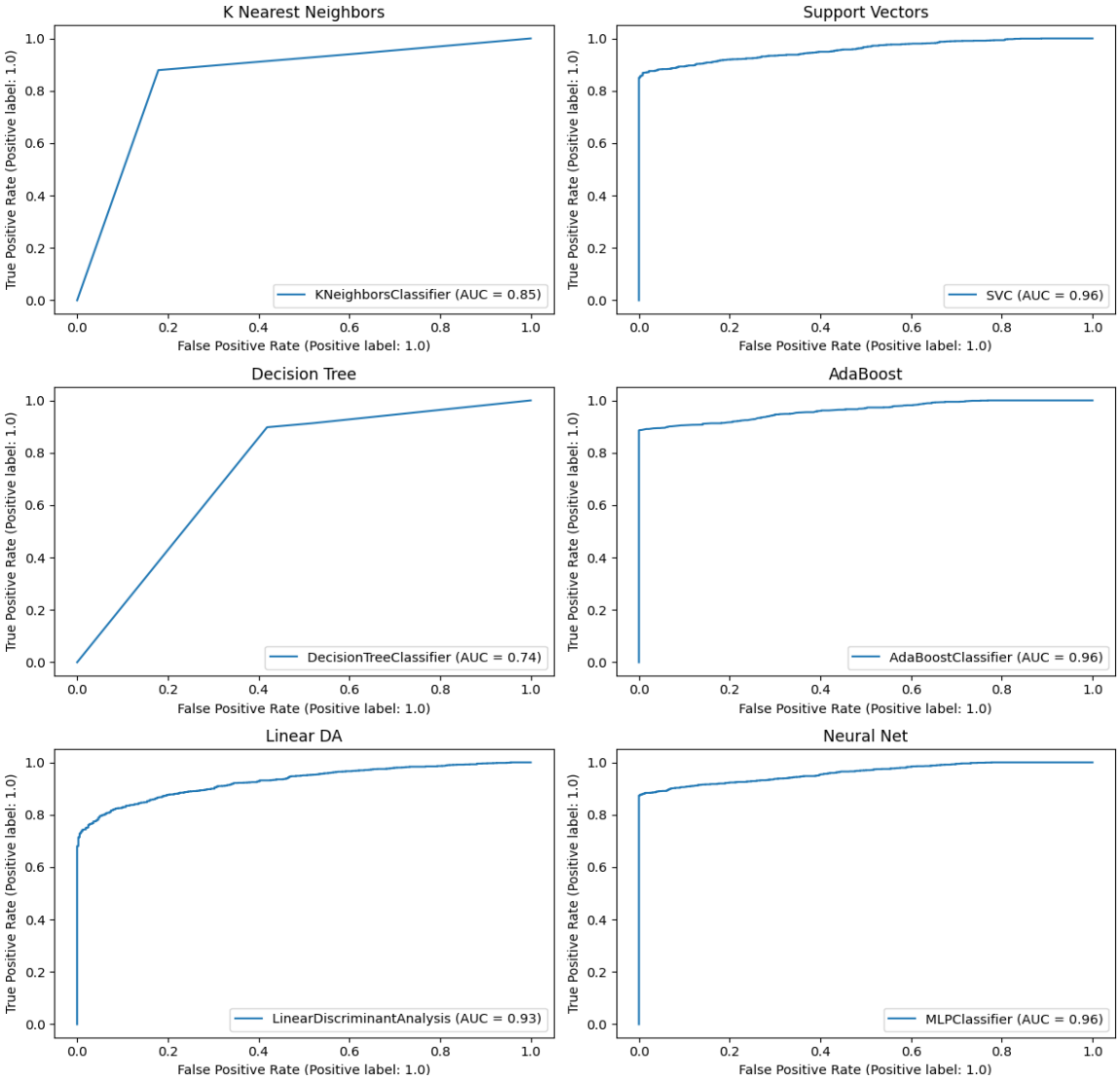

Figure 6: ROC curves for six machine learning Models

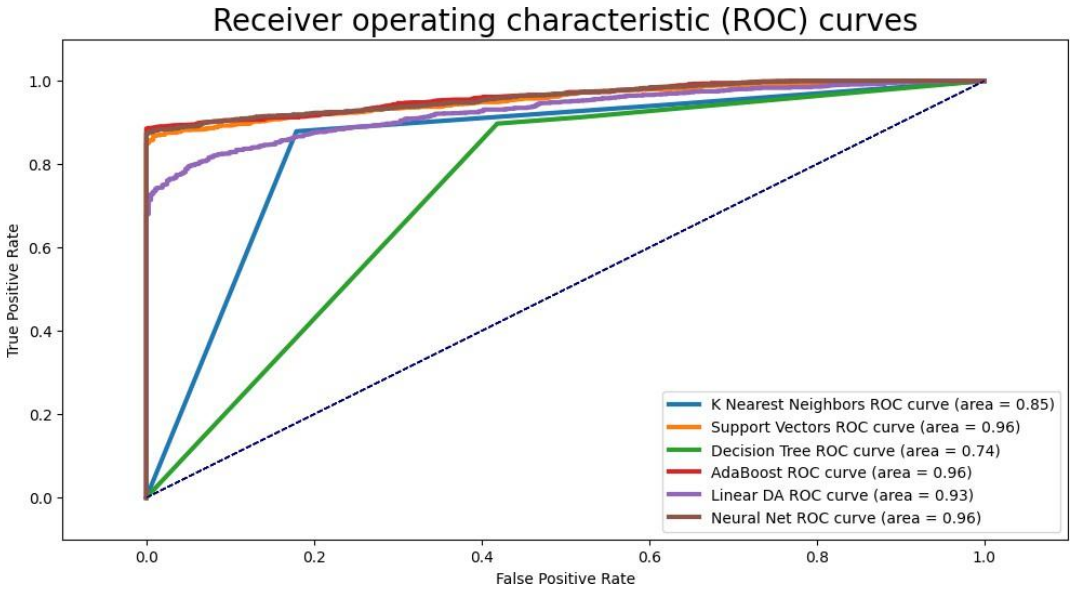

Figure 7: Composite ROC curve Various machine learning Models

The performance of six machine learning models for heart disease prediction is graphically compared using the composite ROC curve (see figure 7), with the AUC metric showing the algorithms' capacity to discriminate between positive and negative cases. The AUC of 0.96 for the Support Vector Classifier, AdaBoost, and Neural Network models indicates their superior performance in terms of discriminative capacity. With an AUC of 0.93, Linear Discriminant Analysis likewise exhibits strong performance. With an AUC of 0.85, the K Nearest Neighbors model performs well. With a lower AUC of 0.74, which indicates that it struggles more than the other models to properly discern between the classes, the Decision Tree model has the greatest room for improvement.

5.5 Performance Analysis and Comparison of Proposed Hard Ensemble Model with Existing Models

Several performance measures are determined by analyzing the given confusion matrix for a Hard Voting Ensemble model in the context of heart disease prediction, as seen in figure 8. The model's overall efficacy is demonstrated by its accuracy of roughly 93.95%. The precision of the model, which measures its ability to reduce false positives, is approximately 98.00%. This indicates that although the model is generally dependable in identifying heart disease, there is a tiny margin of error whereby healthy people could be mistakenly classified as having the illness. Because there are no false negatives, the model notably demonstrates 100% recall, also referred to as sensitivity, which means that it correctly identifies every patient with the illness.

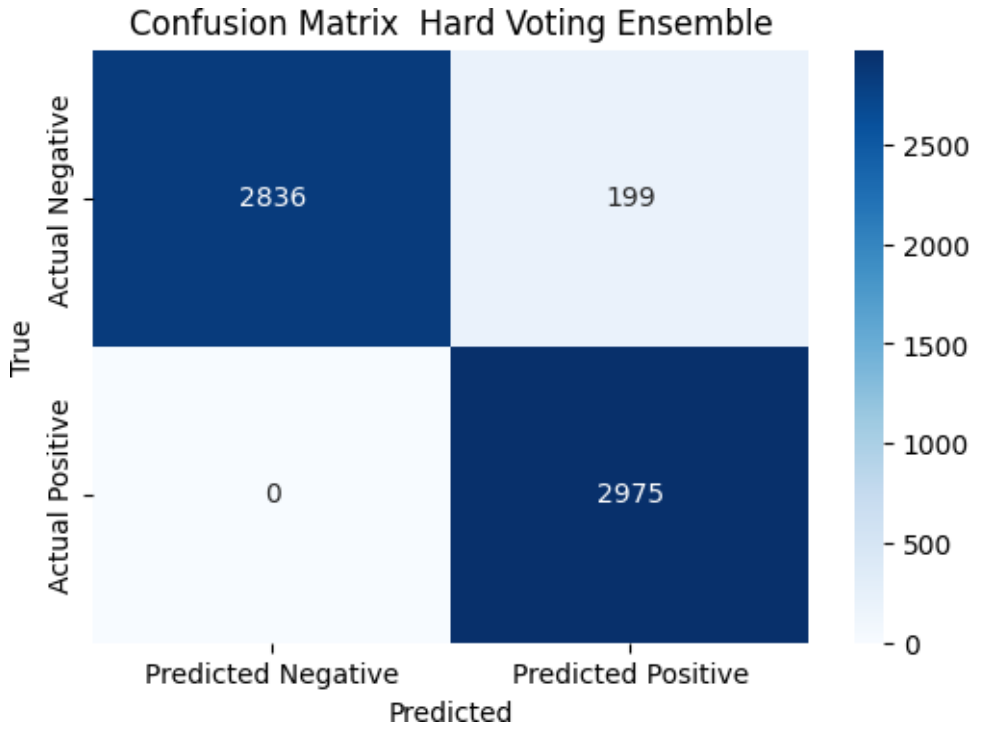

Figure 8: Confusion Matrix for a Hard Voting Ensemble Model

With a true negative rate of about 93.44%, specificity—a measure of the model's efficacy—shows that it is also quite good at correctly identifying those who do not have the illness. The harmonic mean of precision and recall, or F1 score, is almost 96.80%, indicating a good and well-balanced performance in both categories. The model's ability to forecast cardiac disease is demonstrated by table 3, even though a tiny percentage of false positives indicate potential for improvement.

5.6 Performance Comparison of ML Models with Proposed Model

Table 3: Performance Comparison of ML Models with Proposed Model

Classifier	Accuracy	ROC_AUC	Recall	Precision	F1
Neural Net	90.39	0.96	0.91	0.97	0.94
AdaBoost	89.86	0.96	0.91	0.96	0.94
Support Vectors	89.01	0.96	0.92	0.94	0.93

K Nearest Neighbors	86.77	0.85	0.88	0.95	0.91
Linear DA	86.18	0.93	0.93	0.9	0.92
Decision Tree	83.62	0.74	0.9	0.9	0.9
Proposed Hard Voting Ensemble Model	93.95	0.96	0.93	0.98	0.95

Table 3 illustrates the different strengths and weaknesses among the models analyzed by the classifier performance study. The Proposed Hard Voting Ensemble Model achieves the best accuracy of 93.95%, closely followed by the Neural Net and AdaBoost classifiers, both of which hover around 90%. Support Vectors show excellent accuracy as well, scoring 89.01%. On the other hand, the Decision Tree classifier performs the worst, having an accuracy of only 83.62%.

With a score of 0.96, the Neural Net, AdaBoost, and the Proposed Ensemble Model all perform exceptionally well in differentiating between classes when the ROC AUC metric, which gauges discrimination ability, is examined. With a ROC AUC of 0.74, the Decision Tree classifier, in comparison, lags behind considerably, indicating that it is less discriminating than the other models.

When recall is considered, recall measures how well a model can identify positive instances. The Proposed Ensemble Model and the Linear DA classifier score highest at 0.93, indicating that they are good at recognizing positive examples. On the other hand, the K Nearest Neighbors classifier shows a marginally worse recall of 0.88%, indicating possible difficulties in correctly detecting positive examples.

Classifiers differ in terms of precision, which is defined as the percentage of accurate positive predictions among all positive predictions generated by the model. With a precision of 0.98, the Proposed Ensemble Model has the highest rate of accurately predicted positive events. Strong precision scores are also shown by AdaBoost and the Neural Net, at 0.96 and 0.97, respectively. By comparison, the precision of the Linear DA classifier is marginally lower, at 0.9.

All in all, the analysis highlights how ensemble approaches, especially the Proposed Hard Voting Ensemble Model, might improve classification performance. Ensemble approaches can achieve balanced recall rates and considerable improvements in predicted accuracy and precision by utilizing the strengths of several classifiers. This thorough analysis highlights how crucial it is to take into account a variety of performance criteria in order to fully comprehend the efficacy of classifiers in classification tasks.

CONCLUSION

By tackling a significant obstacle in medical diagnostics, this study has effectively illustrated the potential of an advanced ensemble machine learning model to improve the prediction accuracy of cardiovascular disease (CVD). The ensemble model achieves better predictive performance by combining the advantages of six different algorithms: K Nearest Neighbors, Decision Tree, Support Vector Classifier, AdaBoost, Linear Discriminant Analysis, and Multilayer Perceptron. This allows the ensemble model to handle the complexity and diversity of medical data.

The effectiveness of this ensemble strategy is supported by empirical results that show notable measures such as a 95.00% F1-score, a 93.95% accuracy, a 98.00% precision, and a 93.00% recall. These data demonstrate the ensemble's accuracy in detecting the existence of CVD as well as its capacity to reduce false positives, an important feature in medical diagnostics where errors can have very high consequences.

Moreover, the resilient performance of the ensemble model when compared to standalone algorithms emphasizes the benefit of utilizing a variety of machine learning approaches. In addition to improving diagnostic procedures, this integrative approach raises the bar for intelligent diagnostic systems used in clinical settings. This work not only expands the possibilities for using machine learning in healthcare but also creates new paths for future investigations into ensemble approaches for other complicated diseases, which could completely change the field of predictive analytics in medicine.

REFERENCES

1. Ahmad, GhulabNabi, et al. "Mixed machine learning approach for efficient prediction of human heart disease by identifying the numerical and categorical features." Applied Sciences 12.15 (2022): 7449.
2. Ahmad, GhulabNabi, et al. "Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection." ieee access 10 (2022): 23808-23828.
3. Barhoom, Ali MA, et al. "Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms." (2022).
4. Hassan, Ch Anwar ul, et al. "Effectively predicting the presence of coronary heart disease using machine learning classifiers." Sensors 22.19 (2022): 7227.

5. Riyaz, Lubna, et al. "Heart disease prediction using machine learning techniques: a quantitative review." International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3. Springer Singapore, 2022.
6. Uddin, Khandaker Mohammad Mohi, et al. "Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset." Intelligence-Based Medicine 7 (2023): 100100.
7. Ahsan, MdManjurul, and Zahed Siddique. "Machine learning-based heart disease diagnosis: A systematic literature review." Artificial Intelligence in Medicine 128 (2022): 102289.
8. Jothi Prakash, V., and N. K. Karthikeyan. "Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction." Interdisciplinary Sciences: Computational Life Sciences 13.3 (2021): 389-412.
9. Kameswara Rao, B., et al. "Identification of heart failure in early stages using smote-integrated adaboost framework." Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021. Singapore: Springer Nature Singapore, 2022. 537-552.
10. Javeed, Ashir, et al. "Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: A systematic review and future directions." Computational and Mathematical Methods in Medicine 2022 (2022).
11. Garavand, Ali, et al. "Towards diagnostic aided systems in coronary artery disease detection: a comprehensive multiview survey of the state of the art." International Journal of Intelligent Systems 2023 (2023): 1-19.
12. Garavand, Ali, et al. "Towards diagnostic aided systems in coronary artery disease detection: a comprehensive multiview survey of the state of the art." International Journal of Intelligent Systems 2023 (2023): 1-19.
13. Reddy, Karna Vishnu Vardhana, et al. "An efficient prediction system for coronary heart disease risk using selected principal components and hyperparameter optimization." Applied Sciences 13.1 (2022): 118.
14. Rajendran, Rajkamal, and AnithaKarthi. "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers." Expert Systems with Applications 207 (2022): 117882.
15. Pathan, Muhammad Salman, et al. "Analyzing the impact of feature selection on the accuracy of heart disease prediction." Healthcare Analytics 2 (2022): 100060.
16. Mahoto, Naeem Ahmed, et al. "A machine learning based data modeling for medical diagnosis." Biomedical Signal Processing and Control 81 (2023): 104481.
17. Abdar, Moloud, et al. "A new machine learning technique for an accurate diagnosis of coronary artery disease." Computer methods and programs in biomedicine 179 (2019): 104992.
18. Arukonda, Srinivas, and RamalingaswamyCheruku. "A novel stacking framework with PSO optimized SVM for effective disease classification." Journal of Intelligent & Fuzzy Systems Preprint (2023): 1-20.
19. Ahmad, Ahmad Ayid, and HuseyinPolat. "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm." Diagnostics 13.14 (2023): 2392.
20. Chahar, Ravita, and Deepinder Kaur. "A systematic review of the machine learning algorithms for the computational analysis in different domains." International Journal of Advanced Technology and Engineering Exploration 7.71 (2020): 147.
21. Kiran, Siripuri, et al. "A gradient boosted decision tree with binary spotted hyena optimizer for cardiovascular disease detection and classification." Healthcare Analytics 3 (2023): 100173.
22. GhoshRoy, Debasmita, P. A. Alvi, and João Manuel RS Tavares. "Detection of cardiovascular disease using ensemble feature engineering with decision tree." International Journal of Ambient Computing and Intelligence (IJACI) 13.1 (2022): 1-16.