***Article Received:** 20 May 2023 **Revised:** 28 May 2023 **Accepted:**16 August 2023*

Hybrid Ensemble Stacking Techniques for Coronary Artery Disease Prediction using Machine Learning Algorithms

CMM. Mansoor
Senior Lecturer, Department of Infromation Technology,
South Eastern University of Sri Lanka,
Sri Lanka
mansoorr@seu.aclk
Sarat Kumar Chettri
Assistant Professor,
School of Technology,Assam Don Bosco University,India
Sarat.chettri@dbuniversity.ac.in
HMM. Naleer
Professor , Department of Computer Science, Faculty of Applied Sciences, South Eastern University of Sri Lanka,Sri Lanka
hmmnaleer@seu.ac.lk

Abstract— Throughout history, humanity has been plagued by several outbreaks that have claimed numerous lives. Since coronary artery disease is among the most fatal illnesses that humanity has faced in the modern era, it has been recognized in our time. It links several Coronary Artery Disease (CAD) risk factors to the critical requirement for precise, reliable, and workable methods for early identification and management. In light of this, we suggest a technique called Hybrid Ensemble Stacking that combines Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Ada Boosting for the prediction of CAD illnesses. To combine the forecasts of the basis models, a meta-logistic regression model is utilized. According to a quantitative study, the ensemble model and brute force feature selection method together produce a classification accuracy for heart disease of up to 92.66%. The suggested stacking model has demonstrated its effectiveness and outperforms current methods in the categorization of cardiac disorders. Several classification issues have been solved successfully using ensemble techniques. The suggested method was constructed using the Sani dataset, which contains 303 nearly completed records. Using Min-Max Normalization, the data are pre-processed to making it suitable for a Machine Learning (ML) model. SMOTE and SelectKBest technique were applied to increases the accuracy and efficiency of a model. Using the metrics such as accuracy, precision, recall, F1, ROC and log-loss, the outcomes produced by the suggested model had the greatest performance.

Keywords— Machine Learning, Stacking Techniques, Coronary Artery Disease, Ensemble Model

I. INTRODUCTION

The most crucial organ in the heart, body, pumps blood to every part of the human body. The heart is interconnected to the circulatory system's arteries, capillaries, and veins. The heart is prone to illness and injury even though it is one of the most vital organs in the body. Because of the danger it brings to individuals and the diseases and injuries it reasons, its existence cannot be ignored. The heart's pumping mechanisms are affected by cardiac illness, which makes them incompetent. A group of diseases known as cardiovascular illness [1][2] are brought on by heart issues. Shortness of breath, physical weakness, swollen feet, and weariness are signs of cardiovascular disease. A few risk factors for cardiovascular disease include elevated blood pressure, smoking, elevated cholesterol levels, and a lack of exercise [3].

Medical diagnosis is using machine learning (ML) more and more. This might be partly explained by improvements in the categorization and identification of illnesses, data that helps doctors find and diagnose ailments, maintaining people's health, and a decline in deaths. ML classification algorithms are frequently used to estimate the likelihood of a disease occurrence. Recently, writers have improved the functionality of these models in the healthcare domain using ensemble learning. In order to enhance the outcome, ensemble learning combines the judgments of several basic classifiers using a variety of procedures, such as voting or averaging. Boosting, stacking, and bagging are the three branches of ensemble algorithms [4]. Because it is built on a meta-learner that learns from data how to weight the basic classifiers and merge them in the optimal way to optimize the performance of the final model, stacking ensemble is regarded as the best method for creating ensemble models. Using a meta-learner, ensemble stacking optimizes a collection of diverse base models as well as combinations their judgments.

We provide an effective ensemble stacking model in this study to enhance CAD prediction [2]. The suggested model is built using a variety of heterogeneous classifiers as base models, such as K-NN, SVM [4], Naive Bayes, Ada boosting, and Extreme Gradient Boosting (XGBoost). In order to aggregate the forecasts of the base models and arrive at the final choice, a Meta model based on RF is also employed. Sani datasets are used for experimental evaluation [5]. Additionally, choosing features is a crucial step. In data mining methods, the more crucial feature subset that contributes the most to output is chosen in order to minimize the dimensionality of the input features [6]. We utilized the feature selection strategy in this paper. The proposed ensemble model will be constructed and trained using the UCI Cleveland dataset [7].

This is how the remainder of the paper is structured. After reviewing prior research that was pertinent to the CAD forecast in Section II, we addressed methodology in Section III with appropriate suggested methodologies and classification algorithms. In Section IV, the outcomes generated by every model are shown together with the precision of the author's forecast. Section V provides an outline of the conclusion and potential improvements.

II. LITERATURE REVIEW

Knowing a person's risk of heart failure when they have a cardiac condition is challenging. The prediction approach is built with the aid of several longitudinal auto-regression investigations. Healthcare organisations now have to manage vast volumes of data in their databases as a result of technological advancements [8], which makes data analysis incredibly difficult [9]. The heart ailment known as coronary heart disease (CHD) results in myocardial ischemia and hypoxia because to coronary atherosclerosis. The World Health Report [10] states that the mortality rate of cardiovascular illnesses is currently the biggest hazard to human health globally. CAD [11] [12] is the most common kind of cardiovascular disease (CVD). It is the main reason of death in the entire planet [13]**.** The performance of DL techniques is superior to that of extracted features used in traditional ML classifiers since they learn features from the training data. Recurrent neural networks (RNN), CNN, LSTM [14], and gated recurrent units (GRU) are examples of modern designs. The current networks think that every ailment has a single therapy [15]. Researchers may assess how well modern architectures like LSTM [16] and GRU perform when diagnosing patients' heart problems by applying them to the provided dataset. In some ways, coronary angiography is the most accurate method for detecting CAD [17]. The degree of artery stenosis may be identified using the pictures from coronary angiography to assess the severity of CAD [18] [19] and ACS. It is, however, a costly and time-consuming method [20].

Statistics from the World Health Organisation (WHO) show that CAD killed 17.7 million people globally, or 31% of all deaths. Rapid action and accurate CAD detection may have avoided many of these CAD deaths [21]. In this paper, a unique approach to CAD identification is proposed that combines evolutionary algorithms and a variety of well-established machine learning approaches [22]**.** CAD and electrocardiography (ECG), which tracks the electrical activity of the heart, are related. ECG should be investigated as a viable screening tool because to its non-invasiveness, affordability, and wide operational availability [23] . However, using patient ECG recordings to accurately screen for CAD patients might be difficult due to substantial inter- and intra-individual variability [24].

The hybrid method increases the classification algorithm's recognition accuracy even with less features. Using a genetic algorithm, a neural network is optimized to achieve a recognition accuracy of 89%. Combine feature subset selection with optimization techniques to increase forecast accuracy. This technology is further employed in conjunction with classification and clustering algorithms to increase the accuracy of classification approaches[25]. Utilizing the XGBoost Classifier, the specificity and sensitivity were 88.79% and 87.17%, respectively. The stress test, which has a specificity range of 70% to 80% and a sensitivity range of 60% to 70%, and these findings beat screening tests that are now available [7]. An ECG-based ML system is a very practical and flexible screening tool. An ECG may be obtained quickly, painlessly, and affordably in comparison to other screening and triage methods, allowing for its use even at the first point of contact [24].

The heart, the body's most crucial organ, distributes blood to every part of the body. If the heart isn't beating properly, the mind and other organs will stop working, and the person will die in a matter of minutes. Therefore, healthy heart function is crucial. Globally, heart diseases are becoming among the foremost causes of death. Consequently, numerous academics from across the globe initiated to focus on using the massive databases to predict cardio related illnesses. Various ML [26] approaches may examine large datasets and produce insightful results. Since ML models use a variety of methodologies, these algorithms are crucial for properly predicting whether or not cardiac abnormalities would arise[27].

Furthermore, researchers are looking at the best method for identifying this condition. The outcomes of the trials unambiguously show that the ML algorithms LR, RF, MLP, and DLDA are successful for best attributes datasets. The LR algorithm is just somewhat more accurate overall than the others. It offers 90.88% accuracy. It is conceivable to state that the greatest attributes were successfully identified in this circumstance. It is planned to use huge datasets with more features in the future to get better performance and more insightful results. It also focuses on deep learning technique experiments between [28].

III. METHODOLOGY

This section initially gives an overview of the resources (such as the dataset) used in our studies. The feature engineering methods taken into consideration for our research are then briefly discussed. In addition, we provide a description of the suggested hybrid stacking model for CAD diagnosis.

A. Data Description

The UCI repository provided the CAD dataset that were used in this study. The suggested method was constructed using the Sani dataset, which consists of 303 nearly finished entries. Because of its high usage, the dataset may be used to compare identification accuracy results with those from other studies. We have selected 12 features with 303 instances.

TABLE I. CAD DATASET

Feature Name	**Range**	**Data Types**	**Mean**
Typical Chest Pain	[No,Yes]	int64	0.54
Atypical	[No,Yes]	Object	---
Region RWMA	[0,1,2,3,4]	int64	0.62
DM	[No,Yes]	int64	0.30
Tinversion	[No,Yes]	int64	0.30
HTN	[No,Yes]	int64	0.60
Q Wave	[No,Yes]	int64	0.05
St Depression	[No,Yes]	int64	0.23
Nonanginal	[No,Yes]	Object	---
Current Smoker	[No,Yes]	int64	0.20
St Elevation	[No,Yes]	int64	0.05
Sex	[Female,Male]	Object	---

B. Preprocessing

First, we must deal with missing data, which are records containing partial observations that might lead to skewed findings and reduce the overall effectiveness of the model. It has no missing values. Second, we have applied Label Encoder to change format. Depending on how labels are applied, ML models produce superior results. In order for machines to interpret the target labels for better illness diagnosis, the Label Encoder converts them into numeric representations. Then, we used SMOTE [29] and ADASYN [30] techniques are used for data balancing. The data sets were balanced using the SMOTE technique. SMOTE is an oversampling method used to produce fictitious minority samples.

As illustrated in Figure 2, the minority class is oversampled using a method that makes up fake sample classes for KNN and combines informed oversampling of the minority class with random under-sampling of the majority class. By artificially inflating the number of members of the minority class to match the number of members of the majority class, SMOTE balances a data set.

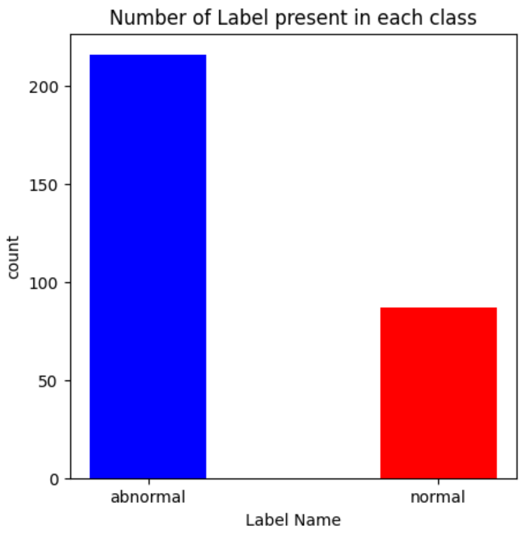


Fig. 1. Number of label present in each class

Fig.1 elaborates the different between normal and abnormal label presented in each class. The majority of labels are abnormal.

C. Feature Engineering Techniques

To decrease the number of dimensions of the input features, enhance model performance, besides minimize model calculation time, feature engineering approaches are preferred. Traditionally, choosing the most essential attributes has an impact on the final model's power and predictability. There are several feature selection techniques available. In this case, we chose 12 features from the sani dataset using the SelectKBest and chi2 approaches to identify more closely related features.

D. Feature Importance

"Feature importance" refers to how every input feature is scored for a particular model; the scores just express the "importance" of each feature. An elevated score signifies that the specific attribute will exert a greater influence on the model utilized to predict a given variable. When developing a predictive model in ML, feature significance scores are used to assess the relative value of each feature in a dataset. Many other methods, including DT, RF, linear models, and neural networks, are used to determine these ratings.

In order to rate the characteristics according to how much they contribute to the final forecast, feature importance can be used. Although this may need subject-matter expertise, it may be utilized for feature selection, which is the process of choosing a subset of pertinent features to be used in developing a model. In this work we have selected 12 features based on scores.

E. Feature selection

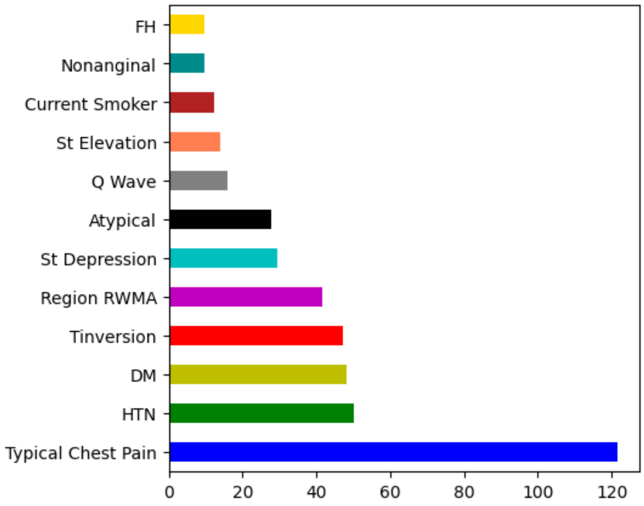


Fig. 2. Selected features

In figure 2 selected features are given with the scores. In order to that there are 12 features were selected. The highest score gained by typical chest pain and lowest was FH. Regarding to this the relatable table given below.

TABLE II. FEATURE AND ITS SCORES

Features	**Scores**
Typical Chest Pain	126.43
DM	52.32
HTN	44.18
Tinversion	43.57
Region RWMA	39.42
Atypical	27.76
St Depression	26.16
Q Wave	16.00
Current Smoker	15.52
St Elevation	14.00
FH	12.00
Poor R Progression	9.00

The above table is expounded accordance to the figure 2. The scores are given precisely. From 126 to 9 score were gained.

F. Correlation

In order to analyze the data, a correlation value between each dataset attribute and the intended diagnosis was established. The characteristic with the strongest correlation to the target feature is Typical Chest Pain, whereas the one with the lowest correlation is Nonanginal. The linked value with the target property is shown in Table II. When comparing the data being handled with an analysis, this is quite beneficial.

Article Received: *20 May 2023* ***Revised:*** *28 May 2023* ***Accepted:****16 August 2023*

TABLE III. CORRELATION SCORE WITH TARGET VARIABLE

1	Cath	1.0000
2	Typical Chest Pain	0.6813
3	HTN	0.4220
4	Region RWMA	0.4143
5	DM	0.3773
6	Tinversion	0.3731
7	Atypical	0.2835
8	St Depression	0.2779
9	Q Wave	0.1961
10	Current Smoker	0.1832
11	St Elevation	0.1830
11	Sex	0.1713
13	Nonanginal	0.1542

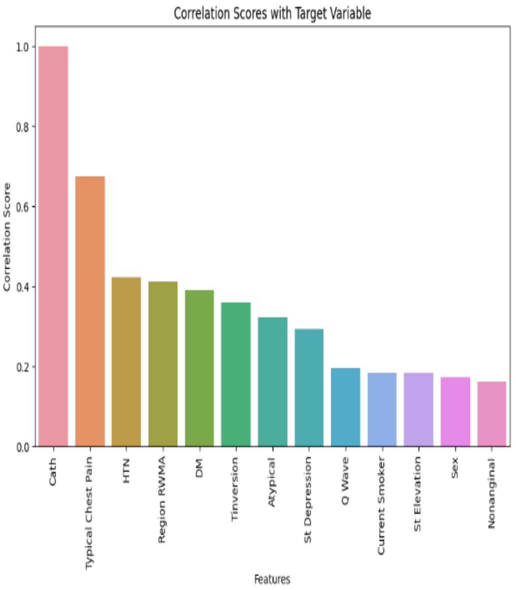
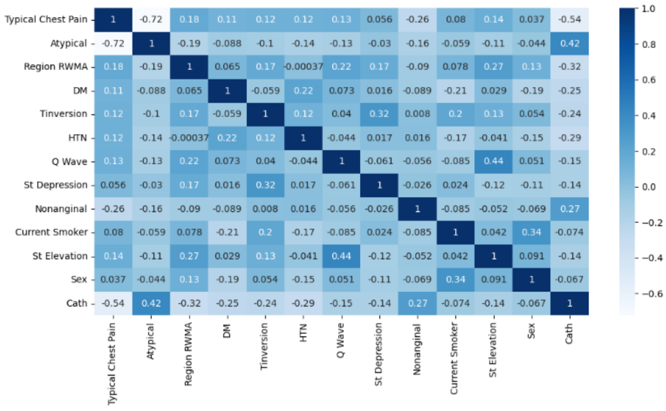


Fig. 3. Correlation score with target variable

The concept of correlation describes the connections between one or more variables. These factors might be characteristics of the raw data that were utilized to forecast our target variable. A statistical approach called correlation shows how one variable moves or changes in connection to additional one. It delivers us with an overall understanding of how thoroughly the two variables are related. This bi-variate analysis measure explains the relationship between many variables. The correlation matrix, which represents the correlation coefficients between sets of variables, is seen in Fig.4.



	Typical Chest Pain	Atypical	Region RWMA	DM	Tinversion	HTN	Q Wave	St Depression	Nonanginal	Current Smoker	St Elevation	Sex	Cath
Typical Chest Pain	1	-0.72	0.18	0.11	0.12	0.12	0.13	0.056	-0.26	0.08	0.14	0.037	-0.54
Atypical	-0.72	1	-0.19	-0.088	-0.1	-0.14	-0.13	-0.03	-0.16	-0.059	-0.11	-0.044	0.42
Region RWMA	0.18	-0.19	1	0.065	0.17	-0.00037	0.22	0.17	-0.09	0.078	0.27	0.13	-0.32
DM	0.11	-0.088	0.065	1	-0.059	0.22	0.073	0.016	-0.089	-0.21	0.029	-0.19	-0.25
Tinversion	0.12	-0.1	0.17	-0.059	1	0.12	0.04	0.32	0.008	0.2	0.13	0.054	-0.24
HTN	0.12	-0.14	-0.00037	0.22	0.12	1	-0.044	0.017	0.016	-0.17	-0.041	-0.15	-0.29
Q Wave	0.13	-0.13	0.22	0.073	0.04	-0.044	1	-0.061	-0.056	-0.085	0.44	0.051	-0.15
St Depression	0.056	-0.03	0.17	0.016	0.32	0.017	-0.061	1	-0.026	0.024	-0.12	-0.11	-0.14
Nonanginal	-0.26	-0.16	-0.09	-0.089	0.008	0.016	-0.056	-0.026	1	-0.085	-0.052	-0.069	0.27
Current Smoker	0.08	-0.059	0.078	-0.21	0.2	-0.17	-0.085	0.024	-0.085	1	0.042	0.34	-0.074
St Elevation	0.14	-0.11	0.27	0.029	0.13	-0.041	0.44	-0.12	-0.052	0.042	1	0.091	-0.14
Sex	0.037	-0.044	0.13	-0.19	0.054	-0.15	0.051	-0.11	-0.069	0.34	0.091	1	-0.067
Cath	-0.54	0.42	-0.32	-0.25	-0.24	-0.29	-0.15	-0.14	0.27	-0.074	-0.14	-0.067	1

Fig. 4. Correlation Metrics

G. Proposed Ensemble Model

The suggested method enhances CD categorization and detection by combining efficient feature selection techniques with a hybrid stacking model. The main concept behind stacking is to use a meta model to aggregate diverse predictions from distinct models. Our suggested method makes use of a number of ML techniques, including KNN, Naive Bayes, XGBoost, SVM, and Ada Boost. Due to the use of numerous classification models with varying aptitudes for resolving classification challenges, this design enhances classification accuracy over individual classifiers. Additionally, figure 3 shows the stacking hybrid machine learning technique. Our stacking model's design is divided into two primary stages: Base Models Stage and Meta Model Stage. Combining the aforementioned methodologies, which have two stages, is known as stacking distinct machine learning approaches. Learners using the stacking approach are at level, whereas beginners are at level 1.Algorithm 2 explains the processes. At the first level, the suggested models are used to train the datasets, and the results from these models are used to create a new dataset.

H. Base Models Stage

In this step, often referred to as "Level-0 Classifiers," we take into account five distinct classification algorithms: K-NN, SVM, Naive Bayes, Ada Boosting, and Extreme Gradient Boosting, all of which have demonstrated success in dealing with the forecast of CAD illness. To prevent equal CAD and Non-CAD predictions for an observation, use an odd number of models (five) in the Base Model Stage

I. Meta Model Stage

In this stage, we employ a meta model classifier Random Forest to combine the predictions of the basis models. The input features and the five base model prediction vectors are combined, and the meta model is then trained using these features. The training process for the suggested hybrid stacking model is shown in Figure 5. Sani dataset is used to train the model.

J. *Proposed methodology*

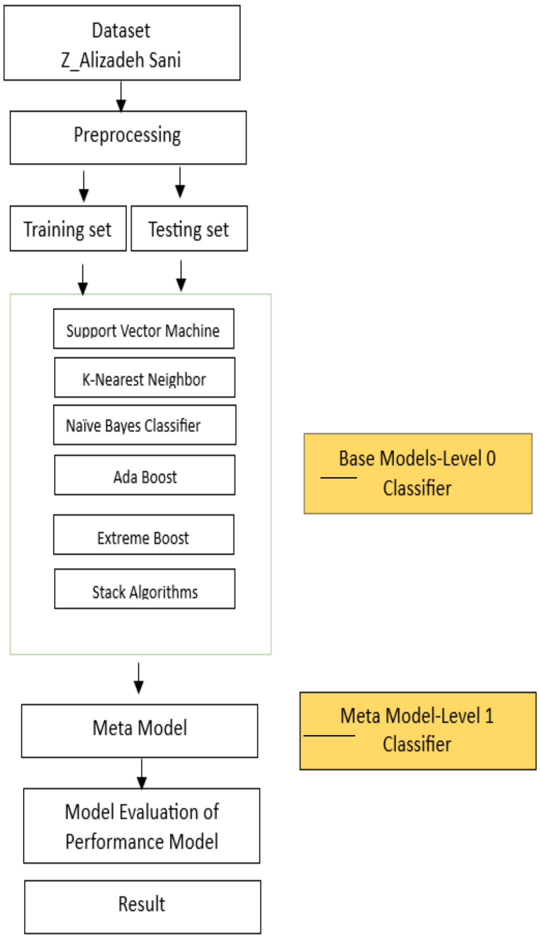


Fig. 5. Proposed Hybrid ML stacked based approach for the cad disease prediction

K. Performance Metrics

Performance metrics such as sensitivity, F-score, specificity, recall, precision, accuracy, and ROC, are used to assess the effectiveness of CAD identification. Its correctness is measured by the accurately categorized class. Precision is defined as the proportion of the overall number of positive instances to the real positive cases (false positive and truly positive cases). A clinical prediction model should retain a high level of specificity, sensitivity, and recall. Recall is a measure of how successfully a capacity to construct true positive occurrences is formed. Specificity is the quantity of real negative cases (false negative and actually negative) that the model accurately classified as negatives.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{1}$$

$$\text{Sensitivitity} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{3}$$

$$\text{Precision} = \frac{\text{TN}}{\text{TP+FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$\text{F} - \text{score} = 2 \text{ x} \frac{\text{TP}}{\text{TP+FP}} \tag{6}$$

L. Confusion Matrix

TABLE IV. CONFUSION METRICS

Actual Class	**Predictive Class**	
	Positive	**Negative**
Negative	FP	TN
Positive	TP	FN

IV. RESULTS AND DISCUSSION

TABLE V. COMPARATION OF DIFFERENT CLASSIFIERS

Model	**Precision**	**Recall**	**F1Score**	**ROC**
KNN	92.19%	83.33%	87.72%	86.41%
SVM	92.04%	87.93%	91.07%	90.26%
NB	90.74%	76.35%	81.20%	75.69%
ADB	91.44%	89.47%	91.89%	90.29%
XGB	92.89%	83.33%	87.72%	88.94%
STACK	93.59%	90.91%	92.74%	92.04%

Based on the most widely used ML algorithms in the field of illness diagnosis and classification, we used six distinct well-known classifiers for the diagnosis of CAD in this research to achieve good performance in the diagnosis of CAD using a UCI dataset. With regard to patient information suspected of having CAD, we evaluated the effectiveness of selected algorithms.

When comparing different classifiers KNN has the 92.79% of precision, 83.33% of recall, 87.72% of F1-score and 86.41% of ROC. Next, SVM has 92.04%, 87.93%, 91.07% and 90.26% of precision, recall, F1Score and ROC respectively. Then, NB has the 90.74% of precision, 76.35% of recall, 81.20% of F1-

score and 75.69% of ROC. Next ADB having a little increased precision, recall, F1-score and ROC. In this sense, it gained 91.44%, 89.47%, 91.89% and 90.29% accordingly. XGB has 92.89% of precision, 83.33% of recall, 87.72% of F1-score and 88.94% of ROC. Finally, STACK has 93.59% precision, 9.91% of recall, 92.74% of F score and 92.04% of ROC. Considering these all classifiers, STACK has the increased percentage among others.

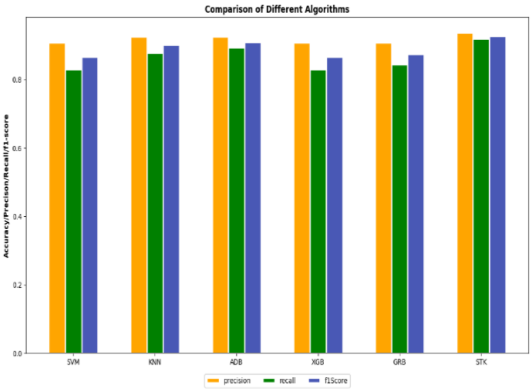


Fig. 6. Comparision of diffreent algorihtms

Fig.6 executes the different algorithms with different metrics. Especially it shows precision, recall and F1-score of each algorithm. This is same as TABLE VI. In this sense, STACK has the highest value of precision, recall and f1-score, which are 93.59%, 90.91% and 92.74% respectively.

A. Accuracy

TABLE VI. ACCURACY OF DIFFERENT ALGORITHMS

Model	Accuracy
KNN	87.03%
SVM	90.74%
NB	76.85%
ADB	90.90%
XGB	87.03%
STACKING	92.66%

TABLE VII showing the accuracy of each classifiers. KNN has 87.03% of accuracy. Then, SVM has 90.74%, NB has 76.85%, ADB has 90.90%, XGB has 87.03% and finally STACKING has 92.66%. According to the table STACKING has the highest accuracy. Meanwhile NB has the lowest level of accuracy. This value's graphical representation is given below.

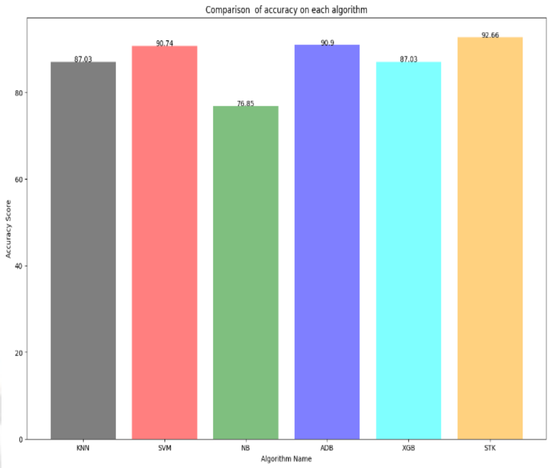


Fig. 7. Comparision of accuracy of diffreent algorihtm

B. ROC

AUC is a metric or level of separateness, whereas ROC is a likelihood curve. It indicates the degree to which the framework can discriminate among classes. The algorithm performs better at identifying 0 courses as 0 and 1 classes as 1, with a greater the AUC. In a similar vein, a model's ability to discriminate between individuals who have the condition and those who do not is indicated by a higher AUC. The AUC of the ROC curve for the stacking model was 0.920, which was greater than the model and indicated good accuracy in differentiating between negative and positive situations.

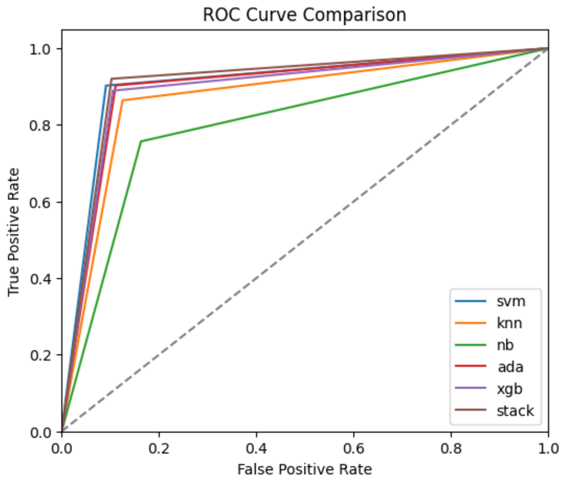


Fig. 8. Comparision of ROC curve

In overall these research articles increased their accuracy level by using various classifiers or algorithms. When compared to existing systems, our unique system had the best accuracy of 93%. Actually, accuracy is the traditional metric in ML. It's a statistic used to evaluate how well models perform in classification tasks. In this regard, our innovative method with the best accuracy outperforms all previous algorithms.

V. CONCLUSION

A hybrid ensemble stacking model is used in this study to enhance CD detection. The suggested model is built using a variety of heterogeneous classifiers as basis models, including RF, DT, K-NN, SVM, and Naive Bayes. The final choice is also produced using a meta model based on LG, which combines the predictions from the basic models. To choose the most significant characteristics with a strong influence on the model's efficiency, a variety of feature selection strategies were used. The UCI Cleveland dataset was used for model construction and the recognition of important characteristics. The UCI dataset was used to assess the effectiveness of the suggested model and characteristics that were retrieved. According to experimental assessment, the ensemble model combined with the basic force feature selection method produces a highest accuracy of 92.66%, outperforming other techniques to CD classification.

REFERENCES

[1] S. Dudakiya, H. Galani, A. Shaikh, D. Thanki, R. A. Late, and S. E. Pawar, “Monitoring mobile patients using predictive analysis by data from wearable sensors,” *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, pp. 332–335, 2016, doi: 10.1109/ICEEOT.2016.7755305.

[2] K. Akyol, “Coronary artery disease classification with support vector machines tuned via randomized search,” pp. 1–15, 2022.

[3] A. Kishor and C. Chakraborty, “Artificial Intelligence and Internet of Things Based Healthcare 4.0 Monitoring System,” *Wireless Personal Communications*, no. 0123456789, 2021, doi: 10.1007/s11277-021-08708-5.

[4] I. C. Dipto, T. Islam, H. M. M. Rahman, and M. A. Rahman, “Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease,” *Journal of Data Analysis and Information Processing*, vol. 08, no. 02, pp. 41–68, 2020, doi: 10.4236/jdaip.2020.82003.

[5] O. Debauche, S. Mahmoudi, P. Manneback, and A. Assila, “Fog iot for health: A new architecture for patients and elderly monitoring.,” *Procedia Computer Science*, vol. 160, pp. 289–297, 2019, doi: 10.1016/j.procs.2019.11.087.

[6] D. C. Yadav and S. Pal, “Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques,” *International Journal of Big Data and Analytics in Healthcare*, vol. 6, no. 1, pp. 40–56, 2021, doi: 10.4018/ijbdah.20210101.oa4.

[7] L. Verma, S. Srivastava, and P. C. Negi, “A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data,” *Journal of Medical Systems*, vol. 40, no. 7, 2016, doi: 10.1007/s10916-016-0536-z.

[8] S. A. Asra, “Various Technologies used in Smart Transportation System in the Direction of a Smart City : A Systematic Review,” 2022.

[9] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, “Enhanced Heart Disease Prediction Based on Machine Learning and χ2 Statistical Optimal Feature Selection Model,” *Designs*, vol. 6, no. 5, 2022, doi: 10.3390/designs6050087.

[10] C. J. Qin, Q. Guan, and X. P. Wang, “Application of ensemble algorithm integrating multiple criteria feature selection in coronary heart disease detection,” *Biomedical Engineering - Applications, Basis and Communications*, vol. 29, no. 6, pp. 1–11, 2017, doi: 10.4015/S1016237217500430.

[11] O. Y. Atkov *et al.*, “Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters,” *Journal of Cardiology*, vol. 59, no. 2, pp. 190–194, 2012, doi: 10.1016/j.jjcc.2011.11.005.

[12] M. Shamsollahi, A. Badiee, and M. Ghazanfari, “Using Combined Descriptive and Predictive Methods of Data Mining for Coronary Artery Disease Prediction: a Case Study Approach,” *Journal of AI and Data Mining*, vol. 7, no. 1, pp. 47–58, 2019, doi: 10.22044/JADM.2017.4992.1599.

[13] B. Balakrishnan, K. Kumaran, A. Sankar, and K. Rao, “Coronary Artery Disease Prediction using Machine Learning Algorithms,” in *Proceedings of the 2022 International Conference on Electronic Systems and Intelligent Computing, ICESIC 2022*, 2022, pp. 95–100, doi: 10.1109/ICESIC53714.2022.9783592.

[14] A. Patel and J. Shah, “Sensor-based activity recognition in the context of ambient assisted living systems: A review,” *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 4, pp. 301–322, 2019, doi: 10.3233/AIS-190529.

[15] M. Hartmann, U. S. Hashmi, and A. Imran, “Edge computing in smart health care systems: Review, challenges, and research directions,” *Transactions on Emerging Telecommunications Technologies*, no. December 2018, pp. 1–25, 2019, doi: 10.1002/ett.3710.

[16] S. Majumder and M. J. Deen, "Smartphone sensors for health monitoring and diagnosis," *Sensors (Switzerland)*, vol. 19, no. 9, pp. 1–45, 2019, doi: 10.3390/s19092164.

[17] I. Javid, A. K. Z. Alsaedi, and R. Ghazali, "Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 540–551, 2020, doi: 10.14569/ijacsa.2020.0110369.

[18] T. Suresh, T. A. Assegie, S. Rajkumar, and N. K. Kumar, "A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 1831–1838, 2022, doi: 10.11591/ijece.v12i2.pp1831-1838.

[19] K. Nazeri, A. Aminpour, and M. Ebrahimi, *Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification*, vol. 10882 LNCS. Springer International Publishing, 2018.

[20] A. D. Jamthikar, D. Gupta, L. E. Mantella, L. Saba, A. M. Johri, and J. S. Suri, "Ensemble Machine Learning and Its Validation for Prediction of Coronary Artery Disease and Acute Coronary Syndrome Using Focused Carotid Ultrasound," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, doi: 10.1109/TIM.2021.3139693.

[21] F. Ali *et al.*, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208–222, 2020, doi: 10.1016/j.inffus.2020.06.008.

[22] N. Kazmi and T. R. Gaunt, "Diagnosis of coronary heart diseases using gene expression profiling; Stable coronary artery disease, cardiac ischemia with and without myocardial necrosis," *PLoS ONE*, vol. 11, no. 3, pp. 1–16, 2016, doi: 10.1371/journal.pone.0149475.

[23] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," *Knowledge-Based Systems*, vol. 132, pp. 62–71, 2017, doi: 10.1016/j.knosys.2017.06.003.

[24] G. Phadke, M. R. Rajati, and L. Phadke, "Prediction of coronary artery disease using electrocardiography: A machine learning approach," *Proceedings - International Conference on Machine Learning and Cybernetics*, vol. 2020-Decem, pp. 175–180, 2020, doi: 10.1109/ICMLC51923.2020.9469585.

[25] M. Sayadi, V. Varadarajan, F. Sadoughi, S. Chopannejad, and M. Langarizadeh, "A Machine Learning Model for Detection of Coronary Artery Disease Using Noninvasive Clinical Parameters," *Life*, vol. 12, no. 11, 2022, doi: 10.3390/life12111933.

[26] L. J. Al'Aref, S. J., Maliakal, G., Singh, G., van Rosendael, A. R., Ma, X., Xu, Z., ... & Shaw, "Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry.," *European heart journal*, vol. 41, no. 3, pp. 359–367, 2020.

[27] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, "Cardiovascular disease prediction using deep learning techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 981, no. 2, 2020, doi: 10.1088/1757-899X/981/2/022006.

[28] K. AKYOL, "Feature Selection Based Data Mining Approach for Coronary Artery Disease Diagnosis," *Academic Platform Journal of Engineering and Science*, vol. 9, no. 3, pp. 451–459, 2021, doi: 10.21541/apjes.899055.

[29] E. Nasarian *et al.*, "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, doi: 10.1016/j.patrec.2020.02.010.

[30] Z. Ahmad, "Network intrusion detection system : A systematic study of machine learning and deep learning approaches," no. May 2020, pp. 1–29, 2021, doi: 10.1002/ett.4150.