



ENSEMBLE LEARNING FOR HEART DISEASE DIAGNOSIS: A VOTING CLASSIFIER APPROACH

Optimizing Cardiovascular Diagnosis with Ensemble Learning Techniques

^[1] YOGESH S, ^[2] PANEER THANU SWAROOP C

^[1] Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India

^[2] Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India

Abstract: Cardiovascular disease remains a serious public health problem internationally, responsible for a considerable number of fatalities. Early and correct detection of cardiovascular illness is crucial for optimal care and control of the condition. In this paper, we present an ensemble learning technique that includes voting classifiers to increase the reliability of cardiovascular disease diagnosis. We obtained a set of data from five cardiology databases, which included the Cleveland, Hungary, Switzerland, Long Beach VA and Statlog (Heart) datasets, which supplied us with a total of 1189 entries. We employed a feature engineering approach to extract relevant features from the dataset, enabling us to acquire vital information to enhance our model's performance. We trained and evaluated several machine learning algorithms, such as Random Forests, MLP, K-Nearest Neighbors, Extra Trees, XGBoost, Support Vector Machines, AdaBoost, Decision Trees, Linear Discriminant Analysis, and Gradient Boosting, and then incorporated these models using voting classifiers to produce more reliable and accurate models. Our findings reveal that the proposed ensemble learning process outperforms standalone models and conventional ensemble approaches, obtaining an accuracy rate of 91.4%. Our technique is likely to benefit clinicians in the early diagnosis of heart problems and improve patient outcomes. This work has major significance for the area of cardiology, indicating the possibility for machine learning approaches to boost both the reliability and accuracy of heart disease identification. The recommended ensemble learning technique may be adopted in hospitals to enhance patient care and eventually lessen the worldwide impact of cardiovascular disease. Further study is required to investigate the uses of predictive modeling in cardiology and other medical domains.

Index Terms - Cardiovascular Disease; Heart Disease Diagnosis; Ensemble Learning; Voting Classifier; Machine Learning; Feature Engineering; Cleveland Dataset; Hungary Dataset; Switzerland Dataset; Long Beach VA Dataset; Statlog (Heart) Dataset

I. INTRODUCTION

Heart disease is a truly serious global medical issue, creating a significant amount of mortality each year. As per to data from the World Health Organization (WHO), coronary artery disease account for around 17.9 million deaths per year [1], making it the top cause of death internationally. Among the various kinds of cardiovascular conditions heart disease ranks as one of the most prevalent and life-threatening. Heart disease involves a variety of conditions that impair the way that the hearts operate including clogged coronary arteries, cardiac arrhythmia, and heart failure [2].

Considering the availability of improved medical technology and therapies, the prompt detection of heart ailments remains a difficulty. Early identification and proper identification of cardiovascular disorders are crucial for effective management and treatment of the illness. The usual techniques of assessment, including electrocardiogram, or ECG, and echocardiogram, have limits in regard to accuracy and reliability. There exists an urge for improved and more successful techniques of identifying cardiac disease [3]. Machine learning approaches have emerged as potential tools for enhancing the accuracy and dependability of cardiac disease detection. Machine learning algorithms may learn from data and detect patterns that are not clearly observable by humans. These algorithms may be trained on enormous amounts of patient data and medical pictures to detect early indications of heart failure and forecast the chance of acquiring cardiac disease in future years [4].

The implementation of machine learning methods in the area of medicine has already demonstrated promising outcomes. Several research have claimed promising findings in the early identification and diagnosis of cardiac conditions using predictive machine learning algorithms [5]. Yet, the precision and dependability of these algorithms may be improved by employing ensemble learning approaches. Ensemble learning is a method for machine learning that mixes different models to enhance the overall efficiency of the system. The concept driving ensemble learning is to construct a set of different approaches that are able to support each other's capabilities and shortcomings. By merging the predictions of different models, the precision as well as the dependability of the system may be increased. [6]

In this paper, we present an ensemble learning technique for heart disease detection that incorporates voting classifiers. We utilized four cardiology datasets from the Machine Learning Repository at UCI - the Cleveland, Hungary, Switzerland, and Long Beach VA data sets [7] - to construct a dataset of 1189 entries. We employed a feature engineering strategy to extract significant properties from the dataset, allowing us to acquire critical information to increase our model's performance. We trained and tested many machine learning methods, such as Random Forests [8], MLP, K-Nearest Neighbors, Extra Trees, XGBoost, Support Vector Machines, AdaBoost, Decision Trees, Linear Discriminant Analysis, and Gradient Boosting, to develop our base models. These models were then combined through voting classifiers to build more reliable and accurate models.

To examine the success of our proposed method of ensemble learning for heart disease detection, we applied different performance criteria. We tested our models using measures such as Accuracy, Precision, Sensitivity, Specificity, F1 Score, ROC, and Matthews Correlational Coefficient. These measures enabled us to assess the accuracy, dependability, and overall success of our method. The dataset utilized in our investigation includes a total of 1189 records and 12 characteristics. The characteristics included patient information such as age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, rest ECG, maximum heart rate reached, exercise-induced angina, ST depression, ST slope, and goal. The target characteristic, which was our major emphasis, had binary values of 0 and 1, where 0 indicated the absence of heart illness and 1 indicated the existence of heart disease. [9]

Overall, our study makes an important improvement in the discipline of predictive modeling in medicine, notably in the domain of heart illness identification. Our technique has the potential to increase the correctness and dependability of existing approaches for diagnosing cardiovascular disease, potentially contributing towards more efficient treatment and management of this life-threatening illness.

II. RELATED WORKS

Up to today, several study inquiries have been taken out on the early diagnosis of coronary artery disease and heart failure. They have utilized several machine learning prediction methods and obtained amazing performance. This section includes a complete literature review of research works in the domain of heart disease diagnostics underpinned by machine learning techniques:

In [10], the authors present an improved use of machine learning techniques that predicts heart disease probability. Their strategy comprises randomly splitting the data set and modelling each partition via a regression as well as classification tree strategy. A homogenous ensemble is built from the multiple models using an accuracy-based weighted aging classifier ensemble. This technique obtained 93% and 91% classification accuracies on the Cleveland and Framingham datasets, respectively, exceeding existing machine learning algorithms and related efforts. The findings illustrate the usefulness of the suggested ensemble learning technique for forecasting heart disease risk.

In [11], the authors employed multiple machine learning techniques including deep learning to examine the UCI Machine Learning Heart Disease dataset. They handled extraneous characteristics utilizing Isolation Forest and normalized the findings. The research yielded encouraging findings, which were confirmed by applying accuracy and confusion matrices. Additionally, the research examined the possible coupling of this study with multimedia technologies on mobile devices. The deep learning method yielded an accuracy rate of 94.2%.

In [12], the authors created an ensemble model for heart disease prediction utilizing the University of California, Irvine ML Repository. The meta-algorithm provided a steadier performance than baseline machine learning algorithms, leading to higher prediction accuracy and analytical output dependability. The Bagging Ensemble classifier was determined to have the greatest prediction probability score and was recommended for deployment. The authors provided a cost-effective and user-friendly graphical user interface for the ensemble heart disease prediction model, which is scalable and extendable. These data imply that the ensemble model is a viable prediction alternative for heart disease.

In [13], the authors developed a deep stacking ensemble model for predicting heart disease, which incorporates two pre-trained blended deep learning models (CNN-LSTM and CNN-GRU) with SVM acting as the meta-learner model. Recursive Feature Elimination (RFE) was utilized for feature selection optimization, and the suggested model was compared with five different machine learning models. The suggested model attained the maximum performance utilizing the whole feature set.

In [14], the authors suggest a StackingCVClassifier ensemble model built of Logistic Regression, K Nearest Neighbors, and Naïve Bayes classifiers as base learners to boost the diagnostic performance of these classifiers for predicting heart disease. They preprocessed the information from Cleveland Clinic Foundation adopting different data extraction and feature selection approaches to increase illness prediction accuracy. The suggested ensemble model attained a prediction accuracy of 90.0% on test data, beating the component models. The authors argue that such models may aid healthcare practitioners in creating medical diagnostic systems, particularly in disadvantaged regions, and offer a noninvasive manner of illness diagnosis.

In [15], the authors present a novel ensemble model, named RSS-KNN, for early prediction of heart disease. The framework merges stochastic subspace and K-nearest neighbor (KNN) technique with an isolation-based outlier elimination approach and an upgraded squirrel optimizer. The suggested technique captures key information through RSS and then feeds it into KNN for accurate classification. The simulation results on the UCI ML dataset demonstrate that the ensemble model that has been suggested attains an accuracy of 98.56% with features and 98.10% specificity, exceeding current modern facilities classifiers.

In [16], the authors apply machine learning approaches to identify heart problems in people. The work combines both heterogeneous as well as homogeneous ensemble classifiers and leverages the Synthetic Minority Oversampling Technique (SMOTE) to solve class imbalance and noise. The suggested technique incorporates two steps: SMOTE preprocessing and classification utilizing Naive Bayes (NB), decision tree (DT), and their ensembles. The findings reveal that the AdaBoost-Random Forest predictor achieves 95.47% accuracy in the early detection of heart disease.

In [17], the authors present a heart disease prediction model that incorporates body signals and an algorithm centered upon the integrated datasets and train-test split approach. They compare their findings with prior efforts and employ accuracy, precision, recall, F1-score, and ROC-AUC curves as performance measures for various classifiers. The Random Forest Classifier exhibits 100% performance on the combined heart disease datasets, whereas the Decision Tree, Random Forest, and Gradient Boosting Classifiers yield a 99-100% ROC-AUC. They also employ five-fold cross-validation and the Stacking CV Classifier to increase the machine learning algorithms' performance. The authors discovered that the RFC classification algorithm's accuracy is excellent, and their proposed approach is efficient and accurate for predicting heart disease.

Overall, the studies demonstrate the effectiveness of ensemble learning and feature selection in enhancing heart disease prediction accuracy. The results of the research can be useful in developing more accurate and robust heart disease prediction models, which can assist in early diagnosis and treatment of the condition.

III. COMPUTATIONAL INTELLIGENCE TECHNIQUES

Computational Intelligence Techniques refer to the set of methods and algorithms that enable computers to learn from data, adapt to changing environments, and perform tasks that traditionally require human intelligence. These techniques encompass a wide range of subfields, including machine learning, artificial intelligence, and data mining. Machine learning algorithms are one of the key components of Computational Intelligence Techniques. These algorithms learn from data to improve performance on a specific task. In this section, we will discuss several popular machine learning algorithms that have been used to develop the ensemble model, including Random Forests, MLP, K-Nearest Neighbors, Extra Trees, XGBoost, Support Vector Machines, AdaBoost, Decision Trees and Gradient Boosting [18].

1. Random Forests

Random Forest is a powerful ensemble learning technique used for both classification and regression tasks. It builds a forest of decision trees on random subsets of data and features to reduce the risk of overfitting and increase the accuracy of the model. In this paper, we utilized the `RandomForestClassifier()` to implement the Random Forest algorithm with two different sets of parameters. The first set included the `criterion='entropy'` parameter, which measures the quality of splits, and `n_estimators=100`, which specifies the number of trees in the forest. The second set of parameters was `criterion='gini'` and `n_estimators=100`. This allowed us to compare the performance of Random Forest with different criteria for measuring the quality of splits.

2. Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a type of artificial neural network that is used for classification tasks. It has an output layer with numerous neurons connected by weighted connections, one or more hidden layers, and an input layer. The MLP classifier is a supervised learning algorithm that adjusts the weights between neurons using a backpropagation algorithm to minimize the error between predicted and actual output. We used the `MLPClassifier()` as one of the base model in this paper.

3. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a classification algorithm commonly used in mining medical databases. It works by measuring the similarity between a new case and other cases in the dataset. The algorithm determines the distance between the new case and every other case in the model, and then assigns the new case to the output that contains the closest neighbors.

The KNN algorithm has two steps:

- 1) finding the K training instances that are closest to the unidentified instance.
- 2) selecting the most frequently occurring classifications for these K instances.

In this paper, we have used the `KNeighborsClassifier()` to implement the KNN algorithm with `K=9`, which means that the algorithm considers the 9 nearest neighbors to the new instance for classification.

4. Extra Trees

Extra Trees is an ensemble learning method that uses a combination of multiple decision trees to generate a forest. It constructs many random decision trees and averages their output to produce the result. This approach helps to reduce overfitting and improve the accuracy of the model. In Python, the `ExtraTreesClassifier()` function is used to implement the Extra Trees algorithm for classification problems. The parameter `n_estimators` specify the number of trees to be used in the forest. In this paper, we used three `ExtraTreesClassifier` models as base models for our ensemble model, with `n_estimators` set to 100, 500, and 1000 respectively.

5. XGBoost

XGBoost, short for eXtreme Gradient Boosting, is a gradient boosting algorithm used for supervised learning tasks. It is designed to enhance the performance of traditional gradient boosting algorithms by incorporating a regularization term to the loss function, which helps to prevent overfitting. XGBoost builds multiple decision trees iteratively, with each tree attempting to correct the errors of the previous tree. The algorithm assigns weights to each instance in the data, with the weights of the misclassified instances being increased in subsequent iterations. In this paper, we utilized the `XGBClassifier()` to create five models. The first model was built without any additional parameters, while the second model had `n_estimators` set to 2000. The third model was created with `n_estimators` set to 500, the fourth with `n_estimators` set to 100, and the fifth with `n_estimators` set

to 1000. These models were used as the base models for our ensemble learning approach, aimed at improving the classification performance on our medical dataset.

6. Support Vector Machines (SVM)

Support Vector Machine (SVM) has shown excellent performance in the medical sector for disease prediction in recent years. SVM is a supervised learning technique designed for regression and classification tasks, with the primary goal of minimizing generalization errors. SVM works by classifying data into two classes over a hyperplane. It is highly effective in high-dimensional spaces, even when the number of dimensions exceeds the number of samples. Mathematically, SVM is represented as follows.

$$\text{If } Y_i = +1; w x_i + b \geq 1 \quad (1)$$

$$\text{If } Y_i = -1; w x_i + b \leq -1 \quad (2)$$

$$\text{For all } i; y_i (w_i + b) \geq 1 \quad (3)$$

In the equation, x is a vector point and w is a weight and a vector. So, to separate the data in Equation (1) should always be greater than zero and the data in Equation (2) should always less than zero. Among all possible hyperplanes, SVM selects the one where the distance of hyperplane is as large as possible. In this paper, we have used the SVC() with kernel='linear', gamma='auto', and probability=True parameters. The SVC (Support Vector Classifier) is a type of SVM algorithm that is used for classification tasks. The kernel parameter specifies the type of kernel function to be used for the decision function, while the gamma parameter controls the width of the kernel function. The probability parameter is set to True to enable probability estimates for each class.

7. AdaBoost

AdaBoost is an ensemble learning method used for classification problems. It combines multiple weak classifiers to create a strong classifier. The weak classifiers are typically decision trees with a depth of 1 and are trained on subsets of the data. During training, the algorithm assigns weights to the samples in the data set and trains a new weak classifier in each iteration. The final classifier is a weighted sum of the weak classifiers, where each classifier is assigned, a weight based on its accuracy. In this paper, we used the AdaBoostClassifier function without any parameters.

8. Decision Tree

Decision trees are powerful and popular tools for predicting and classifying medical data. Decision Tree is a classifier expressed as a recursive part of the information space depending on the values of the attributes. Each internal node splits the instance space into two or more subspaces according to certain function of the input attribute values. Each leaf is allocated to one class that represents the most appropriate value. There are many types of algorithms for DT. In this paper, we used the DecisionTreeClassifier() as one of the base models in our ensemble model. We did not specify any additional parameters for this model, meaning that it used the default parameters to build the decision tree.

9. Gradient Boosting

Gradient Boosting is an ensemble learning technique used for classification and regression problems. It builds trees in a sequential way by giving more weight to the misclassified instances in the previous tree. The GradientBoostingClassifier in Python uses the gradient boosting algorithm for classification problems, with n_estimators and max_features as important parameters. In our research, we used the GradientBoostingClassifier with n_estimators=100 and max_features='sqrt' as one of the base models for our ensemble model.

IV. METHODS AND MATERIALS

In this study, to forecast the coronary artery heart disease, we use the following procedures, (1) Data collection, (2) Exploratory Data Exploration (EDA), (3) Data preprocessing, (4) Base Models Evaluation, (5) Feature Selection, and (6) Voting Ensembling. The following sections address all the processes in depth.

1. Data Collection

The data accumulating method is an important part of every research study as it assures the quality and confidence of the results. In this work, all the five data sets' collection was done using the heart disease dataset accessible via the UCI Machine Learning Repository. The Heart Disease dataset is an extensive database that is commonly utilized in studies relating to heart disease.

The dataset utilized in this research is generated by combining five significant heart disease datasets, including Cleveland, Hungarian, Switzerland, Long Beach VA, and Statlog (Heart) Data Set, over 12 common features. This unification procedure has resulted in a single dataset that is the greatest heart disease dataset available for research reasons. By merging data from numerous sources, the dataset gives a more complete picture of cardiac disease, including its causes, risk factors, and possible therapies.

The Heart Disease dataset covers both male and female patients, giving it a diversified sample for heart disease research. With an aggregate total of 1189 samples divided into 11 categories, the dataset provides a plethora of information for academics to evaluate. The 11 factors presumably include the clinical assessment information of the patient, which may give insights into the development and progression of cardiac disease. The target categorization, which is the presence or absence of cardiac disease, is the 12th characteristic, making it the subject of this investigation. The availability of such a vast and diversified dataset is a huge benefit for researchers investigating cardiac disease. The scale and variety of the dataset enables more accurate and dependable discoveries, which may eventually lead to improved diagnosis, prevention, and treatment of cardiac disease. More information regarding the data set may be accessed in Table 1.

TABLE I. Description of the Heart Disease Dataset

S. No	Attribute	Datatype	Value
(i)	age	Numeric	[28 – 77]
(ii)	sex	Binary	1 = male, 0= female;
(iii)	chest pain type	Nominal	1 indicates typical angina, 2 indicates atypical angina, 3 indicates non-anginal pain, and 4 indicates no symptoms;
(iv)	resting bp s	Numeric	[94–200]
(v)	cholesterol	Numeric	[126–564]
(vi)	fasting blood sugar	Binary	1,0; > 120 mg/dl
(vii)	resting ecg	Numeric	According to Estes' criteria, 0 represents normal, 1 represents an abnormal ST-T wave, and 2 represents either definite or likely left ventricular hypertrophy.
(viii)	max heart rate	Nominal	[71–202]
(ix)	exercise angina	Binary	1 = yes; 0 = no
(x)	oldpeak	Numeric	1 = yes; 0 = no
(xi)	ST slope	Nominal	1 = upsloping; 2 = flat; 3 = downsloping;
(xii)	target	Binary	1 = heart disease; 0 = Normal;

Overall, the Heart Disease dataset utilized in this study is a great resource for researchers investigating heart disease and associated illnesses.

2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial component in every data-driven research endeavor. It includes assessing and summarizing data to acquire insights and knowledge about underlying patterns and relationships between variables. EDA tools, like visualizations and statistical summaries, may assist spot outliers, missing values, and irregularities in the data. By employing EDA, researchers may acquire a better understanding of the data and highlight any difficulties that may need to be resolved before further investigation. The goal of doing EDA on the dataset is to acquire insights into the distribution and relationships between variables. The information comprises numerous demographic, medical, and clinical variables, and examining these features might assist reveal possible risk factors and enhance preventive strategies. Additionally, EDA may assist in uncovering any missing or erroneous data pieces that might impair the accuracy of later research.

Figure 1 depicts the correlation coefficients among every pair of features in the dataset. The graphic was built using Seaborn's heatmap() function with the "PuBuGn" colour map and illustrates the correlation values between each feature. Purple denotes a weaker correlation, whereas green represents a stronger correlation.

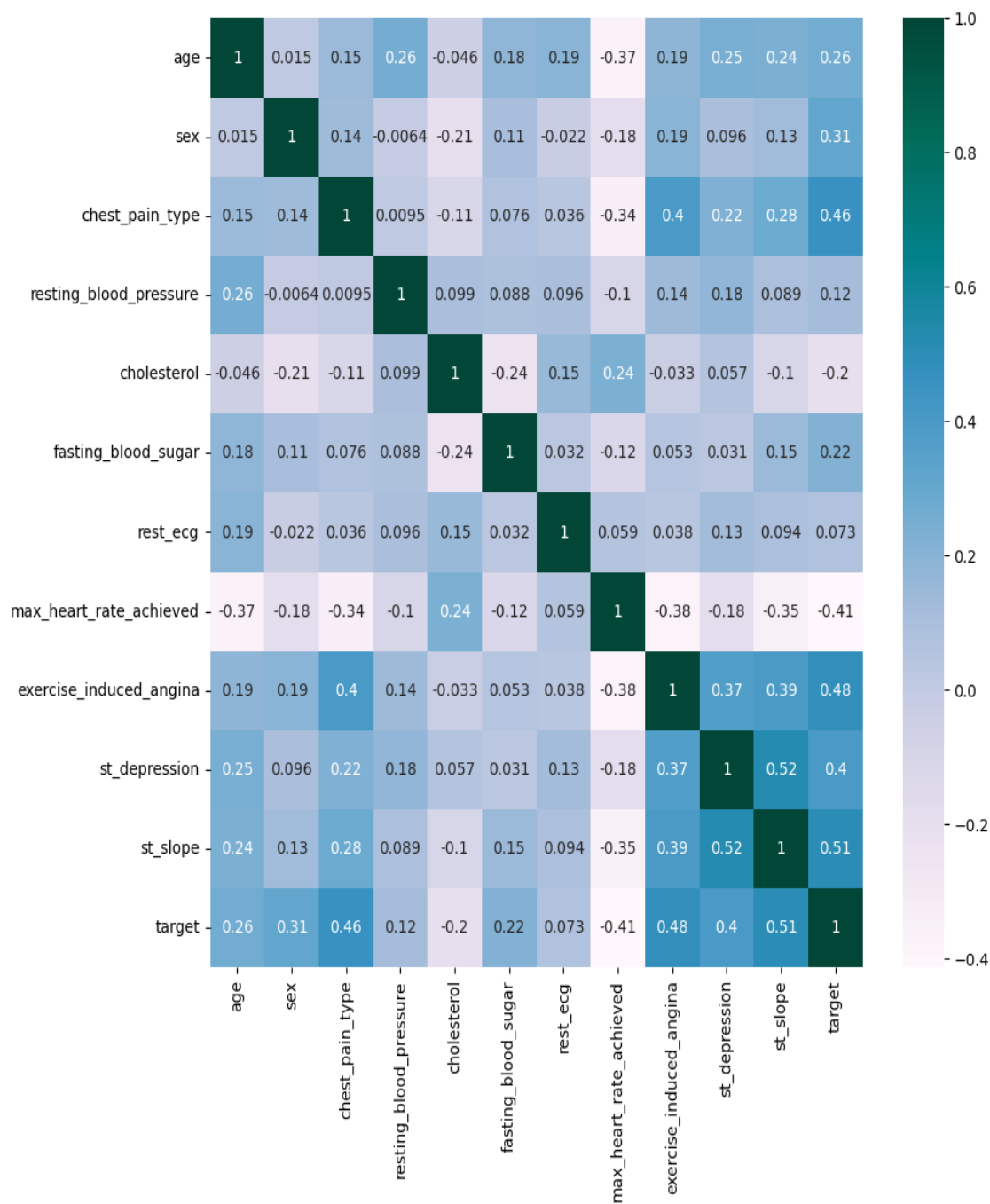


Figure 1 - Correlation heatmap of the dataset

Figure 2 depicts the number of normal and heart disease patients in this dataset. It displays the count of the patients on the x-axis and both categories on the y-axis. The green bars show the normal patients, whereas the red column depicts the heart disease victims. The count for every category is presented at the end of the column.

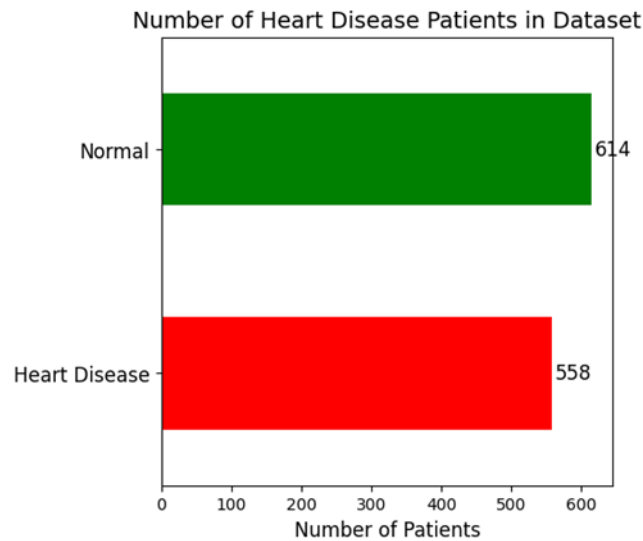


Figure 2 - Number of normal and heart disease patients

Figure 3 shows the distribution of gender in the dataset was investigated and the findings reveal that 76.4% of the people were male while 23.6% were female. This observation shows that the dataset may be biased towards a male population and might possibly affect any gender-related research or conclusions taken from the data.

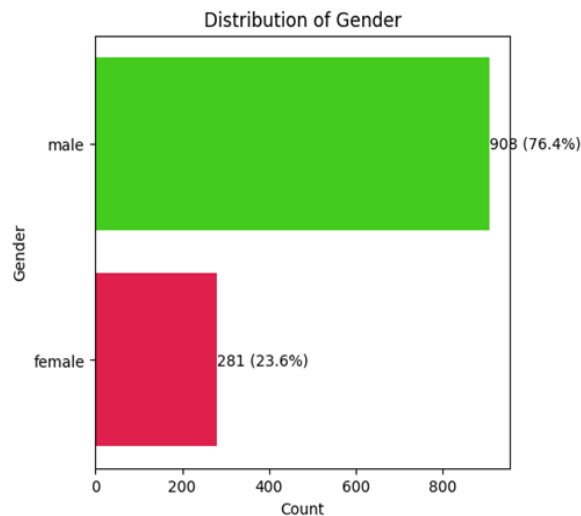


Figure 3- Count of Gender

In this research initiative, we have done an in-depth investigation of the dataset, which has yielded vital insights into the distribution and interactions between the variables. We have identified possible risk factors that could lead to heart disease and underlined the need to correct missing or erroneous data. Furthermore, we have detected a gender bias in the dataset, which may have ramifications for gender-related research or conclusions taken from the data. With these insights in mind, we will go to the data processing step, where we will prepare the dataset for future analysis.

3. Data Preprocessing

Data preprocessing is a vital stage in every machine learning effort. The performance of classification problems is heavily reliant on the quality and standard of the dataset used for training and testing the models. One typical problem that may develop in datasets is missing values, which may impair the accuracy and dependability of the conclusions produced. Therefore, it is vital to detect and manage missing information effectively before continuing with any analysis.

There are numerous techniques for addressing missing values, including altogether disregarding them, replacing them with a numeric value such as zero, replacing them with the most common value for that property, or replacing them with the mean value for that attribute. It is important to emphasize that there are no missing values in the dataset used for this research work. As a result, handling missing values is not necessary, and we can proceed to the following stage of data preparation.

To make the categorical variables more intelligible, we have applied label encoding to the 'chest_pain_type', 'rest_ecg' and 'st_slope' columns.

During the exploratory data analysis phase, it was found that several of the numerical characteristics in the dataset contained outliers. Specifically, the age, resting blood pressure, cholesterol, and maximum heart rate reached had outliers that might possibly impair the accuracy and dependability of the machine learning models. Therefore, it is vital to detect and treat outliers effectively before continuing with any study. Various approaches, such as z-score, interquartile range, or Tukey's fences, may be used to detect and treat outliers. In this work, we employed the z-score approach to detect and eliminate outliers from the dataset to assure the

correctness and reliability of the findings acquired from the machine learning models. To address outliers in the dataset, a threshold value of 3 was specified, which was used to detect data points that were more than 3 standard deviations away from the mean. This was performed using the `np.where()` function to find the indices of all data points that are above the threshold value, and then filtering out those points from the dataset using the `all()` method with `axis=1` option. After eliminating outliers from the sample, the size of the dataset dropped from (1190, 12) to (1172, 12). This shows that a total of 18 data points were recognized as outliers and eliminated from the dataset.

After outlier elimination, categorical variables in the dataset were encoded using one-hot encoding. The generated dataset comprised a total of 15 columns, containing 5 numerical variables and 10 category features. To analyze the performance of our model, we divided the data into training and testing datasets. In this paper, the data is divided into two sets: one for training and one for testing. The `stratify` option is set to `y`, which guarantees that the datasets used for training and testing have the same percentage of target variable classes as the original dataset. The `test_size` option is set to 0.3, which implies that 30% of the data is utilized for testing, while the remaining 70% is used for training. The `shuffle` option is set to `True`, which implies that the data undergoes shuffle before splits, and the `random_state` parameter is set to 5, which assures that the findings are repeatable.

To verify that all numerical features are on the same scale, we conducted normalization on the dataset using the `MinMaxScaler` from the `scikit-learn` module. The scaler stabilizes data that is within the range between 0 to 1. The age, resting blood pressure, cholesterol, maximum heart rate reached, and st depressive characteristics were all adjusted using this procedure. This normalizing method helps guarantee that no one characteristic dominates the others and that all features serve equally to the study.

In summary, the data preparation phase includes different stages including resolving missing values, eliminating outliers, and encoding categorical variables. After all the preparation stages, we ended up with a total of 820 data points in the training set and 352 data points in the test set, both containing 15 columns. These data sets will be utilized to train and assess our machine learning model in the next part of the research work.

4. Base Model Evaluation

In this work, several machine learning approaches were employed to forecast the incidence of heart disease. The implementation specifics of these models were described in Section III. To further increase the prediction accuracy, an ensemble model utilizing the Voting Classifier approach was constructed. The Voting Classifier is a simple approach to aggregate predictions from many models, and it may enhance the overall performance of the model. The ensemble model was trained on the training data and deployed to predict the labels of the test data. The implementation specifics of the Voting Classifier model are as follows: initially, a Voting Classifier object was constructed with the estimators comprising all the individual models. The 'hard' voting approach was utilized, which means the projected class labels of each model were used to create the final forecast. The ensemble model was then trained on the training data and used to predict the labels of the test data.

After building a voting classification model employing the combined results of the models, the model gets evaluated using the confusion matrix. The confusion matrix is used to illustrate the number of accurate and wrong predictions produced by the model. The true negative (TN), false negative (FN), true positive (TP), and false positive (FP) values are computed from the confusion matrix. This assessment helps to determine the usefulness of the model in predicting heart disease outcome. Figure 4 illustrates the confusion matrix associated with the voting classifier model. The TN value is 150, FP value is 22, FN value is 14, and TP value is 170.

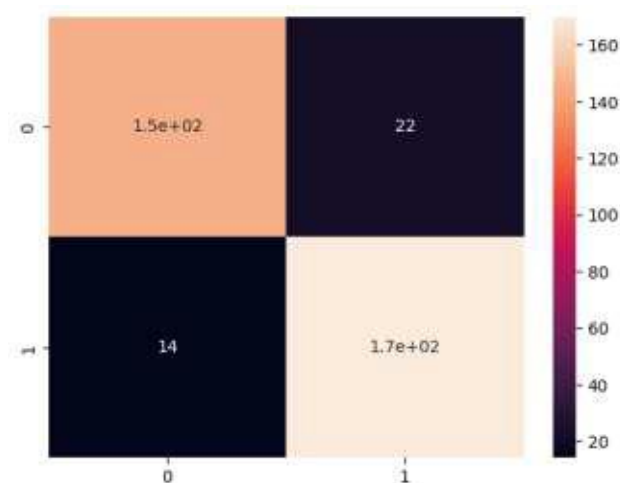


Figure 4 - Confusion matrix for the ensemble model on the test set

5. Feature Selection

The feature selection process involves identifying the most relevant features from a dataset that contribute the most towards a model's predictive performance. It helps to remove redundant or irrelevant features, which can lead to overfitting and reduced model performance. Based on the Random Forest classifier, we have performed feature selection by selecting the most important features. The feature importances have been plotted to identify the important features, and a threshold of 0.05 has been used to select the most important features. The selected features have been used to create new training and testing sets, and the other less

important features have been discarded. Figure 5 shows the feature importance plot generated by the Random Forest classifier. The plot displays the relative importance of each feature, with higher bars indicating greater importance.

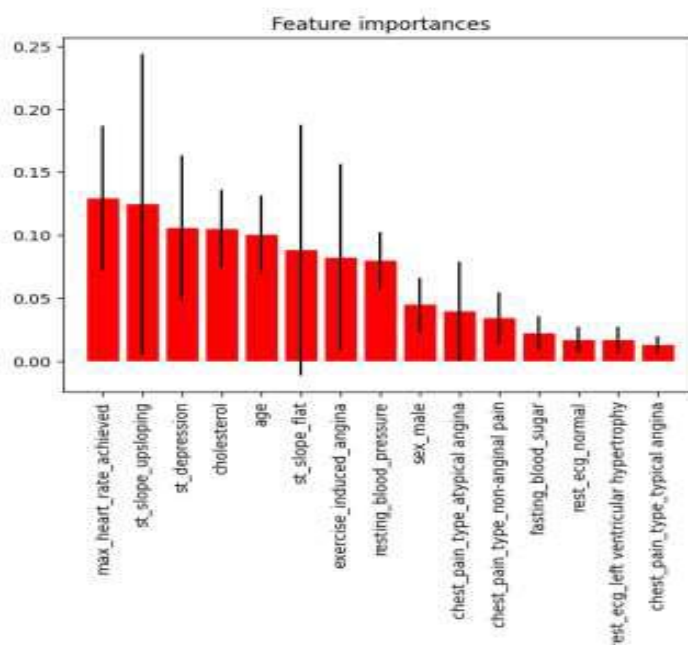


Figure 5 - Feature importance plot generated using Random Forest classifier.

The feature importance plot generated using Random Forest classifier shows that the most important features for predicting heart disease are the maximum heart rate achieved, ST slope during upsloping, ST depression, cholesterol level, age, ST slope during flat, presence of exercise-induced angina, resting blood pressure, and male sex. The least relevant characteristics include chest pain kind (atypical angina, non-anginal discomfort, typical angina), fasting blood sugar level, and resting ECG findings (normal, left ventricular hypertrophy).

Next, these essential attributes will be utilized to develop a new ensemble voting classifier, which is projected to substantially increase the accuracy of the initial ensemble model.

6. Voting Ensembling

In this step, we defined a new voting classifier with soft voting and used the same machine learning models described in section III as estimators. The main difference is that the base voting classifier model has undergone feature engineering, where only the best features selected in the previous step are fed into the model. The voting classifier is then trained using the training set and evaluated using the test set. After training and assessing the voting classifier using the given features, we constructed a confusion matrix and estimated the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). We then showed the confusion matrix using a heatmap with annotations to illustrate the number of correct and wrong predictions for each class. Figure 6 illustrates the confusion matrix of the feature-engineered Voting Classifier on the test set. In this case, the confusion matrix has TN = 98, FP = 13, FN = 7, and TP = 120.

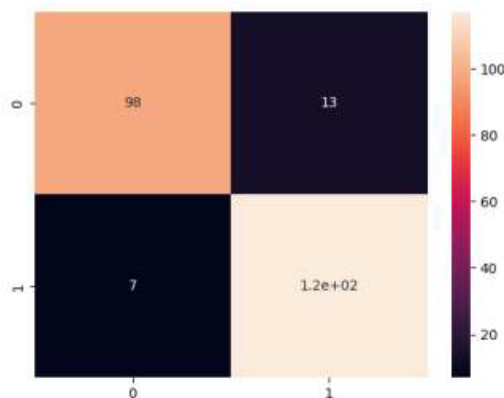


Figure 6 - Confusion matrix of the feature-engineered Voting Classifier.

Figure 7 illustrates the comprehensive system design, providing a concise overview of the proposed ensemble learning approach for heart disease diagnosis. The system is designed to address the critical need for accurate and reliable cardiovascular disease detection.

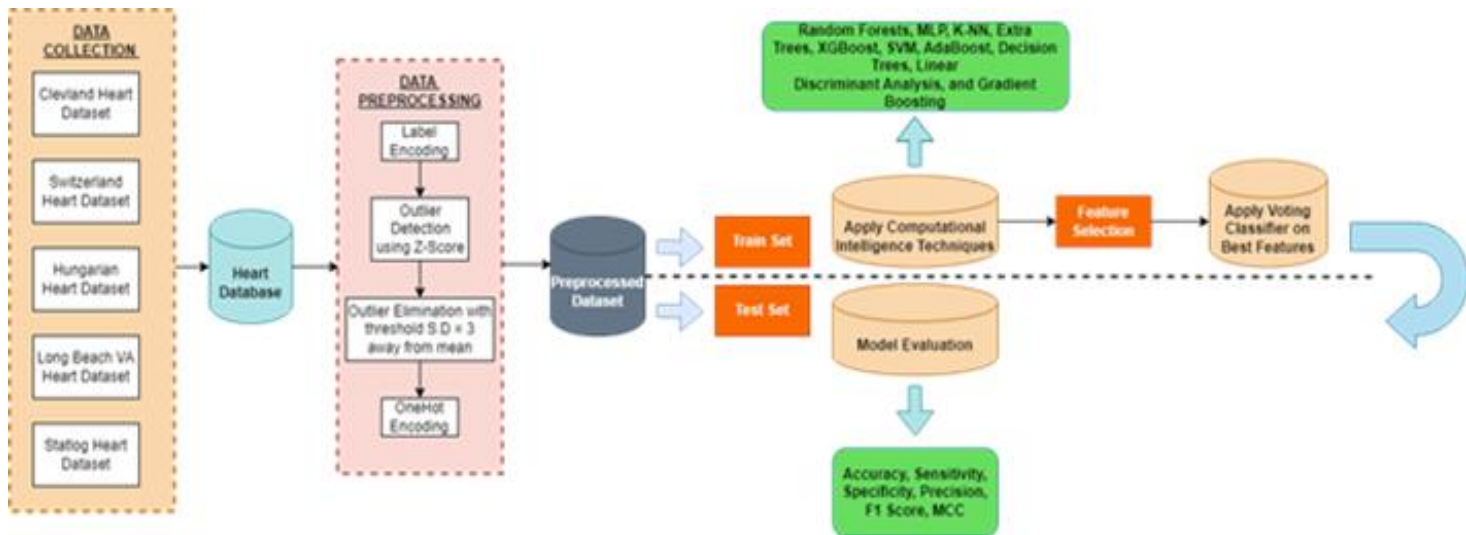


Figure 7 - Full system for coronary heart disease diagnosis prediction

V. PERFORMANCE MEASURE

Using the confusion matrix, we can visualize the performance of both the voting classifier models. In the confusion matrix, four classification performance indices are present. The definition of these is given below:

- TP = True Positive (Correctly Identified)
- TN = True Negative (Correctly Rejected)
- FP = False Positive (Incorrectly Identified)
- FN = False Negative (Incorrectly Rejected)

To evaluate the performance of the system, the following qualities are measured:

- (1) Accuracy: Ratio of correctly classified samples to total samples. It is represented by:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (4)$$

- (2) Sensitivity: Ratio of correctly classified positive samples to total positive instances, given by:

$$\text{Sensitivity} = TP / (TP + FN) \quad (5)$$

- (3) Specificity: Ratio of correctly classified negative instances to total negative instances, given by:

$$\text{Specificity} = TN / (TN + FP) \quad (6)$$

- (4) Precision: Precision is calculated as the number of correctly positive predicts divided by the total number of positively predict, given by:

$$\text{Precision} = TP / (TP + FP) \quad (7)$$

- (5) F1 Score: The range of F1 scores is between 0 and 1. The best value is 1 and the worst value is 0, F1 is represented as follows:

$$\text{F1 score} = 2TP / (2TP + FP + FN) \quad (8)$$

- (6) Mathews Correlation Coefficient (MCC): For binary classification, the range of MCC is +1 to -1.

When the value is +1 the best performance is shown and when the value is -1 the worst performance is shown. It is represented as:

$$\text{MCC} = (TP \times TN) - (FP \times FN) / \sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))} \quad (9)$$

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

1. Experimental Setup

The experiment was done using Python programming language on a single Virtual Machine configured through Microsoft Azure with the following specifications, Intel (R) Xeon(R) Platinum 8272CL CPU @ 2.60GHz and 32 GB RAM, running on Windows 10 operating system. This information is supplied to aid understanding and replication of the findings.

2. Results Analysis

In this part, we give a performance assessment of several computational intelligence methodologies presented in Part III. We trained various machine learning models using the heart disease dataset, with a training-to-testing information ratio of 80:20. The accuracy ratings achieved by the various machine learning models are as follows: Random Forest (90.63%), Multi-Layer Perceptron (82.95%), K-Nearest Neighbors (80.68%), Extra Tree Classifier (89.77%), XGBoost (90.34%), Support Vector Classifier (82.67%), Stochastic Gradient Descent (73.86%), Adaboost (80.97%), and Classification and Regression Trees (82.95%). In addition to accuracy, we also examined the performance of the heart disease prediction model using numerous other measures,

including precision, sensitivity, specificity, F1 score, and Mathew correlation coefficient. The results of these measures are provided in Table 2.

TABLE 2. Evaluation metrics of the heart disease prediction system

Parameters	RF	MLP	KNN	SVC	GBM	CART	XGB	Voting Classifier
Sensitivity (%)	94.56	87.50	83.69	85.32	87.50	82.06	91.30	94.35
F1 Score	91.33	84.29	81.91	83.73	84.96	83.42	90.81	92.49
Accuracy (%)	90.62	82.95	80.68	82.67	83.80	82.95	90.34	91.91
Precision (%)	88.32	81.31	80.20	82.19	82.56	84.83	90.32	90.69
Specificity (%)	86.30	77.97	77.38	79.76	79.76	83.92	89.28	89.18
MCC	0.813	0.659	0.659	0.652	0.675	0.659	0.806	0.838

We then integrate these different models into a hard voting classifier to increase the overall accuracy. The hard voting classifier aggregates the outcomes of each individual model and picks the class that has the highest number of votes. After merging the separate models, we achieved an accuracy score of 90.34%, which is greater than the individual models. This implies that the mix of models has a beneficial influence on the overall accuracy of the classifier.

We sought to increase the accuracy of the vote classifier by doing feature selection. Our feature selection strategy enhanced the accuracy of the voting classifier, with an accuracy score of 91.91% which is greater than the original voting classifier's accuracy of 90.34%. This indicates that feature selection may be a useful strategy to increase the accuracy of the classifier.

Overall, our findings suggest that integrating separate models into an ensemble model and conducting feature selection may greatly increase the accuracy of the heart disease prediction model. These results have crucial implications for the creation of more accurate and reliable heart disease prediction models that may assist in the diagnosis and treatment of heart disease.

To further show the efficacy of our final feature-selected voting classifier model, we have included a Receiver Operating Characteristic (ROC) curve plot in Figure 8. The ROC curve is a graphical depiction of the true positive rate (sensitivity) versus the false positive rate (1-specificity) for various classification thresholds. The area under the ROC curve (AUC) gives a single statistic to assess the performance of the model, with values ranging from 0.5 to 1.0. An AUC value of 0.5 suggests random chance, whereas an AUC value of 1.0 shows flawless categorization.

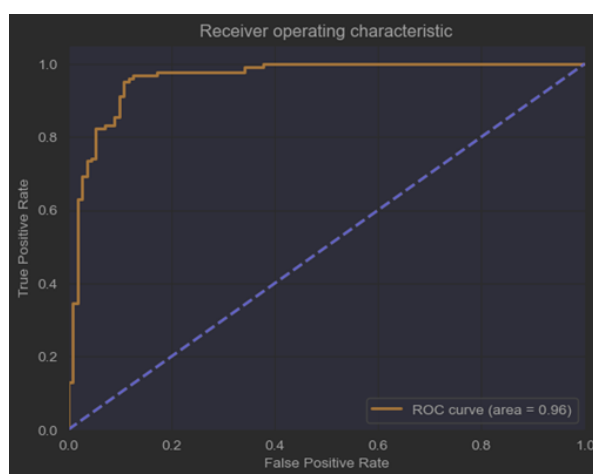


Figure 8 - ROC curve plot for the final feature-selected voting classifier model.

As illustrated in Figure 8, our final feature-selected voting classifier model attained an AUC of 0.96, demonstrating outstanding performance in differentiating between positive and negative instances of heart disease. The steep curve and high AUC value reflect the high sensitivity and specificity of the model, suggesting that it can reliably predict both genuine positive and true negative situations.

Besides this, a ROC curve plot, we additionally included a Precision-Recall curve plot for the final feature-selected voting classifier model in Figure 9. The Precision-Recall curve is another essential assessment measure for binary classification issues, especially in circumstances when the classes are unbalanced. The curve is constructed by graphing the precision (positive predictive value) against the recall (true positive rate) for various categorization thresholds. The area under the Precision-Recall curve (AUC-PR) gives a single statistic to assess the performance of the model, with values ranging from 0 to 1.

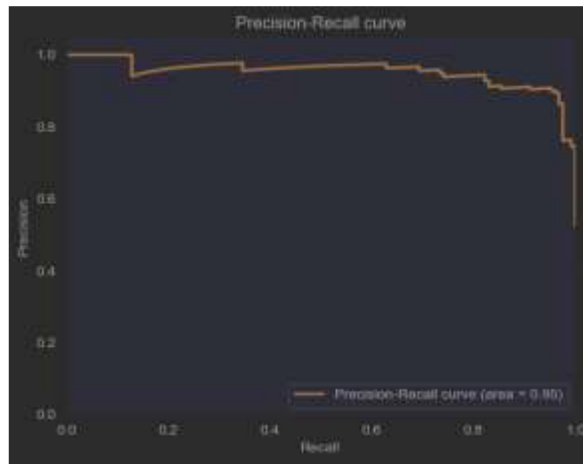


Figure 9 - Precision-Recall curve plot for the final feature-selected voting classifier model.

As illustrated in Figure 9, our final feature-selected voting classifier model obtained an AUC-PR of 0.95, showing good accuracy in detecting positive instances of heart disease while reducing false positive predictions. The high accuracy and recall numbers reflect the model's ability to reliably identify positive instances, while the steep curve and high AUC-PR value imply a low false positive rate.

VII. CONCLUSION

In conclusion, our work reveals that merging different machine learning models into an ensemble model and employing feature selection strategies may increase the accuracy of heart disease prediction models. The individual models we examined, including Random Forest, Multi-Layer Perceptron, and XGBoost, showed great accuracy rates, but integrating them into a hard voting classifier significantly increased the total accuracy. Moreover, adding feature selection methods to identify the most significant characteristics for prediction enhanced the accuracy of the voting classifier even more.

Our results have major implications for improving heart disease detection and therapy. With dependable prediction models, physicians and medical professionals may offer more effective treatment and preventative measures for individuals at risk of heart disease. The integration of several models and feature selection strategies may also be used to other healthcare domains to increase the accuracy of prediction models and help clinical decision-making.

While our study gives encouraging findings, there are still several limitations that need to be addressed in future research. For instance, the dataset utilized in this research only comprises a limited number of characteristics, and it may be interesting to examine other features that might increase the accuracy of the prediction models further. Additionally, it would be helpful to study the generalizability of our technique across diverse datasets and demographics.

Overall, our work illustrates the possibility of combining machine learning models and employing feature selection strategies to increase heart disease prediction accuracy. Our results offer a platform for ongoing study and development of precise and dependable prediction models to enhance clinical decision-making in the battle against heart disease.

VIII. ACKNOWLEDGEMENT

We would like to convey our profound thanks to Dr. K Ruba Soundar, Associate Professor, CSE Dept, Mepco Schlenk Engineering College for his significant help, advice, and mentoring during the study endeavour. We are thankful for his continual encouragement, unshakable support, and useful suggestions, which have contributed to the effective completion of this task.

We would also want to recognize the contribution of the authors of the heart disease dataset utilized in this research, as well as the developers of the numerous machine learning tools and frameworks used in this study. Their effort has made this research feasible and supplied us with the tools and resources required to perform this investigation.

Finally, we would like to thank our colleagues and friends who have given us with vital comments, encouragement, and support throughout the study process. Their suggestions and thoughts have been incredibly helpful in constructing this study article.

REFERENCES

- [1] World (2021). Cardiovascular diseases (CVDs). [online] Who.int. Available at: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) [Accessed 26 Apr. 2023].
- [2] CDC (2022). Heart Disease Facts. [online] Centers for Disease Control and Prevention. Available at: <https://www.cdc.gov/heartdisease/facts.htm> [Accessed 27 Apr. 2023].
- [3] Mayo Clinic. (2022). Electrocardiogram (ECG or EKG) - Mayo Clinic. [online] Available at: <https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983> [Accessed 27 Apr. 2023].

- [4] Park, D.-J., Park, M.S., Lee, H.K., Young Ho Kim, Kim, Y. and Park, Y.-H. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Scientific Reports*, [online] 11(1). doi:<https://doi.org/10.1038/s41598-021-87171-5>.
- [5] May, M.J. (2021). Eight ways machine learning is assisting medicine. *Nature Medicine*, [online] 27(1), pp.2–3. doi:<https://doi.org/10.1038/s41591-020-01197-2>.
- [6] M.A. Ganaie, Hu, M., Tanveer, M. and Ponnuthurai Nagaratnam Suganthan (2021). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, [online] 115, pp.105151–105151. doi:<https://doi.org/10.1016/j.engappai.2022.105151>.
- [7] Uci.edu. (2023). UCI Machine Learning Repository: Heart Disease Data Set. [online] Available at: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease> [Accessed 27 Apr. 2023].
- [8] Google Developers. (2022). Available at: <https://developers.google.com/machine-learning/decision-forests/random-forests> [Accessed 27 Apr. 2023].
- [9] MANU SIDDHARTHA (2020). Heart Disease Dataset (Comprehensive). [online] IEEE DataPort. Available at: <https://ieeedataport.org/open-access/heart-disease-dataset-comprehensive> [Accessed 27 Apr. 2023].
- [10] Ibomoiye Domor Mienye, Sun, Y. and Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked*, [online] 20, pp.100402–100402. doi:<https://doi.org/10.1016/j.imu.2020.100402>.
- [11] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S. and Singh, P. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience*, 2021, pp.1–11. doi:<https://doi.org/10.1155/2021/8387680>.
- [12] Ahmad Bamanga Mahmud, Ahmadu Asabe Sandra, Musa Yusuf Malgwi and Sajoh, D.I. (2021). Ensemble model for Heart Disease Prediction. [online] *Science Progress and Research*. Available at: <https://www.semanticscholar.org/paper/Ensemble-model-for-Heart-Disease-Prediction-Mahmud-Sandra/c941743125ea116d0862012b15330ce03ee8aa2d> [Accessed 26 Apr. 2023].
- [13] Almulihi, A., Saleh, H., Ali Mohamed Hussien, Mostafa, S., Shaker El-Sappagh, Khaled Alnowaiser, Ali, A.A. and Moatamad Refaat Hassan (2022). Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction. *Diagnostics*, [online] 12(12), pp.3215–3215. doi:<https://doi.org/10.3390/diagnostics12123215>.
- [14] Naveed Anjum Itoo and Vijay Kumar Garg (2022). Heart Disease Prediction using a Stacked Ensemble of Supervised Machine Learning Classifiers. 2022 International Mobile and Embedded Technology Conference (MECON). [online] doi:<https://doi.org/10.1109/mecon53876.2022.9751883>.
- [15] Kannan, S. (2022). An automated clinical decision support system for predicting cardiovascular disease using ensemble learning approach. *Concurrency and Computation: Practice and Experience*, [online] 34(18). doi:<https://doi.org/10.1002/cpe.7007>.
- [16] Mahesh, T.R., Kumar, V., Kumar, V., Asghar, J., Oana Geman, G. Arulkumaran and Arun, N. (2022). AdaBoost Ensemble Methods Using K-Fold Cross Validation for Survivability with the Early Detection of Heart Disease. *Computational Intelligence and Neuroscience*, [online] 2022, pp.1–11. doi:<https://doi.org/10.1155/2022/9005278>.
- [17] Ghulab Nabi Ahmad, None Shafiullah, Fatima, H., Abbas, M., Md. Obaidur Rahman, None Imdadullah and Alqahtani, M. (2022). Mixed Machine Learning Approach for Efficient Prediction of Human Heart Disease by Identifying the Numerical and Categorical Features. *Applied sciences*, [online] 12(15), pp.7449–7449. doi:<https://doi.org/10.3390/app12157449>.
- [18] Lubna Riyaz, Muheet Ahmed Butt and Zaman, M. (2022). Ensemble Learning for Coronary Heart Disease Prediction. 2022 2nd International Conference on Intelligent Technologies (CONIT). [online] doi:<https://doi.org/10.1109/conit55038.2022.9848292>.
- [19] Indu Yekkala, Dixit, S. and Makhdoom Abdul Jabbar (2017). Prediction of heart disease using ensemble learning and Particle Swarm Optimization. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). [online] doi:<https://doi.org/10.1109/smarttechcon.2017.8358460>.
- [20] Indu Yekkala, Dixit, S. and Makhdoom Abdul Jabbar (2017). Prediction of heart disease using ensemble learning and Particle Swarm Optimization. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). [online] doi:<https://doi.org/10.1109/smarttechcon.2017.8358460>.
- [21] Gahtani, A. A. . and El-Sayed, H. . (2022) “Building a Novel Ensemble Learning – Based Prediction Framework for Diagnosis of Coronary Heart Disease”, *International Journal of Intelligent Systems and Applications in Engineering*, 10(3), pp. 265–273. Available at: <https://ijisae.org/index.php/IJISAE/article/view/2164> (Accessed: 27 April 2023).

- [22] Raza, K. (2019). Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. *U-Healthcare Monitoring Systems*, pp.179–196. doi:<https://doi.org/10.1016/b978-0-12-815370-3.00008-6>.
- [23] R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.
- [24] Chandra Shekar, K., Chandra, P. and Venugopala Rao, K. (2019). An Ensemble Classifier Characterized by Genetic Algorithm with Decision Tree for the Prophecy of Heart Disease. *Innovations in Computer Science and Engineering*, pp.9–15. doi:https://doi.org/10.1007/978-981-13-7082-3_2.
- [25] Dželila Mehanović, Zerina Mašetić and Kečo, D. (2019). Prediction of Heart Diseases Using Majority Voting Ensemble Method. *IFMBE proceedings*, pp.491–498. doi:https://doi.org/10.1007/978-3-030-17971-7_73.
- [26] Mienye, I.D., Sun, Y. and Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked*, 20, p.100402. doi:<https://doi.org/10.1016/j.imu.2020.100402>.
- [27] Rahim, A., Rasheed, Y., Azam, F., Anwar, M.W., Rahim, M.A. and Muzaffar, A.W. (2021). An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases. *IEEE Access*, 9, pp.106575–106588. doi:<https://doi.org/10.1109/access.2021.3098688>.