

# Multi-model Ensemble Based Approach for Heart Disease Diagnosis

Areesa Ashfaq  
Dept. of Creative Technologies  
Air University  
Islamabad, Pakistan  
areesaashfaq39@gmail.com

Azhar Imran  
Dept. of Creative Technologies  
Air University  
Islamabad, Pakistan  
azhar.imran@mail.au.edu.pk

Inam Ullah  
Dept. of Creative Technologies  
Air University  
Islamabad, Pakistan  
inamullahiufet@gmail.com

Abdulkareem Alzahrani  
Faculty of Computer Science and I.T.  
Al Baha University  
Albaha, Saudi Arabia  
ao.alzahrani@bu.edu.sa

Khattab M. Ali Alheeti  
Computer Networking Systems Dept.  
University of Anbar  
Ramadi, Iraq  
co.khattab.alheeti@uoanbar.edu.iq

Amanullah Yasin  
Dept. of Creative Technologies  
Air University  
Islamabad, Pakistan  
amanyasin@mail.au.edu.pk

**Abstract**—Heart is the core part responsible for blood pumping in the human body. Any disorder that effectuates the functions of the cardiovascular named heart disease or cardiovascular disease. CVD (cardiovascular disease) turns into a major cause of death worldwide. This cardio disease leads to 17.9 million deaths per year according to an estimation by World Health Organization. Machine learning approaches become the effective for prediction of heart disease to overcome the mortality rate. Researchers deployed various techniques to predict heart disease. Proposed ensemble methods that provide the optimal model by combining multiple models. Stacking, bagging, voting, and boosting are the ensemble methods that have been deployed on the UCI dataset of heart disease by feature extraction. Feature scaling is performed using a standard scalar to normalize the feature set or independent variables. Data is segmented into training and testing set in the ratio of 70:30 for machine learning models. Experiments results indicated that bagging-XGB (extreme gradient boosting) achieved 88% which is the highest accuracy among all proposed models. In boosting methods AdaBoost achieved a high accuracy of 86% among all proposed techniques. The soft voting classifier acquired an accuracy of 86% which is more than the hard voting classifier. A combination of Random Forest, Gaussian Naive Bayes, and K-Nearest Classifier achieved the highest accuracy of 87% among all others stacking approaches. All ensemble techniques are evaluated by the ROC curve (receiver operating characteristic curve) and performance measures. Experiments results indicate that heart disease can be effectively predicted by ensemble techniques.

**Index Terms**—cardiovascular disease, machine learning, ensemble methods, feature extraction, AdaBoost, performance measures, random forest, Gaussian naive Bayes, and k-nearest classifier.

## I. INTRODUCTION

Heart is the core part responsible for blood pumping in the human body. Various diseases are associated with the heart. Any disorder that effectuates the functions of the cardiovascular named heart disease or CVD (cardiovascular disease).

CVD turns into a major cause of death worldwide. This cardio disease leads to 17.9 million deaths per year according to the estimation of WHO (world health organization). In 2019, 18.6 million died due to cardiovascular disease. WHO described heart attacks and strokes cause of four-fifth of CVD deaths and one-third rate of the death occur before the age of seventy.

The exponential growth in Cardiovascular disease increases the death rate. Risk factors such as high blood pressure and cholesterol, smoking, deficiency of physical exercise, and hypertension are the main reasons for the growth of heart disease. CVD contributing risk factors identification and diagnosis at an early stage reduces the death rate and can save the lives of many people. Classical work emphasized on influencing the convertible conduct of individuals [1]. Prediction and anticipation steps need to take with the rapid growth of heart disease [2].

Traditional methods like angiography for CVD detection are costly and unbearable expensive for middle-income countries. Several machine learning techniques have been utilized for the medical discipline rather than adopting traditional approaches. Machine learning-based analysis can minimize the traditional human error for the improvement of accuracy [3].

In this research, we proposed composite and enhanced ensemble learning methods. Ensemble methods generate multiple models and aggregate them. Voting classifier, stacking, bagging, and boosting are the ensemble learning methods that have been considered for the prediction. The proposed framework predicts the presence or absence of heart disease based on the feature sets with reasonable accuracy. The utilized models acquired a better accuracy rate as compared to previous results.

## II. LITERATURE REVIEW

Significant research work for the prediction of heart disease has been done. Various machine learning models have been utilized by researchers in order to diagnose cardiovascular heart disease. It can be found that every algorithm has the potency to exhibit determined objectives. In recent years, the focus of researchers is to predict heart disease using ensemble learning which provides better performance by the combination of models as compared to the individual models. Previous researchers mainly focused on the identification of features for their prediction model [2]. Significant attention had not been given to their priority level identification and features' relationship [4].

The author explored the bagged tree, random forest, and AdaBoost ensemble methods as well as the practical swarm optimization feature subset approach to predict heart disease based on heart Stalog dataset from UCI repository. Bagged tree and PSO obtained high accuracy depicted by experiments results [5]. The artificial neural network has been inducted into the domain of medical to attain a better accuracy rate [6]. The author applied a multi-layer perceptron of an artificial neural network on the Cleveland dataset to predict heart disease [7]. The author operated the datasets of ictus, annealing, etc., and found improvement in results [8]. Work on heart disease prediction has been done with an accuracy rate of 81.8% by considering the market dataset [9]. Hybrid techniques have been introduced to predict heart disease in which 14 attributes were utilized from 74 attributes on the UCI dataset of HD [10].

The proposed technique retrieved dimensionality-reduced features by applying probability principal component analysis from the medical results as input data to predict heart disease. This presented methodology attained 86.43% average accuracy for a given dataset [11]. The author worked on the prediction of coronary heart disease based on 'Cardiovascular Disease Dataset'. NB (Naïve Bayes), KNN (k-nearest neighbor), and binary logistic classifier worked as learning mechanisms. Ensemble methods of bagging, stacking, and boosting have been utilized. Boosted methods reached the average accuracy of 73.4%. The stacked ensemble method along with random forest, support vector machine, and k-nearest neighbor acquired the accuracy of 75.1% which was the best model [12].

The author presented a methodology to truly classify the presence of heart disease on the basis of minimal attributes set rather than considering all available features in Cleveland Dataset [13]. Up-to-date models have been explored in which attained accuracy was 81.23% by utilizing XGBoost (extreme gradient boosting) and SMOTE (synthetic minority oversampling technique). This proposed model intended the physical factors and workplace besides the medical features of the coronary artery disease dataset [14].

Author obtained significant feature using correlation coefficient from Cleveland dataset and applied random forest that makes number of trees and combined all the results, this named as ensemble technique in which weak output

combined as strong output by combining all the results [15]. The author performed the comparative analysis of previous studies conducted by enormous researchers on the basis of knowledge acquisition and expert systems for the detection of coronary artery disease. The author found the weakness of previous research to bridge in a subsequent study [16]. Heart disease in the Cleveland dataset has been explored by adopting a random forest model. The author applied a genetic algorithm and attribute election model. They demonstrated the improvement in experimental results as compared to existing work. Evaluation depicts that they acquired higher accuracy than the proposed machine learning models [17].

The author proposed a risk score strategy for heart disease prediction with the help of associative classification. They applied feature election measures of IG, SI, and genetic search using association classification [18]. The clinical features set such as cholesterol, sex, blood pressure, and chest pain have been utilized by the author. The author applied six different machine learning algorithms k-nearest neighbor, support vector machine, decision tree, logistic regression, and Naive Bayes for classification analysis. The author infers the superior approach based on the confusion matrix [19].

## III. PROPOSED METHODOLOGY

This section describes the framework of the proposed methodology as in Fig. 1. It consists of data collection, feature scaling, data segmentation, and classification techniques. Proposed classification techniques based on ensemble learning methods. Ensemble methods involve voting classifiers, stacking, bagging, and boosting models. These ensemble methods identify the chances of disease or not using a given dataset.

### A. Data Collection

The heart disease Cleveland Dataset is retrieved from the heart disease database accessible in the UCI machine learning repository which is extensively operated by researchers. It obtains 303 patient records along with 76 attributes. We considered 14 attributes including the target variable from a total of 76 attributes as these are closely related to heart disease. The attributes set contains age, sex, cp, trestbps, chol, fbs, thalach, exang, oldpeak, slope, ca, thal and target. The target label has two values 0 for the absence of disease and 1 for the presence of disease. Target classes are almost balanced as shown in Fig. 3. Data values have different ranges as indicated in Fig. 2.

### B. Feature Scaling

Feature scaling is performed using a standard scalar to normalize the feature set or independent variables. The standard scaling technique aims to normalize or scale the data to unit variance and zero means which is computed using Eq. (1).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

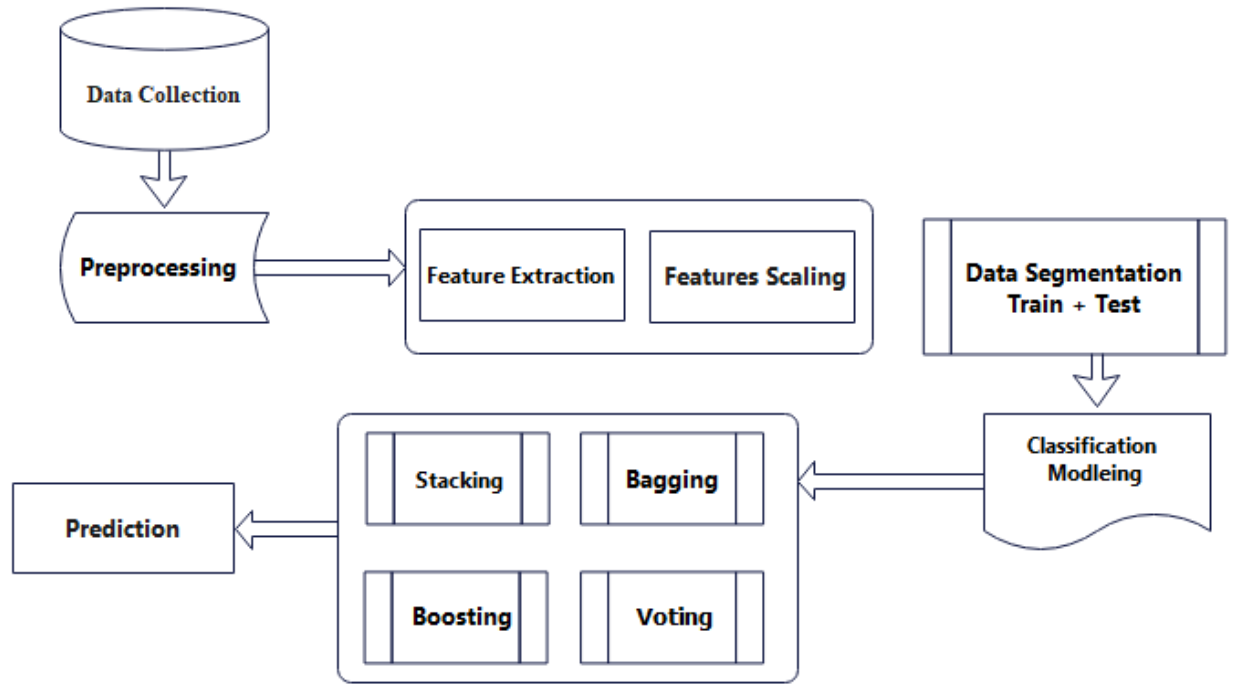


Fig. 1. Proposed framework for heart disease prediction using ensemble learning methods.

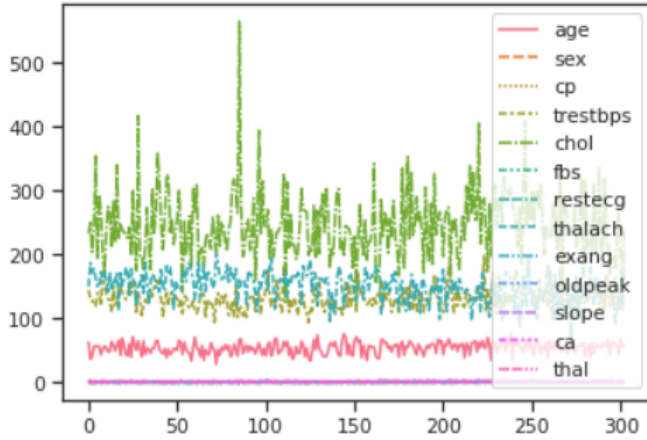


Fig. 2. Dataset values range plotting.

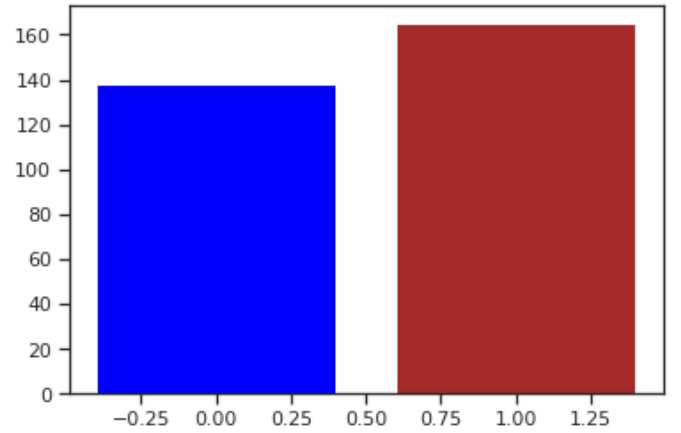


Fig. 3. Visualization of target columns.

### C. Data Segmentation

Data is segmented into training and testing set in the ratio of 70:30 for machine learning models. Training data is utilized to train the model to predict an outcome. The test set is utilized to evaluate the performance in terms of accuracy, precision, recall, and  $f_1$ -score.

Let  $t_1$  be the test data and  $t_2$  be the train data, then data  $d$  is the sum of  $t_1$  and  $t_2$ .

$$d = t_1 + t_2 \quad (2)$$

### D. Ensemble Methods

Ensemble learning is a powerful approach that combines every individual model prediction to attain the best performance. Ensemble techniques are voting, stacking, bagging, and boosting which have been applied to the UCI dataset to estimate the heart disease.

1) *Bagging*: Bagging is a kind of ensemble learning method that generates random samples with replacements from the original training data to set the base classifier. It combines the prediction of each model for final estimation on the basis of voting or averaging. This aggregating approach provides better performance and accuracy as compared to individual models.

The bagging method can be deployed to any machine learning model. Bagging models aim to bring down the variance and abstain from overfitting by drawing multiple training samples [20].

$$S_L(.) = \operatorname{argmax}_k [f(l|w_l(.)) = k] \quad (3)$$

Presented bagged model operated on the Logistic Regression (LR), K Neighbor Classifier(KNC), Extreme Gradient Boosting(XGB), Random Forest(RF), Gaussian Naïve Bayes(GNB), Decision Tree (DT) as base-classifier, and most of predicts better outcome with respect to solitary classifier as shown in Fig. 4.

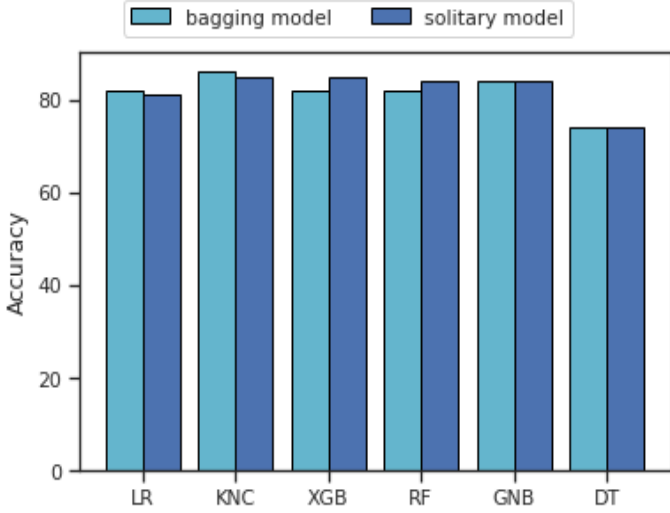


Fig. 4. Accuracy comparison of bagging with solitary models.

Logistic regression enhanced 81% to 82% and k-nearest classifier enhanced 85% to 86% which is best performance among all models. Accuracy of random forest and XGB is decreased. GNB and DT have not enhanced although both approaches sustained their rate of accuracy.

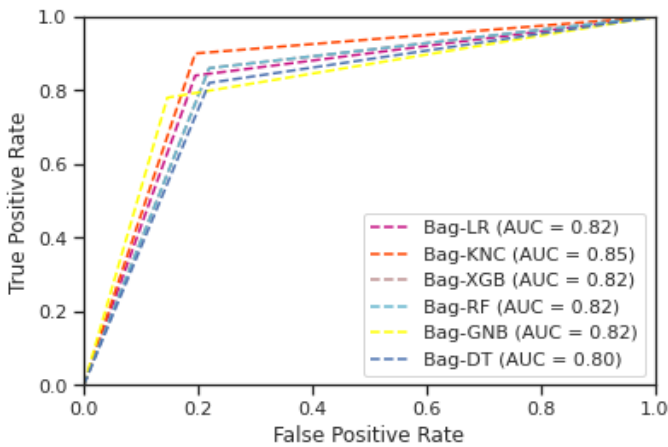


Fig. 5. ROC curve for bagging classifier.

As Fig. 5 depicts the performance of bagging classification models with the receiver operating characteristic (ROC) curve.

ROC curve has been illustrated along with true positive rate with false-positive rate.

2) *Stacking*: A powerful ensemble-based approach that utilizes a meta-learner to train a model. It merges the individual prediction of two or more weak learners to construct the optimal model. Base models are trained for the entire dataset. The output of based models is used as input for meta learner. The meta-learner provides the best performance extracting features from base models.

Suppose  $L_1, L_2, L_3, L_4, L_5, L_6, \dots, L_k$  be the classification models and  $S_n$  be a combination numbers of models for stacking.

$$S_n = \sum_{i=1}^n L_i + L_{i+1} \dots L_{i+n} \quad (4)$$

Stacking methods are proposed with meta learner and different models' combinations of base classifiers as shown in (Table 1). Logistic regression, random forest, k-nearest neighbors' classifier, extreme gradient boosting, decision tree, and Gaussian naive Bayes are the base classifiers.

TABLE I  
STACKING PERFORMANCE OF BASED CLASSIFIERS

Base classifiers	Meta learner	Accuracy
KNC-LR-XGB	DT	81%
RF-GNB-KNC	DT	84%
RF-XGB	DT	84%
LR-GNB-DT	RF	74%

Stack integration of KNC-LR-XGB acquired 81% accuracy, LR-GNB-DT acquired 74% while RF-XGB and RF-GNB-KNC obtained 84% respectively which is indicated in Fig. 6.

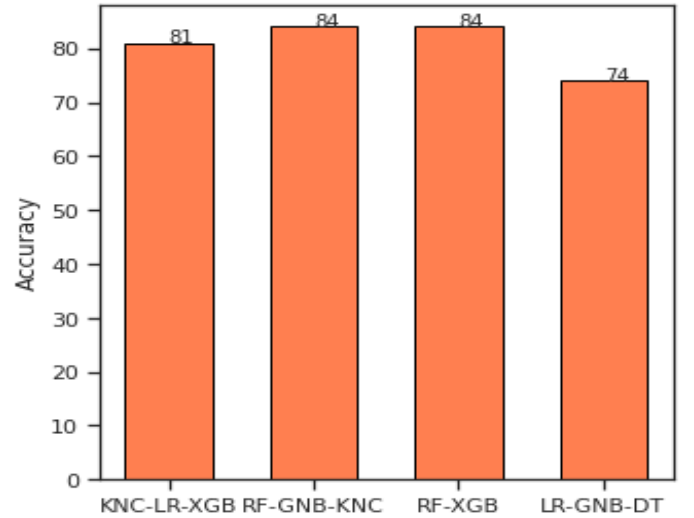


Fig. 6. Accuracy comparison of different stacked classifier.

ROC curve is a powerful tool to measure the performance of models. The stacking ensemble method is evaluated by ROC

curve. For each combined model, the ROC curve is drawn as in Fig. 7.

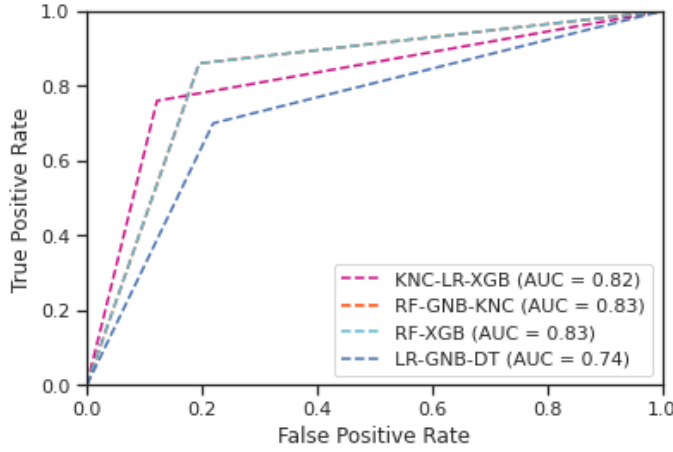


Fig. 7. ROC curve for stacking classifier.

3) *Boosting*: Boosting is a technique of ensemble learning that has the ability to transform weak models into strong models. The data sample is selected randomly to train the model sequentially, each iteration tries to minimize the error of the predecessor.

Light gradient boosted machine (LGBM) is an efficient technique and high-performance framework. LGBM is based on a decision tree structure and generates a leaf-wise tree.

Ada boosting technique stands for adaptive boosting, learning from multiple weak learners known as decision stump on the basis of training data and reducing errors before to move next decision stump. Adaptive boosting re-allocates the higher weights to instances that are not accurately classified.

$$sampleweight = \frac{1}{totalnumberofsamples} \quad (5)$$

Gradient boosting is another boosting type that sums the all-weak learners to build a new model.

$$S_L(.) = \sum_{i=1}^L (w_n * c_l) \quad (6)$$

where  $c_l$ 's are coefficients and  $w_n$  are weak learners.

It tries to minimize error or loss function to construct the best-fitted model based on an iterative approach.

$$y = ax + b + e \quad (7)$$

Three different boosting methods such as AdaBoost, light gradient boosted machine, and gradient boosting classifier (GBC) have been utilized as shown in Fig. 8. Adaboost achieved 82% accuracy, 85% by LGBM, and 78% by GBC.

Boosting ensemble approach performance is illustrated by the ROC curve as shown in Fig. 9.

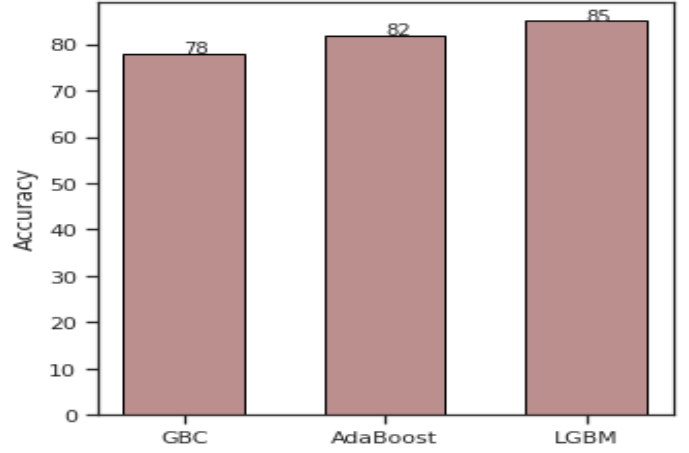


Fig. 8. Accuracy comparison of boosting classifier.

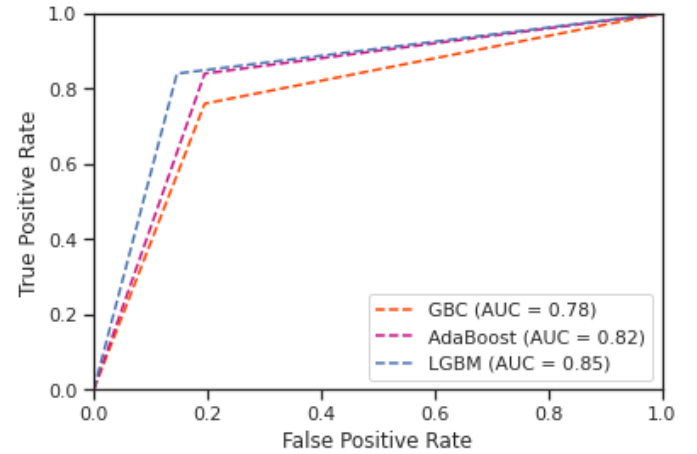


Fig. 9. ROC curve for boosting classifier.

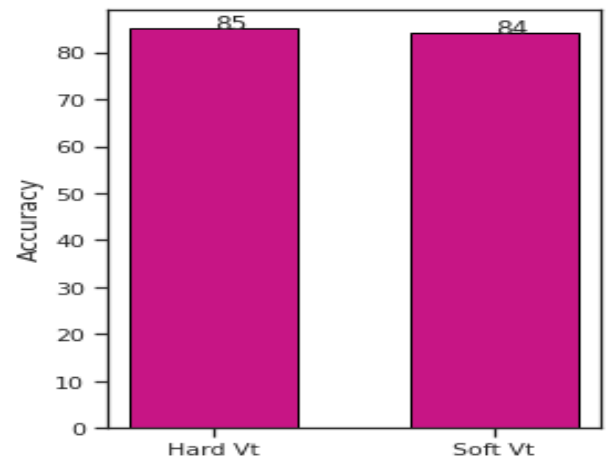


Fig. 10. Accuracy comparison of voting classifier.

4) *Voting Classifier*: Voting is an ensemble learning technique that is used for classification. The voting classifier builds multiple stand-alone models and predicts the final output. The voting classifier implements two approaches that are hard voting and soft voting. Hard voting predicts finalized output based on majority voting while soft voting aggregates the estimated probabilities to return the outcome.

$$S_L(.) = \operatorname{argmax}_k [f(l|w_l(.) = k)] \quad (8)$$

Hard and soft voting classifiers which worked on the basis of majority voting and probability estimation have been adapted. As shown in Fig. 10, hard voting and soft voting classifiers attained 85% and 84% accuracy.

The receiver operating characteristic curve is deployed to evaluate the performance of voting classifiers. As depicted in Fig. 11, the ROC curve is illustrated by a true positive rate and a false-positive rate.

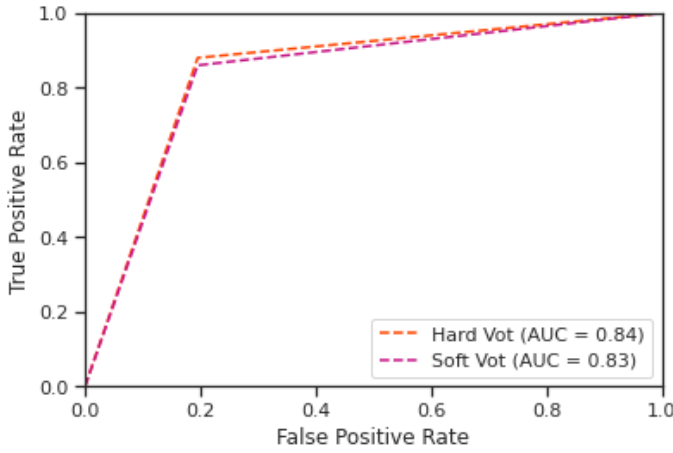


Fig. 11. ROC curve for voting classifier.

#### IV. RESULTS AND COMPARATIVE ANALYSIS

In this section, the results of 15 ensemble methods are presented as shown in Fig. 13. Various types of experiments are conducted to measure the evaluation of ensemble techniques. Published work [12] based on Cleveland dataset is considered for comparative analysis of proposed framework as shown in (Table 2). In this study, to analyze the model, different performance measures such as confusion matrix, accuracy, precision, recall, and f1-score are considered.

TABLE II  
COMPARATIVE ANALYSIS WITH RECENT STUDY

Author	Method	Accuracy	
		Published	Proposed
[12]	Voting	83%	85%
	Bagging	84%	86%
	Stacking	83%	84%
	Boosting	84%	85 %

As (Table 3) depicts the accuracy, precision, recall, and f1-a score of all ensemble methods for both 0 and 1 classes.

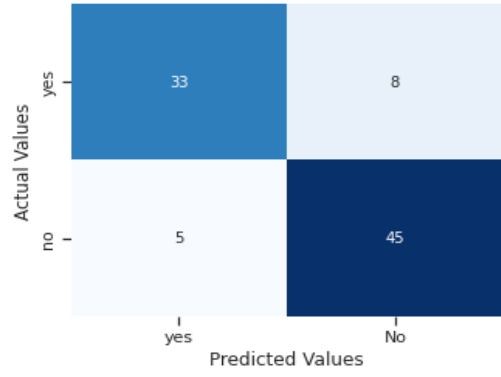


Fig. 12. Confusion matrix for best model (Bag-KNC).

Accuracy is referred to as the correct rate of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Precision is defined as the number of correctly positive classified values that are actually positive.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The recall is defined as the number of positive classified labels out of total examples.

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$f_1$ -score also known as f-score, is the combined score of precision and recall by carrying the harmonic mean.

$$f_1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (12)$$

A confusion matrix is a performance measure of the  $n \times n$  matrix that determine the accurate and inaccurate predictions. The confusion matrix contains the true positive (TP) and true negative (TN) as well as false positive (FP) and false-negative (FN) which are used to compute precision, recall, and accuracy. Fig. 12 depicts the confusion matrix for an approach that attained the highest accuracy between experimented approaches.

#### V. CONCLUSION

In this study ensemble-based approach is proposed to predict the presence and absence of disease. Data have been preprocessed utilizing the feature scaling technique of standard scalar to scale down the variance to a unit vector before moving for classification. Stacking, bagging, voting, and boosting are the techniques that have been utilized for prediction. Experimental result shows that Bagging -XGB achieved more accuracy than all other proposed models. A combination of RF GNB, and KNC achieved the highest accuracy of 87% among all others stacking approaches. In the bagging method, XGB achieved the highest accuracy of 88% which is the best performance with respect to the proposed techniques. The soft

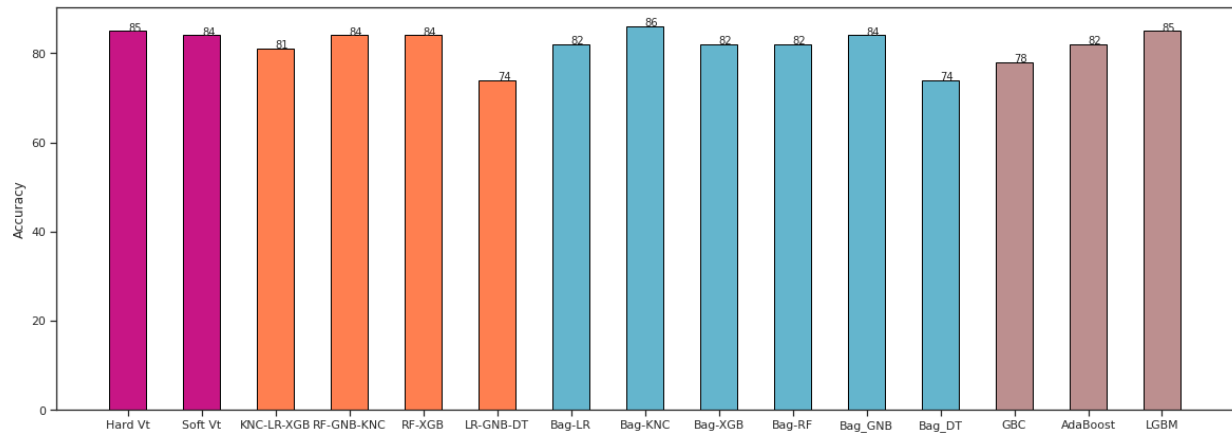


Fig. 13. Accuracy comparison of All-ensemble methods.

TABLE III  
PERFORMANCE MEASURES OF ENSEMBLE TECHNIQUES

Model	Class	Precision	Recall	F1-score	Accuracy
Bag-LR	0	0.80	0.80	0.80	82%
	1	0.84	0.84	0.84	
Bag-KNC	0	0.87	0.80	0.84	86%
	1	0.85	0.90	0.87	
Bag-XGB	0	0.82	0.78	0.80	82%
	1	0.83	0.86	0.84	
Bag-RF	0	0.82	0.78	0.80	82%
	1	0.83	0.86	0.84	
Bag-GNB	0	0.76	0.85	0.80	81%
	1	0.87	0.78	0.82	
Bag-DT	0	0.78	0.78	0.78	80%
	1	0.82	0.82	0.82	
Stack KNC-LR-XGB	0	0.75	0.88	0.81	81%
	1	0.88	0.76	0.82	
Stack RF-GNB-KNC	0	0.82	0.80	0.81	84%
	1	0.84	0.86	0.85	
Stack RF-XGB	0	0.82	0.80	0.81	84%
	1	0.84	0.86	0.85	
Stack LR-GNB-DT	0	0.68	0.78	0.73	74%
	1	0.80	0.70	0.74	
GBC	0	0.73	0.80	0.77	78%
	1	0.83	0.76	0.79	
AdaBoost	0	0.80	0.80	0.80	82%
	1	0.84	0.84	0.84	
LGBM	0	0.81	0.85	0.83	85%
	1	0.88	0.84	0.86	
Hard Voting	0	0.85	0.80	0.83	85%
	1	0.85	0.88	0.86	
Soft Voting	0	0.82	0.80	0.81	84%
	1	0.84	0.86	0.85	

voting classifier acquired the accuracy of 86% which is more than the hard voting. Adaboost acquired the best accuracy of 86% in comparison to the gaussian naïve Bayes and light gradient boosting machine. In future, an ensemble of deep learning models can also be utilized for better results and prediction of complex diseases.

## REFERENCES

- [1] H. S. Buttar, T. Li, and N. Ravi, "Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation," *Experimental and Clinical Cardiology* vol. 10, p. 229, 2005.
- [2] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics and Informatics*, vol. 36, pp. 82-93, 2019.
- [3] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," In: 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, pp.1329-1333, 2021.
- [4] K. I. Mohammed, A. A. Zaidan, B.B. Zaidan, O. S. Albahri, A. S. Albahri, M. A. Alsalem, and A.H. Mohsin, "Novel technique for reorganisation of opinion order to interval levels for solving several instances representing prioritisation in patients with multiple chronic diseases," *Computer Methods and Programs in Biomedicine*, vol. 185, pp. 105151, 2020.
- [5] I. Yekkala, S. Dixit, and M.A. Jabbar, "Prediction of heart disease

using ensemble learning and Particle Swarm Optimization,” In 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon), pp. 691-698, 2017.

- [6] L. Baccour, “Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,” *Expert Systems with Applications*, vol. 99, pp. 115-125, 2018.
- [7] M. Durairaj, and V. Revathi, “Prediction of heart disease using back propagation MLP algorithm,” *International Journal of Scientific and Technology Research*, vol. 4 no. 8, pp. 235-239, 2015.
- [8] G. Felici, B. Simeone, and V. Spinelli, “Classification techniques and error control in logic mining,” In: *Data Mining*, Springer, Boston, MA, pp. 99-119, 2010.
- [9] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale, and A. Lamgunde, “Heart disease prediction using data mining techniques,” In: 2017 International Conference on Intelligent Computing and Control (I2C2), IEEE, pp. 1-8, 2017.
- [10] A. Malav, K. Kadam, and P. Kamat, “Prediction of heart disease using k-means and artificial neural network as Hybrid Approach to Improve Accuracy,” *International Journal of Engineering and Technology*, vol. 9 no. 4, pp. 3081-3085, 2017.
- [11] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, S. A. Hussain, “Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis,” *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 796-807, 2017.
- [12] V. Shorewala, “Early detection of coronary heart disease using ensemble techniques,” *Informatics in Medicine Unlocked*, vol. 26, p.100655, 2021.
- [13] K.M. Almustafa, “Prediction of heart disease and classifiers’ sensitivity analysis,” *BMC Bioinformatics*, vol. 21 no. 1, pp. 1-18, 2020.
- [14] E. Nasarian, M. Abdar, M.A. Fahami, R. Alizadehsani, S. Hussain, M.E. Basiri, and N. Sarrafzadegan, “Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach,” *Pattern Recognition Letters*, vol. 133, pp. 33-40, 2020.
- [15] M.A. Alim, S. Habib, Y. Farooq, and A. Rafay, “Robust heart disease prediction: a novel approach based on significant feature and ensemble learning model,” In 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, pp. 1-5, 2020.
- [16] L.J. Muhammad, E. J. Garba, N. D. Oye, and G. M. Wajiga, “On the problems of knowledge acquisition and representation of expert system for diagnosis of coronary artery disease (CAD),” *International Journal of u-and e-Service, Science and Technology*, vol. 11 no. 3, pp. 49-58, 2018.
- [17] M.A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Intelligent heart disease prediction system using random forest and evolutionary approach,” *Journal of Network and Innovative Computing*, vol. 4 no. 2016, pp. 175-184, 2016.
- [18] P. Chandra, and B. L. Deekshatul, “Prediction of risk score for heart disease using associative classification and hybrid feature subset selection,” In 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE, pp. 628-634, November 2012.
- [19] I. A. Singh, and R. Kumar, “Heart disease prediction using machine learning algorithms,” In 2020 International Conference on Electrical and Electronics Engineering (ICE3), IEEE, pp. 452-457, 2020.
- [20] M. Baydoun, L. Safatly, H. Ghaziri, and A. E. Hajj, “Analysis of heart sound anomalies using ensemble learning,” *Biomedical Signal Processing and Control*, vol. 62, p. 102019, 2020.