



RESEARCH ARTICLE

Ensemble classifiers with hybrid feature selection approach for diagnosis of coronary artery disease

Sowmiya M¹, Banu Rekha B² and Malar E¹

Abstract

Coronary artery disease (CAD) is a common type of cardiovascular disease with a high mortality rate worldwide. As symptoms may not be recognized until, after the cardiac attack, early diagnosis and treatment are critical to lowering mortality. The proposed study focuses on the creation of an intelligent ensemble system for the accurate detection of CAD. This paper presents the hybrid feature selection method based on Lasso, random forest-based boruta, and recursive feature elimination methods. The significance of a feature is determined by the score each approach provides. Machine learning techniques such as random forest, support vector machine, K-nearest neighbor, logistic regression, decision tree, and Naive Bayes are developed as base classifiers. Then, ensemble techniques like bagging and boosting models are created using base classifiers. The Z-Alizadeh Sani dataset was used to build and test the model. The bagged random forest model achieved 97.6% accuracy and 100% recall. The CatBoost model achieved 97.7% accuracy and 99.0% recall. Compared to traditional classifiers, the ensemble models achieved higher accuracy and can be used to assist clinicians in diagnosing coronary artery disease.

Keywords: Coronary artery disease, Artificial Intelligence, Machine learning.

Introduction

Machine Learning (ML), a subset of artificial intelligence (AI), has found widespread use in healthcare applications (Bunyamin *et al.*, 2021). ML approaches have mostly emerged as dominant tools to aid physicians in making decisions, predicting risk factors, discovering hidden patterns, and diagnosing illnesses (Richens *et al.*, 2020; Mirbabaie *et al.*,

2021). Currently, studies demonstrate that ML is useful in the early diagnosis of CAD (Arkadip *et al.*, 2021). CAD occurs due to plaque accumulation in the arteries by obstructing the arterial wall, which is called atherosclerosis (Malakar *et al.*, 2019). CAD is caused by a combination of variables, including age, gender, high blood pressure, obesity, physical inactivity, stress, and alcohol use (Malakar *et al.*, 2019). The provision of appropriate drugs in clinical care, as well as early identification and treatment, are crucial for lowering the mortality rate (Arkadip *et al.*, 2021). In developing countries, CAD diagnosis is seen as a difficulty since it necessitates using well-trained clinical professionals. Coronary Angiography (CA) is the gold standard invasive modality for diagnosing CAD, although it has several risks and limitations, including being extremely expensive and time-demanding (Doris *et al.*, 2016). Many people may experience no symptoms until they experience chest discomfort or a heart attack later. As a result, in healthcare applications, an intelligent decision system will aid physicians (Bunyamin *et al.*, 2021; Malakar *et al.*, 2019). It can predict whether a patient suffers from CAD and advise them on how to proceed with therapy. The importance of an intelligent CAD diagnosis system developed from patient data and risk variables using machine learning algorithms has been highlighted in this research (Muhammad *et al.*, 2020). In developing countries, machine learning can help doctors diagnose CAD early by using data collected from patient data.

Researchers have developed many intelligent models

¹Department of Electronics and Communications Engineering, PSG Institute of Technology and Applied Research, Coimbatore, India.

²Department of Biomedical Engineering, PSG College of Technology, Coimbatore, India.

³Department of Electrical & Electronics Engineering, PSG Institute of Technology and Applied Research, Coimbatore, India.

***Corresponding Author:** Sowmiya M, Department of Electronics and Communications Engineering, PSG Institute of Technology and Applied Research, Coimbatore, India, E-Mail: swm.muruganatham@gmail.com

How to cite this article: Sowmiya, M., Rekha, B.B., Malar, E. (2023). Ensemble classifiers with hybrid feature selection approach for diagnosis of coronary artery disease. *The Scientific Temper*, **14**(3): 726-734.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.3.24

Source of support: Nil

Conflict of interest: None.

to aid in the automated diagnosis of CAD. Clinical data and symptoms, electrocardiogram (ECG), computed tomography, and magnetic resonance imaging are used to assess (Malakar *et al.*, 2019). Artificial neural network (ANN), support vector machines (SVM), decision tree (DT), and random forest (RF) are some of the used classifiers (Alizadehsani *et al.*, 2013, 2016). Several approaches have been developed using the UCI dataset, while a few have been created using the Z-Alizadeh Sani dataset (Alizadehsani *et al.*, 2019).

Yar Muhammad *et al.*, 2019 built an intelligent diagnosis system on the UCI Cleveland dataset using ten machine-learning methods. After using feature selection methods, extra trees and Naive Bayes produced greater accuracy of 94.41 and 93.36%, respectively. Tayefi *et al.*, 2017 created a CAD model incorporating risk variables using decision trees. The model has a prediction accuracy of 94% on average. In addition, the research finds that the tree chooses gender, age, and fasting blood glucose as significant risk variables. Arabasadi *et al.*, 2017 suggested a hybrid machine-learning approach. A genetic algorithm and an artificial neural network created the model. On the Z-Alizadeh Sani data, the approach performed better, with an accuracy of 93.85%, sensitivity of 97%, and specificity of 92%. Alizadehsani *et al.* employed a data mining technique to predict coronary artery stenosis by selecting significant variables based on information gain. Bagging, sequential minimal optimization (SMO) with support vector machine (SVM), neural network, and Naive Bayes were trained on the reduced dataset. The SMO approach was the most accurate, with a 94.08% accuracy (Alizadehsani *et al.*, 2013). Ghiasi *et al.*, 2020 created a decision tree method for CAD diagnosis. The algorithm was created utilizing a variety of features. With five characteristics, it attained an accuracy of 92.41%.

Several researchers have used a single classifier model to identify CAD. Due to the diverse nature of model parameters and the imbalance in datasets, individual models can only provide greater accuracy in some situations. Ensemble approaches have been applied in research for breast cancer detection (Moloud *et al.*, 2020), diabetes (Salani *et al.*, 2021), churn prediction (Wang *et al.*, 2019), and agriculture (Archana *et al.*, 2016). For diagnosing CAD, researchers have begun to use ensemble models.

To stack the ML classifiers, Jikuo Wang *et al.*, 2020 presented an enumeration technique. The suggested model's accuracy and F1 score are 95.43 and 96.77%, respectively. Even though they outperformed the challenge, the training duration was excessive, and the hyperparameters were not tuned. In both the Z-Alizadeh Sani and Cleveland datasets, Moloud Abdar *et al.* used support vector classification (SVC) using nested ensemble base classifiers. In the Z-Alizadeh Sani dataset (Abdar *et al.*, 2019), the model has an accuracy of 94.66%. A stacking

ensemble classifier was proposed by Bayu Adhi *et al.*, 2020. As a metaclassifier, the generalized linear model (GLM) is utilized. The model utilized 27 features and has a 98.13% accuracy rate. According to the literature, analyzing ensemble models for CAD diagnosis is required in order to increase performance on the Z-Alizadeh Sani dataset. The researchers have stressed the relevance of feature selection strategies in improving model performance. Many feature selection approaches were integrated to find the essential properties connected with CAD prediction (Burak *et al.*, 2019). On the Z-Alizadeh Sani dataset, Cleveland dataset, and a mixture of CVD datasets, Kolukisa *et al.*, 2019 used hybrid feature selection approaches. On the Z-Alizadeh Sani dataset with 11 features, they created a bagging ensemble classifier that has a precision of 92.07%.

Non-invasive diagnosis utilizing ensemble ML algorithms for early-stage identification of CAD is described in this work. Because of population variety, sample size, and characteristics, there is no one optimum ML approach for diagnosing CAD, according to the literature (Mirbabaie *et al.*, 2021; Muhammad *et al.*, 2020). This research aims to create an ensemble model that uses typical machine learning classifiers to predict CAD at an earlier stage and increase prediction accuracy. On the Z-Alizadeh Sani dataset with specified features, classifiers such as RF, SVM, KNN, LR, DT, and Naive Bayes (NB) were trained initially. To discover the significant features, Lasso, random forest-based boruta feature selection (RF-BFS), and recursive feature elimination (RFE) were used. According to the literature, only a few research studies have employed the Lasso approach on the Z-Alizadeh Sani dataset to diagnose CAD. To prevent overfitting, the dataset's samples are balanced using the sampling minority over-sampling technique (SMOTE), and 10-fold cross-validation is utilized during model development. The fundamental models are then linked to the framework of ensemble techniques such as bagging and boosting models. SMOTE improves the accuracy of simple classifiers.

The main contributions of the work include,

- Using the Z-Alizadeh Sani dataset, classical classifiers are used to develop bagging and boosting ensemble models for the diagnosis of CAD. Prior research has yet to extensively examine these three models for CAD prediction.
- Feature selection methods were employed to choose the significant features which reduce the training time and improve performance. Common features ranked by two of the Lasso, random forest-based boruta and recursive feature elimination methods, are selected for training the model.
- The performance of classical classifiers is compared to that of ensemble classifiers. The approach included hyperparameter tuning, which analyzed the performance metrics. In addition, the study compares

the performance of homogeneous and heterogeneous ensemble models.

The paper is organized as follows; Section 2 includes a review of previous works related to CAD diagnosis. Section 3 presents the description of the dataset and methods for creating an ensemble model. Experimental results are presented in section 4. The discussion is given in section 5, followed by a conclusion in section 6.

Database Description

UCI Heart Disease and Z-Alizadeh Sani databases are popular for diagnosing CAD. The UCI heart disease dataset is frequently used, although it is flawed because it contains some missing data (Alizadehsani *et al.*, 2019). In this proposed work, the Z-Alizadeh Sani dataset is employed. The Z-Alizadeh Sani dataset includes information on 303 patients. The last feature is the target variable, which is labeled as CAD if the artery diameter narrows by more than 50%; otherwise, it is normal. The dataset is unbalanced because CAD accounts for 71% of patient data. There are 31 categorical and 23 numerical attributes in the dataset. The min-max scaler is used to standardize and eventually normalize all the features. The details about the features of the Z-Alizadeh Sani dataset, along with the range or type of value, are given in Table 1. Features 1 to 16 belong to demographic, 17 to 24 belong to ECG, 25 to 38 belong to the symptom and physical examination, and 39 to 54 belong to laboratory test and echo attributes.

Table 1: Features of Z-Alizadeh Sani CAD dataset (Alizadehsani *et al.*, 2019)

S.No.	Attribute name	Range
1.	Age (years)	30–86
2.	Sex	Male, Female
3.	Length (cm)	140–188
4.	Weight (kg)	48–120
5.	Body Mass Index (BMI, Kg/m ²)	18–41
6.	Diabetes Mellitus (DM)	Yes - 1, No - 0
7.	Hyper Tension (HTN)	Yes - 1, No - 0
8.	Current Smoker	Yes - 1, No - 0
9.	Ex-smoker	Yes - 1, No - 0
10.	Family History (FH)	Yes - 1, No - 0
11.	Obesity	Yes (BMI > 25) - 1, else No - 0
12.	Chronic Renal Failure (CRF)	Yes - 1, No - 0
13.	Cerebrovascular Accident (CVA)	Yes - 1, No - 0
14.	Thyroid disease	Yes - 1, No - 0
15.	Airway disease	Yes - 1, No - 0
16.	Congestive Heart Failure (CHF)	Yes - 1, No - 0
17.	Dyslipidaemia (DLP)	Yes - 1, No - 0
18.	Rhythm	Sin, AF
19.	ST elevation	Yes - 1, No - 0
20.	ST depression	Yes - 1, No - 0
21.	Q-wave	Yes - 1, No - 0
22.	T inversion	Yes - 1, No - 0
23.	Left Ventricular Hypertrophy (LVH)	Yes - 1, No - 0
24.	Poor R-wave progression	Yes - 1, No - 0
25.	Blood Pressure (BP, mmHg)	90-190
26.	Pulse Rate (PR, ppm)	50-110
27.	Edema	Yes - 1, No - 0
28.	Weak peripheral pulse	Yes - 1, No - 0
29.	Lung's rales	Yes - 1, No - 0
30.	Systolic murmur	Yes - 1, No - 0
31.	Diastolic murmur (DM)	Yes - 1, No - 0
32.	Typical Chest Pain	Yes - 1, No - 0
33.	Dyspnea	Yes - 1, No - 0
34.	Function Class	1,2,3,4
35.	Atypical	Yes - 1, No - 0
36.	Nonanginal	Yes - 1, No - 0
37.	Exertional Chest Pain	Yes - 1, No - 0
38.	Low Threshold Angina (LowTH Ang)	Yes - 1, No - 0
39.	Fasting Blood Sugar (FBS, mg/dL)	62–400
40.	Creatine (Cr, mg/dL)	0.5–2.2
41.	Triglyceride (TG, mg/dL)	37-1050
42.	Low-Density Lipoprotein (LDL, mg/dL)	18–232
43.	High-Density Lipoprotein (HDL, mg/dL)	15–111
44.	Blood Urea Nitrogen (BUN, mg/dL)	6–52
45.	Erythrocyte Sedimentation Rate (ESR, mm/h)	1–90
46.	Haemoglobin (HB, g/dL)	8.9–17.6
47.	Potassium (K, mEq/lit)	3.0–6.6
48.	Sodium (Na, mEq/lit)	128–156
49.	White Blood Cell (WBC, cells/mL)	3700–18000
50.	Lymphocyte (Lymph %)	7–60
51.	Neutrophil (Neut %)	32–89

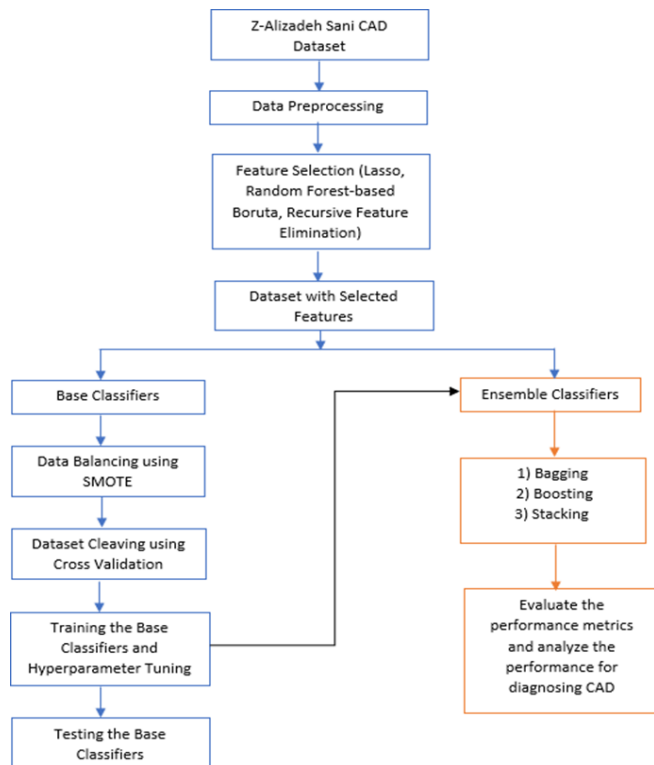


Figure 1: Framework for development of ensemble model for CAD diagnosis

Methodology

To choose the important features for forecasting CAD, Lasso, random forest-based boruta, and recursive feature elimination approaches are employed in this work. The dataset was split using 10-fold cross-validation after feature selection. To remedy the class imbalance problem, SMOTE was applied to the training set. Machine learning-based classifiers were used to train a balanced dataset, and classifiers were blended using the ensemble approach. The SVM, RF, K-Nearest Neighbours (KNN), LR, NB, and DT are the classifiers. A grid search strategy was used to optimize classifier hyperparameters. Ensemble algorithms such as bagging and boosting are created in this study to forecast the end outcomes utilizing basic classifiers. Figure 1 depicts the prototype of the proposed model.

Feature Selection

The accuracy of the detection system depends on the significant features correlated with the target response. The Lasso technique, RF-BFS, and recursive feature elimination (RFE) are applied to select the features of high importance in predicting CAD response. Initially, the top 20 features were preferred, which are ranked high in the three feature selection algorithms. Then, 16 common features selected by at least two algorithms were used in the training stage.

Lasso

The least absolute shrinkage and selection operator (Lasso) is a regularized regression model that eliminates the less contributing features by shrinking the coefficients, reducing the variance (Chen *et al.*, 2021). In each iteration, features that contribute the foremost to the training on that iteration were chosen, ignoring the less important ones. Features with zero coefficients were eliminated, so automatic feature selection was achieved. This method chooses 14 features based on top ranking and eliminates the remaining ones with an alpha score of 0.07. Figure 2 shows the important features extracted using Lasso.

Random Forest-based Boruta Feature Selection (RF-BFS)

The Boruta feature selection method ranks features based on the random forest algorithm (Enas *et al.*, 2021). It adds randomization to the training set by producing shadow features, which are duplicates of all features. The random forest classifier was then trained on the dataset with all attributes, with features chosen based on mean decrease accuracy. The z-score of the original and synthetic characteristics was computed at each iteration. If an attribute's z-score is greater than the maximum z-score of its shadow, it is deemed important. Finally, the algorithm will generate three lists of confirmed, tentative and rejected attributes.

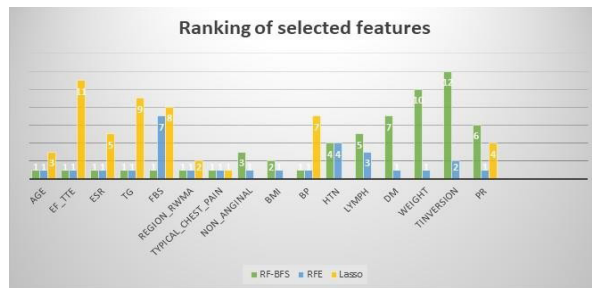


Figure 2: Top features selected by three methods

Recursive Feature Elimination (RFE)

RFE iteratively considers the characteristics in order to find a subset of them. The estimator is trained using the features in the training dataset and assesses each attribute's relevance score. The least significant characteristics in the current set are pruned, and the procedure is repeated until the required number of features is obtained (Enas *et al.*, 2021).

Classification

For model development, the dataset was separated into training and testing sets. To minimize overfitting, 10-fold cross-validation was performed three times during the training procedure. The purpose was to divide the training dataset into 10 files randomly, then build the model using 9 subsets of the dataset and one subset for testing the model for each fold. The method is continued until the test set of each 10-fold has been served. The grid search technique (Alizadehsani *et al.*, 2012) was used to tune hyperparameters. Cross-validation strengthens the model's resiliency to mistakes in this manner.

Base Classifier

Different machine-learning methods were used to create the ensemble classifier models in this work. The SVM, RF, KNN, LR, Naive Bayes, and DT are some of them. The features used in the training phase were 'Age,' 'EF TTE,' 'ESR,' 'Region RWMA,' 'Typical Chest Pain,' 'BMI,' 'BP,' 'TG,' 'HTN,' 'FBS,' 'Weight,' 'PR,' 'Lymph,' 'Nonanginal,' 'Tinversion.' To increase prediction accuracy, hyperparameter tuning of the models utilizing grid search was performed. For the final performance evaluation of the various models, the performance metrics of accuracy, precision, recall, F1 score, and ROC were calculated in each fold and then averaged.

Algorithm 1 lays out a step-by-step approach for creating the base models. For CAD diagnosis, a dataset with selected features was used.

Algorithm 1 Procédure to create the base model

1. Inputs: Balanced dataset DN consisting of i features acquired from the three feature selection algorithms based on top rankings of the attributes and L Base classifier models $M_1, M_2, M_3, \dots, M_L$.
2. for $i = 1, 2, 3, \dots, L$ do

1. Perform 10–fold cross-validation (10–fold) to split DN into a training set (DTR) and testing set (DTS).
 2. For each Nth fold, apply the model on the training set DTR.
 3. Predict the result of M_i on DTS.
 4. Tune the hyperparameters using a grid search for each M_i and repeat.
 5. Finally, calculate the performance metrics in each fold and take the average of all folds for the final prediction of the performance of the model M_i .
2. for $i = 1, 2, 3, \dots L$ do
 - a. Train and evaluate the models on the testing set DTS.
 - b. Predict the score P_i of the model and cost function.
 - c. Identify the misclassified samples.
 end
 3. Assign more weights to misclassified samples.
 4. Repeat steps 2 and 3 until the error is less.
 5. Obtain the result of the boosting ensemble

End

Output: Trained base model.

Ensemble Classifiers

Bagging

Bagging is an ensemble technique in which a random subset of the dataset is used to train the base model simultaneously. The averaging procedure aggregates the predictions from each base model. It lowers the estimator's variance, making the model more generalized (Kaushik *et al.*, 2019). Each base classifier is compared to the original and bagged with its equivalents in this study.

Algorithm 2 explains in detail to generate a weighted average voting ensemble. The output class is a forecast based on the average of the probabilities assigned to that class, and soft voting is applied.

Algorithm 2 Procedure to build bagged ensemble model

1. Input: Trained L base-classifier models $M_1, M_2, M_3, \dots, M_L$, Testing set (DTS).
2. for $i = 1, 2, 3, \dots L$ do
 - a. Evaluate each model M_i on the testing set DTS.
 - b. Predict the score P_i of each model.

end

3. Aggregate the score.

Output: Bagged model

Boosting

Boosting technique builds stronger learners from several weak learners sequentially. Initially, a series of base models are constructed, assigning equal weights to each observation in the dataset. With each next iteration, each model gives more weightage to the observations in the dataset that the earlier models wrongly classified. This way of fitting the difficult samples helps the algorithm identify the parameters it should focus on and produce a strong learner with a lower bias. The boosting algorithms used in this study are AdaBoost, Gradient Boost, XGBoost, LightGBM, and CatBoost. The base estimator model used was a decision tree with variations in hyperparameters. The major hyperparameters are a number of estimators, the learning rate, and the number of times the model is boosted.

Algorithm 3 Procedure to build boosted ensemble model

1. Input: Training set (DTR), Testing set (DTS), Choose the base-classifier estimator, number of iterations.

Results and Discussions

This section discusses the results of different base classifiers and ensemble methods on the Z-Alizadeh Sani dataset. The performance of the base classifier was evaluated on both the imbalanced and balanced datasets. Jupiter python notebook, along with the Scikit package, was used to develop the models.

Performance Metrics

The measures used to assess the performance classifiers in diagnosing CAD are presented in this section. Because the dataset is unbalanced, the study includes parameters including accuracy, precision, sensitivity, F1 score, and receiver operating characteristics (ROC). The metrics' expression is provided below,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{F1 score} = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Where TP – number of CAD records predicted as CAD

TN – number of normal records predicted as normal

FP – number of normal records predicted as CAD

FN – number of CAD records predicted as normal

The ROC curve is a plot of true positive rate (TPR) vs false positive rate (FPR), where Sensitivity is TPR and (1-Specificity) is FPR.

Results of Base Classifiers

Several classic classifiers were used to construct the decision support system for CAD diagnosis, including SVM, RF, KNN, NB, DT, and LR. The classifiers were trained using a 10-fold cross-validation of the training set with a collection of 16 features. SMOTE is used to construct samples to balance the dataset, which contains 216 records belonging to CAD and 87 records belonging to normal. It is clear from the data that class balancing improves the prediction of basic classifiers.

Tables 2 and 3 indicate that random forest has a better performance in distinguishing between CAD and non-CAD persons. When compared to LR, RF has the best F1 score and recall but lacks ROC. The higher the recall score, the

Table 2: Performance of base classifiers on an original dataset with selected features

Classifiers	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
DT	83.7	88.5	88.5	89.2
KNN	86.5	90.1	95.0	86.0
NB	87.5	91.2	91.9	91.9
SVM	89.1	92.2	93.7	91.1
LR	90.0	93.2	95.2	91.6
RF	90.4	93.4	91.5	95.6

Table 3: Performance of base classifiers on a balanced dataset with selected features

Classifier	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
DT	81.5	87.5	85.2	90.7
KNN	85.6	89.2	94.7	84.7
NB	87.0	91.0	90.1	92.5
SVM	88.4	91.7	94.0	89.9
LR	87.9	91.5	90.5	92.7
RF	89.4	92.8	90.0	96.2

Table 4: Parameters of base classifiers

Classifier	Parameters
SVM	kernel= rbf, C=1, gamma= scale, random state = 42, probability = 1
RF	max depth = 10, max features = 2, min samples leaf = 3, min samples split = 3, estimators = 1000
KNN	neighbors = 10, weights = distance, metric = manhattan
DT	criterion= gini, max depth = 10, min samples leaf = 50

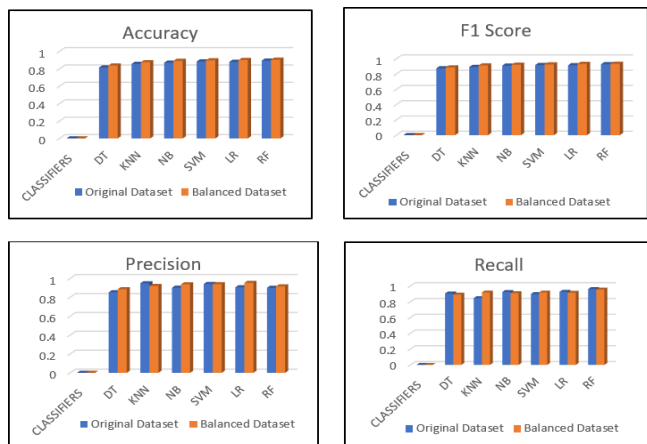


Figure 3: Performance metrics of base classifiers on original and balanced datasets

more accurately the real class of CAD is predicted. The hyperparameters utilized in the base models are provided in Table 4.

Figures 3 and 4 show the base models’ performance measurements and ROC plots. The ROC curve demonstrates the model’s capacity to discriminate between CAD and

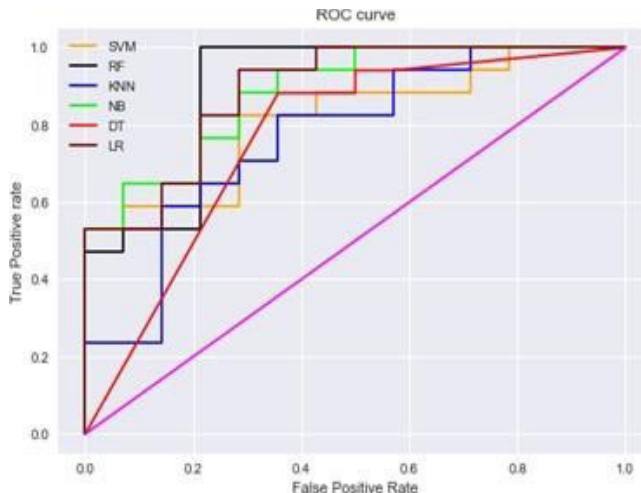


Figure 4: ROC of base classifiers on a balanced dataset

normal people. It is graphed against the true positive rate and the false positive rate. The greater the area under curve (AUC), the greater the diagnostic accuracy.

Results of Ensemble Classifiers

Bagging

The bagging model is created using trained base models and tested on a data set. The base classifiers LR, DT, RF, SVM, ET, and KNN were utilized on a balanced dataset to generate a bagged model. The results suggest that the bagging ensemble approach can increase overall performance in predicting the CAD when compared to individual learners. From Table 5, random forest produced superior performance compared to other bagged models (Figure 5).

Boosting

Five boosting methods are created with a decision tree as the base classifier in this study: adaptive boosting (ADB), gradient boosting (GB), extreme gradient boosting (XGB), lightGBM (LGBM), and CatBoost (Figure 6). The number of estimators, items, and the iterative procedure to be repeated are the major hyperparameters tuned in the models. CatBoost had the greatest accuracy, F1 score, precision, recall, and ROC in identifying CAD patients, according to Table 6 and Figure 7.

This study investigated the ensemble methodologies used to diagnose coronary heart disease. Additionally, it depicts the evolution of bagging and boosting models

Table 5: Results of bagged models

Bagged models	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
DT	87.5	90.0	89.3	92.8
KNN	87.4	89.4	95.2	85.4
LR	93.0	95.4	95.8	95.1
NB	95.6	97.2	95.9	98.6
SVM	96.3	97.6	96.0	99.4
RF	97.6	98.3	96.7	100.0

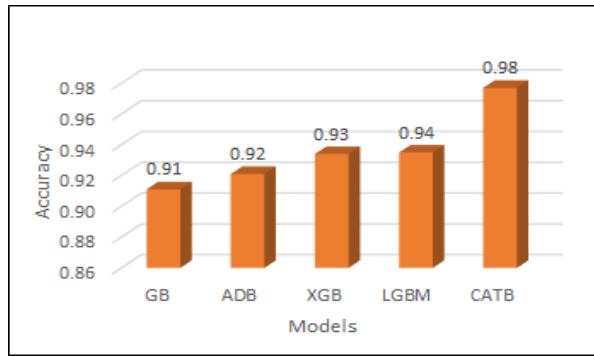


Figure 5: Accuracy of bagging models

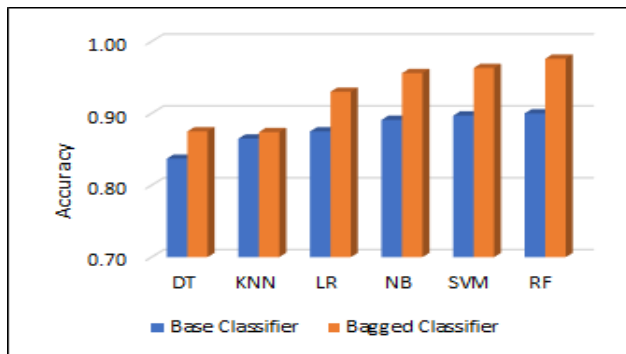


Figure 6: Accuracy of boosting models

Table 6: Results of boosted models

Boosted models	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
GB	91.1	93.7	90.0	97.9
ADB	92.1	94.2	93.2	95.5
XGB	93.4	95.3	91.3	99.6
LGBM	93.5	95.3	93.1	97.8
CATB	97.7	98.3	97.7	99.0

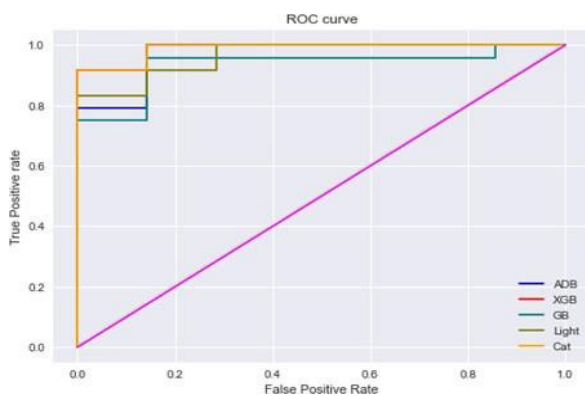


Figure 7: ROC of Boosting models

from classical classifiers. In this study, a consistent level of accuracy, F1 score, precision, and recall in predicting CAD or non-CAD patients was achieved. The main objective is to create an ensemble model from the base models on the Z-Alizadeh Sani dataset using the given characteristics (Table 7). Feature selection techniques were used to choose relatively essential characteristics, and 16 features were chosen. The performance of conventional and ensemble classifiers was tested on the balanced data set. The results show that ensemble approaches outperformed

Table 7: Comparison of the ensemble models with another study from the literature on the Z-Alizadeh Sani CAD data

Study	Algorithm	No. of Features	Accuracy (%)	F1 Score (%)	Recall (%)
Jikuo Wang <i>et al.</i> (2020) [27]	Stacking	20	95.43	96.77	95.84
Roohallah Alizadehsani <i>et al.</i> (2012) [37]	Bagging (SVM and NB)	16	88.52	-	91
Moloud Abdar <i>et al.</i> (2019) [28]	Nested Ensemble	16	94.66	-	-
Moloud Abdar <i>et al.</i> (2019) [38]	SVM with genetic algorithm	29	93.08	91.51	-
Qin <i>et al.</i> (2017) [31]	Bagging	16	93.7	95.53	-
Bayu Adhi <i>et al.</i> (2020) [30]	Two level ensemble	27	98.13	96.6	-
Proposed work	Bagging - RF	16	97.6	98.3	100
Proposed work	Boosting - CatBoost	16	97.7	98.3	99.0

conventional algorithms in terms of performance. When compared to bagging, boosting models offered the best results. The accuracy, recall, and F1 score measurements show that accurately classifying CAD patients is more important than correctly categorizing a normal individual. The findings of bagging reveal that bagged models outperform their conventional counterparts in terms of performance. The results show that the bagged and boosted models worked well, with the CatBoost model achieving the highest accuracy. Age, hypertension, blood pressure, diabetes, typical chest discomfort, BMI, region with regional wall motion, and pulse rate features all played a role in the diagnosis of coronary artery disease.

Conclusion

The effectiveness of ensemble models for diagnosing CAD was investigated in this study. The Z- Alizadeh Sani dataset, which has 16 features, was initially used to train common machine learning techniques. In order to develop a bagging and boosting model, base classifiers were employed. The bagged random forest model had a 97.6% accuracy rate. With a precision of 97.7%, the CatBoost model surpassed the other boosting methods. The results make it very evident that the ensemble model performs better on the balanced dataset. Since the boosting model aims to reduce the error on each observation sequentially, it performs better than other ensembles. Even though the literature has documented several advancements in CAD diagnosis, an ensemble technique has improved performance greatly.

Data Availability

The data presented in this study are publicly available

Acknowledgments

The authors are thankful to PSG institutions for the support provided in completing the paper.

References

- Abdar, M., Acharya, U. R., Sarrafzadegan, N., and Makarenkov, V., (2019). NE-nu-SVC: A New Nested Ensemble Clinical Decision Support System for Effective Diagnosis of Coronary Artery Disease. *IEEE Access*, 7: pp. 167605-167620.
- Alizadehsani, R., Habibia, J., Hosseini, M.J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B., Alizadeh Sani, Z. (2013). a data mining approach for diagnosis of coronary artery disease, *Comput. Methods Programs Biomed.* 111:52–61.
- Alizadehsani, R., Zangooei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., Khozeimeh, F., Sarrafzadegan, N., Nahavandi, S. (2016). Coronary artery disease detection using computational intelligence methods, *Knowl. Based Syst.* 109:187–197.
- Alizadehsani, Roohallah Habibi, Jafar Hosseini, Mohammad Javad Boghrati, Reihane Ghan- deharioun, Asma Bahadorian, Behdad Sani, Zahra. (2012). Diagnosis of Coronary Artery Disease Using Data Mining Techniques Based on Symptoms and ECG Features. *European Journal of Scientific Research.* 82.
- Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., and Yarifard, A. A. (2017). Computer-aided decision making for heart disease detection using hybrid neural network genetic algorithm, *Comput. Methods Programs Biomed.* 141: pp. 19–26.
- Archana Chaudhary, Savita Kolhe, Raj Kamal,(2016). A hybrid ensemble for classification in multiclass datasets: An application to oilseed disease dataset. *Computers and Electronics in Agriculture*, 124: pp.65-72.
- Arkadip Ray, Avijit Kumar Chaudhuri. (2021). Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development. *Machine Learning with Applications*, 3, 100011.
- Bayu Adhi Tama, Sun Im, Seungchul Lee. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. *BioMed Research International*.
- Bunyamin Ozaydin, Eta S. Berner, James J. Cimino. (2021). Appropriate use of machine learning in healthcare. *Intelligence-Based Medicine, Elsevier*, 5, 100041.
- Burak Kolukisa1, Hilal Hacilar, Gokhan Goy, Mustafa Kus , Burcu Bakir-Gungor, Atilla Aral, Vehbi Cagri Gungor, (2019). Diagnosis of Coronary Heart Disease via Classification Algorithms and a New Feature Selection Methodology. *International Journal of Data Mining Science*, 1(1).
- Cai-Jie Qin, Qiang Guan, & Xin-Pei Wang (2017). Application Of Ensemble Algorithm Integrating Multiple Criteria Feature Selection In Coronary Heart Disease Detection. *Biomedical Engineering: Applications, Basis and Communications*, 29(06),1750043
- Chen, R., Dewi, C., Huang, S. (2020). Selecting critical features for data classification based on machine learning methods. *J Big Data*, 52. <https://doi.org/10.1186/s40537-020-00327-4>.
- Detrano, R., V.A.M. Center, (2019). Long Beach and Cleveland clinic foundation and the University of California, Irvine machine learning repository, Cleveland Heart Disease Dataset.
- Dietterich T.G. (2000). Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems. Lecture Notes in Computer Science*, Springer, 1857. <https://doi.org/10.1007/3-540-45014-91>
- Doris, M., Newby, DE. (2016). Coronary CT Angiography as a Diagnostic and Prognostic Tool: Perspectives from the SCOT-HEART Trial. *Curr Cardiol Rep*, 18(2):18.
- Ebrahime Mohammed Senan, Mosleh Hmoud Al-Adhaileh, Fawaz Waselallah Alsaade, Theyazn H. H. Aldhyani, Ahmed Abdullah Alqarni, Nizar Alsharif, M. Irfan Uddin, Ahmed H. Alahmadi, Mukti E Jadhav, Mohammed Y. Alzahrani. (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. *Journal of Healthcare Engineering*. <https://doi.org/10.1155/2021/1004767>
- Enas Elgeldawi, Awny Sayed, Ahmed, R., Galal and Alaa, M. (2021). Zaki Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis, *Informatics*.
- Ghosh, Pronab & Azam, Sami & Jonkman, Mirjam & Karim, Asif & Shamrat, F. M. & Ignatious, Eva & Shultana, Shahana & Beeravolu, Abhijith & De Boer, Friso. (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access*. pp. 1-1. 10.1109/ACCESS.2021.3053759.
- Kaushik, Shruti Choudhury, Abhinav Jatav, Ashutosh Dasgupta, Nataraj Natarajan, Sayee Pickett, Larry Dutt, Varun. (2019). Comparative Analysis of Features Selection Techniques for Classification in Healthcare. *Lecture Notes in Computer Science*. In press.
- Karaolis, M. A., Moutiris, J. A., Hadjipanayi, D. and Pattichis, C. S. (2010). Assessment of the risk factors of coronary heart events based on data mining with decision trees. *IEEE Trans. Inf. Technol. Biomed.*, 14(3), pp. 559–566.
- Kursa, M.B., Rudnicki, W.R. (2010). Feature selection with the Boruta package, *J. Stat. Softw.* 36 (11) : 1–13.
- Malakar, A.K., Choudhury, D., Halder, B., Paul, P., Uddin, A. Chakraborty, S. A. (2019). Review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol.* 234(10):16812-16823.
- Mirbabaie, M., Stieglitz, S. Frick, N.R.J. (2021). Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. *Health Technol*, 11, 693–731, <https://doi.org/10.1007/s12553-021-00555-5>
- Muhammad, Y., Tahir, M., Hayat, M.(2020). Early and accurate detection and diagnosis of heart disease using an intelligent computational model. *Sci Rep* 10, 19747. <https://doi.org/10.1038/s41598-020-76635-9>
- Mohammad M. Ghiasi, Sohrab Zendeheboudi, Ali Asghar Mohsenipour. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer Methods and Programs in Biomedicine*, 192.
- Moloud Abdar, Wojciech Ksiazek, Rajendra Acharya, Ru-San Tan, Vladimir Makarenkov, Pawel Plawiak,. (2019). A new machine learning technique for an accurate diagnosis of coronary artery disease, *Computer Methods and Programs in Biomedicine*, 179.
- Moloud Abdar, Mariam Zomorodi-Moghadam, Xujuan Zhou, Raj Gururajan, Xiaohui Tao, Prabal D Barua, Rashmi Gururajan, (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*, 132: pp.123-131.
- Richens, J.G., Lee, C.M. Johri, S. (2020). Improving the accuracy

- of medical diagnosis with causal machine learning. *Nat Communication*, 11, 3923, <https://doi.org/10.1038/s41467-020-17419-7>
- Roohallah Alizadehsani, Moloud Abdar, Mohamad Roshanzamir, Abbas Khosravi, Parham M. Kebria, Fahime Khozeimeh, Saeid Nahavandi, Nizal Sarrafzadegan, Rajendra Acharya, U. (2019). Machine learning- based coronary artery disease diagnosis: A comprehensive review, *Computers in Biology and Medicine*, 111.
- Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, Hoda Mashayekhi, Reihane Boghrati, Asma Ghandeharioun, Behdad Bahadorian, Zahra Alizadeh Sani. (2013). A data mining approach for diagnosis of coronary artery disease, *Computer Methods, and Programs in Biomedicine*, 111(I): pp.52-61.
- Roohallah Alizadehsani, Mohammad Hossein Zangooei, Mohammad Javad Hosseini, Jafar Habibi, Abbas Khosravi, Mohamad Roshanzamir, Fahime Khozeimeh, Nizal Sarrafzadegan, Saeid Nahavandi. (2016). Coronary artery disease detection using computational intelligence methods, *Knowledge-Based Systems*, 109: pp. 187-197.
- Sani, Z. A., Alizadehsani, R., Roshanzamir, M. (2020). Z-Alizadeh Sani data set. <https://archive.ics.uci.edu/ml/datasets/Z-AlizadehSani>
- Saloni Kumari, Deepika Kumar, Mamta Mittal. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 2:pp. 40-46.
- SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*. (2002).16:321–357.
- Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily, H., Taghipour, A., Ferns, G. A., Moohebbati, M., and GhayourMobarhan, M. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using a decision tree algorithm. *Comput. Methods Programs Biomed*, 141: pp. 105–109.
- Wang, Q.F., Xu, M., Hussain, A. (2019). Large-scale Ensemble Model for Customer Churn Prediction in Search Ads. *Cogn Comput* 11: pp.262–270 <https://doi.org/10.1007/s12559-018-9608-3>.
- Wang, J. (2020). A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease. *IEEE Access*, 8: pp. 37124–37133.