Biomedical Signal Processing and Control 70 (2021) 103033

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Heart disease prediction using hyper parameter optimization (HPO) tuning

R. Valarmathi [a,*], T. Sheela [b]

[a] Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India
[b] Department of Information Technology, Sri Sairam Engineering College, Chennai, India

ARTICLE INFO

Keywords:
Hyper parameter tuning
Heart disease
Grid search
Randomized search
TPOT classifier

ABSTRACT

Coronary artery disease prediction is considered to be one of the most challenging tasks in the health care industry. In our research, we propose a prediction system to detect the heart disease. Three Hyper Parameter Optimization (HPO) techniques Grid Search, Randomized Search and Genetic programming (TPOT Classifier) were proposed to optimize the performance of Random forest classifier and XG Boost classifier model. The performance of the two models Random Forest and XG Boost were compared with the existing studies. The performance of the models is evaluated with the publicly available datasets Cleveland Heart disease Dataset (CHD) and Z-Alizadeh Sani dataset. Random Forest along with TPOT Classifier achieved the highest accuracy of 97.52%for CHD Dataset. Random Forest with Randomized Search achieved the highest accuracy of 80.2%, 73.6% and 76.9% for the diagnosis of the stenos is of three vessels LAD, LCX and RCA respectively with Z-Alizadeh Sani Dataset. The results were compared with the existing studies focusing on prediction of heart disease that were found to outperform their results significantly.

1. Introduction

By 2030, the deaths due to cardiovascular disease is expected to increase to 23.3 million [1]. The blood vessels in the heart supplies the oxygen and when these vessels get blocked or narrowed, it can lead to any heart disease or stroke [2]. According to WHO, every year around 12 million people fall as a victim to the heart disease and 80% of the people dies due to heart ailment [3]. High Blood pressure, High Cholesterol, stress, tension, consumption of alcohols, sedentary lifestyle, obesity, diabetes are the major factors that affects the heart. These attributes helps in the prediction of heart disease. Due to increased blood pressure the walls of the arteries get thickened that causes blockage, which can increase the mortality rate [4]. Early diagnosis of heart disease, proper treatment can prevent and can also reduce the mortality rate of the patients. One of the common method to diagnose the abnormal narrowing of heart vessel is angiography. The symptoms, examination and ECG features were investigated with SMO, Naïve Bayes and Ensemble method and reached an accuracy of 88.5% to predict the presence of CAD [31]. Numerous Supervised and Unsupervised Machine learning algorithms have been applied by a number of researchers in the medical field for diagnosis and prediction of the heart disease [5].

The ability to detect patterns, turning data into information, data mining serves as a strong base for analysis [6]. Machine learning techniques helps to derive useful knowledge to take decision from vast datasets [7]. These algorithms have been broadly used in the area of Health Care, Computer vision, Speech recognition, Social Science, Cosmology and in the Education field [8]. They provide a variety of algorithms to identify the different patterns in large dataset [9]. ML has been widely used in the health care industry for identifying the disease and making effective decisions. It helps to classify the patients with significant risk factors [10–12]. Massive amounts of data is collected by the healthcare industry and ML provides different models to train and analyze the data quickly [13]. These algorithms search through a large search space of solutions and finds an optimal solution by training the dataset. The performance of the models can be examined from the various performance metrics such as Accuracy, Sensitivity, Specificity, Precision and F1-Score. By tuning the hyper parameters, the best hyper parameters are selected and applied to the ML models and the performance is improved. Robust models can be built by adjusting the hyperparameters. Overfitting or Underfitting can be prevented by tuning the hyperparameters. The major contributions of our proposed model is as follows

1. The performance of the models are tested on a subset of features selected by Sequential Forward Selection (SFS) method with 10-fold cross validation for Cleave land Heart disease dataset.

* Corresponding author.
E-mail addresses: valarmathisankar2014@gmail.com (R. Valarmathi), hod.it@sairam.edu.in (T. Sheela).

https://doi.org/10.1016/j.bspc.2021.103033
Received 9 March 2021; Received in revised form 12 July 2021; Accepted 30 July 2021
Available online 12 August 2021
1746-8094/

2. SMOTE technique is applied to balance the dataset in Z-Alizadeh Sani dataset
3. Next, the model performance are tuned and tested with Grid Search CV, Randomized Search CV and Genetic Programming (TPOT Classifier) with 10-fold cross validation.
4. Our work suggests the combination of the ML models and optimization technique that predicts the heart disease with highest accuracy.

The remainder of this paper is structured as follows: Section 2 covers related work on the previous studies employed in heart disease prediction system using different machine learning algorithms. Section 3 demonstrates the methodology of the proposed work in more detail. Section 4 describes the experimental results, comparative analysis of the previous studies and techniques used. Section 5 outlines our findings and future research directions.

2. Literature review

The performance of different datasets were analyzed using Bayesian Optimization based on Gaussian process [14]. The performance of 6 different machine learning models were examined on the heart dataset and found logistic regression predicted the heart disease with the highest accuracy [15]. Least significant features are eliminated using the backward feature elimination. The association rules were mined using frequent item sets and the genetic algorithms is applied to predict the heart disease [16]. Fitness function was used to remove the redundant rules and in the optimization of association rules were generated.

Three neural network model to were used to construct an ensemble model to diagnose the heart disease [17]. SAS enterprise miner 5.2 was used to evaluate the performance. The number of NN was increased but no improvement was observed in the performance.270 Patient records were trained and tested using Cascaded neural network, a Self Organizing network and Support Vector machine with RBF function [18]. Naïve Bayes machine learning model was developed to predict the heart disease [19].

An enhanced SVM classifier was presented to classify the linear and nonlinear inputs. PSO was used in feature extraction and Fuzzy C-means Clustering was applied to improve the accuracy [20]. Bhatla and Jyoti [21] employed Weka Tool on different data mining techniques for heart disease prediction and observations showed that Neural network showed good results compared to other data mining techniques. Krishnaiah et al. [22] incorporated a fuzzy approach to remove the uncertainty in the data and applied a KNN Classifier to classify the heart patients.

Amin et al. [23] identified the risk factors of 50 patients and implemented an integrated model of genetic algorithm and neural network to predict the presence of heart disease. Abdeldjouadet al. [24] used two different software's Weka and Keel tool to build two different models. First model was built by applying the PCA feature selection method to extract the significant features and 3 different classification algorithms Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), Logistic Regression (LR), Adaptive Boosting (AdaBoostM1) using Weka tool for classification. The second model was built by applying the Wrapper method for feature extraction and Genetic Fuzzy System-LogitBoost (GFS-LB), Fuzzy Unordered Rule Induction Algorithm (FURIA) and Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML) for classification under Keel tool. The two models were completely trained and the performance were evaluated.

Purusothaman and Krishnakumari [25] surveyed the different research findings based on single model approach and hybrid model and concluded hybrid model are better in prediction of disease compared to a single model. Khourdifiet al. [26] optimized KNN, RF, SVM, Naïve Bayes and ANN with the combination of Particle Swarm optimization and ant colony optimization. Kalaiselvi and Nasira [27] used PSO for extraction of data and ANFIS with AGKNN was used in classification. Santhanam and Ephzibah [28] has taken genetic algorithm for feature extraction and fuzzy logic for prediction. Mohan et al. [29] developed a hybrid approach of random forest and Linear method for classification of heart disease. KaanUyar et al. [30] proposed a genetic algorithm based recurrent fuzzy neural networks (RFNN) to classify the data. None of the aforementioned studies have implemented hyperparameter optimization (HPO) techniques to boost the accuracy of the heart disease prediction system. Thus, in our proposed model we used HPO techniques to improve the accuracy of the model.

Alizadehsani et al. [32] employed the rule based classifier and cost sensitive algorithm along with Sequential minimal optimization (SMO) to diagnose CAD. Alizadehsani et al. [33] handled data uncertainty. Different evolutionary algorithms were used for feature selection [33-38,65]. Some of the existing studies presented the Real time predictions and the performance of detection of heart disease using hardware. Based on five characteristics Cardiac arrhythmias were differentiated using Multi-Level Support vector machine classifiers [66]. Patient Specific SCAD processor [67], Smart ECG processor [68], Wearable ECG Processor [69] was designed to discriminate the CA in real time. An ECG processor and STAC algorithm was presented to improve the accuracy of heart rate detection [70].

3. Materials and methods

3.1. Proposed methodology

Original datasets were collected and data preprocessing is done on the collected data. Relevant features were selected using sequential forward selection method. The parameters of Random forest and XGBoost were tuned using the hyper parameter optimization techniques Grid search, Randomized search and Genetic Programming (TPOT Classifier). Finally, the models were validated and analyzed to predict the heart disease.

In our proposed model, 10-cross validation is used to validate the data. Fig. 1 shows the flowchart of the proposed model.

3.2. Dataset description

3.2.1. Cleveland heart disease (CHD)

The Cleveland Heart Disease (CHD) dataset is a heart disease prediction dataset available online in UCI repository [45]. The actual dataset contains 76 attributes but most of the published articles used only 14 attributes. The heart disease prediction dataset takencontains13 independent variables and one dependent target variable, a total of 14 columns. The target variable has two classes that represent the presence and absence of heart disease. It has 303 rows. The dataset is found to have no missing values. The dataset description is given in Table 1

3.2.2. Z-Alizadeh Sani Dataset

Z-Alizadeh Sani dataset contains 303 samples and 54 attributes. There are two classes for diagnosis that represent normal and patient affected with coronary artery disease (CAD). Among 303 samples, 216 samples indicate normal patients and 87 samples indicate the presence of heart disease [46]. The dataset description is given in Table 2.

3.3. Feature selection

The features that are irrelevant decreases the performance of the model. In our paper, we have used Sequential Forward Selection with ten-fold cross validation to remove the irrelevant features. Sequential Forward Selection (SFS) selects one best single feature, then best pair is selected, then best triplet of features selected and this procedure is continued until n number of relevant features are selected. Table 2 shows the number of features and the highest accuracy obtained with irrelevant features removed. Random Forest and XGBoost is tested with SFS. Table 3 shows the model and their highest accuracies obtained with

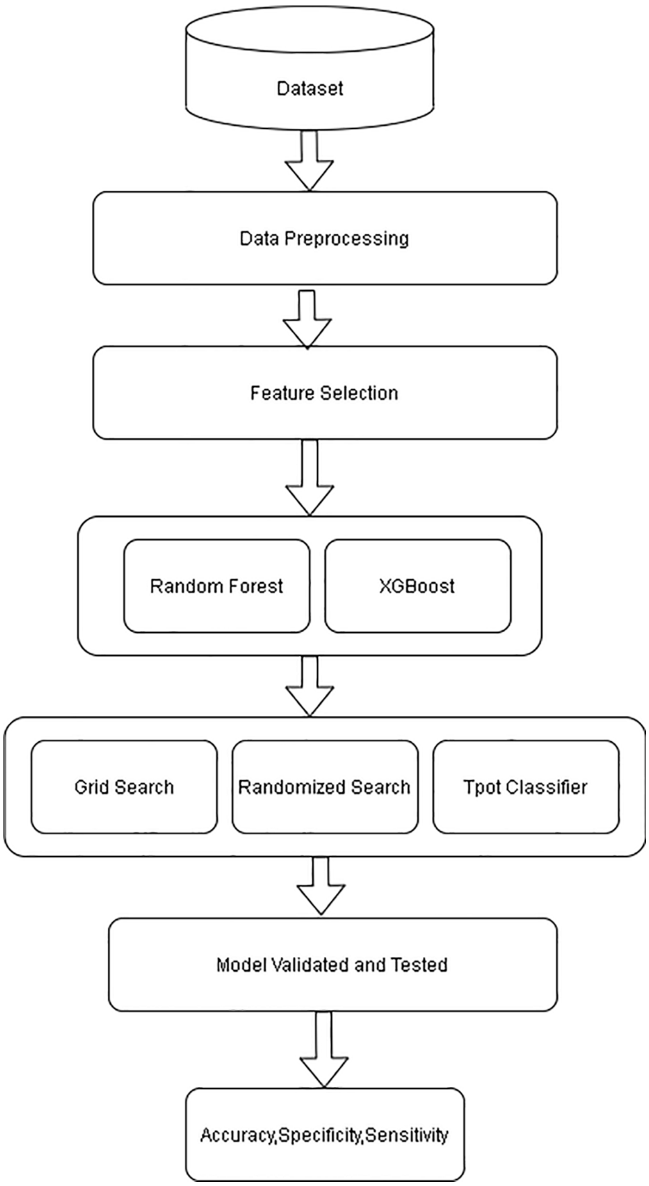

Fig. 1. The Proposed Model of the Heart Disease Prediction system.

Table 1
Dataset description(CHD).

S.No	Feature	Description
1	age	Age
2	sex	male, female
3	cp	chest pain type
4	trestbps	resting blood pressure
5	chol	serum cholesterol
6	fbs	fasting blood sugar
7	restecg	resting electrocardiographic results
8	thalach	maximum heart rate achieved
9	exang	exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	slope of the peak exercise ST segment
12	ca	number of major vessels
13	thal	Type of Defect
14	target	Risk, No risk

reduced number of features. Random forest achieved the highest accuracy of 83.96% with 9 relevant features. XGBoostobtained the highest accuracy of 82.7% with 11 relevant features. Selecting relevant features

Table 2
Dataset Description(Z-Alizadeh Sani Dataset).

Feature Type	Attribute(Values)
Demographic	Age(30–86)
	Weight(48–120)
	Length(140–188)
	Sex(M,F)
	Body Mass Index(BMI in 18.1–40.9)
	Diabetes Milletus(DM)(Y,N)
	Hypertension(HTN)(Y,N)
	Current Smoker(Y,N)
	Ex-Smoker(Y,N)
	Family History(FH) (Y,N)
	Obesity(Y,N)
	Chronic Renal Failure(CRF) (Y,N)
	Cerebrovascular Accident(CVA) (Y,N)
	Airway Disease(Y,N)
	Thyroid Disease(Y,N)
	Congestive Heart Failure(CHF) (Y,N)
	Dyslipidemia(DLP) (Y,N)
Symptom and Examination	Blood Pressure(BP in 90–190)
	Pulse Rate(PR in 50–110)
	Edema(Y,N)
	Weak peripheral pulse(Y,N)
	Lung Rales(Y,N)
	Systolic murmur(Y,N)
	Diastolic murmur(Y,N)
	Typical Chest Pain(Y,N)
	Dyspnea(Y,N)
	Function Class(1,2,3,4)
	Atypical(Y,N)
	Nonanginal(Y,N)
	Exertional CP(Y,N)
	LowTHAng(low Threshold angina) (Y,N)
	Rhythm(Y,N)
ECG	QWave(0,1)
	ST Elevation(0,1)
	ST Depression(0,1)
	T inversion(0,1)
	Left Ventricular Hypertrophy(LVH) (Y,N)
	Poor R Progression(Y,N)
Laboratory and Echo	Fasting Blood Sugar(FBS in 62–100 mg/dl)
	Creatine(Cr in 0.5–2.2 mg/dl)
	Triglyceride(TG in 37–1050 mg/dl)
	Low density lipoprotein(LDL in 18–232 mg/dl)
	High density lipoprotein(HDL in 15.9–111 mg/dl)
	Blood Urea Nitrogen(BUN in 6–52 mg/dl)
	Erythrocyte Sedimentation rate(ESR in 1–90 mm/h)
	Hemoglobin(HB in8.9–17.6 g/dl)
	Potassium(K in 3.0–6.6 mEq/lit)
	Sodium(Nain 128–156 mEq/lit)
	White Blood Cells(WBC in 3700-18000cells/ml)
	Lymphocyte (Lymph in 7–60 %)
	Neutrophil(Neut in 32–89%)
	Platelet(PLT in 25–742/ml)
	Ejection Fraction(EF in 15–60%)
	Regional Wall Motion Abnormality(RWMA)(0,1,2,3,4)
	Valvular Heart Disease(Normal,Mild,Moderate, Severe)

Table 3
Model, Number of features and their accuracy Using Sequential Forward Selection (SFS).

Model	No of Features	Accuracy	No of Features	Accuracy
Random Forest	13	82.68%	9	83.96%
XGBoost	13	79.36%	11	82.7%

improves the accuracy of the proposed system. Figs. 2 and 3 shows the accuracy obtained with Random Forest and XGBoost applied with Sequential Feature Selection(SFS). Table 4 and 5 shows the subset of features selected with Random Forest and XGboost.

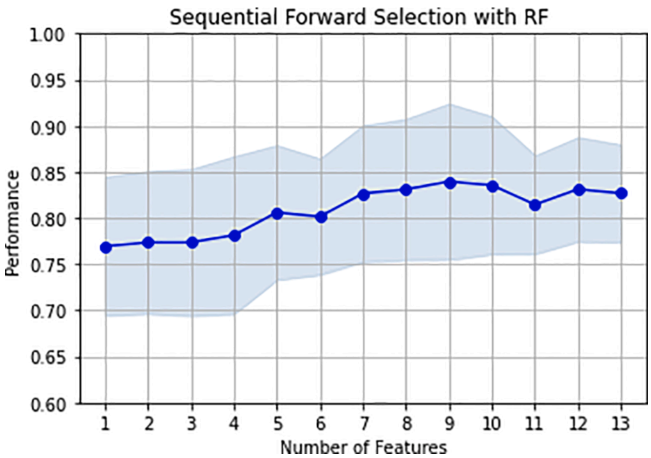

Fig. 2. Performance of RF with Sequential Forward Selection (SFS).

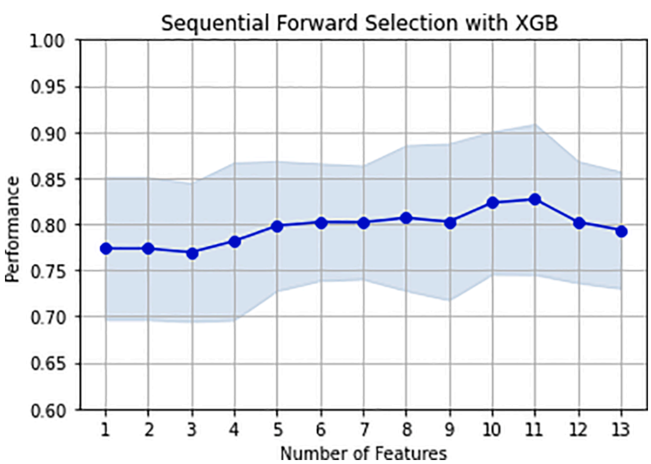

Fig. 3. Performance of XGBoost with Sequential Forward Selection (SFS).

Table 4
Features selected with Random Forest Model.

Feature no.	Feature Name
0	age
1	sex
3	trestbps
4	chol
5	fbs
8	exang
10	slope
11	ca
12	thal

Table 5
Features selected with XGBoost Model.

Feature no.	Feature Name
0	age
1	sex
3	trestbps
4	chol
5	fbs
6	restecg
7	thalach
8	exang
10	slope
11	ca
12	thal

3.4. Building Machine learning model

3.4.1. Random Forest (RF)

Random forest is a combination of tree predictors proposed by Breiman [39].He defines random forest as "A random forest is a classifier consisting of a collection of trees structured classifiers {h(x,Θk), k = 1, …} where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x". In Random forest, several decision tree are built on selecting the features and observations randomly and averaging the predictions. Many different trees are grown and depending upon how the trees are built and randomness introduced, a variety of random forest exists. Most commonly used hyperparameters to optimize the RF model aren_estimators, max_features, min_sample_leaf, max_features, max_depth and criterion. The parameters n_estimators, max_features and min_sample_leaf influences the accuracy prediction value.

3.4.2. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an ensemble technique [40], in which a set of weak learners are combined to improve the accuracy. The training speed and the learning effect of the XGBoost model created wide attention towards research community. XGBoost is an enhancement of Gradient Boosting Decision Tree (GBDT) algorithm. It uses the CART model for classification and regression problems. The model is trained by adding a tree and splitting the features in each iteration to grow the tree. At the end of the training, a score of each leaf node is obtained by multiplying the weight with the predicted value of the tree. The hyperparameters min_samples_split, min_samples_leaf and max_depth controls overfitting. These parameters influences each individual tree in the model. The parameters learning_rate, n_estimators, subsample enhances the boosting operation. These hyperparameter are tuned using the hyper optimization techniques Grid Search CV, Randomized Search CV and Genetic Programming (TPOT Classifier) and the best hyperparameters are selected to improve the performance of the proposed system.

3.5. Hyper parameter optimization (HPO)

Selecting the best hyper parameters has a significant impact on the performance model. Various optimization techniques exist and they have their own advantages and disadvantages. Experiments were performed on different optimization techniques to discover the best combination of hyper parameter and then applied to the Random Forest and XG Boost. Tuning the machine learning model is considered to be one of the optimization problem. The machine learning model consists of two types of parameters.1. Hyperparameter and 2. Model Parameters. Hyper parameters must be set randomly by the User before training the Machine learning Model. The parameters are used to control the process of learning. Different learning rates or weights are used to control the process of learning and to discover the patterns hidden in the data for the same type of machine learning model. These hyper parameters are tuned to minimize the error and maximize the accuracy of the model. Based on trial and error method, these parameters are tuned and the best hyper parameters are determined. The best hyper parameters balances the Over fitting and Under fitting. Choosing good hyper parameters helps in exploring the search space efficiently. The performance of the RF and XG Boost model can be improved by tuning the Hyper parameters. Model parameters learns during the training phase.

Grid and Random Search approaches are often used in hyper parameter optimization. Grid Search is a traditional technique, in which all hyper parameter combinations are evaluated. Grid Search uses learning rate and number of layers as hyper parameters. Initially a subset of values are defined for each hyper parameter. In each iteration, the combination of hyper parameters are estimated. Finally, the best hyper parameter combination are taken and applied for the learning process. The search space is restricted to a grid shaped subset and is not suitable for high dimensional space. Random search samples the search space from the equally distributed search space [41].

The researcher [42] stated that only few parameters had a large impact in the optimization of model score. Grid Search spends more time to find an unimportant parameter. Random search provided better choice of hyper parameter combination compared to Grid Search. Random search focuses on exploring the hyper parameter that has greater impact in improving the model score. Random search works better under the assumption that all parameters are not equally important. In random search the experiments were conducted separately, but there was no way to use the information obtained in one experiment to the next. These two approaches avoid the model falling in Local optima [43]. The disadvantages are these two approaches are time consuming and not suitable for data having high dimensional space. Fig. 4 shows the comparison between grid layout and random layout.

The Tree-Based Pipeline Optimization Tool (**TPOT**) is a Genetic Programming (GP) based Auto ML system that optimizes the ML models automatically [44]. TPOT uses meta learning techniques to optimize the machine learning pipelines using GP primitives to solve a particular problem. Auto ML tool is proposed in which feature selection, feature preprocessing, feature construction, model selection and parameter optimization takes place automatically. The genetic algorithm finds the best parameters. This AutoML considers multiple machine learning algorithms, multiple preprocessing steps and multiple ways to ensemble.

3.6. Performance metrics

Confusion matrix was employed to evaluate the performance of the two models. True Positive (TP) is defined as the count of predicted values correctly identified the presence of disease. True negative is defined as the count of predicted values correctly identified the absence of disease. False Positive (FP) is defined as the count of predicted values incorrectly classified as positive (actually when it was negative). False Negative (FN) is defined as the count of predicted values incorrectly classified as negative (actually when it was positive).

Once the model is trained, the risk of heart disease is predicted and evaluated with ten-fold cross validation. The analysis were done with the Performance metrics Accuracy, Specificity, Sensitivity, Precision and ROC-AUC values.

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

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

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

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

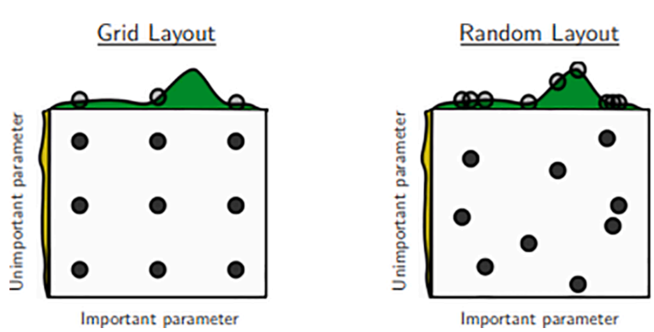

Fig. 4. Comparison between Grid and Random search Layout [42].

4. Experimental results

4.1. Experimental setup

The experiments were implemented in Python 3.0 on DELL, Intel (R) Core (TM) i5-8250U CPU @ 1.60 GHz, RAM 8 GB withWindows10.

4.2. Results and analysis

In this study, the effect of hyper parameter on the predictive performance of two different machine learning models Random Forest and XGBoost were examined. Three different hyper parameter optimization methods: Grid Search, Randomized Search and Genetic programming (TPOT Classifier) were used to optimize the ML models. A comparative analysis of the predictive performance of the 2 algorithms RF and XG Boost with 3 hyper parameter techniques were carried out in the experiments. Each algorithm is analyzed by selecting different hyper parameters. We compared the different hyper parameter optimization results of Randomized Search, Grid Search, Genetic Programming algorithms with the results of existing techniques.80% and 20% of data were taken for training and testing respectively. In our study we used 10-fold Cross Validation. Previous studies demonstrated 10-fold cross validation provides generalized model and avoids over fitting [47,48].

In our research, presence of heart disease is represented as 1 and absence of heart disease as 0. We implemented the three hyper parameter optimization techniques using Scikit Python library. Scikit-learn comes with the built-in functionality for Hyperparameter tuning techniques.

4.3. Experimental results with Random Forest model (Cleave land Dataset)

The Random Forest model were evaluated with different hyper parameter values. Table 6 shows the best hyper parameters obtained with different optimization techniques for Random Forest.

To validate the performance of the model, confusion matrix is used. The correct and incorrect prediction of the classifier is represented in a 2 × 2 matrix. Confusion matrix is depicted in Tables 7–9 for the Random forest Classifier with ten fold cross validation.

Table 10 shows the performance analysis of Random Forest with different optimization techniques.

4.4. Result analysis with XG Boost Model (Cleave land Dataset)

The XG Boost model were evaluated with different hyper parameter values. Table 11 shows the best hyperparameters obtained with different optimization techniques for XG Boost model.

Confusion matrix for the XG Boost classifier is depicted in Tables 12–14 for ten fold cross validation.

Table 15 shows the experimental results of Extreme Gradient Boosting (XGBoost) model.

Table 6
Best hyper parameters obtained with different optimization techniques for Random Forest for Cleave land Dataset.

Model	Parameters	Grid Search	Randomized Search	Genetic Programming (TPOT)
Random Forest	n_estimators	200	555	1333
	min_samples_split	3	2	2
	min_samples_leaf	3	2	2
	max_features	sqrt	log2	sqrt
	max_depth	890	670	780
	criterion	gini	gini	gini

Table 7
Confusion Matrix for RF Using Grid Search for Cleaveland Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	98(88.28%)	8	106
	Present	13	123	136
	Total predicted	111	131(93.89%)	

Table 8
Confusion Matrix for RF Using Random Search for Cleave land Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	103(92.79%)	4	107
	Present	8	127(96.95%)	135
	Total predicted	111	131	

Table 9
Confusion Matrix for RF Using Genetic Programming(TPOT Classifier)for Cleave land Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	108(97.29%)	3	111
	Present	3	128(97.70%)	131
	Total predicted	111	131	

Table 10
Performance analysis of Random Forest with different optimization techniques for Cleave land Dataset.

	Grid Search	Randomized Search	Genetic Programming (TPOT)
AUC-ROC(%)	91.09	94.80	97.50
Accuracy(%)	91.32	95.04	97.52
Sensitivity (%)	88.28	92.79	97.29
Specificity (%)	93.89	96.95	97.70
Precision (%)	92	96	97
F1-Score(%)	90	94	97

Table 11
Best Hyper parameters of XG Boost model with different optimization techniques for Cleave land Dataset.

Model	Parameters	Grid Search	Randomized Search	Genetic Programming (TPOT)
XGBoost	Learning rate	0.01	0.01	0.01
	gamma	1	0.5	2
	max_depth	3	4	3
	min_child_weight	8	1	1
	n_estimators	600	266	1111
	subsample	0.8	0.8	0.70

Table 12
Confusion Matrix for XG Boost Using Grid Search for Cleave land Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	92(82.88%)	14	106
	Present	19	117(89.31%)	136
	Total predicted	111	131	

Table 13
Confusion Matrix for XGBoost Using Random Search for Cleave land Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	100(90.09%)	8	108
	Present	11	123(93.89%)	134
	Total predicted	111	131	

Table 14
Confusion Matrix for XGBoost Using Genetic Programming (TPOT Classifier) for Cleave land Dataset.

	Predicted Class			
Actual Class		Absent	Present	Actual Value
	Absent	97(87.38%)	9	108
	Present	14	122(93.12%)	136
	Total predicted	111	131	

Table 15
Performance analysis of XG Boost Model with different optimization techniques for Cleave land Dataset.

	Grid Search	Randomized Search	Genetic Programming (TPOT)
AUC-ROC(%)	86.07	91.99	88.21
Accuracy(%)	86.36	92.14	90.50
Sensitivity (%)	82.88	90.09	87.38
Specificity (%)	89.31	93.89	93.12
Precision (%)	87	93	92
F1-Score(%)	85	91	89

4.5. *Experimental results with Random Forest (Z-Alizadeh Sani Dataset)*

The extension of Z-Alizadeh Sani dataset consists of 3 attributes LAD, LCX and RCA. The abnormal narrowing of these vessels in the human are considered as stenotic and others as normal. Synthetic Minority Oversampling Technique (SMOTE) is applied to overcome the imbalance problem in the dataset. Table 16 shows the Performance analysis of Random forest for Z-Alizadeh Sani Dataset.

4.6. *Experimental results with XGBoost (Z-Alizadeh Sani Dataset)*

Table 17 shows the experimental results of XGBoost Model.

4.7. *Comparative analysis of the models with previous studies*

Verma et al. [55] classified the heart disease by using a hybrid model (CFS, PSO, K-means, and MLP) and obtained an accuracy of 90.28%. Chadha and Mayank [50] carried out the experiments with Decision Tree (DT) and Naïve Bayes (NB) Model and obtained an accuracy of 88.03% and 85.86% respectively. Mohan et al. [60] developed a hybrid

Table 16
Performance analysis of Random forest for Z-Alizadeh Sani Dataset.

	Artery	Accuracy (%)	Sensitivity* %)	Specificity (%)
Grid Search	LAD	78.0	75.6	80.0
	LCX	70.3	79.6	56.8
	RCA	79.1	92.20	61.5
Randomized Search	LAD	80.2	75.6	84
	LCX	73.6	85.1	56.8
	RCA	76.9	94.2	74.0
TPot	LAD	75.8	73.2	78
	LCX	68.1	72.2	62.1
	RCA	74.7	88.4	56.4

Table 17
Performance of XGBoost for Z-Alizadeh Sani Dataset.

	Artery	Accuracy	Sensitivity	Specificity
Grid Search	LAD	74.7	78.0	72
	LCX	65.9	75.9	51.3
	RCA	70.3	80.76	56.4
Randomized Search	LAD	75.8	75.6	76
	LCX	71.4	79.6	59.5
	RCA	78.0	86.5	66.6
TPot	LAD	74.7	73.1	76
	LCX	69.2	75.9	59.4
	RCA	78.0	90.3	61.5

model (RF with Linear) and obtained an accuracy of 88.4%.Haq et al. [57] carried out the experiments on the combination of various feature selection methods and different Machine Learning models. Their research concluded the hybrid of Relief-based feature selection and Logistic regression achieved the highest accuracy of 89% compared to other models. Saqlain et al. [58] selected the significant features using Fisher Score algorithm and the subset of features were given to SVM and validated. The combination of MFSFSA and SVM obtained an accuracy of 81.19%.

Soni et al. [52] applied association rules to classify the disease and obtained an accuracy of 81.51%.Latha and Jeeva [59] employed a voting model with NB, BN, RF and MLP and found an accuracy of 85.48%.Leema et al. [54] proposed a Computer-Aided Diagnostic system which uses Differential Evolution for global search and Back propagation for local search. The accuracy obtained from the system is found to be 86.6%. Kumari and Godara [53] proposed a SVM model and obtained an accuracy of 84.12%.

Long et el. [51] proposed a chaos firefly algorithm to predict the disease and removed the irrelevant features using rough sets. The highest accuracy obtained from CFARS-AR is 88.3%.Six different Machine models were compared by Dwivedi et al. [49] and LR model obtained the highest accuracy of 85% compared to other models. Amin et al. [56] developed a hybrid voting model with Naïve Bayes and Logistic regression. The model was trained with the significant features. The results of the hybrid model showed an accuracy of 87.41%.Ayon et al [61] made a comparative study of seven machine learning models and found DNN performed better compared to all other algorithms. The outcome of the study showed random forest obtained an accuracy of 87.45% with 10-fold cross validation. Table 18 shows our proposed models is effective in classifying the heart disease when compared to the previous studies.

Table 18
Comparison of the Model from previous studies for Cleave land Dataset.

Authors	Method	Results
Chadha and Mayank [50]	DT	88.03%
	NB	85.86%
Long et al [51]	CFARS-AR	88.3%
Soni et al [52]	Association rules	81.51%
Kumari and Godara [53]	SVM	84.12%
Leema et al [54]	DE + BP	86.6%
Verma et al [55]	CFS + PSO + K-means + MLP	90.28%
Amin et al [56]	Hybrid(NB + LR)	87.41%
A.K.Diwivedi [49]	SVM	82.00%
	LR	85.00%
Haq et al [57]	Relief -based feature selection + Logistic regression	89%
Saqlain et al [58]	MFSFSA + SupportVector Machine	81.19%
Latha and Jeeva [59]	Naïve Bayes + BN + RandomForest + MLP	85.48%
Mohan et al [60]	RF + Linear Model	88.4%
Ayon et al [61]	RF	87.45%
Proposed Model	RF With Grid Search	91.32%
	RF with Randomized Search	95.04
	RF With TPOT Classifier	97.52%
	XGBoost with Grid Search	86.36%
	XGBoost with Randomized Search	92.14
	XGBoost With TPOT classifier	90.50

Babaoglu et al. [62] employed ANN and achieved an accuracy of 73%,64.8% and 69.4%. Alizadehsani et al. [63] studied the presence of CAD by applying different feature selection information gain and Gini Index to extract the effective features and evaluated the performance with C4.5 and Bagging algorithm. Alizadeh sani et al. [64] examined the demographic, examination and ECG features and obtained an accuracy of 74.20%,63.76% and 68.33% for LAD, LCX and RCA respectively. Table 19 shows our proposed models is effective when compared to the previous studies

4.8. Performance analysis with ROC_AUC for Cleave land Dataset

Receiver operating characteristics is a visualization curve to compare the “true positive rate” and “false positive rate”. Area under the receiver operating characteristics is compared for both models for all the three hyper optimization techniques. The best model is found when the AUC value is closer or equal to 1. Fig. 5 shows the AUC-ROC curve for the experiments conducted with Grid Search. When experiments conducted with Grid Search, Random forest and XGBoost achieved an AUC of 91.09% and 86.09%

AUC-ROC curve for the experiments conducted with Randomized Search is illustrated in Fig. 6.When experiments were conducted with Randomized Search, Random forest and XGBoost achieved an AUC of 94.80% and 91.99% respectively.

Fig. 7 shows the AUC-ROC curve for the experiments conducted with Genetic Programming (TPOT Classifier). When experiments were conducted with Genetic Programming (TPOT Classifier), Random forest and XGBoost achieved an AUC of 97.50% and 88.21% respectively.

4.9. Performance Analysis with ROC_AUC for Z-Alizadeh Sani Dataset

Fig. 8–10 shows the AUC-ROC curve for the experiments conducted with Z-Alizadeh sani Dataset. The diagnosis of the stenos is with the vessel LAD achieved 77.8%, 79.8%,75.5% with Random forest and 75.0%,75.8% and 74.5% with XGBoost. The diagnosis of the stenos is with the vessel LCX achieved 68.1%,71%,67.2% with Random forest and 63.6%,69.5% and 67.7% with XGBoost. The diagnosis of the stenos is with the vessel RCA achieved 76.9%,74.0%,72.4% with Random forest and 68.5%,76.6% and 75.9% with XGBoost.

5. Conclusion

Every year, a lot of deaths happen due to heart diseases. Heart disease, if predicted earlier, can save many lives. In this study, Sequential forward selection is applied to remove the insignificant features. Removing the irrelevant features had a huge impact in the performance improving it significantly. The machine learning models: Random forest and XGBoost were tuned and tested with three different hyperparameter tuning techniques and their accuracies were compared to the existing techniques. By tuning the parameters of Random forest with Grid Search, Randomized search and Genetic Programming we obtained

Table 19
Comparison of the model from previous studies for Z-Alizadeh Sani Dataset.

Authors	Method	Accuracy (%)		
		LAD	LCX	RCA
Babaogluetal.[62]	ANN	73.0	64.8	69.4
Alizadehsani,Habibi, Alizadehsaniet al.[63]	Bagging	79.5	65.1	68.0
Alizadehsani,Habibi, Bahadorianet al.[64]	Decision Tree	74.2	63.8	68.3
Proposed Method	Random Forest with Randomized Search	80.2	73.6	76.9

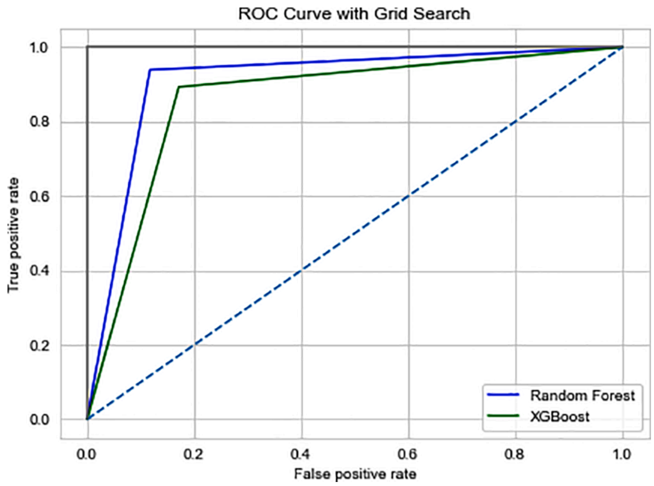

Fig. 5. ROC curve with Grid Search.

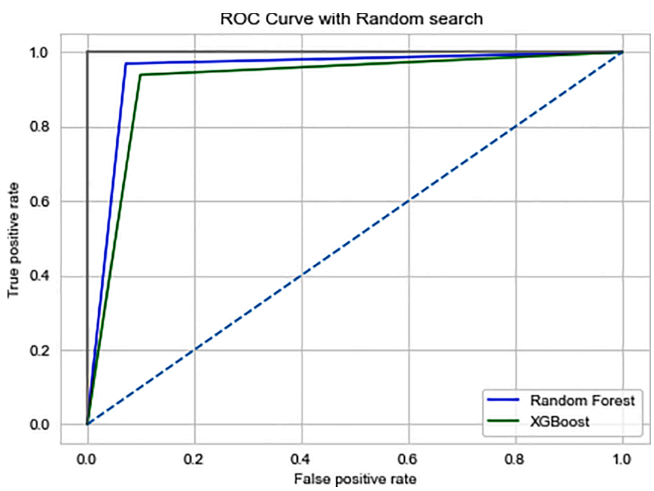

Fig. 6. ROC curve with Randomized Search.

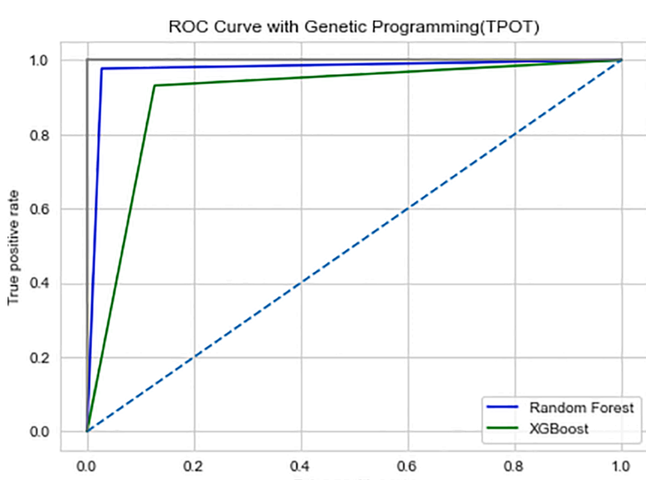

Fig. 7. ROC curve with Genetic Programming (TPOT Classifier).

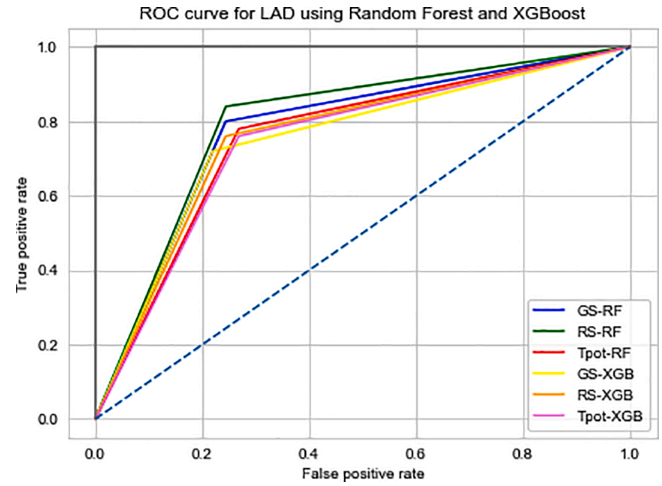

Fig. 8. ROC curve for LAD using Random Forest and XGBoost.

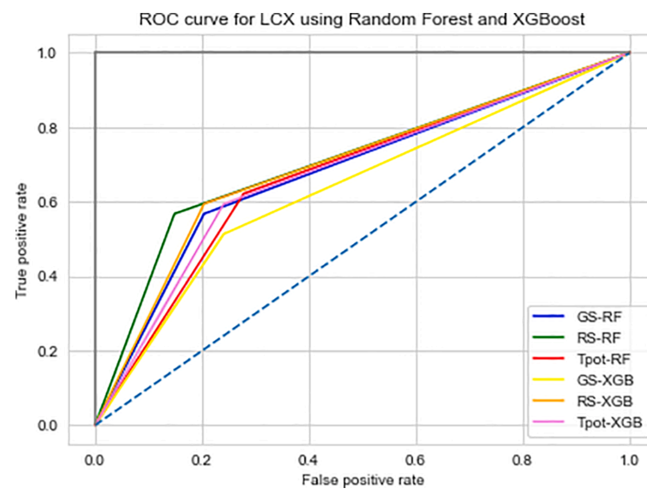

Fig. 9. ROC curve for LCX using Random Forest and XGBoost.

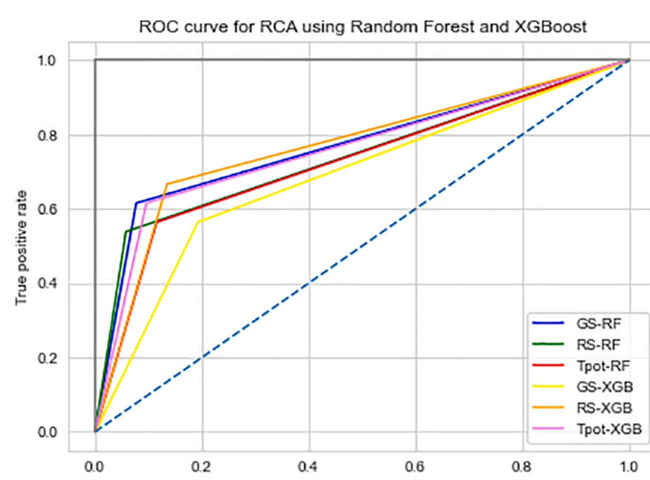

Fig. 10. ROC curve for RCA using Random Forest and XGBoost.

better results with 91.32%, 95.04% and 97.52% compared to XGBoost and other state of the art algorithms for CHD. The proposed algorithms shows the highest accuracy for random forest with randomized search with 80.2%, 73.6% and 76.9% for the diagnosis of abnormality in LAD, LCX and RCA vessels respectively. In the future, heart disease predictions can be done in real time and the performance can be evaluated in hardware.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.D. Mathers, D. Loncar, Projections of global mortality and burden of disease from 2002 to 2030, Plosmed 3 (11) (2006).
[2] M. Akhil, Dr. Priti Chandra, Dr. B.L Deekshatulu, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT, 2012.
[3] Ibrahim Umar Said, Abdullahi Haruna Adam, Dr. Ahmed BaitaGarko, "Association Rule Mining On Medical Data To Predict Heart Disease", International Journal of Science Technology and Management, August 2015, pp. 26-35.
[4] A. Chauhan, A. Jain, P. Sharma, V. Deep, Heart disease prediction using evolutionary rule learning, in: 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 2018, pp. 1–4.
[5] Ilaria Castelli, Edmondo Trentin, Combination of supervised and unsupervised learning for training the activation functions of neural networks, Pattern Recogn. Lett. 37 (2014) 178–191.
[6] R. Safdari, T. Samad-Soltani, M. GhaziSaeedi, M. Zolnoori, Evaluation of classification algorithms vs knowledge-based methods for differential diagnosis of asthma in Iranian patients, Int. J. Inform. Syst. Serv. Sect. 10 (2) (2018) 22–26.
[7] F.S. Alotaibi, Implementation of machine learning model to predict heart failure disease, Int. J. Adv. Comput. Sci. Appl. 10 (6) (2019) 261–268.
[8] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260, https://doi.org/10.1126/science.aaa8415.
[9] M. Motwani, D. Dey, D.S. Berman, G. Germano, S. Achenbach, M.H. Al-Mallah, D. Andreini, M.J. Budoff, F. Cademartiri, T.Q. Callister,"Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease": a 5-year multicentre prospective registry analysis, Eur. Heart J. 38(7), pp. 500–507,2016.
[10] Sani A," Machine Learning for Decision Making", Universitéde Lille 1, 2015,.
[11] W. Raghupathi, V. Raghupathi, "Big data analytics in healthcare: promise and potential", Health Inf. Sci. Syst. 2(3), 2014.
[12] P. Groves, B. Kayyali, D. Knott, S.V. Kuiken, "The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation", 2016.
[13] T. Condie, P. Mineiro, N. Polyzotis, M. Weimer, Machine learning on big data: data engineering (ICDE), in: 2013 IEEE 29th International Conference on, IEEE, 2013, pp. 1242–1244.
[14] Wu. Jia, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, Si-Hao Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (1) (2019).
[15] Debabrata Swain, Preeti Ballal, Vishal Dolase, Banchhanidhi Dash, Jayasri Santhappan, An efficient heart disease prediction system using machine learning. Advances in Intelligent Systems and Computing, Springer Nature Singapore Pvt Ltd, 2020.
[16] M.A. Jabbar, B.L. Deekshatulu, P. Chandra, "An Evolutionary Algorithm for Heart Disease Prediction. In: K.R. Venugopal, L.M. Patnaik (eds.) Wireless Networks and Computational Intelligence". ICIP 2012. Communications in Computer and Information Science, vol 292. Springer, Berlin, Heidelberg. DOI:10.1007/978-3-642-31686-9_44.
[17] R. Das, I. Turkoglu, A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, Expert Syst. Appl. 36 (4) (2009) 7675–7680.
[18] R. Chitra, V. Seenivasagam, Heart disease prediction system using supervised learning classifier, Bonfring Int. J. Software Eng. Soft Comput. 3 (1) (2013) 1.
[19] K. Vembandasamy, R. Sasipriya, E. Deepa, Heart diseases detection using Naive Bayes Algorithm, Int. J. Innov. Sci. Eng. Technol. 2 (9) (2015).
[20] R. Kavitha, T. Christopher, "An effective classification of heart rate data using PSO-FCM clustering and enhanced support vector machine," Indian Journal of Science and Technology, 8(30), 2015.
[21] N. Bhatla, K. Jyoti, An analysis of heart disease prediction using different data mining techniques, Int. J. Eng. Res. Technol. 1 (8) (2012) 1–4.
[22] V. Krishnaiah, G. Narsimha, N.S. Chandra, "Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach. In: Satapathy S., Govardhan A., Raju K., Mandal J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham,2015. DOI:10.1007/978-3-319-13728-5_42.
[23] S. U. Amin, K. Agarwal and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," in: 2013 IEEE Conference on Information & Communication Technologies, Thuckalay, Tamil Nadu, India, 2013, pp. 1227-1231, doi: 10.1109/CICT.2013.6558288.
[24] F.Z. Abdeldjouad, M. Brahami, N. Matta. " A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques.", In: M. Jmaiel, M. Mokhtari, B. Abdulrazak, H. Aloulou, S. Kallel (eds.) The Impact of Digital Technologies on Public Health in Developed and Developing Countries. ICOST 2020. Lecture Notes in Computer Science, vol 12157. Springer, Cham, 2020. DOI: 10.1007/978-3-030-51517-1_26.
[25] G. Purusothaman, P. Krishnakumari, "A survey of data mining techniques on risk prediction: heart disease", Indian J. Sci. Technol. 8(12), 2015.
[26] Youness Khourdifi, Mohamed Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst. 12 (1) (2018) 242–252.
[27] C. Kalaiselvi, G. Nasira, Prediction of heart diseases and cancer in diabetic patients using data mining techniques. Indian J. Sci. Technol. 8(14), 2015.
[28] T. Santhanam, E. Ephzibah, Heart disease prediction using hybrid genetic fuzzy model", Indian, J. Sci. Technol. 8 (9) (2015) 797–803.
[29] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554, https://doi.org/10.1109/ACCESS.2019.2923707.
[30] Kaan Uyar, Ahmet İlhan, Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks, Procedia Comput. Sci. 120 (2017) 588–593.
[31] Roohallah Alizadehsani, Jafar Habibi, Javad Hosseini Mohammad, Reihane Boghrati, Asma Ghandeharioun, Behdad Bahadorian, et al., Diagnosis of Coronary Artery Disease Using Data Mining Techniques Based on Symptoms and ECG Features, European Journal of Scientific Research 82 (4) (2012) 542–553.
[32] Alizadehsani Roohallah et al. "Exerting Cost-Sensitive and Feature Creation Algorithms for Coronary Artery Disease Diagnosis."IJKDB 3.1 (2012): 59-79. Web. 1 May. 2021. doi:10.4018/jkdb.2012010104.
[33] Roohallah Alizadehsani, Mohamad Roshanzamir, Moloud Abdar, Adham Beykikhoshk, Mohammad Hossein Zangooei, Abbas Khosravi, Saeid Nahavandi, Ru San Tan, U. Rajendra Acharya, Model uncertainty quantification for diagnosis of each main coronary artery stenosis, Soft Computing 24 (13) (2020) 10149–10160, https://doi.org/10.1007/s00500-019-04531-0.
[34] Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, Hoda Mashayekhi, Reihane Boghrati, Asma Ghandeharioun, Behdad Bahadorian, Zahra Alizadeh Sani, A data mining approach for diagnosis of coronary artery disease, Comput. Methods Programs Biomed. 111 (1) (2013) 52–61, https://doi.org/10.1016/j.cmpb.2013.03.004.
[35] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, M. Panahiazar, A. Koohestani, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, A database for using machine learning and data mining techniques for coronary artery disease diagnosis, Scientific Data 6 (1) (2019), https://doi.org/10.1038/s41597-019-0206-3.
[36] Roohallah Alizadehsani, Moloud Abdar, Mohamad Roshanzamir, Abbas Khosravi, Parham M. Kebria, Fahime Khozeimeh, Saeid Nahavandi, Nizal Sarrafzadegan, U. Rajendra Acharya, Machine learning-based coronary artery disease diagnosis: a comprehensive review, Comput. Biol. Med. 111 (2019) 103346, https://doi.org/10.1016/j.compbiomed.2019.103346.
[37] Roohallah Alizadehsani, Mohammad Hossein Zangooei, Mohammad Javad Hosseini, Jafar Habibi, Abbas Khosravi, Mohamad Roshanzamir, Fahime Khozeimeh, Nizal Sarrafzadegan, Saeid Nahavandi, Coronary artery disease detection using computational intelligence methods, Knowledge-Based Syst. 109 (2016) 187–197, https://doi.org/10.1016/j.knosys.2016.07.004.
[38] R. Alizadehsani, M.J. Hosseini, Z.A. Sani, A. Ghandeharioun, R. Boghrati. Diagnosis of coronary artery disease using cost-sensitive algorithms. in: Paper presented at the 2012 IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium 2012.
[39] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32.
[40] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system,''in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785_794, doi:10.1145/2939672.2939785.
[41] Douglas C Montgomery, Design and Analysis of Experiments, John Wiley & Sons, 2017.
[42] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (2012) 281–305.
[43] R. Liessner, J. Schmitt, A. Dietermann, and B. Bäker, "Hyperparameter Optimization for Deep Reinforcement Learning in Vehicle Energy Management", in: Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019), pages 134-144 ISBN: 978-989-758-350-6.
[44] R.S. Olson, J.H. Moore, TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning, in: F. Hutter, L. Kotthoff, J. Vanschoren (Eds.), Automated Machine Learning. The Springer Series on Challenges in Machine Learning, Springer, Cham, 2019, https://doi.org/10.1007/978-3-030-05318-5_8.
[45] https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.names.
[46] http://archive.ics.uci.edu/ml/datasets/extention+of+Z-Alizadeh+sani+dataset.
[47] R. Kohavi, D. Wolpert Bias plus variance decomposition for zero-oneloss functions. in: Proc 13th Int. 1996 Conf. Mach. Learn., San Francisco, CA, USA pp. 275_283.
[48] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, San Diego, CA, USA, 2012.
[49] Ashok Kumar Dwivedi, Performance evaluation of differentmachine learning techniques for prediction of heartdisease, Neural Comput. Appl. 29 (10) (2018) 685–693.
[50] Ritika Chadha, Shubhankar Mayank, Prediction of heart diseaseusing data mining techniques, CSI Trans. ICT 4 (2-4) (2016) 193–198.
[51] N.C. Long, P. Meesad, H. Unger, A highly accuratefirefly-based algorithm for heart disease prediction, Expert. Syst. Appl. 42 (2015) 8221–8231.
[52] J. Soni, U. Ansari, D. Sharma, S. Soni, Intelligent andeffective heart disease prediction system using weightedassociative classifiers, Int. J. Comput. Sci. Eng. 3 (6) (2011) 2385–2392.
[53] M. Kumari, S. Godara, Comparative study of datamining classification methods in cardiovascular diseaseprediction, IJCST 2 (2) (2011) 304–308.

[54] N. Leema, H. Khanna Nehemiah, A. Kannan, Neural network classifier optimization using differential evolutionwith global information and back propagation algorithm for clinical datasets, Appl. Soft. Comput. 49 (2016) 834–844.
[55] L. Verma, S. Srivastava, P.C. Negi, A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data, J. Med. Syst. 40 (7) (2016) 178, https://doi.org/10.1007/s10916-016-0536-z.
[56] M. S. Amin, Y. K. Chiam, and K. D. Varathan, ''Identication of signicant features and data mining techniques in predicting heart disease,'' Telematics Inform., vol. 36, pp. 82_93, Mar. 2019, doi: 10.1016/j.tele.2018.11.007.
[57] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, ''A hybrid intelligent system framework for the prediction of heart disease using machinelearning algorithms,'' Mobile Inf. Syst., vol. 2018, pp. 1_21, Dec. 2018,doi: 10.1155/2018/3860146.
[58] S. M. Saqlain, M. Sher, F. A. Shah, I. Khan, M. U. Ashraf, M. Awais,and A. Ghani, ''Fisher score and matthews correlation coef_cient-basedfeature subset selection for heart disease diagnosis using support vectormachines,'' Knowl. Inf. Syst., 58(1), pp. 139_167, Jan. 2019, doi:10.1007/s10115-018-1185-y.
[59] C. B. C. Latha and S. C. Jeeva, ''Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,'' Inform. Med. Unlocked, 16, Jan. 2019, Art. no. 100203, doi:10.1016/j.imu.2019.100203.
[60] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava, Effective heart diseaseprediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554, https://doi.org/10.1109/ACCESS.2019.2923707.
[61] Safial Islam Ayon, Md. Milon Islam, Md. Rahat Hossain, CoronaryArtery heart disease prediction: a comparative study of computational intelligence techniques, IETE J. Res. (2020), https://doi.org/10.1080/03772063.2020.1713916.
[62] I. Babaoglu, O. Fındık, M. Bayrak, Effects of principle component analysis on assessment of coronary artery diseases using support vector machine, Expert Syst. Appl. 37 (3) (2010) 2182–2185, https://doi.org/10.1016/j.eswa.2009.07.055.
[63] Roohallah Alizadehsani, Jafar Habibi, ZahraAlizadeh Sani, Hoda Mashayekhi, Reihane Boghrati, Asma Ghandeharioun, Fahime Khozeimeh, Fariba Alizadeh-Sani, Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features, Res Cardiovasc Med. 2 (3) (2013) 133, https://doi.org/10.5812/cardiovascmed.10888.
[64] Habibi Jafar Alizadehsani Roohallah, MashayekhiHoda Bahadorian Behdad, Ghandeharioun Asma, Reihane Boghrati, et al., Diagnosis of coronary arteries stenosis using data mining, J. Medical Signals Sens. 2 (3) (2012) 57–65.
[65] Jalali, Seyed Mohammad Jafar, Khosravi, Abbas, Alizadehsani, Roohallah, Salaken, Syed Moshfeq, Kebria, Parham Mohsenzadeh, Puri, Rishi and Nahavandi, Saeid 2019, Parsimonious evolutionary-based model development for detecting artery disease, in: ICIT 2019 : Proceedings of the IEEE International Conference on Industrial Technology, IEEE, Piscataway, N.J., pp. 800-805, doi: 10.1109/ICIT.2019.8755107.
[66] M. A. Sohail, Z. Taufique, S. M. Abubakar, W. Saadeh and M. A. Bin Altaf, "An ECG Processor for the Detection of Eight Cardiac Arrhythmias with Minimum False Alarms," in: 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2019, pp. 1-4, doi: 10.1109/BIOCAS.2019.8919053.
[67] S.M. Abubakar, M. Rizwan Khan, W. Saadeh, M. A. Bin Altaf, "A Wearable Auto-Patient Adaptive ECG Processor for Shockable Cardiac Arrhythmia," in: 2018 IEEE Asian Solid-State Circuits Conference (A-SSCC) 2018 267 268 10.1109/ASSCC.2018.8579263.
[68] Shihui Yin, Minkyu Kim, Deepak Kadetotad, Yang Liu, Chisung Bae, Sang Joon Kim, Yu Cao, Jae-Sun Seo, A 1.06- μ W Smart ECG Processor in 65-nm CMOS for Real-Time Biometric Authentication and Personal Cardiac Monitoring, IEEE J. Solid-State Circ. 54 (8) (2019) 2316–2326, https://doi.org/10.1109/JSSC.2019.2912304.
[69] S. M. Abubakar, W. Saadeh and M. A. B. Altaf, "A wearable long-term single-lead ECG processor for early detection of cardiac arrhythmia," in: 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2018, pp. 961-966, doi: 10.23919/DATE.2018.8342148.
[70] S. Izumi, et al., A 14 μA ECG processor with robust heart rate monitor for a wearable healthcare system, in: 2013 Proceedings of the ESSCIRC (ESSCIRC), 2013, pp. 145–148, https://doi.org/10.1109/ESSCIRC.2013.6649093.