An Intelligent Decision Support Ensemble Voting Model for Coronary Artery Disease Prediction in Smart Healthcare Monitoring Environments

ANAS MAACH[1]
LASTIMI, High School of Technology
Mohammed V University in Rabat
Sale, Morocco

JAMILA ELALAMI[2]
LASTIMI, High School of Technology
Mohammed V University in Rabat
Sale, Morocco

NOUREDDINE ELALAMI[3]
LASTIMI, Mohammedia School of Engineers
Mohammed V University in Rabat
Rabat, Morocco

EL HOUSSINE EL MAZOUDI[4]
CISIEV. Faculty of Science and Technology in Marrakech
Cadi Ayyad University
Marrakech, Morocco

Abstract—Coronary Artery Disease (CAD) is one of the most common cardiac diseases worldwide and causes disability and economic burden. It is the world's leading and most serious cause of mortality, with approximately 80% of deaths reported in low- and middle-income countries. The preferred and most precise diagnostic tool for CAD is angiography, but it is invasive, expensive, and technically demanding. However, the research community is increasingly interested in the computer-aided diagnosis of CAD via the utilization of machine learning (ML) methods. The purpose of this work is to present an e-diagnosis tool based on ML algorithms that can be used in a smart healthcare monitoring system. We applied the most accurate machine learning methods that have shown superior results in the literature to different medical datasets such as RandomForest, XGboost, MultilayerPerceptron, J48, AdaBoost, NaiveBayes, LogitBoost, KNN. Every single classifier can be efficient on a different dataset. Thus, an ensemble model using majority voting was designed to take advantage of the well-performed single classifiers, Ensemble learning aims to combine the forecasts of multiple individual classifiers to achieve higher performance than individual classifiers in terms of precision, specificity, sensitivity, and accuracy; furthermore, we have benchmarked our proposed model with the most efficient and well-known ensemble models, such as Bagging, Stacking methods based on the cross-validation technique, The experimental results confirm that the ensemble majority voting approach based on the top three classifiers: MultilayerPerceptron, RandomForest, and AdaBoost, achieves the highest accuracy of 88,12% and outperforms all other classifiers. This study demonstrates that the majority voting ensemble approach proposed above is the most accurate machine learning classification approach for the prediction and detection of coronary artery disease.

Keywords—*Machine learning; smart healthcare; coronary artery disease*

I. INTRODUCTION

No cure exists for Coronary Artery Disease (CAD), as a combination of environmental and inherited factors is thought to be associated with several risk factors, including a family history of heart disease, age, overweight, inactivity, poor diet, and tobacco usage. The diagnosis of coronary artery disease is very challenging for the General Physician (GP). When a patient experiences chest pain, he consults the GP. The chest pain is the main reason for consultation in approximately 4% of cases and in only 15% of cases [1], Coronary Artery Disease (CAD) ultimately will be diagnosed as the reason for the symptoms, The difficulty for the GP is to identify CAD on the basis of symptoms, age, and gender. Distinguishing a life-threatening disease from a non-life-threatening disease is crucial for the effective prevention and management of the disease, but it can be difficult, especially in cases of atypical blood pressure or non-specific chest complaints [1], [2]. Currently, approximately 105,000 people 53% of whom are women are being recommended to a cardiologist, each year in the Netherlands. In fact, only 5% of males and 1% of females have coronary disease needing invasive therapy. Therefore, a clinical need exists in the population suffering from chest pain for optimizing the diagnosis and orientation to the cardiologist [3]. Thus, the development of an accurate diagnostic tool based on ML algorithms would assist GPs in identifying the likelihood of coronary artery disease and in guiding the management of patients. In addition, early diagnosis of chronic diseases saves the expense of medical care and reduces the likelihood of more complex health problems, especially considering the lack of doctors in underserved areas and developing countries. In this case, the association of Wireless Body Area Networks (WBANs) and machine learning methods should be used to assist practitioners in the early diagnosis and identification of CAD by offering predictive models for better and faster decision support. Nevertheless, it is worth noting that machine learning tends to be looked upon with suspicion by some due to what can be termed a "black box" [4]: being unable to reveal its inner decision-making mechanism. However, this inability to explain its inner decision-making tends to lead to skepticism among consumers and slow adoption by end-users in the domain of health care. It is crucial to build trust, especially in healthcare, where errors can be fatal, to be able to convey both the underlying reasoning and the process required to obtain a machine learning prediction. This paper aims to develop a CAD detection, classification, and prediction tool that can be integrated into a smart healthcare system. Through the use of ensemble machine learning approaches combining the best

classifiers Multilayer Perceptron, Adaboost and RandomForest, such a system would be capable of predicting whether a person is likely to have CAD on the basis of various relevant indicators, supplying physicians with an advance diagnostic assessment. The classification models were evaluated using various metrics, namely, F-measure, accuracy, recall, precision, sensitivity, and the ROC curve (receiver operating characteristic curve) to select the most efficient classifier. Several relevant features that can potentially be utilized to predict coronary artery disease have been taken from the best classifier scheme. The Z-Alizadeh Sani dataset [5] is used for the purposes of this work. The Z-Alizadeh Sani dataset comprises 303 records of patients, with 55 features. All the features can be regarded as CAD indicators, as stated in the literature [5]. Features are categorized into four categories: demographics, symptoms, laboratory, and ECG features. Accordingly, every patient may be classified in two different possible classes: CAD or normal. The patient is classified as having CAD if his or her narrowing diameter is greater than 50% and if not, he is normal.

A. An Overview of the Smart Healthcare Monitoring System

A Smart Healthcare Monitoring system based on Wireless Body Area Networks (WBANs) is organized into a three-layer telemedicine system as shown in Fig. 1. This system consists of a network of wireless sensors that continuously track the health parameters of patients [6], [7], Furthermore, this healthcare monitoring system is interconnected to the high-level biomedical server via an internet-based network system.

1) Tier 1: is also called the WBAN level, in which, each patient under healthcare monitoring is connected to several small Body Sensor Networks (BSNs), These inhomogeneous sensors are placed either in the body or in wearable devices As shown in Fig. 1, the BSN detects different physiological body parameters. These include electroencephalogram (EEG), pulse rate, electromyography (EMG), blood pressure, electrocardiogram (ECG), and so forth [6], [8], [7]. To communicate within the WBAN level, those BSNs are using radio waves to communicate with themselves and with the coordinator, A sink node is operating as a hub for all the BSNs .

2) Tier 2 : The second level is implemented in a PC/laptop, mobile phone, or PDA. The data collected from the BSNs in various formats, including graphics, digital, audio and so forth [6], [8], are transferred to a healthcare server. It employs several technologies, such as 4G/5G or WiFi, to communicate with a remotely located medical server.

3) Tier 3: It consists of a large network of various devices, services, healthcare practitioners, and healthcare services providers that are interconnected. This layer delivers many services to potentially thousands of clients through the use of healthcare systems as a centralized point of contact. These medical servers store the health data of patients and deliver various additional services to these patients and other related stakeholders [6], [7]. The tasks of the health server involve authentication of patients, acceptance, and submission of their medical data, and formatting and analysis of the data to identify the severity of health problems. If the analyzed data shows that the patient's potential medical condition is of a life-threatening type, the health server alerts emergency caregivers. Patients and their doctors can access the analyzed data at their location via the Internet. Patient data is reviewed by doctors to assess whether it is in accordance with the desired healthy ranges (e.g. pulse pressure, heart rate, etc.) and whether the given or prescribed medical treatment is working. The rest of the article is organized in the following sections: Section 2 provides a review of the state-of-the-art literature on coronary heart disease and heart disease research, while Section 3 presents the proposed methodology, Section 4 reports the results of the experiments and the discussion, The limitations of this article and future work are described in Section 5, and Section 6 provides a conclusion to the article.

II. RELATED WORKS

Up to now, various research studies have been carried out on the early diagnosis of coronary artery disease and heart disease. They have utilized various machine learning prediction approaches and achieved remarkable performance. This section provides an extensive literature review of research studies in the field of heart disease diagnosis supported by machine learning techniques: In [9] Yadav et al. presented a novel method for ensemble machine learning utilizing Pearson correlation and chi-square feature selection-based algorithms for the correlation strength of heart disease attributes and the Random Forest ensemble method for the diagnosis of heart disease. The authors performed experiments with their proposed system on the CHDD dataset, and they were able to achieve the best performance considering many evaluation metrics Correctly Classified Instances, Mean absolute error, Incorrectly Classified Instances, Kappa statistic, Root relative squared error, Relative absolute error, and root mean squared error, the Random Forest ensemble method outperforms various machine learning techniques RF, AdaBoostM1, Gradient Boosting. In [10] Li et al. proposed a high-performance and intelligent approach for detecting cardiac diseases, and the model is based on a feature selection method (FCMIM) with a support vector machine classifier (SVM). They conducted experiments with their proposed method on the CHDD dataset and were able to achieve the best performance considering many evaluation parameters: accuracy, MCC specificity, processing time, and sensitivity against various machine learning techniques SVM, LR, ANN, kNN, NB, and DT. In [11] Javeed et al. introduced a new heart failure prediction diagnostic method using a random search algorithm (RSA), which is applied for feature selection, and a random forest model to perform classification and prediction. They carried out experiments with their proposed system on the CHDD dataset and were able to achieve the best accuracy, sensitivity, specificity, and MCC. The proposed method outperforms various machine learning techniques, including the random tree model, the Adaboost model, SVM with a linear kernel function, the additional tree ensemble model, and the support vector machine (RBF kernel). In [12] Saxena et al. presented an innovative automated system that combines Generalized Discriminant Analysis (GDA) as an effective feature reduction algorithm together with a Radial Basis Function (RBF) kernel and Online Sequential Extreme Learning Machine (OSELM) based on a Sigmoid activation function, Hardlim, RBF and Sine as a binary classifier for the detection of congestive heart failure (CHF) and coronary artery disease (CAD). They performed experiments with their proposed system on the NSR-CAD, NSR-CHF, and CAD-CHF datasets and were able to obtain the best results in terms of
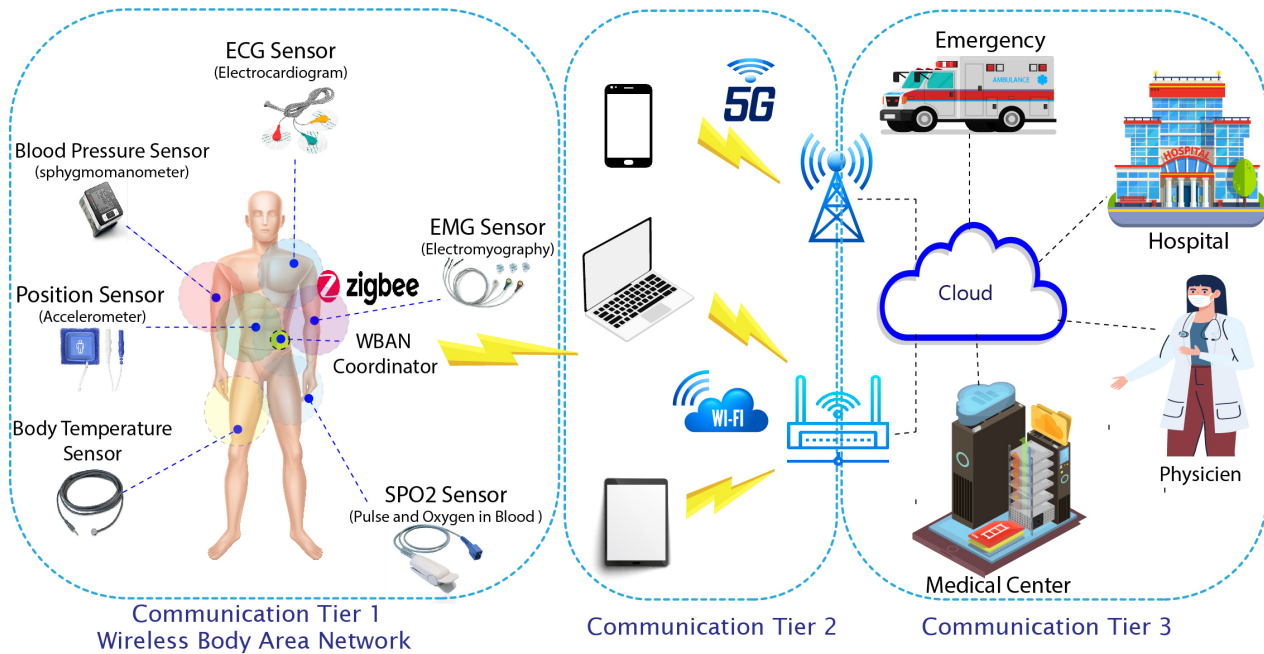
Fig. 1. General Architecture of a Smart Healthcare Monitoring System.

accuracy, sensitivity, specificity, mean ± SD and p-value. The proposed method outperforms the different machine learning techniques GDA-Kernel Function (Gaussian, Polynomial, and RBF) and OS-ELM (RBF activation, Sigmoid, Hardlim, and Sine). In [13] Dwivedi et al. proposed an approach for accurately diagnosing heart disease using the logistic regression method. They conducted experiments with their proposed technique on the StatLog heart disease dataset and were able to achieve the best performance considering many evaluation parameters: classification accuracy, precision, F1-measure, false positive rate (FPR), sensitivity, specificity, negative predictive value (NPV), misclassification rate (MRR), compared to five different data mining techniques: ANN, SVM, kNN, CT, and NB. In [14] Gupta et al. presented an intelligent decision-support model that can help medical experts in predicting heart disease through an advanced ensemble classifier. They conducted experiments with their proposed system on the CHDD dataset and were able to achieve the best performance considering many evaluation parameters: classification accuracy, specificity, F-measure, recall accuracy, MAE, ROC, and RMSE against various machine learning techniques: MLP, NB, J48, RF, SVM, AB, boosted tree, and binary discriminant. In [15] Verma et al. presented a new hybrid method for the diagnosis of CHD, including the identification of risk factors using correlation-based feature subset selection (CFS) with particle optimization search (PSO) and K-means clustering approaches. They conducted experiments with their proposed system on the CHDD dataset and they were able to obtain the best model performance against five different machine learning techniques: MLP, MLG, FURI, and DT (C4.5). In [16] Miao et al. proposed an improved ensemble machine learning scheme using an adaptive boosting algorithm for accurately diagnosing Coronary Artery Disease (CAD). They have performed experiments with their proposed method on

CHDD, HHDD, LBMC, and SUH datasets and they were able to achieve the best performance considering many evaluation parameters: accuracy, ROC, classification error, precision, sensitivity (or recall), F-score, K-S measure, specificity, and AUC, and to outperform other machine learning techniques. In [17] Long et al. proposed a heart disease diagnostic system using rough set-based feature reduction and type 2 fuzzy logic systems (IT2FLS) for early stage heart disease detection, in which the authors implemented the BPSORS-AR Binary Particle Swarm Optimization and the rough set-based feature selection technique. They conducted experiments with their proposed model on the heart disease dataset and the SPECTF dataset, and they were able to achieve the best performance compared to the different data mining techniques, NB, SVM, and ANN. In [18] Nilashi et al. proposed a new methodology for heart disease diagnosis using machine learning algorithms. Such a model was built with unsupervised and supervised machine learning methods. using the implementation of Fuzzy Logic and the Support Vector Machine (SVM)-based ensemble model, and Principal Component Analysis (PCA) was employed with two processes for imputation. Both imputation methods were essentially used for missing value imputation. In addition, they have implemented the Augmented FSVM and Augmented PCA for augmented learning of the data. This was done to reduce the computational time. This was associated with the prediction of the disease. According to the results, it was deduced that the ensemble model showed high accuracy in classifying heart disease and also decreased the computational time required for disease diagnosis. From this brief state of the art, it can be deduced that there is a great interest in the scientific community for the prediction and detection of heart diseases, but in reality, machine learning tools are not yet widely applied in diagnostic systems for heart diseases, especially in developing countries, where the mortality rate

from heart disease is very high. This is because many proposed schemes are too complex to be implemented in a smart health-care monitoring system for heart disease; hence, there is still space for improvement. In this work, a new approach based on a majority voting ensemble model that combines the prediction of three classifiers (Adaboost, Multilayer Perceptron, Random Forest) is proposed. Unlike other approaches, this approach is simple to implement and gives excellent results in the detection and prediction of coronary artery disease.

III. PROPOSED METHODOLOGY

In Section 3, we provide a description of the proposed methodology and also explain that the proposed approach is a process defined by the following steps, as shown in Fig. 3.

A. Dataset Pre-Processing

1) Dataset Description: A variety of experiments were performed utilizing the Z-Alizadeh Sani dataset. Originally, this dataset was provided by the Shaheed Rajaei Cardio-vascular Medical and Research Center. It was constructed from the records of 303 random visitors, The reason for choosing this dataset is that it includes clinical and non-clinical features that can be gathered remotely via WBANS, such as ECG features (EF-TTE, RWMA, Q-wave, and T-inversion), in contrast to other datasets and which have a significant impact on the prediction of coronary artery disease. The predictor variables are Age, Diabetes Milletus(DM), Hypertension(HTN), Blood Pressure(BP), Typical Chest Pain, Atypical, Nonanginal, T-inversion, Fasting Blood Sugar(FBS), Erythrocyte Sed rate(ESR), Potassium(K), Ejection Fraction(EF-TTE), Regional Abnormality(Region RWMA), as shown in Table I The dataset consists of 303 instances, divided into 216 CAD instances and 87 healthy instances The target variable identifies whether a person has CAD, represented by 1, or not, represented by 0 The description of the parameters of each attribute in the dataset, including the mean, the median, the maximum and minimum, and the value of the standard deviation, is presented in Table II.

2) Data Cleaning : Data cleaning is the following phase of the machine learning process. It is regarded as a key step in the workflow process of our approach because it either builds the model or breaks it. Different aspects of data cleaning need to be considered:

- Noise, Duplicates, Invalid or missing data
- Normalization
- Filter unwanted outliers
- Deal with imbalanced datasets.

a) Dealing with Outliers : It is essential to identify faulty measurements (outliers) that are divergent from other measurements and to detect sensor faults in emergency situations in order to minimize false alarms. Anomalous measurements should be excluded in order to minimize unnecessary false alarms and interventions by health professionals. As shown in Fig. 7 and Fig. 6, there are abnormal measurements (outliers) in the Fasting Blood Sugar (FBS), Erythrocyte Sed Rate (ESR), and Potassium (K), or extreme values in the Fasting Blood Sugar (FBS). Therefore, to extract the outliers and

extreme values, we have proceeded to apply the interquartile range filter. The Fig. 2 shows the number of outliers and extreme values found in the dataset. In order to solve the problem of outlier values and extreme values, and since there are not many instances in the dataset (only 303), we proceeded to standardize the features.
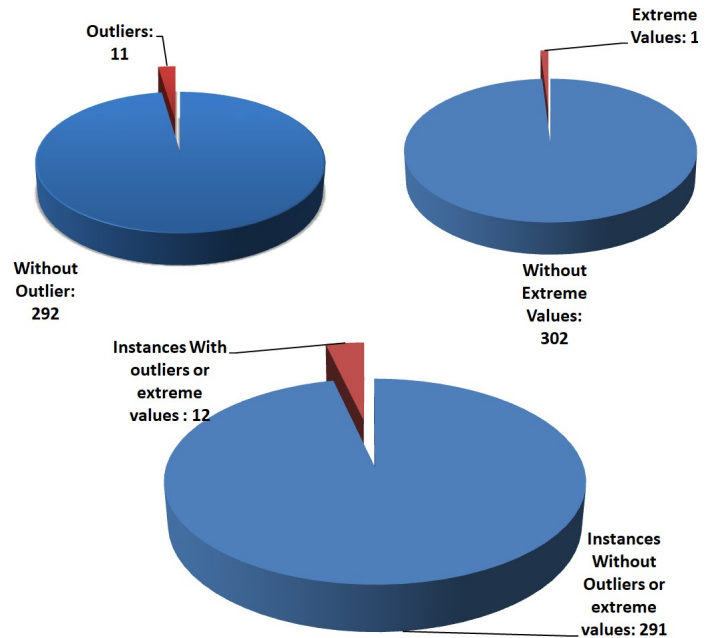

Fig. 2. Outliers and Extremes Values Percentages in the Data.

b) Dealing with Imbalanced Dataset : The unbalanced class instances in the health dataset are a critical issue. In fact, the dataset that was employed in our classification experiment was unbalanced since the instances of the first class exceeded the instances of the second class by a significant ratio, which means that the instances are not adequately distributed among the different classes. Therefore, the results of the classification from unbalanced class data produce a biased outcome in favor of the dominant class as shown in Fig. 4. In order to balance the unbalanced dataset, there exist two main techniques, namely, oversampling and undersampling. In the Z-Alizadeh Sani dataset used for this work, positive instances exceeded negative ones, which were solved using the SMOTE method as shown in Fig. 5, Out of the variety of oversampling techniques that exist, SMOTE has demonstrated tremendous potential [19] and is thus widely used by scientists in the medical research community. SMOTE is a technique of oversampling proposed to prevent the issue of class imbalance in the dataset. It improves the classifier's performance and joins the lesser class points to the line segments with the unreal points positioned on these lines. With SMOTE, newly created instances are generated by synthetically resampling the minor class data points, as has been performed in the conventional oversampling method [20], [21] This varies from the conventional approach by the fact that it is carried out in the space of features rather than data space, by regard to the instance of the smaller class at its nearest vector [20], [21] The newly created synthetic data parameters may be generated by applying two distinct approaches, One approach employs the ratio of oversampling, whereas another approach employs the k-nearest neighbors.
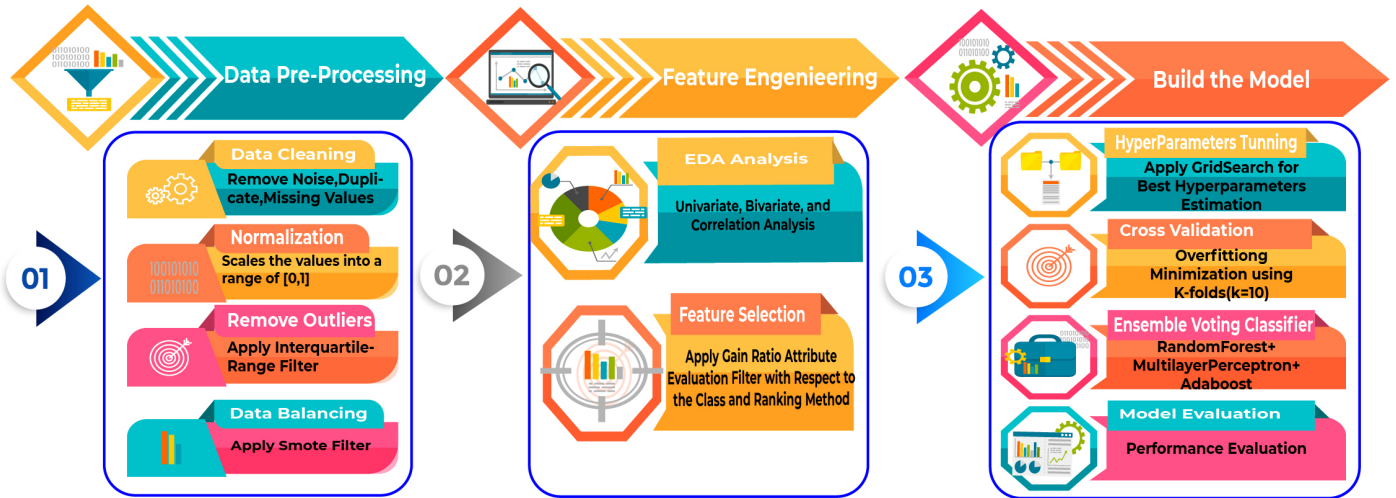
Fig. 3. The Proposed Approach.

This means that SMOTE generates the synthetic data points for the minority class [19] in order to shift the bias of the classifier's learning from the dominant class to the minor class.
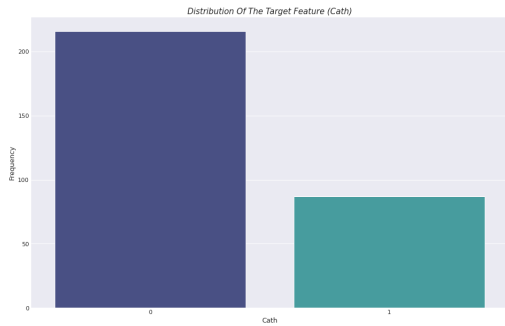

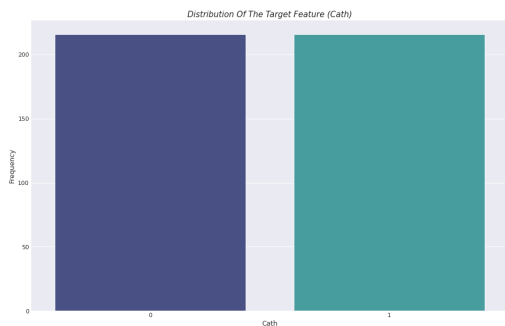Fig. 4. Distribution of Classes before Applying the Smote Technique.

Fig. 5. Distribution of Classes after Applying the Smote Technique.

3) Features Selection: In this paper, we employ the Gain Ratio method [22] to identify the most pertinent and useful features. The gain ratio method allows us to check the closeness of features by different methods. The gain ratio provides one of these techniques. It identifies the pertinence of every feature and selects the attributes that have the maximum gain ratio with regard to the likelihood of each feature value. The chosen test must acquire a large gain of information, which

should be inclusive of or larger than the average of the gains of the assessed tests, with the aim of penalizing the spread of the nodes, and must be large when the data is uniformly distributed as well as small if the data belongs to a single branch. Each attribute's Gain Ratio is computed according to the formula:

$$GainRatio(Attr) = \frac{IG(Attr)}{H(attr)} \qquad (1)$$

where

$$H(attr) = \sum -P(val_i)log_2 P(val_i) \qquad (2)$$

and $P(val_i)$ is defined as the probability of having the value val_i as a factor of t global values for a given attribute i

The dataset used contained 55 features; we applied the gain ratio algorithm with various thresholds regarding the number of most relevant attributes that should be utilized in these experiments, and in fact, by using the 12 features, we found the greatest accuracies.

B. Exploratory Data Analysis

The following section presents a statistical overview of the CAD dataset, described in the Table II. Pair plots provide a simple mechanism to examine how two attributes correlate with each other. Every variable from the dataset is presented in a correlation matrix, which can be immediately visualized. It also provides an effective way to determine the appropriate classification method that should be conducted, Fig. 10 also illustrates the feature distribution in the Coronary Artery Disease dataset, providing a useful representation of the distribution of attributes in the dataset. Fig. 10 represents the plot of all attributes in the dataset (12 attributes). One can observe that three of the attributes, specifically age, K, and BP, are normally distributed. In addition, the dominating value of Tinversion, DM, Atypical, Nonanginal, and RegionRWMA is 0, whereas for the attributes HTN and TypicalChestPain, the

most frequently occurring values are 1 and the least occurring value is 3. Furthermore, Fig. 10 depicts that there were six categorical and six numerical attributes.

1) Analysis of the Correlation Heatmap: The Correlation Heatmap is presented in Fig. 8, it is defined as a graphical plot of a cross-correlation matrix that reveals the interrelationship of various attributes Within the -1 to 1 range, the coefficient of correlation can be given any value, If there is a direct linear relationship between two variables, this relationship is statistically termed a correlation, One can also describe this as a correlation measurement involving two variables, The aim in this case is to identify a correlation between multiple variables and to arrange the results, In this context, a matricial structure of data has been used to store the information, The Fig. 8 shows the correlation feature by feature. In the Fig. 8 we can see various elements of information, First of all, the three features that present the strongest dependency between class and feature are Typical ChestPain, Atypical, Age, RegionR-WMA , HTN, Nonanginal and EF-TTE with corresponding correlations of -0.54, 0.42, -0.36, -0.32, -0.29, 0.29, 0.27, and 0.23, respectively. The next fact points out the correlation between two features in HTN–BP, DM–FBS, Atypical–Typical Chest Pain, and RegionRWMA-EF-TTE with corresponding correlations of 0.57, 0.68, -0.72, -0.45, respectively, whereas FBS, K, ESR, BP, and DM have the weakest correlation with the class.

2) Analysis of The Scatter Plot Matrix: The scatter plot matrix shown in Fig. 9 is useful for finding pairwise relationships of features From this, we can deduce the relationships between the features in advance: The more scattered the points, the weaker the relationship, and the more clustered they are, the stronger the relationship Referring to the scatter plot matrix As shown in Fig. 9 we deduce that there is a relationship between the selected features, such as between DM and FBS, between BP and HTN, and Age.

C. Methods

In this paper 10, high-level classifiers that showed superior performance in detecting cardiac diseases were selected based on the literature and two ensembles of voting classifiers that we designed by combining a set of three high-level classifiers, the novelty is that this combination, to our knowledge, has never been done before in the literature the Ensemble voting classifiers are based on the majority voting method for predicting coronary artery disease, the parameters of the Random forest MultilayerPerceptron and Adaboost classifiers have been optimized using the Grid search and CVParameterSelection hyperparameter techniques and eventually, 10 folds cross-validation technique have been utilized to validate the models the description of the three classifiers that compose our model is presented below:

1) Adaboost Classifier: Using the ADAboost classifier is a well-known boosting method. This classifier aids in the consolidation of several weak classifiers into a single effective classifier. Initially, a classifier is fitted on the initial dataset, and then repeated duplicates from the classifier are fitted to the similar dataset, with the weights of erroneously categorized instances modified such that later classifiers concentrate more on challenging situations.

2) Random Forest (RF): Researchers are paying more and more attention to Random Forest, it is an advanced machine learning scheme that demonstrates the overall ensemble learning abilities and ease of use, Both regression and the creation of random subsets require the RD approach Classification is the principal application of the concept of "bagging" which boosts accuracy rates by mixing learning models, numerous decision trees, of which each is employed in the RF method, make up an ensemble classifier. Since every decision tree is built separately, subsequent trees are intended to be independent of preceding trees [23], each tree in the forest is created to depend on a random vector's values selected separately using a boot-strapped data set sample, and all the forest trees use the same distribution. In the RF-produced model, random sampling with substitutions is implemented [24]. A random subgroup of the entire set of predictors is used to create the best classifier for each node [25]. The fact that RF uses more computing resources—such as storage spaces—than other algorithms is one of the key shortcomings [26] it addresses, but because of its outstanding prediction accuracy, overfitting avoidance, and scalability it is favored by many researchers.

3) Multi-Layer Perceptron: Instead of learning by observation, supervised learning procedures use "learning by example". To build a learning model, a trained data set has been produced. The learning model is used to test the current input, and predictions, are made. The MLP approach allows for the training of a back propagation-based multilayer feed-forward neural network, which calculates the associated network weights based on the intended outcomes and training patterns. MLP belongs to the class of supervised neural networks that iteratively learn a set of weights for categorical variable prediction [27]. An MLP network's components are represented by the layers in Fig. 11 input and output layers and several hidden layers [28] The three parameters of the MLP network may be altered depending on the kind and amount of data. The best prediction should be found by optimizing the parameter's momentum, learning rate and the number of hidden layers. The learning rate is a measure of how quickly the network is being trained. In other words, when learning rates rise, networks train more quickly but at the risk of creating networks that are unstable. By balancing the network [29], the momentum avoids potential issues that might arise from choosing a fast learning rate that renders instability on the network [30]. Various objective functions and characteristics of input are represented by including the hidden layers.

IV. Experiment and Results

A. Method of Validation of the Models

This research paper utilized the cross-validation technique with ten folds and four performance evaluation measures More detail is provided in the subsections below:

1) Cross-Validation (CV): In the present study, a 10-fold cross-validation method [31] is employed to validate the classification model. Aiming to minimize the bias related to selecting random sets from the training data samples while making a comparison of the predictive accuracies of at least two different methods, a k-fold crossvalidation technique was used. In the k-fold cross-validation technique the training dataset S is partitioned randomly into k mutual subsets folds
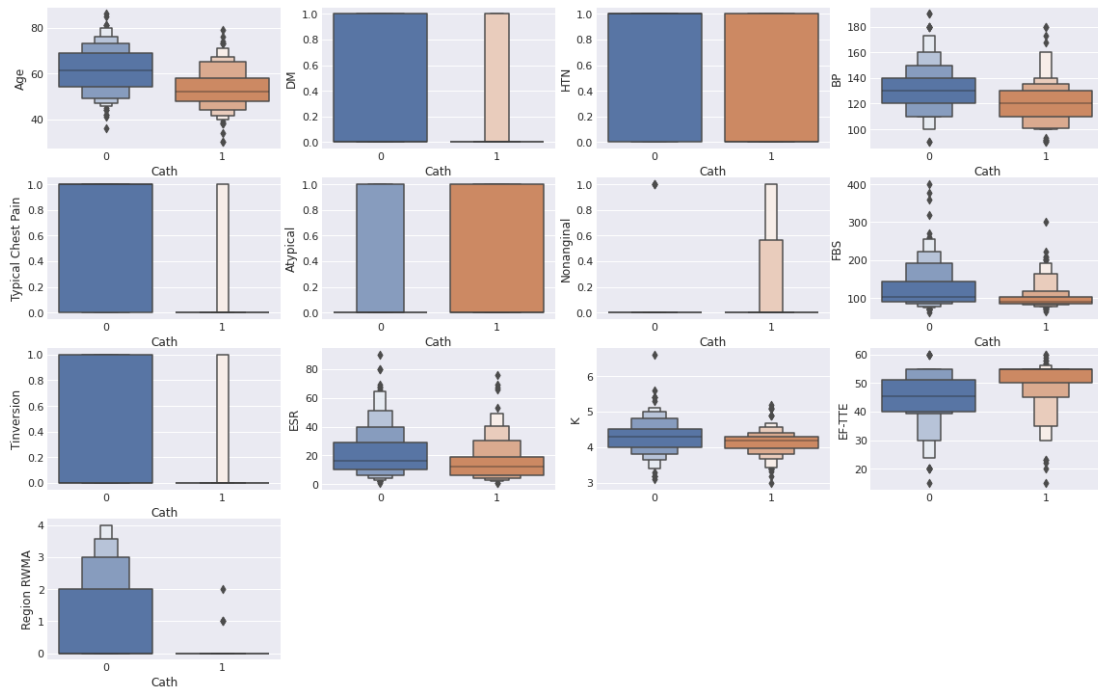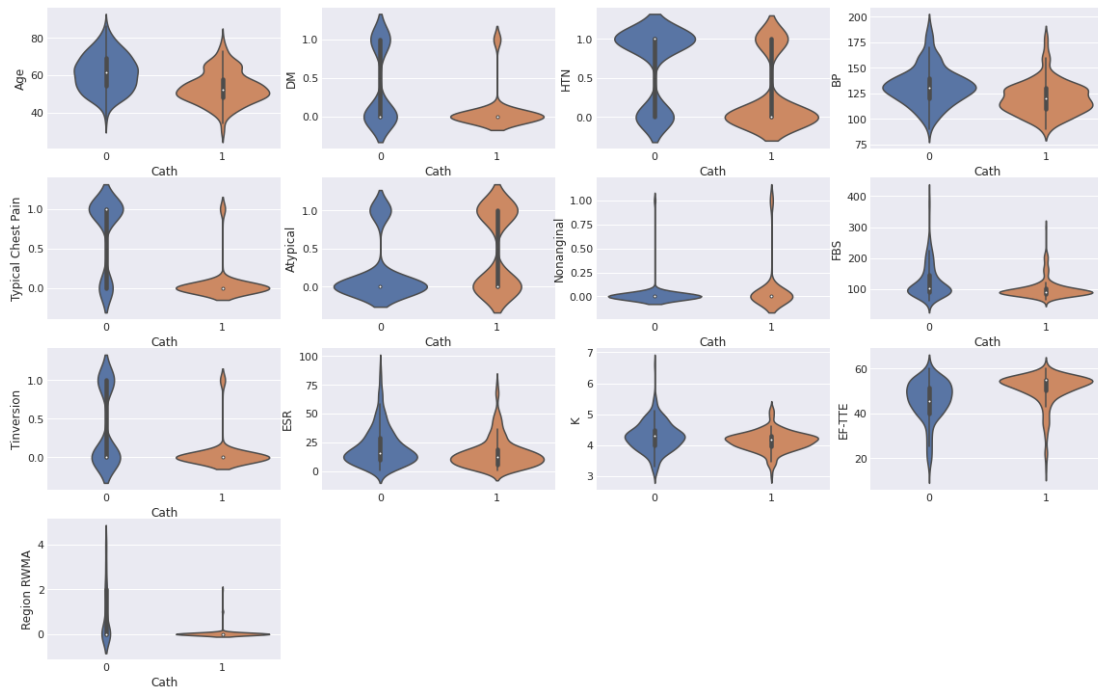
Fig. 6. The Box Plot of the Features.

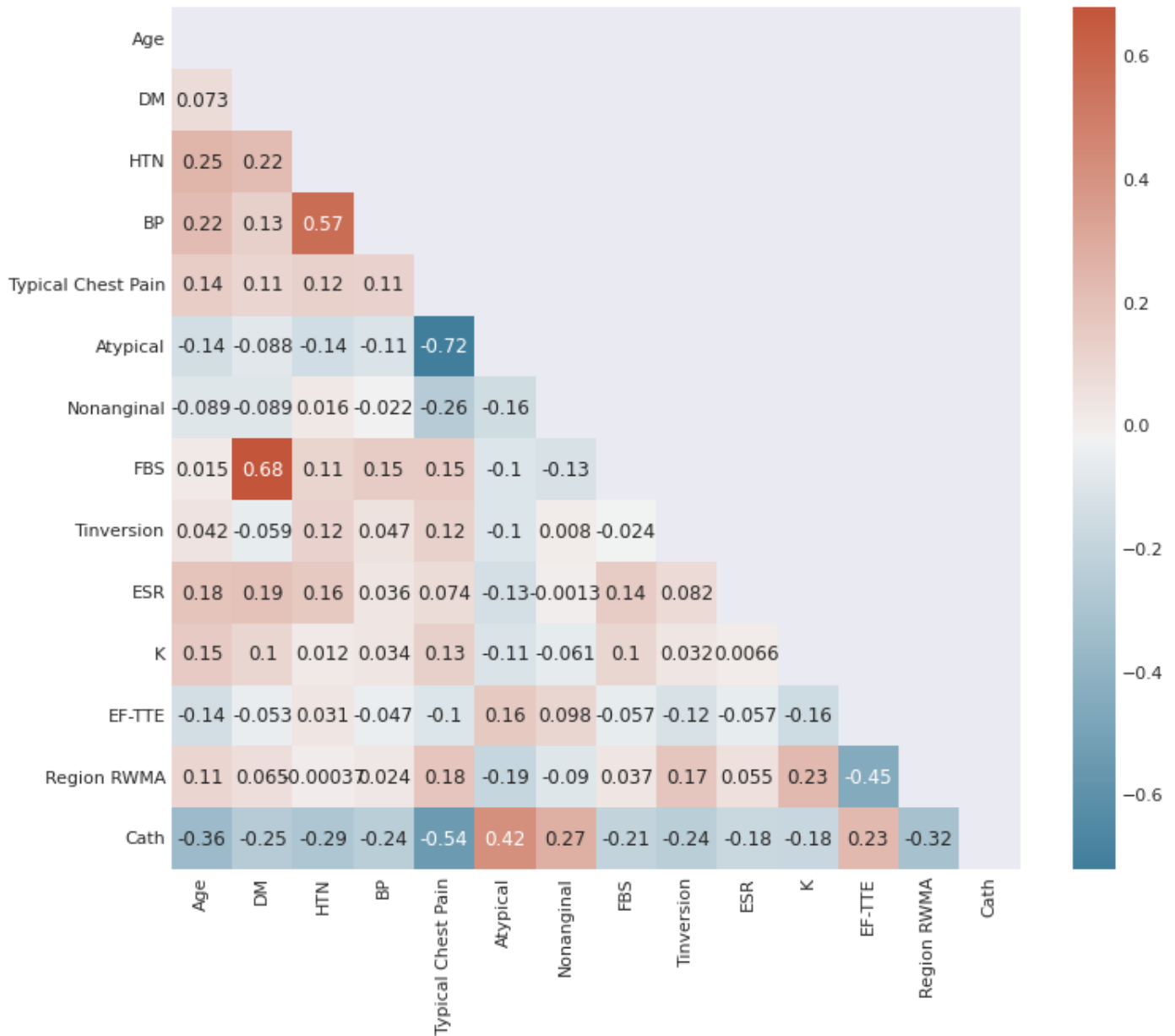Fig. 7. The Violin Plot of the Features.

Fig. 8. Correlation Heatmap.

$Sa_1, Sa_2, ..., Sa_k$ of roughly the same sizes The estimator will be trained k folds and then tested every time $\eta 1, 2, 3...k$ it is trained on Sa_t and then tested again on Sa_t The accuracy of the cross validation technique is calculated as the number of classifications that are correct subdivided by the total number of records in the dataset, Thus formally we can state that Sa_i is the test dataset containing the instance $m_i=(r_i, p_i)$ and therefore the accuracy of the cross validation is

$$accuracy_{cv} = \frac{1}{n} \sum_{(r_i, m_i) \in S} \sigma(I(Sa\xi_{(i)}, r_{je}), p_i) \quad (3)$$

2) Confusion Matrix: The Confusion Matrix typically assesses the outcome of the classification model for a given request. This summarizes the count values of the correct and incorrect hypotheses by effective class. Table III illustrates the confusion matrix. For the purpose of this study, the negative class is the 0 class and the positive class is the 1 class. With True Positive (TP) showing the positives that are correctly classified and True Negative (TN) showing the negatives that are correctly classified, False Positive (FP) shows misclassified instances that are positive, and False Negative (FN) represents misclassified negative instances, respectively.

3) Accuracy: The accuracy of the model is the proportion of correctly classified prediction points divided by the number of total predictions evaluated, as follows:

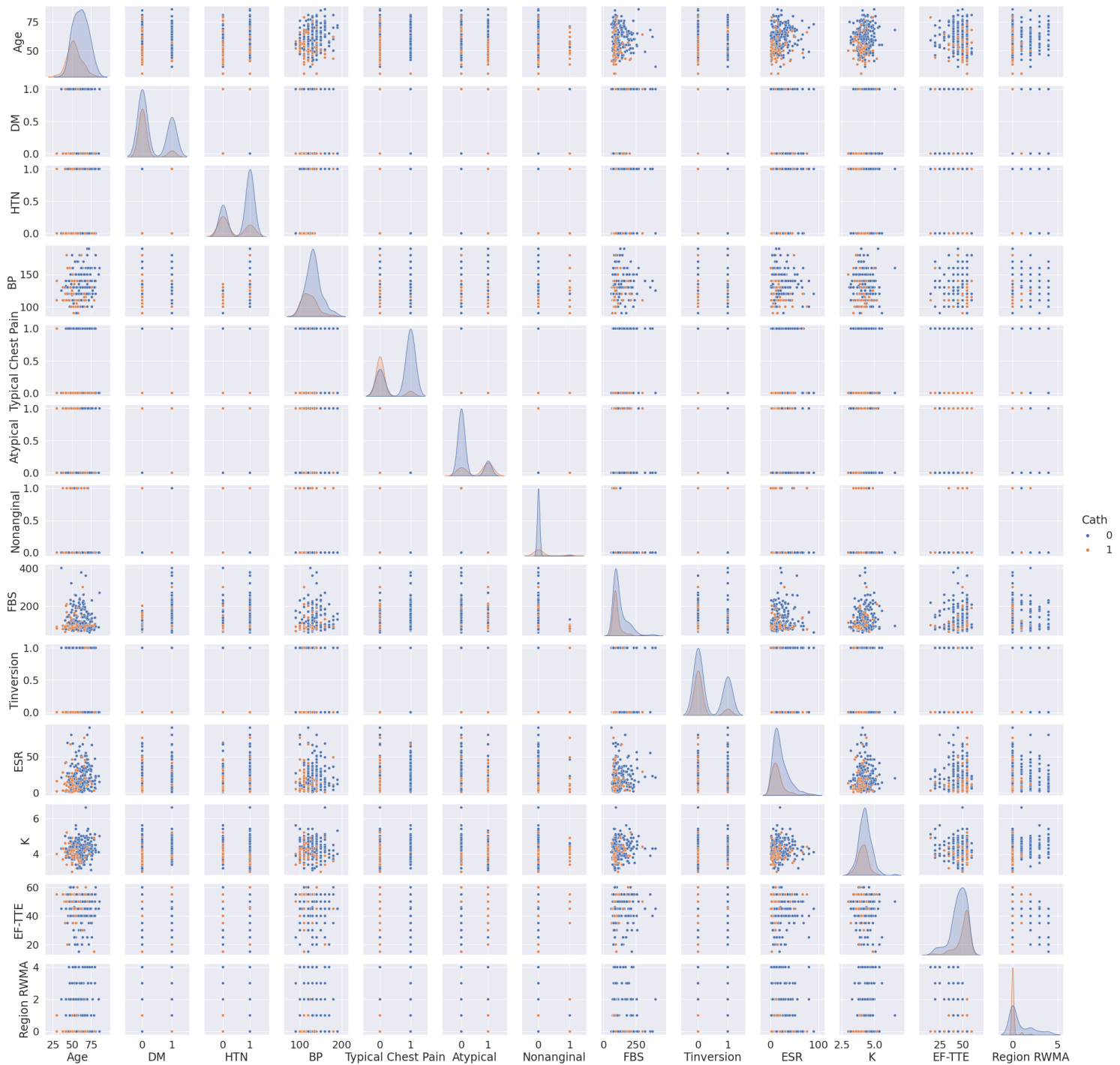$$Accuracy = \frac{(TP + TN)}{(TP + FN + FP + TN)} \quad (4)$$

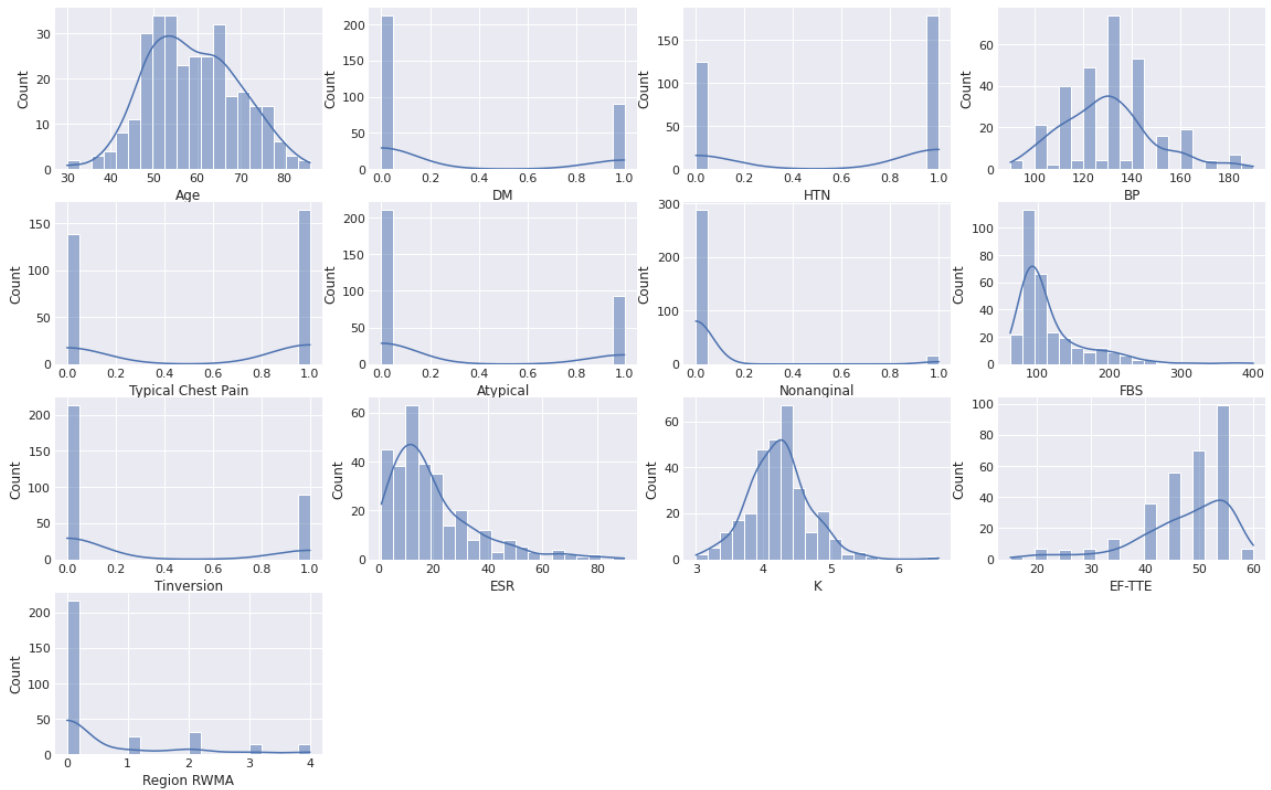Fig. 9. The ScatterPlot Matrix of the Features.

Fig. 10. Histogram of each Feature.

TABLE I. RANGE VALUE OF SELECTED FEATURES FROM THE Z-ALIZADEH SANI DATASET

Feature Type	Attribute	Values
Demographic	Age	30–86
	Diabetes Milletus(DM)	Y,N
	Hypertension(HTN)	Y,N
Symptom and examination	Blood Pressure(BP)	90–190
	Typical Chest Pain	Y,N
	Atypical	Y,N
	Nonanginal	Y,N
ECG	T inversion	0,1
Laboratory tests	Fasting Blood Sugar(FBS)	62–100 mg/dl
	Erythrocyte Sed rate(ESR)	1–90 mm/h
	Potassium(K)	3.0–6.6 mEq/lit
	Ejection Fraction(EF-TTE)	15–60%
	Regional Abnormality(Region RWMA)	0,1,2,3,4

4) Sensitivity: This is calculated by dividing the ratio of the number of coronary patients diagnosed as true positives by the total number of patients with coronary artery disease It, or the true positive rate, is also called recall It is assessed as following:

$$Recall = \frac{(TP)}{(TP + FN)} \quad (5)$$

5) Specificity: The specificity, or "True Negative" TN rate, is the percentage of reported diseases that are correctly diagnosed. It is assessed as follows:

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

B. Results of the Machine Learning Algorithms

We implemented a variety of models and used the cross-validation technique with 10 folds, in order to select the best performing models, the more accurate models are employed in the voting ensemble, and the resulting accuracies of the models are shown in the Table IV and as it is represented

TABLE II. Statistical Summary of Selected Features from Z-Alizadeh Sani Dataset

Feature	Count	Mean	Std	Min	25%	50%	75%	Max
Age	303.0	58.897690	10.392278	30.0	51.0	58.0	66.0	86.0
DM	303.0	0.297030	0.457706	0.0	0.0	0.0	1.0	1.0
HTN	303.0	0.590759	0.492507	0.0	0.0	1.0	1.0	1.0
BP	303.0	129.554455	18.938105	90.0	120.0	130.0	140.0	190.0
Typical Chest Pain	303.0	0.541254	0.499120	0.0	0.0	1.0	1.0	1.0
Atypical	303.0	0.306931	0.461983	0.0	0.0	0.0	1.0	1.0
Nonanginal	303.0	0.052805	0.224015	0.0	0.0	0.0	0.0	1.0
FBS	303.0	119.184818	52.079653	62.0	88.5	98.0	130.0	400.0
Tinversion	303.0	0.297030	0.457706	0.0	0.0	0.0	1.0	1.0
ESR	303.0	19.462046	15.936475	1.0	9.0	15.0	26.0	90.0
K	303.0	4.230693	0.458202	3.0	3.9	4.2	4.5	6.6
EF-TTE	303.0	47.231023	8.927194	15.0	45.0	50.0	55.0	60.0
Region RWMA	303.0	0.620462	1.132531	0.0	0.0	0.0	1.0	4.0
Cath	303.0	0.287129	0.453171	0.0	0.0	0.0	1.0	1.0
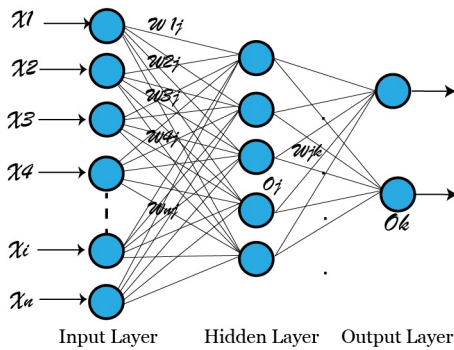
Fig. 11. The Multilayer Feed Forward Neural Network.

TABLE III. Confusion Matrix

	CAD	NORMAL
Actual CAD	TP	FP
Actual Normal	FN	TN

graphically in Fig. 13, the best-performing machine learning classifiers are the RandomForest, Multilayer Perceptron, Stacking, Bagging and Adaboost. In addition, different ensembles were constructed and tested by combining these classifiers, as shown in Fig. 13 and detailed in Table IV, We evaluated the classifiers in terms of accuracy, sensitivity, specificity, F-measure, and Matthew's correlation coefficient (MCC) to measure performance. As shown in Table IV, the ensemble voting classifier has the greatest classification accuracy of 88.12% compared to the other classifiers. Taking into account the other factors, the voting classifier has the greatest F-measure and MCC with values of 88.12 and 73.4, respectively, as illustrated in Fig. 14. The ensemble voting classifier has the best precision of 89.4% and the best recall of 88.1%, while the Multilayer Perceptron has the second-best precision of 87.79%. Once again, the voting classifiers have the best ROC and the precision values of 93.2/% and 89.4/% respectively, as shown in Fig. 14. The diagnostic ability of the classifier is shown in Fig. 12 by the calculated and presented ROC curves. The better the diagnostic ability of the model, the closer the ROC curve area value is to one.

V. Limitations and Future Work

The Z-Alizadeh Sani dataset contains the records of 303 patients from a nearby population of the Department of Cardiovascular Imaging, Rajaei Cardiovascular Medical Research Center, University of Iran, Tehran, Iran. Some limitations of the Z-Alizadeh Sani dataset are that patients under 30 years of age are not presented, as well as people from developing or low-income countries who are at high risk of developing CAD. This is to allow generalization of the proposed approach to a larger population with Coronary Artery Disease, To overcome this limitation, we suggest extending this research beyond the Z-Alizadeh Sani dataset to other CAD datasets and then investigating its generalizability to state-of-the-art machine learning models, The aim will be to design a one-time diagnostic system for Coronary Artery Disease, regardless of age or origin.

VI. Conclusion

The aim of this paper is to design a more accurate classification model that predicts coronary artery disease by taking advantage of clinical and non-clinical features such as symptoms, examination, ECG, and laboratory tests. This will support remote monitoring and diagnosis of patients using vital signs and gathering features. In order to enhance the classification results with respect to accuracy, sensitivity, specificity, and Matthews correlation coefficient, however, the accuracy is improved by incorporating the gain ratio feature selection method. In addition, the benchmark dataset is experimented with to check whether there is a meaningful enhancement in prediction using the feature selection methods among the twelve classifier models. The proposed ensemble voting classifier outperforms the State-of-the-Art Machine Learning Models in terms of precision, accuracy, recall, and F-measure. The results of the proposed ensemble voting classifier are even more encouraging, as it achieved a prediction accuracy of 88.12% compared to the other classifiers. Therefore, an e-diagnosis tool based on an Ensemble Voting classifier (RF + Adaboost + MLP) would be beneficial to remote patients through cost-effective diagnosis and monitoring. Furthermore, the research can be extended by using other datasets to predict other diseases.
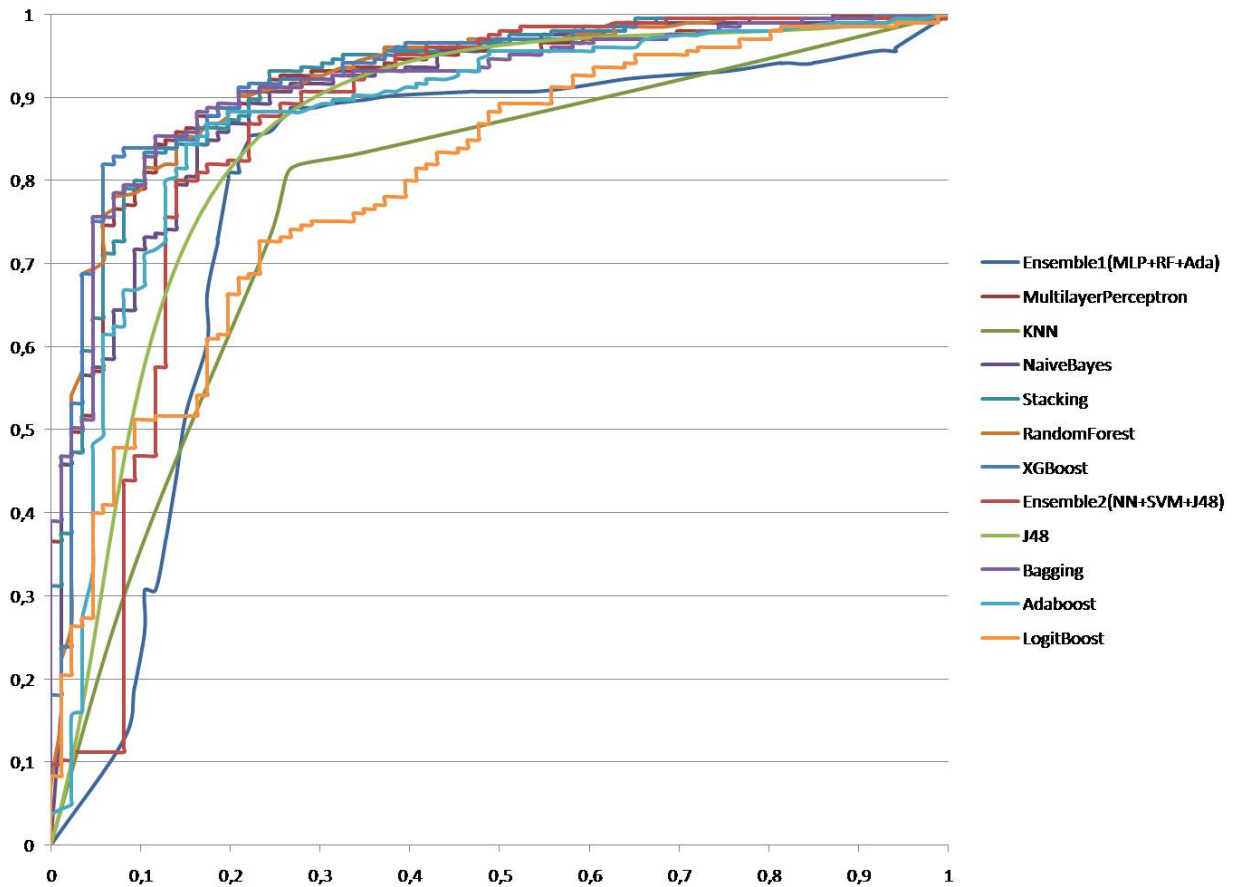
Fig. 12. Comparison of the ROC Curves of the Proposed Ensemble Voting Model with State-of-the-Art Machine Learning Models

TABLE IV. COMPARISON OF THE PROPOSED MODEL WITH STATE-OF-THE-ART MACHINE LEARNING MODELS

Model	Accuracy	Precision	Recall	F-Measure	MCC	ROC Area	Kappa	RMSE
Ensemble 1(MLP +RF+Adaboost)	88,12%	89,40%	88,10%	88,40%	0,734	0,932	0,7267	0,3137
MultilayerPerceptron	87,79%	88,00%	87,80%	87,90%	0,707	0,927	0,7067	0,3033
Stacking	87,13%	86,90%	87,10%	87,00%	0,679	0,927	0,6778	0,3101
Bagging	87,13%	88,00%	87,10%	87,40%	0,703	0,932	0,699	0,315
RandomForest	86,47%	86,40%	86,50%	86,40%	0,668	0,918	0,6683	0,3133
Ensemble2(SVM+KNN+J48)	86,47%	87,00%	86,50%	86,70%	0,681	0,85	0,6794	0,3678
AdaBoost	85,15%	85,90%	85,10%	85,40%	0,653	0,917	0,6504	0,3504
J48	84,82%	85,10%	84,80%	84,90%	0,634	0,848	0,6342	0,3634
XGboost	84,82%	85,10%	84,80%	84,90%	0,634	0,871	0,6342	0,4668
NaiveBayes	82,84%	83,60%	82,80%	83,10%	0,597	0,899	0,5947	0,3731
LogitBoost	81,19%	82,40%	81,20%	81,60%	0,567	0,827	0,563	0,4143
KNN	78,88%	79,90%	78,90%	79,30%	0,507	0,781	0,5044	0,4583

REFERENCES

[1] M. Bouma, M. Bouma, G. de Grooth *et al.*, "Standaard stabiele angina pectoris versie 4.0," *M43: Nederlands huisartsen genootschap*, 2020.

[2] B. B. Hoorweg, R. T. Willemsen, L. E. Cleef, T. Boogaerts, F. Buntinx, J. F. Glatz, and G. J. Dinant, "Frequency of chest pain in primary care, diagnostic tests performed and final diagnoses," *Heart*, vol. 103, no. 21, pp. 1727–1732, 2017.

[3] M. Bouma, F. Rutten, A. Bohnen, and T. Wiersma, "Samenvatting van de nhg-standaard stabiele angina pectoris (tweede herziening)," *Nederlands tijdschrift voor geneeskunde*, vol. 148, no. 45, pp. 2221–2225, 2004.

[4] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[5] R. Alizadehsani, J. Habibi, M. J. Hosseini, H. Mashayekhi, R. Boghrati, A. Ghandeharioun, B. Bahadorian, and Z. A. Sani, "A data mining approach for diagnosis of coronary artery disease," *Computer methods and programs in biomedicine*, vol. 111, no. 1, pp. 52–61, 2013.

[6] M. M. Dhanvijay and S. C. Patil, "Internet of things: A survey of enabling technologies in healthcare and its applications," *Computer Networks*, vol. 153, pp. 113–131, 2019.

[7] F. R. Yazdi, M. Hosseinzadeh, and S. Jabbehdari, "A review of state-of-the-art on wireless body area networks," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 11, 2017.

[8] J. Y. Khan, M. R. Yuce, G. Bulger, and B. Harding, "Wireless body area network (wban) design techniques and performance evaluation," *Journal of medical systems*, vol. 36, no. 3, pp. 1441–1457, 2012.

[9] D. C. Yadav and S. Pal, "Analysis of heart disease using parallel and sequential ensemble methods with feature selection techniques: heart disease prediction," *International Journal of Big Data and Analytics in Healthcare (IJBDAH)*, vol. 6, no. 1, pp. 40–56, 2021.

[10] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107 562–107 582, 2020.

[11] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, pp. 180 235–180 243, 2019.

[12] S. Saxena, V. K. Gupta, and P. Hrisheekesha, "Coronary heart disease detection using nonlinear features and online sequential extreme learning machine," *Biomedical Engineering: Applications, Basis and Communications*, vol. 31, no. 06, p. 1950046, 2019.

[13] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Computing and Applications*, vol. 29, no. 10, pp. 685–693, 2018.

[14] N. Gupta, N. Ahuja, S. Malhotra, A. Bala, and G. Kaur, "Intelligent heart disease prediction in cloud environment through ensembling," *Expert Systems*, vol. 34, no. 3, p. e12207, 2017.

[15] L. Verma, S. Srivastava, and P. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *Journal of medical systems*, vol. 40, no. 7, pp. 1–7, 2016.

[16] K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing coronary heart disease using ensemble machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 10, 2016.

[17] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8221–8231, 2015.

[18] M. Nilashi, H. Ahmadi, A. A. Manaf, T. A. Rashid, S. Samad, L. Shahmoradi, N. Aljojo, and E. Akbari, "Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates," *International Journal of Fuzzy Systems*, vol. 22, no. 4, pp. 1376–1388, 2020.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[20] P. Radivojac, N. V. Chawla, A. K. Dunker, and Z. Obradovic, "Classification and knowledge discovery in protein databases," *Journal of Biomedical Informatics*, vol. 37, no. 4, pp. 224–239, 2004.

[21] E. H. Houssein, M. Kilany, and A. E. Hassanien, "Ecg signals classification: A review," *Int.J.Intell.Eng.Inform.*, vol. 5, no. 4, pp. 376–396, 2017.

[22] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, and M. DATA, "Practical machine learning tools and techniques," in *Data Mining*, vol. 2, no. 4, 2005.

[23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[24] ——, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[26] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognition*, vol. 44, no. 2, pp. 330–349, 2011.

[27] R. Arora and S. Suman, "Comparative analysis of classification algorithms on different datasets using weka," *International Journal of Computer Applications*, vol. 54, no. 13, pp. 21–25, 2012.

[28] J. Han and M. Kamber, "Data mining: Concepts and techniques," *Data Mining: Concepts and Techniques*, 2001.

[29] M. Minsky and S. Papert, "An introduction to computational geometry," *Cambridge tiass., HIT*, vol. 479, p. 480, 1969.

[30] N. Baba, "A new approach for finding the global minimum of error function of neural networks," *Neural Networks*, vol. 2, no. 5, pp. 367–373, 1989.

[31] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.
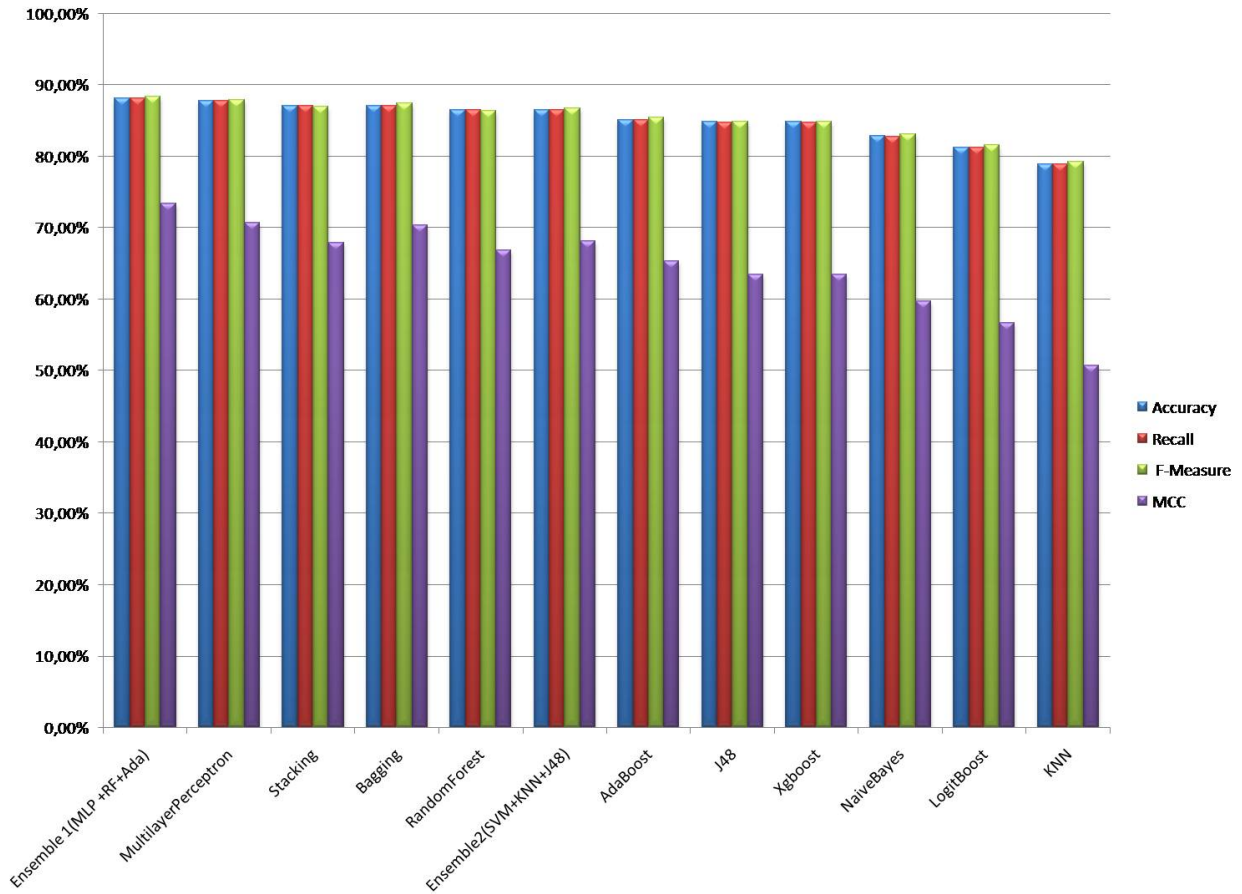
Fig. 13. Comparison of the Accuracy, Recall, F-Measure, and MCC of the Proposed Ensemble Voting Model with State-of-the-Art Machine Learning Models.
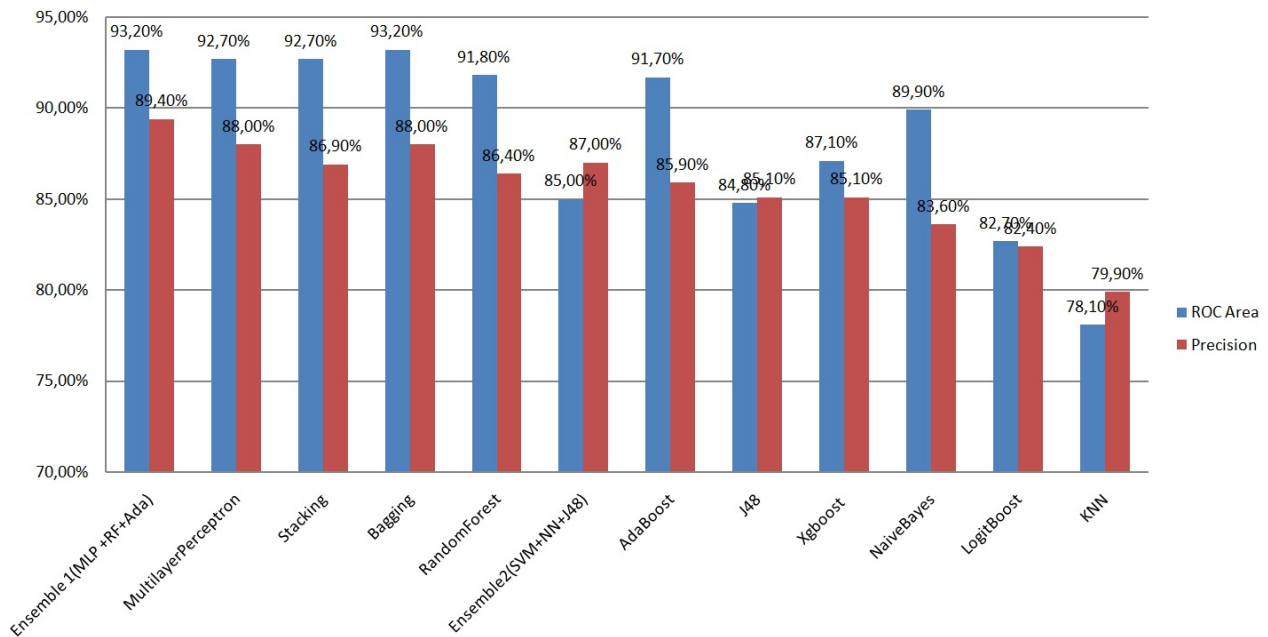

Fig. 14. Comparison of the ROC and the Precision of the Proposed Ensemble Voting Model with State-of-the-Art Machine Learning Models.