



Ensemble Stacking Approach with Heterogeneous Base Classifiers to Predict Cardiovascular disease

Veena Kumari H M¹, Suresh D S²

Article History:

Received: 25.05.2023

Revised: 27.06.2023

Accepted: 20.07.2023

Abstract:

Health research concentrates on quality of life. Computational biology converts clinical data into practice which helps in the diagnosis of disease. According to WHO, heart disease is the principal cause of death. The diagnosis and analysis of diseases using artificial intelligence and determination of optimal methods to provide a solution to medical issues are of considerable attention in the research domain. This study concentrates on major Noncommunicable diseases like cardiovascular diseases. Medical expert systems use machine learning algorithms K-Nearest Neighbor, Naive Bayes, Linear Regression, Decision tree, Neural Networks, Support Vector Machine and Random Forest algorithm for the prediction of diseases. This study gives the proper visualization, analysis, and prediction of which concentrates of 14 features of cardiovascular diseases using four different heart disease datasets from UCI repositories. The research objective is the accurate prediction of the disease. The diagnosis depends on the new techniques which is a combination of classifiers. The proposed work uses an ensemble stacking approach for the particular disease prediction, where in, the end prediction is the result of each classifiers output. Further, the performance evaluations of the individual classifiers, as well as ensemble methods, are carried out by metrics accuracy, specificity, sensitivity, precision, recall, and F1 score from the heart disease datasets. The comparative study on different datasets proves that the slightly imbalanced datasets also give improved accuracy.

Keywords: Ensemble, Features, UCI repositories, Cardio Vascular Diseases, Stacking

DOI: 10.48047/ecb/2023.12.Si8.604

Research Scholar, VTU, Belagavi¹

Department of ECE Channabasaveshwara Institute of Technology, Gubbi, Tumkur, Karnataka, India

veenakumari.hm@cittumkur.org

Professor, Dept. of ECE²

Channabasaveshwara Institute of Technology, Gubbi, Tumkur, India

director@cittumkur.org²

INTRODUCTION

Heart is the central part cardiovascular system that pumps blood into the body. The blood vessels and heart diseases are referred to as cardiovascular disease (CVD), which includes coronary heart diseases like heart attack, heart failure, and stroke [1,2]. From the last few decades cardiovascular diseases are the primary cause of death worldwide in developed, developing, and also in underdeveloped countries. The main cause of heart disease is because of lifestyle changes, the stress in the workplace, and bad food habits. Early detection and continuous monitoring of cardiac diseases may help to reduce the death rate. However, for proper treatment, there is a need for an accurate and reliable system to diagnose. ML algorithms have been used for the analysis of medical datasets.

In the medical domain, Data mining extracting the hidden patterns from large medical datasets to analyze them for prediction of the disease [3]. The essential objective is to extract and then convert it into logical manner for further analysis. Data mining techniques develop a numerous predictive model which enables classification and prediction. The output data is labeled and the model is trained in supervised learning, e.g., SVM, Decision Tree. No class label in unsupervised learning, e.g., K-means clustering. Ensemble approach outperforms the individual ML techniques by combining the results of each classifier. The Ensemble model can be constructed by multiple techniques with each may be of different accuracy. The ensemble approaches are boosting, bagging and stacking.

I. BACKGROUND STUDY

Already research has been carried out to extract the efficient methods for medical diagnoses like heart diseases. Papers related to data mining techniques used for the classification as well as prediction related cardiovascular diseases have studied the dataset, different algorithms, attributes, results, and also future work. El-bialy et al [4] proposed a fast decision tree algorithm and performance comparison done by using a different dataset. Miao et al [5] developed an adaptive boosting algorithm for accurate diagnosis. Vijayashree & Iyengar [6] explained various symptoms and risk factors for the cause of heart attack. Singh et al [7] used the prediction of heart disease system by MLP by back propagation algorithm. Bharati et al [8] compared the deep learning performance with other algorithms. L Zhi Min et al [9] experimented on Cleveland and statlog datasets and proved that selected 6 to 8 features for the heart disease prediction. Nashif et al [10] used WEKA, Open Access Java-Based Data Mining Platform for the accurate diagnosis of heart disease. Harnoor Kaur et al [11] proposed a hybrid approach (HPCC) used for the prediction of diabetes using Pima Indian dataset is used. Bashir et al [12] proposed a HMM Bagmoov model, which overcomes the conventional performance bottlenecks. The model is evaluated on four different CVD datasets. The model achieved the highest accuracy. Our research is an attempt to predict and classify cardiac diseases efficiently using stacked ensemble technique.

METHODOLOGY

The proposed ensemble technique is a combination of seven heterogeneous classifiers called Naïve Bayes, Linear Regression, Support Vector Machine, K Nearest Neighbor, Decision Tree, Random Forest, Extended Boost. The ensemble techniques are based on the base classifier's performance [9]. In the proposed an ensemble technique stacking, the end prediction is based on the performance of the base classifiers and it achieves better accuracy.

Ensemble method

Ensemble means combining multiple models. This approach gives better prediction performance compared to an individual model. Thus, a collection of base models is used to make predictions [12]. The main challenge is to obtain base model which gives different kinds of error. If the ensemble technique bagging, boosting and stacking are used for classification, then high accuracies can be obtained [13]. Random forest is a bagging ensemble technique and each classifier in the ensemble model is a decision tree. Each tree votes and their majority vote is used for classification and the most popular class is returned. RF algorithm can handle the data set containing binary, continuous variables as well as categorical variables in case of regression and classification respectively. RF gives better results for classification problems. Random forest is a simple, fast, flexible, and robust model and it can handle missing values. Boosting is a broadly used and highly effective machine learning algorithm. The important factor of XGBoost its scalability for better accuracy. The system is ten times faster than existing conventional methods. The scalability of XGBoost is due to several algorithm optimizations. Parallel and distributed computing will make learning faster [15]. Stacking is an ensemble ML technique, in which meta classifiers are used to learn from the prediction of two or more base ML classifiers. The base or level 0 classifiers consists of different ML algorithms and therefore stacking ensembles are generally heterogeneous classifiers as shown in Fig. 1. Any classifier can be used as a meta classifier [16].

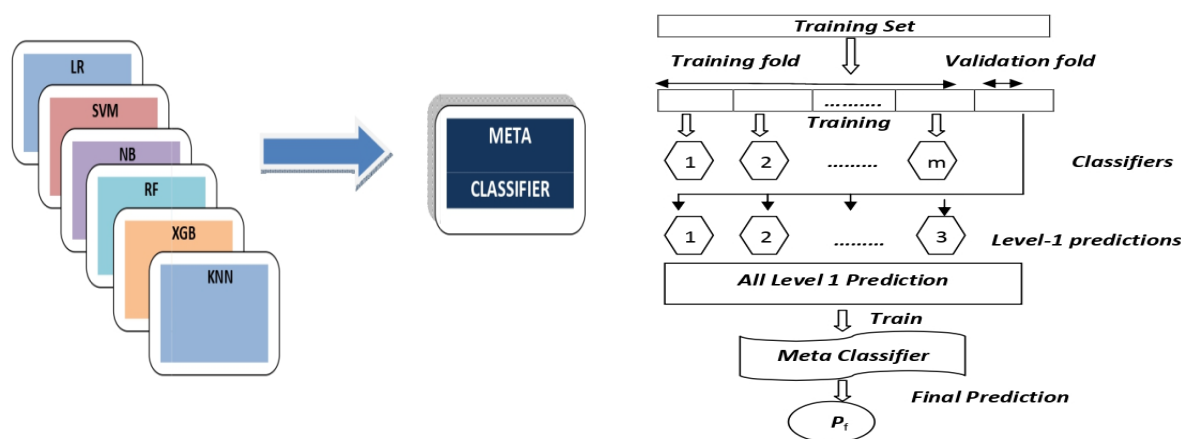


Fig. 1: A model stacking with layers and ensemble procedure

In the stacking algorithm, the base (Level 0) classifiers are trained by the same set of samples which prepare the inputs to the meta (Level 1) classifier. Sometimes it leads to over fitting. The stackingCVclassifier uses the cross-validation method. The dataset is split into k folds, and k-1 folds are used to fit the level-1 classifier in k successive rounds. In every iteration, the level-1 classifiers are then applied to the remaining subset. The predictions from base classifiers are then stacked and that is an input to the level-2 classifier.

Data acquisition and Preprocessing

The main aim of the module is to get cardiovascular disease data from different repositories, process them into a suitable form for further analysis. Datasets has features/attributes which will finally distinguish the data into patient sick and healthy. Every dataset consists of set of features and also of different data types. The dataset is then spilt into training dataset and testing dataset. The pre-processing comprises selection of features, imputation, and removal of noise. The proposed model combines seven ML classifiers. Each base classifier predicted output is used in the Ensemble framework by training the data to make it useful for the prediction of cardiovascular diseases. Each dataset's features and target values are known to each classifier, which in turn can predict whether the disease is present or not [17]. The proposed model flow chart is shown in fig 2. The datasets were collected from the University of California, Irvine (UCI repository). Datasets are normally stored in .CSV format and the available data is generally homogeneous. The dataset contains 76 attributes, but among them 13 input features and one prediction called target class is used for the analysis. Data can be missed due to various reasons. It may reduce the performance of the model and then the whole column is replaced with the median or mean. Categorical data are numerical values rather than label values. The dataset is divided into two cases: 0 corresponding to the absence of disease, while 1 corresponding to the presence. Correlation Matrix explains that few features have a negative correlation and some of them have a positive with the target value. The histogram tells that each of these is a categorical variable and feature is distributed in different ranges. In the proposed research four different heart disease datasets heart_disease, statlog_cleveland, heart_statlog_cleveland_hungary_final (hsch_final) have been used [19]. Table I shows four heart disease datasets used for performance analysis [5,18].

TABLE I: Statistics of Four Heart Disease Datasets

SI No	Dataset	Cases	No.of features	Features/Attributes
1	heart_disease	303	13 + 1 (target)	age, chol, fbs , cp, sex, trestbps, oldpeak, restecg, ca, thalach, thal, exang, target slope,
2	statlog	271	13 + 1 (target)	age, trestbps, sex, chol, cp, restecg, fbs, exang, oldpeak, thalach, thal, ca, slop, target
3	cleveland	298	13+ 1 (target)	age, trestbps, chol, sex, cp, fbs, restecg, exang, thalach, slop, ca, thal, oldpeak, target
4	hsch_final	1190	11+ 1 (target)	Age, chest pain type, sex, resting bps, cholesteral, resting ecg, old peak, max heart rate, ST slope, target.

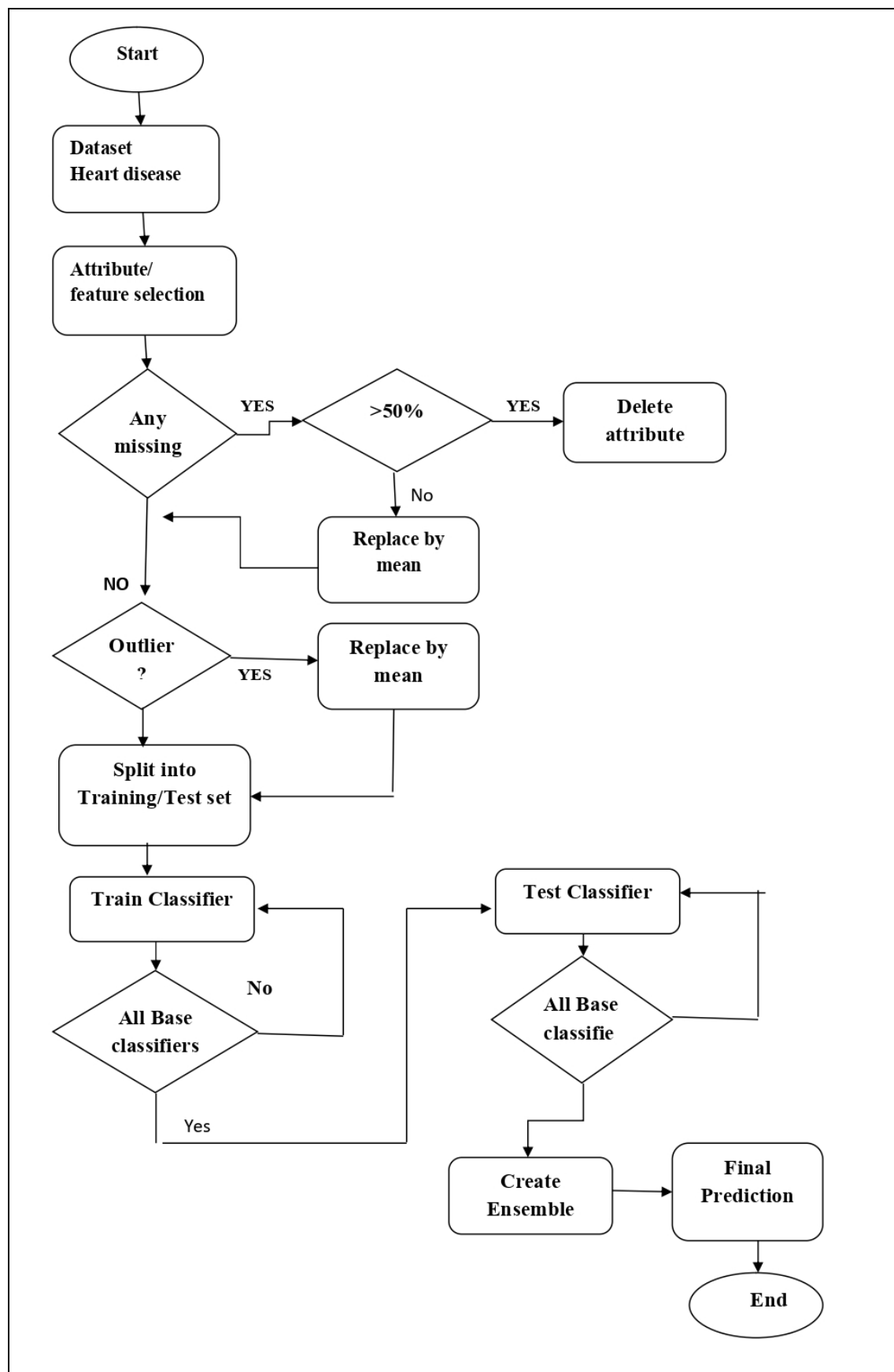


Fig.2. Flowchart of the proposed ensemble framework

IV. DATA ANALYSIS AND INTERPRETATION

The performance evaluation of the ensemble framework has experimented on four cardiovascular disease datasets. The datasets used for the analysis are heart_disease, statlog, cleveland, hsch_final from the UCI repository. The class labels are replaced with class 0 and class 1 to maintain uniformity of each dataset. Each dataset is split into training and testing set. Cross validations of 10-fold is applied and then obtain the confusion matrices. The performance measure of each base classifier as well as ensemble model is calculated using confusion matrix. The Seven classifiers LR, NB, KNN, XGB, DT, SVM & RF are trained first and then they are tested. The metrics are accuracy, specificity, sensitivity, precision, recall, and f1-score. The confusion matrix is a 2x2 matrix which illustrates the actual and predicted classification [20,21]. The equations (1), (2), (3), (4), and (5) are used to calculate the accuracy, specificity, recall, precision, and F1 score respectively [20,21].

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \dots \dots \dots (1)$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} \dots \dots \dots (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \dots \dots \dots (3)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \dots \dots \dots (4)$$

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

TABLE II: Performance comparison of Proposed Ensemble Vs Other Machine Learning Classifiers

heart_disease				statlog		
Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
LR	85.24	77.77	91.17	72.22	73.33	70.83
NB	85.24	77.77	91.17	70.37	73.33	66.66
DT	81.96	85.18	79.41	70.37	63.33	79.16
KNN	88.52	88.88	88.23	75.00	80.00	70.83
SVM	88.00	85.18	91.17	74.07	76.66	70.83
RF	88.57	85.18	91.17	70.37	70.00	70.83
XGB	88.10	85.18	91.17	68.51	70.00	66.66
Stacking	90.16	85.18	94.11	75.92	76.66	75.00
cleveland				hsch_final		
LR	83.30	90.9	78.94	79.83	77.98	81.39
NB	83.33	95.45	76.31	85.29	85.32	85.27
DT	65.10	86.36	52.63	81.51	78.89	83.72
KNN	76.66	95.45	65.78	84.45	84.4	84.49
SVM	80.60	86.36	78.94	86.00	81.65	89.14
RF	76.66	86.36	71.10	86.90	86.23	87.59
XGB	78.33	95.45	68.42	85.71	83.48	87.59
Stacking	81.10	95.45	71.05	86.97	86.23	87.59

The accuracy, sensitivity, and specificity comparison of the proposed ensemble Vs other machine learning classifiers for different heart disease datasets are shown in Table II. The experimental results specify that the proposed method has attained the highest accuracy. It has achieved accuracy, sensitivity, and specificity 90.16 % ,85.18 % , and 94.11% for heart _disease dataset, 75.92 % , 76.66 % , and 75.00 % for statlog dataset, 86.97 % , 86.23 % , and 87.59% for hsch _final compared with other classifiers, whereas 80.00 % , 95.45 % , and 71.05 % for cleveland dataset.

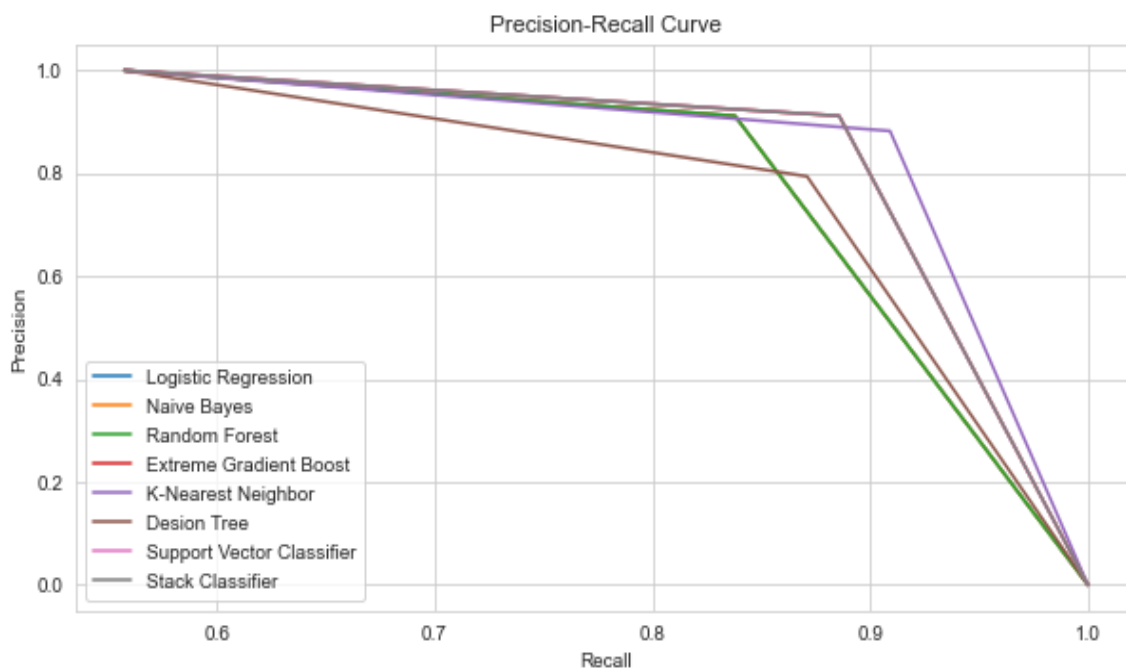
TABLE III: The Precision, Recall and F1_Score comparison of Proposed Ensemble Vs Other Machine Learning Classifiers

heart_disease							statlog				
Classifiers		Class 0	Class 1	Precision (%)	Recall (%)	f1-score (%)	Class 0	Class 1	Precision (%)	Recall (%)	f1-score (%)
LR	Class 0	21	6	88	78	82	20	2	71	91	80
	Class 1	3	31	84	91	87	8	30	94	79	86
NB	Class 0	21	6	88	78	82	21	1	70	95	81
	Class 1	3	31	84	91	87	9	29	97	76	85
DT	Class 0	23	4	77	85	81	19	3	51	86	64
	Class 1	7	27	87	79	83	18	20	87	53	66
KNN	Class 0	24	3	86	89	87	21	1	62	95	75
	Class 1	4	30	91	88	90	13	25	96	66	78
SVM	Class 0	23	4	88	85	87	19	3	70	86	78
	Class 1	3	31	89	91	90	8	30	91	79	85
RF	Class 0	23	4	88	85	87	19	3	63	86	73
	Class 1	3	31	89	91	90	11	27	90	71	79
XGB	Class 0	23	4	88	85	87	21	1	64	95	76
	Class 1	3	31	89	91	90	12	26	96	68	80
Stacking	Class 0	23	4	92	85	88	21	1	66	95	78
	Class 1	2	32	89	94	91	11	27	96	71	82
cleveland							hsch_final				
LR	Class 0	20	2	71	91	80	85	24	78	78	78
	Class 1	8	30	93	79	86	24	105	81	81	81
NB	Class 0	21	1	70	93	81	93	16	83	85	84
	Class 1	9	29	96	76	85	19	110	87	85	86
DT	Class 0	19	3	51	86	64	86	23	80	79	80
	Class 1	18	20	87	53	66	21	108	82	84	83
KNN	Class 0	21	1	62	94	75	92	17	82	84	83
	Class 1	13	25	96	66	78	20	109	87	84	85
SVM	Class 0	19	3	70	86	78	89	20	80	79	80
	Class 1	8	30	91	79	85	14	115	82	84	83
RF	Class 0	19	3	63	86	73	94	15	85	86	86
	Class 1	11	27	90	71	79	16	113	88	88	88
XGB	Class 0	21	1	64	95	76	91	18	85	83	84
	Class 1	12	26	96	68	80	16	113	86	88	87
Stacking	Class 0	21	1	66	95	78	94	15	85	86	86
	Class 1	11	27	96	71	82	16	113	88	88	88

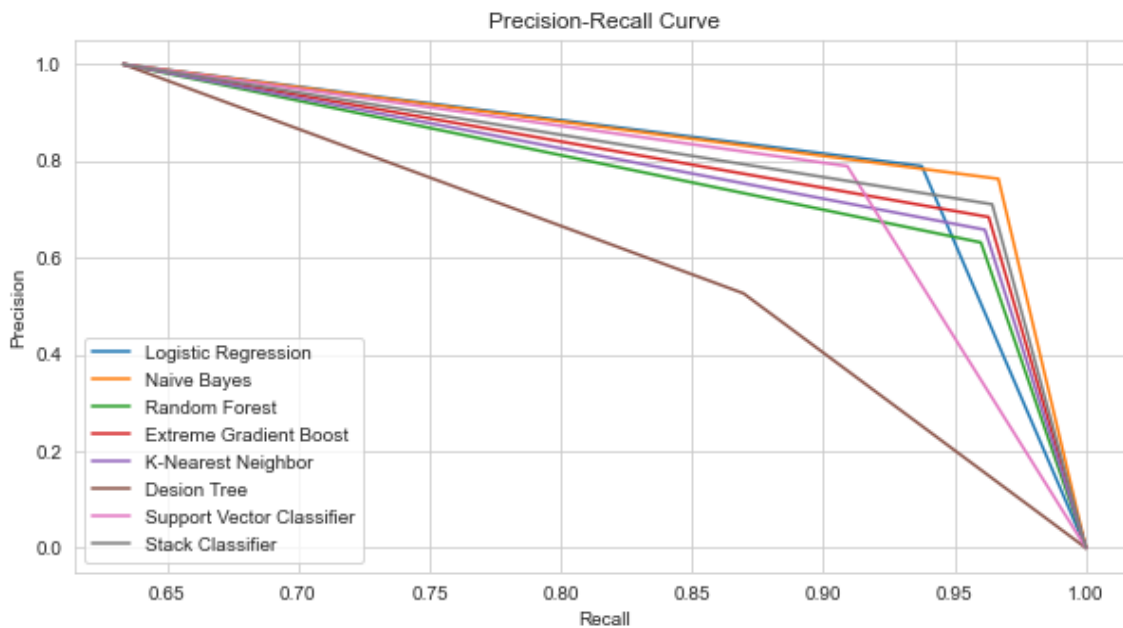
The class 0 and class 1 precision, recall, and f1 score comparison of proposed ensemble Vs other machine learning classifiers for different CVD datasets are shown in Table III. The analysis shows that the stacking ensemble has attained 92% and 89% precision, 85% and 94% recall, and 88% and 91% f1 score for the heart_disease dataset for class 0 and class 1 respectively. Similarly, 66% and 96% precision, 95% and 71% recall, and 78% and 82% f1 score for cleveland dataset, 79% and 72% precision, 77% and 75% recall, and 78% and 73% f1 score for statlog, 85% and 88% precision,

86% and 88% recall, and 86% and 88% f1 score hsch_final compared with other ML classifiers. The confusion matrix of each classifier is tabulated. It is observed that only 4 instances misclassify as false positives and 2 as false negative for heart disease, 7 as false positive and 6 as false negative for statlog, 1 as false positive and 11 as false negative for cleveland, and 15 as false positive and 16 as false negative hsch_final datasets. As per the observation from the table, the bagging and boosting ensemble techniques random forest and XGB gives better results compared to other classifiers.

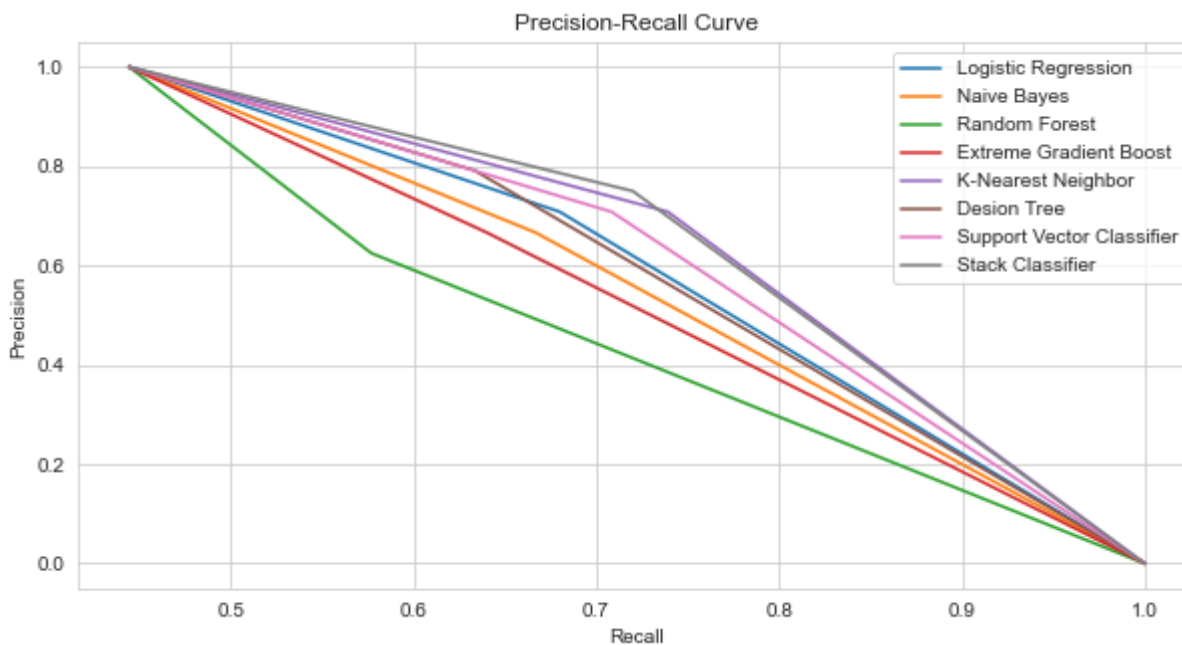
The Receiver Operating Characteristic curve and the Precision - Recall curve is a graphical representation of a, by calculating and plotting the false positive rate (FPR) Vs the true positive rate (TPR) and precision Vs recall for each classifier at various threshold values. The precision and recall curve plotted for (a) heart _disease (b) cleveland (c) statlog and (d) hsch_final dataset shown in Fig.3. In Fig 3(a), the precision recall curve for LR, NB, XGB, SVM classifiers is overlapped with the another represents similar classifier's performance.



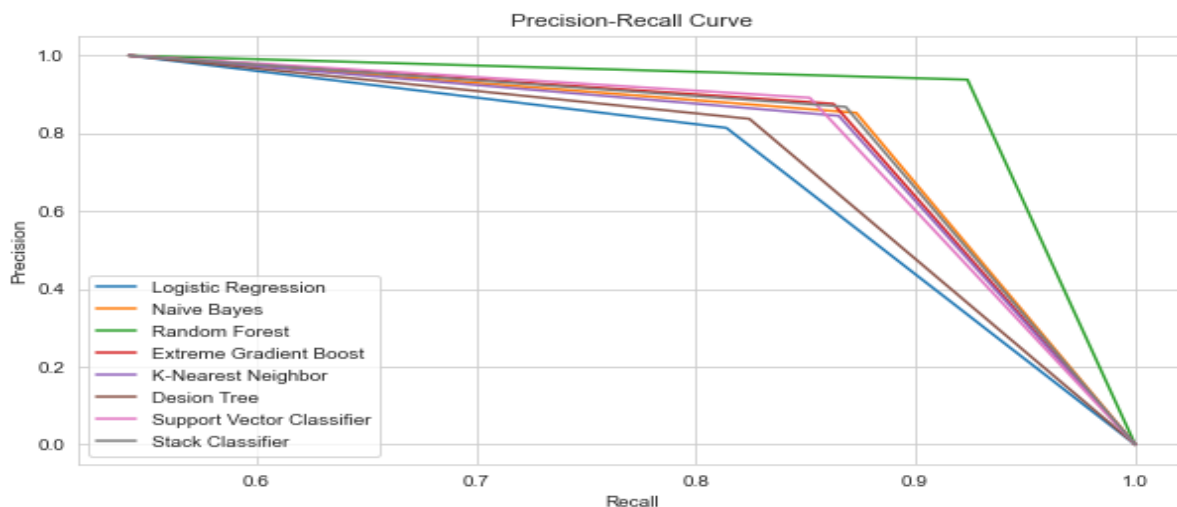
(a)



(b)



(c)



(d)

Fig 9. Precision and Recall curve of (a) heart_disease (b) cleveland (c) statlog, and (d) hsch_final dataset

TABLE IV: AUC comparison of proposed ensemble Vs other machine learning classifiers, heart_disease and cleveland, statlog and hsch_final respectively.

Classifier	heart_disease	cleveland	statlog	hsch_final
auc_lr	0.8401	0.8492	0.7208	0.8492
auc_nb	0.8440	0.8325	0.7000	0.8325
auc_rf	0.8447	0.7930	0.6291	0.7930
auc_knn	0.8756	0.8062	0.7541	0.8062
auc_xgb	0.8718	0.8193	0.6833	0.8193
auc_svc	0.8610	0.8265	0.7375	0.8265
auc_scv	0.8818	0.8588	0.7583	0.8568

The ability of the classifier can be measured by Area Under the (AUC) Curve. It is the summary of the ROC curve. High AUC indicates that performance of the model is better, wherein differentiating between the positive and negative groups. AUC comparison with other classifiers is listed in TABLE IV. The AUC for the proposed method is 88.18% for heart_disease, 75.83 % for statlog, and 83.25% for cleveland and hsch_final datasets. As shown in Table III, the AUC values for the datasets lie between 0.8 to 0.9, indicates “Good” that means the positive class values correctly distinguished from the negative class values.

V. CONCLUSION

Cardiovascular diseases are the leading cause of death and it require early-stage diagnosis. In medical domain, machine learning algorithms used to diagnose to find probable solutions based on the past clinical electronics health records. The accuracy plays a major role in ML techniques. The aim is to apply an ensemble model for disease prediction and classification and also the outcome of the ML algorithms applied on datasets. The proposed model utilizes seven heterogeneous classifiers. Linear Regression, Naïve Bayes, Support Vector Machine, K Nearest Neighbor, Decision tree, Random Forest, XGB are used to improve prediction. The ensemble stacking evaluation is performed on four different heart disease datasets. Different parameters, confusion matrices, accuracy, specificity, sensitivity, and f1 score are used to show the significance. The proposed ensemble classifier is compared with other individual classifiers. The experimental analysis signifies that proposed ensemble stacking has achieved promising and competitive performance of cardiac disease classification and prediction for all heart disease datasets. The study proves that the ensemble stacking can be a good classification algorithm for the accurate prediction of heart disease with an improved accuracy of 91 % for balanced dataset, 81.10 % & 86.97 % for moderately imbalanced dataset and 75.92 % for imbalanced dataset. The comparative study on different dataset proves that the proposed method gives better accuracy than other machine learning techniques. The high value of AUC represents the proper distinguish between positive and negative classes. The research may help medical practitioners for proper diagnosis based on symptoms

REFERENCES

- [1] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review," 2017. [Online]. Available: <http://www.ripublication.com>.
- [2] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, "Feature Analysis of Coronary Artery Heart Disease Data Sets," in *Procedia Computer Science*, 2015, vol. 65, pp. 459–468, doi: 10.1016/j.procs.2015.09.132.
- [3] S. Sahin, M. R. Tolun, and R. Hassanpour, "Hybrid expert systems: A survey of current approaches and applications," *Expert Systems with Applications*, vol. 39, no. 4. pp. 4609–4617, Mar. 2012, doi: 10.1016/j.eswa.2011.08.130.
- [4] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, "Feature Analysis of Coronary Artery Heart Disease Data Sets," in *Procedia Computer Science*, 2015, vol. 65, pp. 459–468, doi: 10.1016/j.procs.2015.09.132.
- [5] K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing Coronary Heart Disease Using Ensemble Machine Learning," 2016. [Online]. Available: www.ijacsa.thesai.org.
- [6] J. Vijayashree and N. C. S. N. Iyengar, "Heart disease prediction system using data mining and hybrid intelligent techniques: A review," *International Journal of Bio-Science and Bio-*

- Technology, vol. 8, no. 4. Science and Engineering Research Support Society, pp. 139–148, 2016, doi: 10.14257/ijbsbt.2016.8.4.16.
- [7] S. Bashir, U. Qamar, and F. H. Khan, “IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework,” *J. Biomed. Inform.*, vol. 59, pp. 185–200, Feb. 2016, doi: 10.1016/j.jbi.2015.12.001.
- [8] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning,” *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/8387680.
- [9] N. S. Chandra Reddy, S. Shue Nee, L. Zhi Min, and C. Xin Ying, “Classification and Feature Selection Approaches by Machine Learning Techniques: Heart Disease Prediction,” *Int. J. Innov. Comput.*, vol. 9, no. 1, May 2019, doi: 10.11113/ijic.v9n1.210.
- [10] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, “Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System,” *World J. Eng. Technol.*, vol. 06, no. 04, pp. 854–873, 2018, doi: 10.4236/wjet.2018.64057.
- [11] H. Kaur and S. Batra, “HPCC: An Ensembled Framework for the Prediction of the Onset of Diabetes.”
- [12] S. Bashir, U. Qamar, F. H. Khan, and L. Naseem, “HMF: A medical decision support framework using multi-layer classifiers for disease prediction,” *J. Comput. Sci.*, vol. 13, pp. 10–25, Mar. 2016, doi: 10.1016/j.jocs.2016.01.001.
- [13] E. Nikookar and E. Naderi, “Hybrid Ensemble Framework for Heart Disease Detection and Prediction,” 2018. [Online]. Available: www.ijacsa.thesai.org.
- [14] S. Bashir, U. Qamar, and F. Hassan, “Bagmoov: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting,” *Australas. Phys. Eng. Sci. Med.*, vol. 38, no. 2, pp. 305–323, Jun. 2015, doi: 10.1007/s13246-015-0337-6.
- [15] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-Aug, pp. 785–794, doi: 10.1145/2939672.2939785.
- [16] M. Jan, A. A. Awan, M. S. Khalid, and S. Nisar, “Ensemble approach for developing a smart heart disease prediction system using classification algorithms,” *Res. Reports Clin. Cardiol.*, vol. Volume 9, pp. 33–45, Dec. 2018, doi: 10.2147/rrcc.s172035.
- [17] M. Ghosh and G. Sanyal, “Performance Assessment of Multiple Classifiers Based on Ensemble Feature Selection Scheme for Sentiment Analysis,” *Appl. Comput. Intell. Soft Comput.*, vol. 2018, 2018, doi: 10.1155/2018/8909357.

- [18] S. B. Patel, P. Kumar Yadav, and D. P. Shukla, "Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques." [Online]. Available: www.iosrjournals.org
- [19] S. A. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment," *Appl. Intell.*, vol. 48, no. 5, pp. 1233–1250, May 2018, doi: 10.1007/s10489-017-1102-1.
- [20] X. Y. Gao, A. Amin Ali, H. Shaban Hassan, and E. M. Anwar, "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6663455.
- [21] H.M.Veena kumari and D.S. Suresh, "Comparative Study of Classification algorithms used for the Prediction of Non-communicable diseases," *Int. J. Emerg. Trends Eng. Res.*, vol. 9, no. 7, pp. 917–920, 2021, doi: 10.30534/ijeter/2021/14972021.