ISSN (Online): 3139-7689

Available online: www.ijsrmes.com

International Journal for Scientific Research in Modern Engineering and Science

Published in Conference Special Issue: Vol. 1, Special Issue 1, pp. 187–199, 2026

HEART DISEASE RISK PREDICTION USING MACHINE LEARNING

Chaitanya Kumar Sahu[1*], Rohan Chandra [2], B. Bharadwaj [3], Shankar Saran Tripathi [4]

[123]Student, Computer Science and Engineering, Shri Shankaracharya Technical Campus (SSTC), Bhilai

[4]Assistant Professor, Computer Science and Engineering, Shri Shankaracharya Technical Campus

Abstract

Cardiovascular diseases (CVDs) kill about 20.5 million people every year. Early prediction can help people to change their lifestyles and to ensure proper medical treatment if necessary. In this research, ten machine learning (ML) classifiers from different categories, such as Bayes, functions, lazy, meta, rules, and trees, were trained for efficient heart disease risk prediction using the full set of attributes of the Cleveland heart dataset and the optimal attribute sets obtained from three attribute evaluators. The performance of the algorithms was appraised using a 10-fold cross-validation testing option. Finally, we performed tuning of the hyperparameter number of nearest neighbors, namely, 'k' in the instance-based (IBk) classifier. The sequential minimal optimization (SMO) achieved an accuracy of 85.148% using the full set of attributes and 86.468% was the highest accuracy value using the optimal attribute set obtained from the chi-squared attribute evaluator. Meanwhile, the meta classifier bagging with logistic regression (LR) provided the highest ROC area of 0.91 using both the full and optimal attribute sets obtained from the ReliefF attribute evaluator. Overall, the SMO classifier stood as the best prediction method compared to other techniques, and IBk achieved an 8.25% accuracy improvement by tuning the hyperparameter 'k' to 9 with the chi-squared attribute set.

Keywords: heart disease; data pre-processing; attribute evaluation; machine learning classifiers; hyperparameter tuning

1. Introduction

Cardiovascular disease (CVD) is one of the most serious health challenges in the world today. According to global health reports, millions of people lose their lives every year due to heart-related conditions. Common causes include high blood pressure, cholesterol, smoking, obesity, and lack of physical activity. Traditional diagnosis methods depend heavily on a doctor's experience and patient history, which may not always provide highly accurate predictions. With the rapid growth of healthcare data and advancements in machine learning, automated systems can now assist medical professionals in predicting heart disease risk more effectively. This research focuses on building an intelligent prediction system by applying various machine learning algorithms and feature selection techniques to improve early-stage detection accuracy.

20.5 million people die every year due to cardiovascular disease, which is approximately 31.5% of all deaths globally. It is also estimated that the number of annual deaths will rise to 24.2 million by 2030. About 85% of cardiovascular disease deaths are due to heart attack and strokes [1]. A heart attack is mainly caused when the

blood flow to the heart is blocked due to the build-up of plaque in the arteries. Stroke is caused by a blood clot in an artery within the brain, which cuts off blood circulation to the brain [2]. Heart disease is triggered mostly when the heart is unable to provide enough blood supply to parts of the body [3,4]. It results in early symptoms, such as an irregular heartbeat, shortness of breath, chest discomfort, sudden dizziness, nausea, swollen feet, and a cold sweat. The accurate prediction and proper diagnosis of heart disease in time are indispensable for improving the survival rate of patients. The risk factors that cause CVD include high BP, cholesterol,alcohol intake, and tobacco consumption, as well as obesity, physical inactivity, and genetic mutations. The early detection of signs and changes in lifestyle, such as physical activity, avoiding smoking, and appropriate medical examination by clinicians, can help to reduce mortality [5].

2. Literature Review

In recent years, Artificial Intelligence has risen to become one of the biggest technology advancements of the 21st century with some real-life applications now being used to aid in decision making processes in areas such as banking, insurance, medical diagnostics, cyber security and recommenders. Machine learning algorithms can take in large quantities of information and identify complex relationships that may be hard for humans to find easily. Most modern machine learning algorithms, however, are referred to as black boxes (e.g., deep learning neural networks, ensemble learning, gradient boosting). While these black box type machine learning algorithms can predict the output well, they do not give a user any insight into how the prediction was made.

Another important technique called SHAP (Shapley Additive Explanations) uses concepts from cooperative game theory to measure how much each feature contributes towards a final prediction by measuring the effect the feature has on the model's output, providing both global & local interpretations of a model. Numerous studies have demonstrated through numerous studies the usefulness of Explainable Artificial Intelligence through a variety of different applications including: Health Care Diagnostics, Fraud Detection & Financial Decision Making.

In one example of scientific research, scientists have applied techniques of Explainable AI (XAI) to provide an interpreter of a medical imaging model to explain to physicians what a model predicted for diagnoses. While these advancements have been made, a challenge still exists in striking the right balance between accurate and interpretable models. The model with the best interpretability, such as a decision tree, may not perform very well when it comes to predicting outcomes compared to more complex ensemble models. Therefore, it is important for researchers to look at combining XAI techniques with the highest performing models to continue to further their research.

The techniques that are currently used to predict and diagnose heart disease are primarily based on the analysis of a patient's medical history, symptoms, and physical examination reports by doctors. Most of the time, it is difficult for medical experts to accurately predict a patient's heart disease, where they can predict with up to 67% accuracy [6] because, currently, the diagnosis of any disease is done concerning the similar symptoms observed from previously diagnosed patients [7]. Hence, the medical field requires an automated intelligent system for the accurate prediction of heart disease. This can be achieved by utilizing the huge amount of patient data that is available in the medical sector, along with machine learning algorithms [8]. In recent times, data science research groups have

paid much attention to disease prediction. This is owing to the rapid development of advanced computer technologies in the healthcare sector, as well as the availability of massive health databases [9]. The combination of new deep-learning and intelligent decision-making systems has great potential to improve healthcare assistance in our society [10]. Data is the most valuable resource for obtaining new or additional knowledge and collecting important information. There is an enormous amount of data (big data) in various sectors, such as science, technology, agriculture, business, education, and health. This is completely unprocessed data, either in a structured or unstructured form [11]. It is necessary to extract valuable information from big data to store, process, analyze, manage, and visualize this data via performing data analysis [12].The main contribution of this research work was to implement an intuitive medical prediction system for the diagnosis of heart disease using contemporary machine learning techniques. In this work, different kinds of machine learning classifier algorithms, such as naïve Bayes (NB), logistic regression (LR), sequential minimal optimization (SMO), instance-based classifier (IBk), AdaBoostM1 with decision stump (DS), AdaBoostM1 with LR, bagging with REPTree, bagging with LR, JRip, and random forest (RF) were trained to select the best predictive model for the accurate heart disease detection at an initial stage. Three attribute selection techniques, such as correlation-based feature subset evaluator, chi-squared attribute evaluator, and ReliefF attribute evaluator, were utilized to obtain the optimal set of attributes that greatly influenced the performance of the classifiers when predicting the target class. Finally, tuning the hyperparameter "number of nearest neighbors" in the IBk classifier was performed on both the full attribute set and optimal sets obtained from attribute evaluators.

Related Works

This section discusses the state-of-the-art methods for heart disease diagnosis using machine learning techniques that were accomplished by various effective research works.R. Perumal et al. [18] developed a heart disease prediction model using the Cleveland dataset of 303 data instances through feature standardization and feature reduction using PCA, where they identified and utilized seven principal components to train the ML classifiers. They concluded that LR and SVM provided almost similar accuracy values (87% and 85%, respectively) compared to that of k-NN with 69%. C. B. C. Latha et al. [19] performed a comparative analysis to improve the predictive accuracy of heart disease risk using ensemble techniques on the Cleveland dataset of 303 observations. They applied the brute force method to obtain all possible attribute set combinations and trained the classifiers. They achieved a maximum increase in the accuracy of a weak classifier of 7.26% based on ensemble algorithm, and produced an accuracy of 85.48% using majority vote with NB, BN, RF, and MLP classifiers using an attribute set of nine attributes. D. Ananey-Obiri et al. [20] developed three classification models, namely, LR, DT, and Gaussian naïve Bayes (GNB), for heart disease prediction based on the Cleveland dataset. Feature reduction was performed using single value decomposition, which reduced the features from 13 to 4. They concluded that both LR and GNB had predictive scores of 82.75% and AUC of 0.87. It was suggested that other models, such as SVM, k-NN, and random forest, be included.

3. Materials and Methods

This section discusses the proposed methodology, which comprises the dataset description, data pre-processing, machine learning classifiers, attribute evaluators, and performance metrics.

The proposed methodology follows a systematic workflow for predicting heart disease risk using the Cleveland heart disease dataset obtained in CSV format from the UCI Machine Learning Repository. Initially, the dataset was imported into a software tool to examine its attributes, data types, value ranges, and other statistical characteristics. Data preprocessing was then performed to enhance model reliability.

For model evaluation, 10-fold cross-validation was applied. The dataset was randomly divided into ten equal subsets, where nine folds were used for training and the remaining fold for testing. This process was repeated until each fold served once as the test set, and the average of all results was considered the final performance measure. Multiple machine learning algorithms were implemented, including Naïve Bayes, Logistic Regression, SMO, IBk, AdaBoost-based models, bagging with REPTree and Logistic Regression, JRip, and Random Forest. Additionally, feature selection methods—Correlation-based Feature Selection with BestFirst, Chi-squared with Ranker, and ReliefF with Ranker—were used to determine the most relevant attributes. Finally, the IBk classifier's hyperparameter 'k' was tuned to further improve predictive performance.

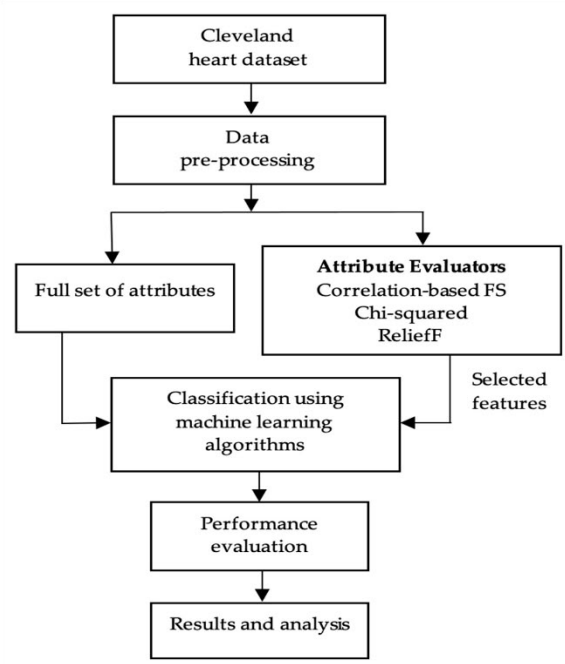


Figure 1. Experimental workflow of the proposed methodology.

Dataset Description and Statistics

The Cleveland heart dataset consists of 303 instances with 76 attributes, but only 14 attributes are considered more suitable for research experimental purposes. The attribute descriptions for the Cleveland heart dataset are given in Table 1.

Table 1. Attribute descriptions for the Cleveland heart dataset from the UCI machine learning repository [33].

Attribute	Description	Type of Attribute	Attribute Value Range
Age	Age in years	Numeric	29 to 77
Sex	Gender	Nominal	0 = female, 1 = male
cp	Chest pain type	Nominal	1. = typical angina, 2. = atypical angina, 3. = non-angina pain, 4. = asymptomatic
trestbps	Resting blood pressure in mm Hg on admission to the hospital	Numeric	94 to 200
chol	Serum cholesterol in mg/dL	Numeric	126 to 564
Fbs	Fasting blood sugar > 120 mg/dL	Nominal	0 = false, 1 = true
restecg	Resting electrocardiographic results	Nominal	0 = normal, 1 = ST-T wave abnormality, = definite left ventricular hypertrophy by Estes' criteria
thalach	Maximum heart rate achieved	Numeric	71 to 202
exang	Exercise induces angina	Nominal	0 = no 1 = yes
oldpeak	ST depression induced by exercise relative to rest	Numeric	0 to 6.2
slope	The slope of the peak exercise ST segment	Nominal	1 = upsloping, 2 = flat, = downsloping
ca	Number of major vessels colored by fluoroscopy	Nominal	0–3
thal	The heart status	Nominal	3 = normal, 6 = fixed defect, = reversible defect
target	Prediction attribute	Nominal	0 = no risk of heart disease, to 4 = risk of heart disease

In the Cleveland heart disease dataset, attributes with fewer than 10 classes are treated as nominal or categorical variables. The attribute 'sex' has two classes: male (1) and female (0). 'cp' includes four chest pain types: typical angina, atypical angina, non-anginal pain, and asymptomatic. 'fbs' indicates whether fasting blood sugar exceeds 120 mg/dL (1 = true, 0 = false). 'restecg' has three categories describing resting ECG results: normal, ST-

T wave abnormality, and left ventricular hypertrophy. 'exang' represents exercise-induced angina (1 = yes, 0 = no), while 'slope' describes the peak exercise ST segment slope (upslope, flat, downslope). 'ca' indicates the number of major vessels (0–3), and 'thal' reflects heart condition (normal, fixed defect, reversible defect). The 'target' attribute originally had five classes, but values 1–4 were combined into 1 to create a binary classification (0 = no risk, 1 = risk). Numeric attributes include age, trestbps, chol, thalach, and oldpeak, with no missing values reported.

Table 2. (a) The statistical outline of the numeric attributes. (b) The statistical outline of the nominal attributes.

(a)

Attribute	Min.	Max.	Mean	StdDev	Missing	Distinct	Unique
age	29	77	54.439	9.039	0	41	4 (1%)
trestbps	94	200	131.69	17.6	0	50	17 (6%)
chol	126	564	246.693	51.777	0	152	61 (20%)
thalach	71	202	149.607	22.875	0	91	28 (9%)
oldpeak	0	6.2	1.04	1.161	0	40	10 (3%)

(b)

Attribute	Label	Count	Proportion	Missing	Distinct
sex	0	97	32%	0	2
	1	206	68%		
cp	1	23	7.6%	0	4
	2	50	16.5%		
	3	86	28.4%		
	4	144	47.5%		
fbs	0	258	85.15%	0	2
	1	45	14.85%		
restecg	0	151 4	49.83% 1.32%	0	3
	1				
	2	148	48.84%		
exang	0	204	67.33%	0	2
	1	99	32.67%		
slope	1	142	46.86%	0	3
	2	140	46.20%		
	3	21	6.93%		
ca	0	176	58.08%	4 (1.32%)	4
	1	65	21.45%		
	2	38	12.54%		

	3	20	6.6%		
thal	3 6	166 18	54.79% 5.95%	2 (0.66%)	3
	7	117	38.6%		
target	0 1	164 139	54% 46%	0	2

Min.—minimum, Max.—maximum, StdDev—standard deviation.

Table 2(b) presents the statistical details of nominal attributes, including label counts, missing, and distinct values. Out of 303 instances, six (2%) contained missing values—four in 'ca' and two in 'thal'. The dataset includes 164 (54%) no-risk cases and 139 (46%) heart disease risk cases.

3.2. Pre-Processing of Dataset

Having missing data means that the dataset is incomplete. In statistics, missing values or missing data occur when no data value is stored for the variable in an observation. These missing values are represented by blank/dashes. The main reason for having missing values is that respondents forget/refuse/fail to answer certain questions. Other reasons include sensor failure, loss of data while transferring, internet connection disruption, and wrong mathematical calculations, such as dividing by zero. It is always hard to predict when missing values are present in the dataset because sometimes, they affect results and sometimes not. In a dataset, each variable may only have a small number of missing responses, but in combination, the missing data could be numerous. The analysis might run but the results may not be statistically significant because of the missing data. For research purposes, replacing missing values either by a user constant or the mean value will be more effective than removing those observations from the dataset.

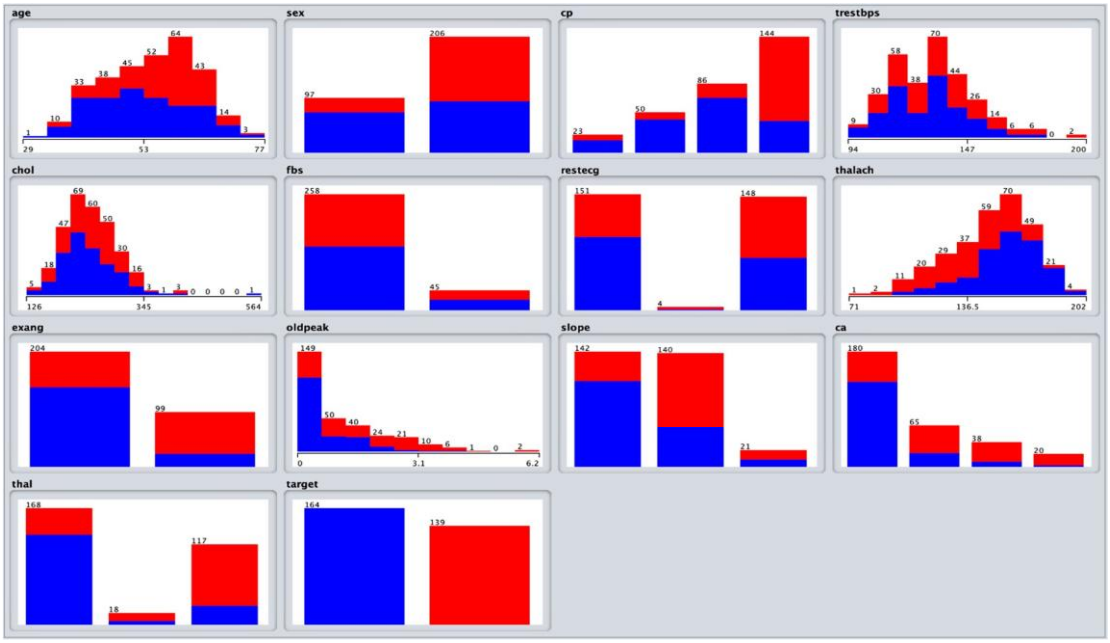


Figure 2. Visualization of attributes of the Cleveland heart dataset.

3.3. Machine Learning Models for Classification

Researchers applied multiple supervised machine learning algorithms to determine the most effective classifier for heart disease prediction. Naïve Bayes uses Bayes' theorem to calculate class probabilities based on statistical distributions. Logistic Regression combines weighted predictors with a logistic function for classification. SMO efficiently solves optimization problems for Support Vector Machines. IB (k-nearest neighbours) classifies instances based on majority voting among nearest data points. Bagging reduces variance by training models on bootstrapped samples, while AdaBoostM1 strengthens weak classifiers to improve accuracy. Rip generates classification rules using incremental pruning. Random Forest builds multiple randomized decision trees to enhance accuracy and control overfitting.

3.4. Attribute Evaluators

This study used three attribute evaluators: correlation-based feature selection with Best First, chi-squared evaluation with Ranker, and Relief with Ranker. Correlation-based selection assesses each attribute's predictive ability and redundancy, favoring subsets with high class correlation and low intercorrelation. The selected attribute set is presented in Table 3.

Table 3. Attribute sets that were obtained from the correlation-based feature selection.

Attribute No.	Attribute Name
3	cp
7	restecg
8	thalach
9	exang
10	oldpeak
12	ca
13	thal

The chi-squared attribute evaluation method ranks attributes by computing their chi-squared statistic relative to the class using the Ranker search method. Based on the ranking results, the three lowest-ranked attributes—fbs, trestbps, and chol—were removed, and the top 10 predictors were used to retrain the machine learning models.

Table 4. Attribute sets that were obtained from the chi-squared attribute evaluation.

Attribute No.	Attribute Name	Rank
13	thal	82.6845
3	cp	81.8158
12	ca	72.6169
10	oldpeak	61.5234
9	exang	56.5193

8	thalach	51.5870
11	slope	45.7846
1	age	24.8856
2	sex	23.2181
7	restecg	10.0515
6	fbs	0.1934
4	trestbps	0
5	chol	0

Relief is an attribute ranking filter that evaluates features by repeatedly sampling instances and comparing nearest neighbours from the same and different classes. Using k = 10 neighbours and Ranker search, it ranked the Cleveland attributes. The four lowest-ranked features—age, trestbps, fbs, and chol—were removed, and classifiers were trained on the top nine attributes.

Table 5. Attribute sets that were obtained from the ReliefF attribute evaluation.

Attribute No.	Attribute Name	Rank
12	ca	0.18812
3	cp	0.17789
13	thal	0.11452
2	sex	0.09307
11	slope	0.06898
9	exang	0.06667
7	restecg	0.05842
10	oldpeak	0.02350
8	thalach	0.02118
1	age	0.01786
4	trestbps	0.01577
6	fbs	0.01386
5	chol	0.00181

3.5. Performance Metrics

The performance metrics used in this research work, namely, accuracy, mean absolute error (MAE), sensitivity (recall), fallout, precision, F-measure, specificity, and ROC area, are discussed here. The confusion matrix shown in Table 6 depicts various performance metrics for evaluating a classifier. True positives are the responses equal to the positive class that are correctly predicted as positive. True negatives are the responses equal to the negative class that are correctly predicted as negative. False positives are the responses equal to the

negative class but are predicted as positive. False negatives are the responses equal to the positive class but are predicted as negative.

Table 6. Confusion matrix.

		Predicted Class High Risk (1)	**Low Risk (0)**
Actual class	High risk (1) Low risk (0)	True Positive (TP) False Positive (FP)	False Negative (FN) True Negative (TN)

Accuracy = TP + TN TP + FP + TN + FN × 100% (2)

MAE = ∑ |Predicted value − Actual value| Number of predictions (3)

Sensitivity = TP TP + FN × 100% (4)

Specificity = TN TN + FP × 100% (5)

Fallout = FP TN + FP × 100% (6)

Precision = TP TP + FP × 100% (7)

F − Measure = 2 × Precision × Recall Precision + Recall (8)

ROC Area: The area under the ROC curve measures the quality of a model's predictions regardless of what classification threshold is chosen. The ROC curve represents the true positive rate (sensitivity or recall) vs. the false positive rate (fallout) at every 0 → 1 threshold.

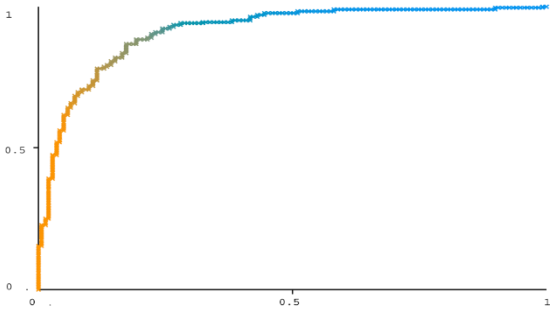
4. Results

The results of the machine learning classifiers using the full set of attributes and optimal set that was obtained from attribute evaluators, tuning the parameter 'k' in the IBk method, and comparison with related works are discussed in the following.

As shown in Table 7, the highest accuracy of 85.148% was attained with the sequential minimal optimization (SMO) algorithm, followed by logistic regression (LR) with an accuracy of 84.818% using the full set of attributes from the Cleveland dataset using the 10-fold cross-validation test option. The SMO algorithm also provided the best MAE of 0.148, sensitivity of 0.851, fallout of 0.157, precision of 0.852, F-measure of 0.851, and specificity of 0.90 compared to other machine learning algorithms. The meta classifier bagging with LR achieved a high ROC area value of 0.91, followed by the single classifier LR with a ROC area of 0.909. The LR and AdaBoostM1 with LR classifiers also reached a specificity of 0.90. NB provided the second-best MAE of 0.184. The visualization of the threshold curve for the target class provided an ROC area of 0.91 using the bagging with LR meta classifier using the full set of attributes, as shown in Figure 3. The bar plot of the performance metrics using the full set of attributes of the Cleveland heart dataset is shown in Figure 4.

Table 7. Performance of machine learning classifiers based on the full set of attributes using 10-fold cross-validation.

Classifier	Accuracy	MAE	Sensitivity	Fallout	Precision	F-Measure	ROC Area	Specificity
NB	83.828	0.184	0.838	0.167	0.838	0.838	0.907	0.870
LR	84.818	0.210	0.848	0.162	0.85	0.847	0.909	0.900
SMO	85.148	0.148	0.851	0.157	0.852	0.851	0.847	0.900
IBk/KNN	76.897	0.233	0.769	0.235	0.769	0.769	0.764	0.790
AdaBoostM1 + DS	82.838	0.227	0.828	0.178	0.829	0.828	0.888	0.870
AdaBoostM1 + LR	84.818	0.204	0.848	0.162	0.850	0.848	0.860	0.900
Bagging + REP-Tree	80.858	0.290	0.809	0.198	0.809	0.809	0.878	0.850
Bagging + LR	84.488	0.214	0.845	0.162	0.845	0.845	0.910	0.880
JRip	74.917	0.325	0.749	0.256	0.749	0.749	0.755	0.780
RF	81.848	0.276	0.818	0.188	0.818	0.818	0.897	0.850



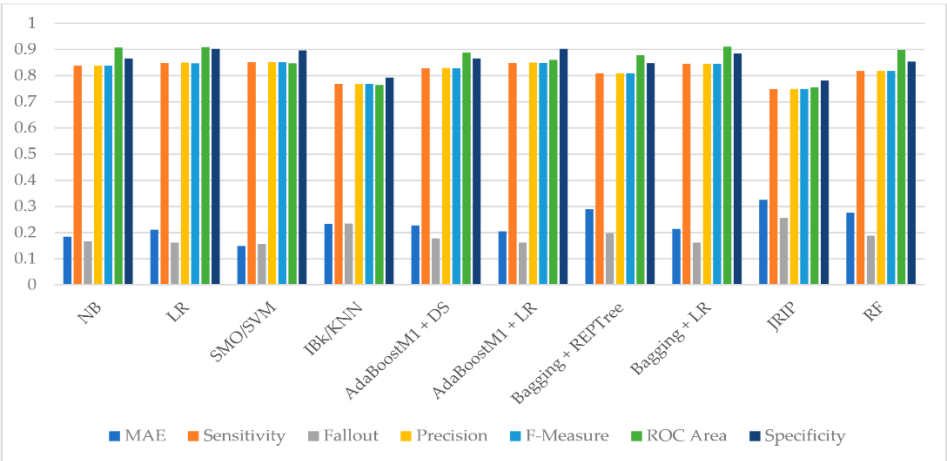


Figure 3. ROC curve of bagging with LR meta classifier using the full set of features.

5. Conclusions

In conclusion, this study highlights the importance of selecting appropriate features and machine learning models for accurate heart disease risk prediction. Among all classifiers tested, SMO demonstrated the best overall performance. The study also showed that tuning model parameters and applying feature selection techniques can significantly enhance predictive accuracy. Although the dataset used was relatively small, the findings indicate strong potential for future research using larger datasets and advanced algorithms to further improve prediction systems.

10 machine learning classifiers, and 3 feature selection methods were used in this research. There is a huge scope to explore various machine learning algorithms and feature selection techniques. In the future, we intend to combine multiple datasets to obtain a higher number of observations and conduct more experiments by selecting appropriate attributes to improve the classifier's predictive performance

References

[1] 1. WHO. Available online: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 (accessed on 9 February 2021). 2. Healthline. Available online: https://www.healthline.com/health/stroke-vs-heart-attack#treatment (accessed on 20 February 2021).

[2] Chicco, D.; Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med. Inform. Decis. Mak. 2020, 20, 1–16.

[3] Karthick, D.; Priyadharshini, B. Predicting the chances of occurrence of Cardio Vascular Disease (CVD) in people using classification techniques within fifty years of age. In Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018, Coimbatore, India, 19–20 January 2018; pp. 1182–1186.

[4] Obasi, T.; Shafiq, M.O. Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases. In Proceedings of the 2019 IEEE International Conference on Big Data, Big Data 2019, Los Angeles, CA, USA, 9–12 December 2019; pp. 2393–2402.

[5] Sharma, H.; Rizvi, M.A. Prediction of Heart Disease using Machine Learning Algorithms: A Survey. Int. J. Recent Innov. Trends Comput. Commun. 2017, 5, 99–104.

[6] Ramalingam, V.V.; Dandapath, A.; Raja, M.K. Heart disease prediction using machine learning techniques: A survey. Int. J. Eng. Technol. 2018, 7, 684–687.

[7] Alaa, A.M.; Bolton, T.; Di Angelantonio, E.; Rudd, J.H.F.; Van Der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. PLoS ONE 2019, 14, e0213653.

[8] Uddin, S.; Khan, A.; Hossain, E.; Moni, M.A. Comparing different supervised machine learning algorithms for disease prediction. BMC Med. Inform. Decis. Mak. 2019, 19, 1–16.

[9] Song, Q.; Zheng, Y.-J.; Yang, J. Effects of Food Contamination on Gastrointestinal Morbidity: Comparison of Different MachineLearning Methods. Int. J. Environ. Res. Public Heal. 2019, 16, 838.

[10] Chen, M.; Hao, Y.; Hwang, K.; Wang, L.; Wang, L. Disease Prediction by Machine Learning Over Big Data from Healthcare Communities. IEEE Access 2017, 5, 8869–8879.

[11] Aljanabi, M.; Qutqut, M.; Hijjawi, M. Machine Learning Classification Techniques for Heart Disease Prediction: A Review. Int. J. Eng. Technol. 2018, 7, 5373–5379.

[12] Pasha, S.J.; Mohamed, E.S. Novel Feature Reduction (NFR) Model with Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction. IEEE Access 2020, 8, 184087–184108.

[13] Swain, D.; Pani, S.K.; Swain, D. A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning. In Proceedings of the 2018 International Conference on Advanced Computation and Telecommunication, ICACAT, Bhopal, India, 28–29 December 2018; pp. 1–6.

[14] Weng, S.F.; Reps, J.M.; Kai, J.; Garibaldi, J.M.; Qureshi, N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE 2017, 12, e0174944.

[15] Khan, Y.; Qamar, U.; Yousaf, N.; Khan, A. Machine Learning Techniques for Heart Disease Datasets: A Survey. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing, Zhuhai, China, 22–24 February 2019; pp. 27–35.

[16] Goel, S.; Deep, A.; Srivastava, S.; Tripathi, A. Comparative Analysis of various Techniques for Heart Disease Prediction. In Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019, Mathura, India, 21–22 November 2019; pp. 88–94.

[17] Perumal, R. Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques. Int. J. Adv. Sci. Technol. 2020, 29, 4225–4234.

[18] Latha, C.B.C.; Jeeva, S.C. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Inform. Med. Unlocked 2019, 16, 100203.

[19] Ananey-Obiri, D.; Sarku, E. Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms. Int. J. Comput. Appl. 2020, 176, 17–21.