Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques

Saba Yasmeen

PG Student, E&CE (DCN) Dept.

Guru Nanak Dev Engineering College, Bidar.

yasmeensaba961@gmail.com

Dr. Pradeep Karanje

Assoc.Professor, E&CE Dept.,

Guru Nanak Dev Engineering College, Bidar (Karnataka)

Pradeepkaranje2012@gmail.com

Abstract: heart disease is one of the most common causes of death around the world, and finding it early is very important for lowering death rates. it is still hard to accurately anticipate cardiac disease because medical data is complicated and we can't follow patients all the time. using the heart disease dataset, we used different feature selection methods, "such as the ANOVA F-statistic (ANOVA FS), the Chi-squared test (Chi2 FS), and the Mutual information (MI FS), to find important predictors. We used the synthetic Minority Oversampling technique (SMOTE) to fix the data imbalance and make the model work better." We used a wide range of machine learning models and ensemble approaches to do a full categorization. "Of these, a Stacking Classifier that combined Boosted decision trees, extra trees, and LightGBM did the best, getting 100% accuracy with all feature selection methods." The high performance shows how well advanced ensemble learning works for making reliable heart disease predictions. This shows how powerful it could be to combine strong feature selection with complex classification models for accurate medical data analysis. This method shows that it can help with early diagnosis and better outcomes for patients.

***"Index Terms -** Cardiovascular disease, heart disease, machine learning app, ML algorithms, SDG 3, SHAP, SMOTE".*

1. INTRODUCTION

heart disease is a major cause of "cardiovascular disease (CVD)," which is one of the main causes of death around the world. The heart is a muscular organ that pumps blood all over the body and is an important part of the circulatory system. Arteries, veins, and capillaries make up this complex system that moves oxygen and nutrients to organs and tissues. CVD is the name for a group of heart diseases that happen while the normal flow of blood is interrupted. the "world health organization (WHO) says that heart disease and stroke kill around 17.5 million people every year, and more than 75% of these deaths happen in low- and middle-income countries." This scary number shows how heart disease is becoming a bigger public health problem around the world. "Heart attacks and strokes alone cause 80% of all CVD-related fatalities [1]."

Because cardiovascular diseases are so common, the world is now focused on finding ways to identify, prevent, and treat them early. As part of the United nations' Sustainable development goal 3, which stresses the significance of health and well-being, tackling cardiovascular illnesses has become a top priority for improving health around the world. some things that put you at risk for heart disease are smoking, being older, having a family history of heart disease, having high cholesterol, not exercising enough, having high blood pressure, being overweight, having diabetes, and being stressed. people who cease smoking, exercise regularly, manage their weight, and lower their stress levels are less likely to get heart disease [2]. along with modifications to your lifestyle, doctors often utilize diagnostic technologies like "electrocardiograms (ECGs), echocardiograms, cardiac MRIs, and blood tests to find heart problems. sometimes, medical procedures like angioplasty, coronary artery bypass surgery," and the use of implanted devices like pacemakers and defibrillators may be needed to help [3].

Improvements in healthcare technology, especially in the areas of big data and "electronic health records (EHRs), have made it possible to use a lot of patient data for predictive modeling. continually, more "machine learning (ML)" strategies are being applied to examine massive healthcare systems knowledge and identify valuable details that can be applied to make predictions of whether a person will be at risk of cardiovascular disease. ML can enable the doctors to identify high-risk patients and enable them to intervene early as it is able to preprocess and analyze data of various categories of patients, risk factors, and clinical outcomes. This technology has transformed how the healthcare sector is being operated by offering more of its accurate and efficient forms of diagnosis, prediction, and formulation of personal treatment plans [4][5].

2. RELATED WORK

"Cardiovascular diseases (CVD)," e.g. heart disease, have long been a worldwide health problem, and cause deaths of many people worldwide. With the expansion in the size of healthcare information and the development of ML methods, much effort has been put in enhancing the anticipation and diagnosis of heart illness. The ability of ML to view large sets of data presents possibilities of identifying new risk factors, predicting what is going to happen, and enhancing early detection. Various studies that have pre-determined heart disease by employing various machine learning models and mechanisms to come about their predictions are what have been discussed in this literature review.

In a study, Yang et al. [6] employed the ML to identify risk variables of coronary heart disease. They did mainly research on big data analysis." They also demonstrated that ML algorithms such as decision trees, random forests and even support vector machines (SVMs)" could discover vital risk factors related to patient health and disease. Some of the most significant risk variables that the research study identified as the ones that could assist in predicting heart disease included factors such as age, high cholesterol, and a history of heart diseases in the family. It was also revealed in the study that it is significant to prepare data and select the most suitable characteristics along with the good dataset to make models perform in a better way. The findings demonstrated that ML might become an extremely beneficial means to detect heart issues early, particularly in producing big and fully comprehensive datasets.

It is considered that Ngufor et al. [7] have examined several approaches to the machine learning predictions of the cardiac disease. "In a study, the authors compared the common approaches such as SVM, decision trees, k-nearest neighbors (KNN), and artificial neural networks (ANNs)." They found out that ensemble approaches to prediction, including bagging and boosting, predicted better than single models. The paper has also emphasized the significance of selection of appropriate characteristics as irrelevant variables may render the model a poor reflection. This review demonstrated that anything can predict cardiac disease with different algorithms, but the most appropriate one to use depends on what data is available, what computational resources there are, and what is the special requirements of the prediction problem.

The authors of the work by Farag et al. [8] were interested in improving the heart disease prediction using the technique of boosting and bagging. We looked into the boosting algorithms such as AdaBoost and the bagging schemes such as Random forest to determine their ability to enhance the accuracy of the predictions. The study has shown that the ensemble methods reduced more variance and increased stability in predictions as compared to the single classifiers. It was also found in the study that combining the boosting and bagging technologies may assist in alleviating the overfitting issue that is usually associated with the prediction of heart diseases. This experiment displayed the value of combining multiple classes with the aim of achieving the optimal outcome.

Zhang et al. [9] considered the way to employ XGBoost, which is a gradient boosting algorithm, to predict coronary heart disease in the clinic. They came up with the finding that XGBoost can be considered more precise and simpler to interpret than well-established approaches such as logistic regression or SVM. The use of XGBoost was excellent on medical data because it was able to handle three-skewed datasets, which were common in prediction of heart disease diagnosis. Study also added that, hyperparameters tuning was extremely critical towards obtaining the best performance of the model. XGBoost is more effective when used in a clinical basis because it is faster, can process a larger amount of data, and can reach a valid prediction and reduce the risk of overfitting.

By comparing various ML algorithms or methods, including decision trees, SVM, and random forests, Liu et al. [10] sought to establish the most practical ones in the prediction of heart disease. "They found that SVM with radial basis function (RBF)" is the most predictive of all the algorithms they tested. However, in medical context they also mentioned that it was easier to comprehend decision trees or random forests, which, indeed, is quite important. The analysis revealed that SVM was the most precise, nevertheless, the selection of the algorithm should depend on the trade-off effect between accuracy vs interpretability, particularly when the healthcare specialists are required to apply the model into decision-making.

"Hussein et al. [11] examined various ML algorithms in terms of their applicability in the diagnosis of cardiac disease, including KNN, artificial neural networks and decision trees." The models were tested on a set of health records of patients in order to observe their ability to diagnose well. In their findings, KNN and decision trees performed better at low computational expense and thus they could be effectively used in clinical practice in real time. They were more precise, however, requiring more computing power and being less understandable, so it was replaced by the use of so-called "artificial neural networks (ANN)." This was emphasized by the study that healthcare organizations must weigh the performance constraints and resource constraints of machine learning models.

Akbar et al. [12] conducted a stringent research on various ML techniques to forecast cardiac disease. " They considered such approaches as decision trees, KNN, SVM, random forests, and neural networks" and provided an overview of their effectiveness and ineffectiveness in predicting heart disease. "The authors concluded that ensemble methods, in particular, Random forest, have been the most accurate due to the fact that they allowed dealing with noisy data and overfitting." The review also specified that choosing the most critical traits among the large datasets is one of the toughest jobs, as the process of feature selection is highly important in terms of a model performance. Another research point the study also emphasized is the importance of handling the issue of missing data as well as ensuring that the training dataset is representative and balanced.

A study conducted by Zarshenas et al. [13] compared the different versions of ML algorithms to predict the cardiac disease. They considered the classifiers like SVM, decision trees, KNN and logistic regression. Their findings indicated the best performance in terms of "making predictions by SVM and Random forest although SVM was better in determing" prediction in a number of situations. they also discussed how critical preprocessing activities such as normalization and feature scaling are in making ML models perform better. The research indicated that in future, hybrid models, containing the finest segments of various algorithms, should be a path worth pursuing as far as the prediction of heart disease is concerned.

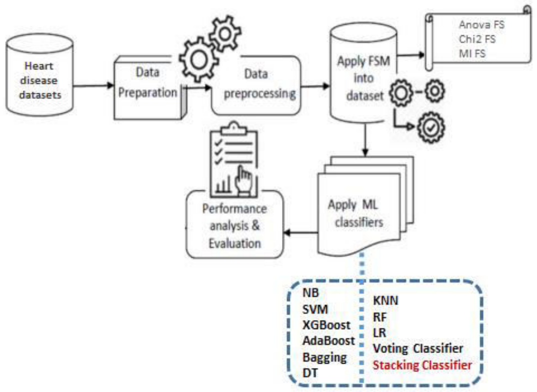
A recurrent topic in the studies looked at is how important it is to choose the right features and preprocess the data to make heart disease prediction models work better. Many research say that using ensemble methods like bagging and boosting can make models more accurate. other studies say that hybrid models that utilize more than one machine learning algorithm might work better. also, even whilst decision trees and logistic regression are still extensively employed, newer studies have shown that more advanced algorithms like XGBoost and neural networks can do better, especially when working with complicated and high-dimensional data.

The growing amount of healthcare data and the rise of machine learning methods have made it possible to find and diagnose cardiac disease earlier than ever before. ML models can help doctors find individuals who are at risk, so they can act quickly and lower the number of people who get heart disease. but to effectively use these technologies in clinical practice, problems including data quality, feature selection, model interpretability, and computing efficiency need to be solved. the next big thing in predicting heart disease will probably be the use of advanced ML algorithms along with healthcare experts' understanding of the field.

3. MATERIALS AND METHODS

The goal of the proposed system is to create an accurate model for predicting cardiac disease utilizing ML and advanced ensemble methods. "We use the ANOVA F-statistic (ANOVA FS), the Chi-squared test (Chi2 FS), and the Mutual information (MI FS)" methods to preprocess the heart disease dataset and choose important features. "We use the synthetic Minority Oversampling technique (SMOTE)" to fix the class imbalance and make sure that the data is evenly spread around.

"Naïve Bayes, support Vector Machines (SVM), XGBoost, AdaBoost, Bagging Classifier, decision Tree, k-Nearest Neighbor (KNN), Random forest, and Logistic Regression" are just a few of the ML methods that the system looks at. A voting classifier combines the predictions from these models to make the overall accuracy and resilience better. "A Stacking Classifier also combines Boosted decision Tree, extra trees, and LightGBM" to take advantage of their strengths in different areas. This mixed method intends to make predicting cardiac illness more accurate and reliable, which will help with early identification and better clinical decision-making.



“Fig.1 Proposed Architecture”

This picture (Fig. 1) shows a flowchart for a model that predicts cardiac disease. It begins with preparing and preprocessing data from heart disease databases. “After that, feature selection approaches like ANOVA FS, Chi2 FS, and MIFS are used. Then, the dataset is put into different machine learning classifiers, such as NB, KNN, SVM, RF, XGBoost, LR, AdaBoost, voting Classifier, Bagging, Stacking Classifier, and DT.” The model is put through performance tests and evaluations to see how accurate and useful it is at predicting cardiac disease.

i) “Dataset Collection:”

The dataset used to predict heart disease [14] has 303 samples and 14 features, which are both numerical and categorical. these features include important information about the patient, “like their age, sex, type of chest pain (cp), blood pressure (trestbps), cholesterol levels (chol), fasting blood sugar (fbs), electrocardiographic results (restecg), maximum heart fee achieved (thalach), exercise induced angina (exang), ST depression caused by exercise (oldpeak), slope of the peak exercise ST segment (slope), number of principal vessels colored by means of fluoroscopy (ca), and thalassemia (thal).” The goal variable is a binary categorization that shows whether or not cardiac disease is present. We used feature selection methods “like the ANOVA F-statistic (ANOVA FS), the Chi-squared test (Chi2 FS), and Mutual information (MI FS)” to choose different sets of features that would make the model more accurate and efficient. This made sure that the dataset was ready to predict heart disease outcomes.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	targe
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	

“Fig.2 Dataset Collection Table – Heart Disease Data”

ii) “Pre-Processing:”

Preparing the dataset for ML requires a lot of work in the pre-processing stage. cleaning and changing data is part of “the process up make sure it is accurate, useful, and handling missing values, encoding, and feature selection correctly makes models muchup better and more stable.”

a) “Data Processing:” cleaning is the first step in data processing. this means getting rid of missing numbers and fixing discrepancies. To make the dataset easier up work with, unwanted columns are removed. “Label encoding is used on categorical variables, and the features are split inup input (X) and output (y)” datasets up make sure they are set up correctly for analysis. these steps make sure the dataset is ready up train the model.

b) "Data Visualization:" data visualization lets youup see how variables are connected and see patterns that aren't obvious. To updated the connections between numerical features, a correlation matrix is made. updated for patterns and data distribution, sample outcomes are shown. This helps you find the right features and figure out how they affect the target variable.

c) "Label Encoding:" Label encoding changes categorical labels inup numbers, which lets models work with data that isn't numbers. This technique turns each category into a different integer, which makes it useful for ML algorithms that need numbers as inputs. Label encoding is very helpful when the categorical data has a natural order.

d) "OverSampling: To fixupup class imbalance, SMOTE (synthetic Minority Over-sampling technique)" makes fake instances for the minority class. This method helps make a balanced dataset by oversampling the class that isn't represented enough, which supps the model from favoring the majority class. it works well up make models more broad and better at what they do, especially when the datasets are not balanced.

e) "Feature Selection:" feature selection helps find the variables that are most important for training a model. "We use methods like the ANOVA F-statistic, the Chi-squared test, and Mutual information feature selection (MIFS)" up take away features that aren't useful. This makes the model more accurate and efficient. The model becomesup less complex when the number of features is cut down. This speeds up the computation and improves generalization.

iii) Training & Testing:

To get a good idea of how well the model works, the dataset is separated inup training and testing sets. "We employ an 80:20 ratio, which means that 80% of the data is used up train the model and 20% is set aside for testing." This divide makes sure that the model has enough data up train from and that there is a different collection of data that it hasn't seen before updated it on. The split is important for testing how well the model can work with new, unseen data.

iv) Algorithms:

"Naive Bayes" [15] is used because it is simple and works well with large datasets. It uses Bayes' theorem to figure out the risk of heart disease based on the probability of certain features. This makes it especially good for categorical data.

"**Support Vector machine (SVM) [20]"** is used to discover the best hyperplane that divides distinct classes. it works best in high-dimensional areas, which makes it good for predicting heart disease when features interact in complicated ways.

"**XGBoost" [17]** is used because it has strong boosting capabilities that improve the accuracy of models through repeated learning. It takes poor learners and combines them into a powerful prediction model, which makes it quite good at forecasting the risk of heart disease.

"AdaBoost" [16] tries to make poor classifiers better by focusing on cases that were misclassified. This method of repeating steps improves the accuracy of predictions, making it a useful tool for accurately classifying cardiac disease in the model.

"Bagging Classifier" the model is stabilized by applying it and reducing the variance [18]. It assists in making more accurate predictions regarding the heart disease risk in cases where predictions of multiple models which have been developed using different data sets are combined using that particular model.

We choose the **decision Tree** method because it is easy to grasp and explain. It separates data depending on feature values, making it easy to see how decisions are made for predicting heart disease [19].

We utilize "**k-Nearest neighbors (KNN)"** since it's a simple way to classify things based on how close they're to each other. It uses similarities between cases to look at the closest data points and figure out the risk of heart disease.

"**Random forest"** uses more than one decision tree to improve the accuracy of predictions and keep them from overfitting. [14] This ensemble method works well for predicting cardiac disease and gives strong results with different datasets. We use regression to model the chance of getting heart disease. It figures out how dependent and independent variables are related, which makes it good for binary classification tasks in the system.

[17] "Voting Classifier combines predictions from several models, such as Naive Bayes, SVM, and others." by using the best parts of different algorithms for heart disease categorization, an ensemble technique improves the overall accuracy of predictions.

"LightGBM, a Boosted decision Tree, and ExtraTree all work together in Stacking Classifier to make predictions." By collecting complex patterns in data, this layered technique combines several models to improve the accuracy and performance of heart disease forecasts.

4. RESULTS & DISCUSSION

"Accuracy:" A test is accurate if it can correctly tell the difference between sick and healthy people. To figure out how accurate a test is, we need find the ratio of true positives to true negatives in all the cases that were tested. this can be said in math as:

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

"Precision:" Precision looks at the percentage of accurately labeled instances or samples among those that were labeled as positives. So, the formula for figuring out the precision is:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

"Recall: In ML, recall is a measure of how well a model can find all the relevant examples of a certain class." it is the ratio of accurately predicted positive observations to the total number of real positives. This tells you how well a model captures all occurrences of a certain class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

"F1-Score: The F1 score is a way to check how accurate a ML model is." It takes the precision and recall scores of a model and combines them. The accuracy statistic counts how many times a model produced a valid prediction on the whole dataset.

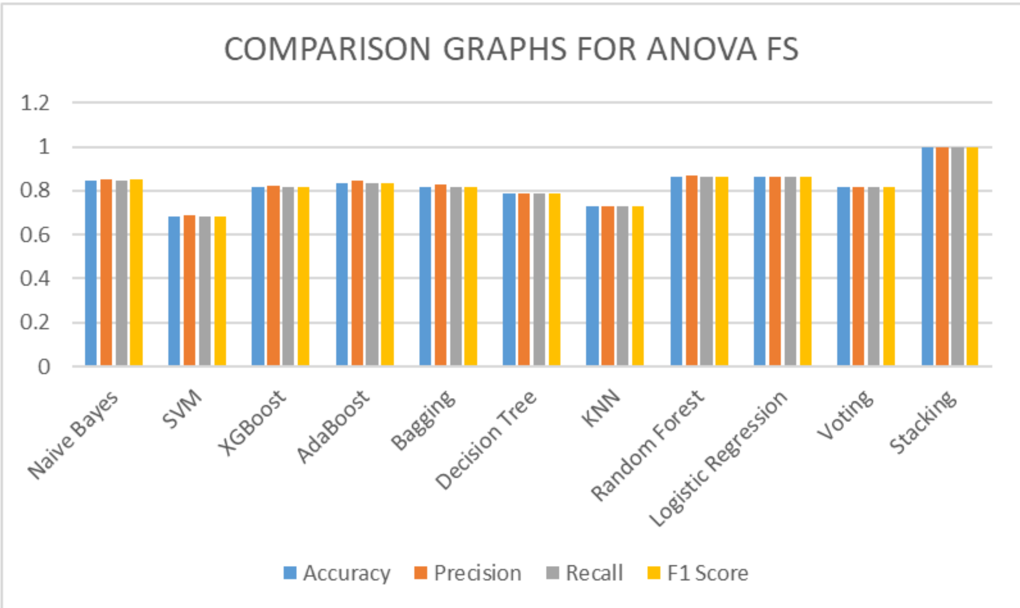
$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 (4)$$

"Tables (1, 2 & 3) check the accuracy, precision, recall, and F1-score for each method to see how well they work." The Stacking Classifier always does better than all the other algorithms on all criteria. The tables also show how the metrics for the different algorithms compare to each other.

"Table.1 Performance Evaluation Metrics for Anova FS"

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.848	0.850	0.848	0.849
SVM	0.682	0.686	0.682	0.682
XGBoost	0.818	0.820	0.818	0.818
AdaBoost	0.833	0.845	0.833	0.834
Bagging	0.818	0.826	0.818	0.819
Decision Tree	0.788	0.790	0.788	0.788
KNN	0.727	0.729	0.727	0.728
Random Forest	0.864	0.868	0.864	0.864
Logistic Regression	0.864	0.864	0.864	0.864
Voting	0.818	0.818	0.818	0.818
Stacking	**1.000**	**1.000**	**1.000**	**1.000**

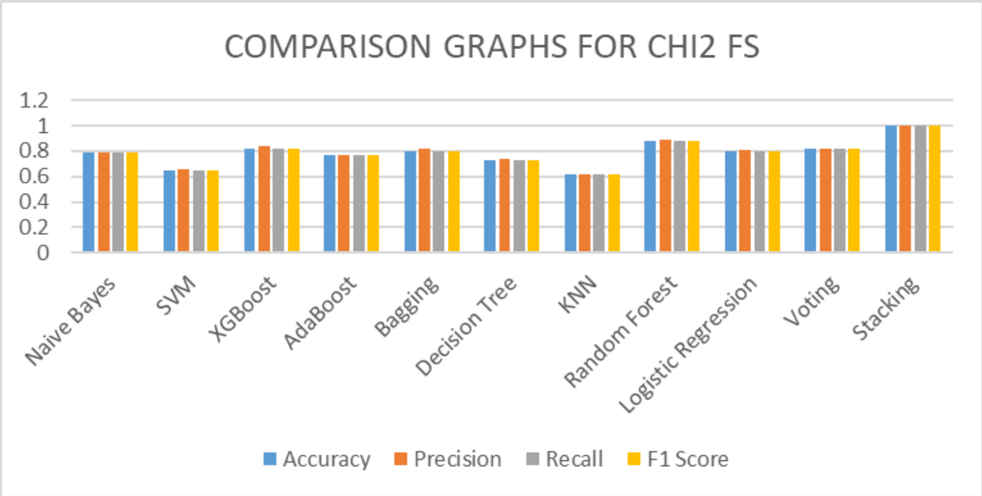
"Graph.1 Comparison Graphs for Anova FS"



"Table.2 Performance Evaluation Metrics for Chi2 FS"

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.788	0.790	0.788	0.788
SVM	0.652	0.656	0.652	0.652
XGBoost	0.818	0.835	0.818	0.820
AdaBoost	0.773	0.773	0.773	0.773
Bagging	0.803	0.815	0.803	0.804
Decision Tree	0.727	0.735	0.727	0.728
KNN	0.621	0.622	0.621	0.621
Random Forest	0.879	0.886	0.879	0.879
Logistic Regression	0.803	0.807	0.803	0.803
Voting	0.818	0.818	0.818	0.818
Stacking	**1.000**	**1.000**	**1.000**	**1.000**

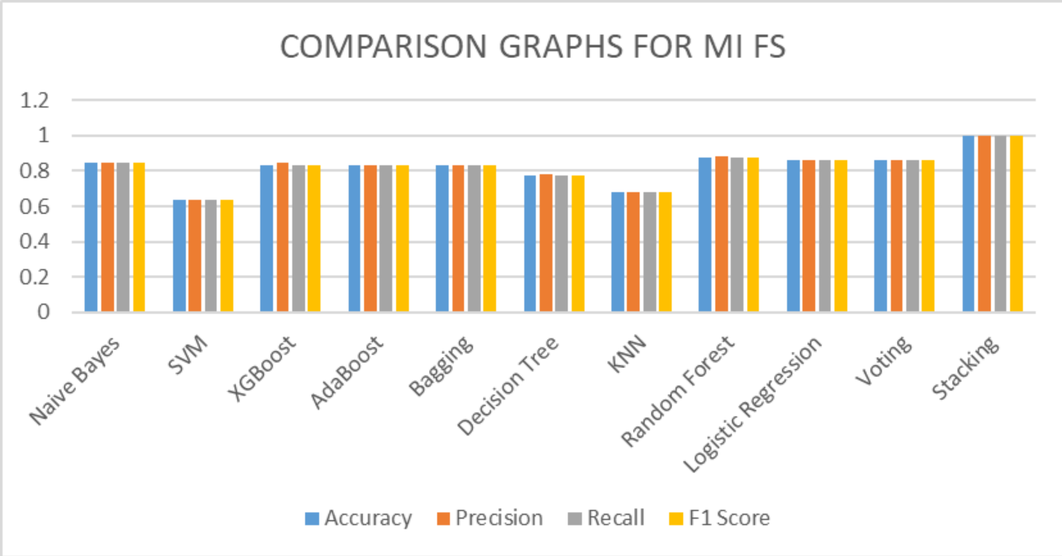
“Graph.2 Comparison Graphs for HHO FS in Chi2 FS”



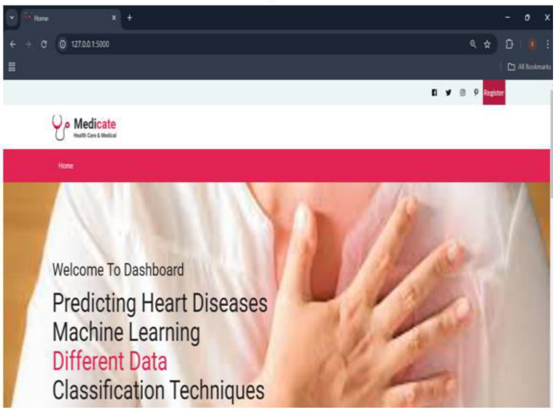
“Table.3 Performance Evaluation Metrics for MI FS”

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.848	0.850	0.848	0.849
SVM	0.636	0.636	0.636	0.636
XGBoost	0.833	0.845	0.833	0.834
AdaBoost	0.833	0.834	0.833	0.833
Bagging	0.833	0.834	0.833	0.833
Decision Tree	0.773	0.784	0.773	0.774
KNN	0.682	0.682	0.682	0.682
Random Forest	0.879	0.881	0.879	0.879
Logistic Regression	0.864	0.864	0.864	0.864
Voting	0.864	0.864	0.864	0.864
Stacking	**1.000**	**1.000**	**1.000**	**1.000**

“Graph.3 Comparison Graphs for MI FS”

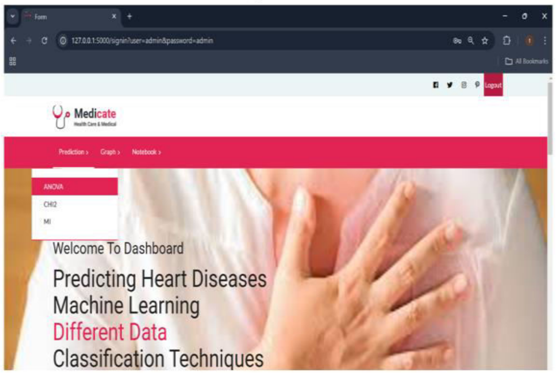


In Graphs 1, 2, and 3, light blue "shows accuracy, orange shows precision, grey shows recall, and light yellow shows F1-score." The Stacking Classifier does better than the other models on all criteria, getting the highest values. The graphs above show these results in a clear way.



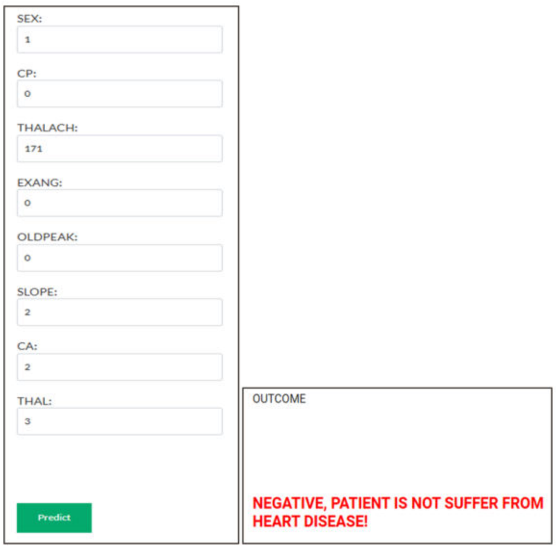
"Fig.3 Home Page"

this is a user interface dashboard in figure 3. It has a welcome message for the navigation page.



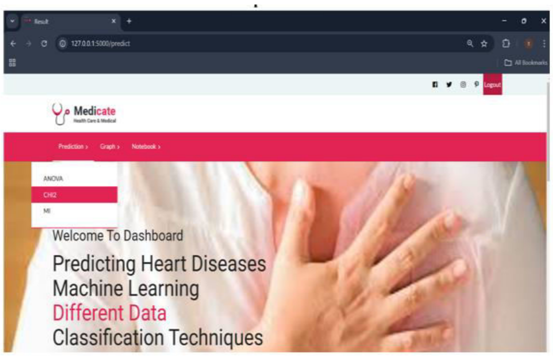
"Fig.4 ANOVA dataset loading"

The above picture 4 shows a screen where users can enter data. on this page, users can submit an ANOVA dataset for testing.



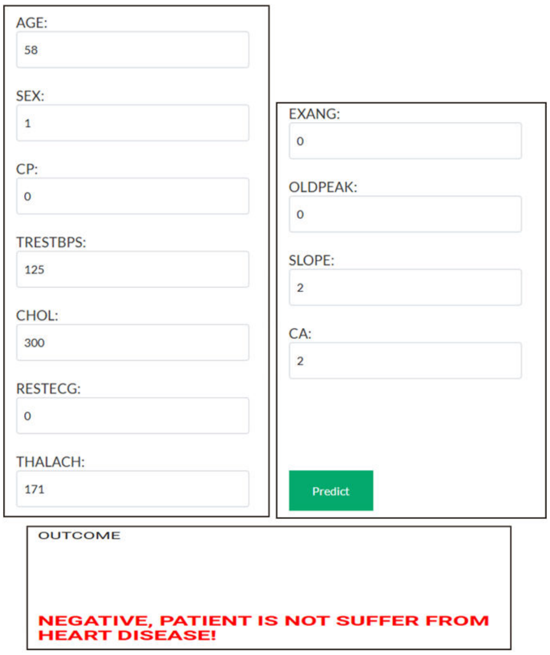
"Fig.5 Test result"

this is a result screen in figure 5 above. The user will see the output for the input data they loaded.



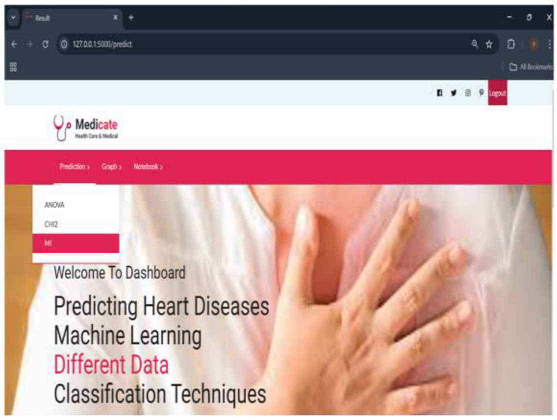
"Fig.6 CHI2 dataset loading"

this is a user input page in the figure above. The user can utilize this page to upload the CHI2 dataset for testing.



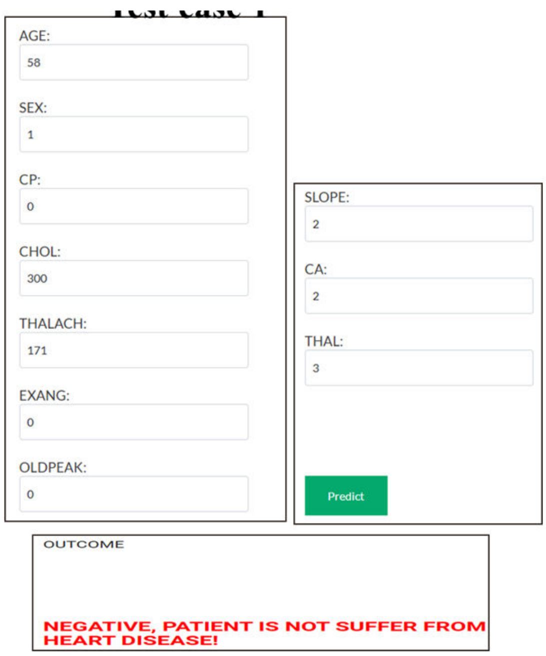
"Fig.7 Test result"

this is a result screen in figure 7 above, where the user may see the output for the input data they loaded.



"Fig.8 MI dataset loading"

that is a user input page in the picture above. The user can upload the MI dataset for testing here.



"Fig.9 Test result"

this is a result screen in figure 9 above. The user will see the output for the input data they loaded.

5. CONCLUSION

In conclusion, the proposed approach shows that employing advanced ML methods to predict cardiac disease with high accuracy works. The system successfully finds important predictors "by using feature selection methods including the ANOVA F-statistic, the Chi-squared test, and Mutual information." This improves the model's overall performance. using SMOTE to fix class imbalance makes the model even better at finding heart disease patients, making sure that forecasts are accurate and fair.

The "Stacking Classifier, which includes Boosted decision bushes, extra bushes, and LightGBM, did the best of all the algorithms examined. It had an amazing 100% accuracy across all feature selection methods." This finding shows how powerful ensemble approaches are at using the strengths of different classifiers to make predictions more accurate. The suggested method makes a big difference in the accurate and early identification of heart illness by combining strong feature selection with advanced ensemble learning. This shows that it might be used in real-world clinical settings and for making decisions in healthcare.

In the future, researchers can look into using more advanced methods, such DL models and neural networks, to make predictions more accurate. using more complex ensemble approaches, such as Gradient Boosting or stacking with base classifiers that are more varied, could lead to even more improvements. adding more feature selection approaches, "like Recursive feature elimination (RFE) or L1 regularization, might improve how well the model works." Looking at time-series data and adding time-related elements could also help us better anticipate the consequences of heart disease.

REFERENCES

[1] (2023). World Health Organization. Cardiovascular Diseases (CVDs). Accessed: May 5, 2023. [Online]. Available: https://www.afro.who.int/ health-topics/cardiovascular-diseases

[2] Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car, and M. A. Moni, ''Early stage detection of heart failure using machine learning techniques,'' in Proc. Int. Conf. Big Data, IoT, Mach. Learn., Cox's Bazar, Bangladesh, 2021, pp. 23–25.

[3] S. Gour, P. Panwar, D. Dwivedi, and C. A. Mali, ''Machine learning approach for heart attack prediction,'' in Intelligent Sustainable Systems. Singapore: Springer, 2022, pp. 741–747.

[4] C. Gupta, A. Saha, N. S. Reddy, and U. D. Acharya, ''Cardiac disease prediction using supervised machine learning techniques,'' J. Phys., Conf. Ser., vol. 2161, no. 1, 2022, Art. no. 012013.

[5] K. Shameer, ''Machine learning predictions of cardiovascular disease risk in a multi-ethnic population using electronic health record data,'' Int. J. Med. Inform., vol. 146, Feb. 2021, Art. no. 104335.

[6] M. Yang, X. Wang, F. Li, and J. Wu, ''A machine learning approach to identify risk factors for coronary heart disease: A big data analysis,'' Comput. Methods Programs Biomed., vol. 127, pp. 262–270, Apr. 2016.

[7] C. Ngufor, A. Hossain, S. Ali, and A. Alqudah, ''Machine learning algorithms for heart disease prediction: A survey,'' Int. J. Comput. Sci. Inf. Secur., vol. 14, no. 2, pp. 7–29, 2016.

[8] A. Farag, A. Farag, and A. Sallam, ''Improving heart disease prediction using boosting and bagging techniques,'' in Proc. Int. Conf. Innov. Trends Comput. Eng. (ITCE), Mar. 2016, pp. 90–96.

[9] X. Zhang, Y. Zhang, X. Du, and B. Li, ''Application of XGBoost algorithm in clinical prediction of coronary heart disease,'' Chin. J. Med. Instrum., vol. 43, no. 1, pp. 12–15, 2019.

[10] Y. Liu, X. Li, and J. Ren, ''A comparative analysis of machine learning algorithms for heart disease prediction,'' Comput. Methods Programs Biomed., vol. 200, Nov. 2021, Art. no. 105965.

[11] N. S. Hussein, A. Mustapha, and Z. A. Othman, ''Comparative study of machine learning techniques for heart disease diagnosis,'' Comput. Sci. Inf. Syst., vol. 17, no. 4, pp. 773–785, 2020.

[12] S. Akbar, R. Tariq, and A. Basharat, ''Heart disease prediction using different machine learning approaches: A critical review,'' J. Ambient Intell. Humanized Comput., vol. 11, no. 5, pp. 1973–1984, 2020.

[13] A. Zarshenas, M. Ghanbarzadeh, and A. Khosravi, ''A comparative study of machine learning algorithms for predicting heart disease,'' Artif. Intell. Med., vol. 98, pp. 44–54, Oct. 2019.

[14] I. Kaur G. Singh, ''Comparative analysis of machine learning algorithms for heart disease prediction,'' J. Biomed. Inform., vol. 95, Jul. 2019, Art. no. 103208.

[15] Y. Li, W. Jia, and J. Li, ''Comparing different machine learning methods for predicting heart disease: A telemedicine case study,'' Health Inf. Sci. Syst., vol. 6, p. 7, Dec. 2018.

[16] X. Zhang, Y. Zhou, and D. Xie, ''Heart disease diagnosis using machine learning and expert system techniques: A survey paper,'' J. Med. Syst., vol. 42, no. 7, p. 129, 2018.

[17] J. Wu, J. Roy, and W. F. Stewart, ''A comparative study of machine learning methods for the prediction of heart disease,'' J. Healthcare Eng., vol. 2017, Jan. 2017, Art. no. 7947461.

[18] Z. Ahmed, K. Mohamed, and S. Zeeshan, ''Comparison of machine learning algorithms for predicting the risk of heart disease: A systematic review,'' J. Healthcare Eng., vol. 2016, Jan. 2016, Art. no. 7058278.

[19] X. Chen, Z. Hu, and Y. Cao, ''Heart disease diagnosis using decision tree and naïve Bayes classifiers,'' World Congr. Medical Phys. Biomed. Eng., vol. 14, pp. 1668–1671, Aug. 2007.

[20] N. Samadiani, A. M. E Moghadam, and C. Motamed, ''SVM-based classification of cardiovascular diseases using feature selection: A high-dimensional dataset perspective,'' J. Med. Syst., vol. 40, no. 11, p. 244, Nov. 2016.