Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 21s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques

Arun Babu Allamudi[1], Dr. S. HrushiKesava Raju[2]

[1, 2] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India

Email ID: arunbabuallamudi7@gmail.com

Email ID, hkesavaraju@kluniversity.in

Cite this paper as: Arun Babu Allamudi, Dr. S. HrushiKesava Raju, (2025). Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques. *Journal of Neonatal Surgery*, 14 (21s), 695-706.

ABSTRACT

The main cause of worldwide dying may be prevented in the first -class element in the first identity of its symptoms and symptoms. Due to complicated clinical records and difficulty offering continuous monitoring, the exact prognosis of heart disease is still hard. Significant predictors have been discovered by many strategies of features that include Anova F-statistical (Anova FS), Chi-Squared Test (Chi2 FS) and Mutual Information (Mi FS) Using Data Set of Heart Diseases. The data imbalance has become solved and the overall performance of the version was improved using the Synthetic Minority (SMOTE) technique. Several system studies of fashion and file strategies have been used in the radical classification strategy. Among them, a stacking classifier involving reinforced decision -making trees, extra trees and LightGBM, which led to huge results that achieved 100% accuracy in all approaches to selecting functions. Excessive performance, emphasizing the opportunity to mix robust selections with state -of -the -art category modes for accurate medical facts, illustrates the effectiveness of the superior set that has been recognized in generating stable forecasts of heart disease. This strategy shows how it is able to support early prognosis and results with a higher affected person.

Keyword: - *Cardiovascular disease, heart disease, machine learning app, ML algorithms, SDG 3, SHAP, SMOTE..*

1. 1. INTRODUCTION

Cardiovascular disease (CVD) is highly among the main causes of deaths around the world; The main factor here is heart disease. The muscle organ that draws blood throughout the body is necessary for the circulatory system. Tesky, veins and capillaries - which supply oxygen and nutrients to organs and tissues - this complex system increases. When disorders arise in normal blood flow, they lead to several types of heart disease, all known as cardiovascular disease (CVD). The World Health Organization (WHO) estimates that stroke and heart disease represent approximately 17.5 million deaths annually; Most of these deaths in countries with low and medium incomes. With heart attacks and strikes, which represent 80% of all CVDs related to CVD, this emphasizes that the world is a heart disease in the field of public health

[1].The weight of cardiovascular diseases caused early identification, prevention and treatment plans to become worldwide. Fastening to the target of sustainable development of 3 UN, which emphasizes the need for health and well -being, solving cardiovascular diseases, has increased to increase world health health results. The combination of heart disease in family lineage alongside smoking coupled with aging and elevated cholesterol levels and inactivity and high blood pressure and obesity and diabetes and stress constitutes frequent factors that increase heart

disease risks. Research shows that this intervention effectively minimizes heart disease threats amongst patients. The main heart disease risks for patients include weight changes such as weightFamily history with heart disease, smoking, age, high cholesterol, physical inactivity, high blood pressure, obesity, diabetes and stress. Medical research shows that weight modifications including weight decrease will decrease heart disease susceptibility. control, stress reduction, regular exercise and smoking cessation

[2]. In addition to lifestyle changes, heart disease usually occurs using diagnostic tools including electrocardiograms (ECG), echocardiograms, heart MRI and blood tests. Sometimes treatment requires medical procedures including angioplasty, surgery of the bypass of coronary arteries and implanted devices such as pacemakers and defibrillators

[3].Thanks to the development of medical technologies - especially with regard to large data and electronic health records (EHRS), it is now possible to use huge volumes of patients for predictive modeling. Analysis of a huge amount of data from medical systems using machine learning methods (ML) is becoming increasingly important that it can be inferred by bright information to predict the likelihood of heart disease. Machine learning can help medical physicians to identify high -risk

patients and enable timely intervention by processing and analyzing data from many demographic data on patients, r factors and diagnostic results. With a more accurate and more effective diagnosis, prediction and treatment strategy adapted to this approach changes the health care scene [4]. [5] [5]

2. 2. RELATED WORK

With a large proportion of world mortality, "cardiovascular disease (CVD)", including coronary heart diseases, remains a global health problem. Efforts for more accurate anticipation and diagnosis of heart disease have increased sharply in the tandem with the growth of health care data and the methods of machine learning (ML). The use of machine learning to evaluate huge quantities opens new ways for early detection, prediction of results and identification of the risk factor. This overview of literature deals with several research using various machine learning models and approaches to predict heart disease and emphasizes their results.

Yang et al. [6] They were looking for variable risks for coronary heart disease by machine learning. Their work focused on a major analysis of data, which shows that the use of machine learning models of patients, including decision -making trees, random forests and supportive vector machines (SVM), could effectively detect important risk factors. Among the most important risk factors affecting the prognosis of heart disease, the study has proven to be high cholesterol, age and family history. In addition, the post emphasized the need for selection of functions, data preparation and a strong data file to increase the performance of the model. The finding has verified that, in conjunction with large, thorough data sets, machine learning could be a very effective technique for early identification of heart disease.

In the same line of Ngufor et al. [7] They examined numerous machine learning techniques to predict heart disease. Popular methods such as SVM, decision-making trees, the "K nearest neighbors (KNN) and artificial neural networks (Anns)" were compared by authors. Their results have shown that they offered improved predictive performance within individual file technology models, including bags and increasing. The study also emphasized the need for elements, as unnecessary elements can reduce the accuracy of the model. This overview stressed that although several algorithms could predict heart disease, the selection of technology mostly relied on the properties of the data file, available computing resources and specific requirements of the prediction task.

Farag et al. [8] They focused on the methods of improving the prediction of heart disease. To evaluate their ability to improve the accuracy of prediction, approaches to strengthen approaches such as random forest and adaboost, as well as bagging methods in terms of reduction in scattering and increase the stability of prediction, the study revealed that file methods were more reliable than individual classes. In addition, the study indicated that the combination of bags and strengthening could help reduce the problem with the excess that is sometimes found in models of heart disease. This work emphasized the need to combine several classifiers to achieve the best performance.

Using XGBOOST, Methods of Gradient, Zhang et al. [9] They looked at the best way to clinically predict coronary heart disease. Regarding accuracy and interpretability, XGBOOST has defeated conventional techniques, including logistics regression and SVM according to their research. Medical data were particularly suitable for XGBOOST because of their ability to drive unbalanced data files - a significant problem in the prediction of heart disease: the study also noted that the maximization of the model's performance critically depends on the adaptation of the hyperparameter. XGBOOST, scalability, ability to generate consistent prediction, while reducing the risk of helping to explain its excellent success in the clinical environment.

For the purposes of prediction of heart disease Liu et al. [10] They conducted a comparative study of numerous machine learning techniques including random forests, SVM and decision -making trees. Between the examined methods, the best accuracy of prediction according to their analysis has provided the best accuracy of prediction. Nevertheless, they also pointed out that random forests and decision -making trees provided excellent interpretability - which is absolutely essential in the medical environment. The study concluded that although SVM showed the best accuracy, the choice of the method should be based on a compromise between accuracy and interpretability, especially if the model should use healthcare professionals for decision -making.

Hussein et al. [11] They evaluated artificial neural networks, decision -making trees and KNN, among other things, machine learning methods for diagnosis of heart disease. The study tried to evaluate the diagnostic performance of models from the patient's health. Their finding showed that with reduced computing costs, KNN and decision -making trees, they worked well and therefore suitable for real -time applications in clinical environments. Artificial neural networks were difficult to understand, more computational resources were needed, although they offered better accuracy. The study emphasized that the acceptance of machine learning models in health systems must consider both performance and financial restrictions.

Akbar et al. [12] They reviewed several machine learning techniques critically for the prediction of heart disease. They assessed methods including decision -making trees, KNN, SVM, random forests and neural networks, then offered a summary of their strengths and shortcomings in the prediction of heart disease. Because they can reduce excessive and manage noisy data, the writers have found that the technicians of the file - especially random forest - have acquired the best accuracy. But because the selection of features is so important for the success of the model, the evaluation also emphasized the difficulty of choosing the most suitable elements from large data sets. The study also emphasized the need to manage missing data and guarantee a balanced and representative data set used for training.

Zarshenas et al. [13] They made a comparison of machine learning algorithms to predict heart disease. Given that SVM provides greater accuracy in some situations, their data revealed that random forest and SVM have achieved the best prediction performance. They also emphasized how fundamental preliminary actions processing, including scaling of functions and normalization, are to increase the power of the machine learning model. According to a study of hybrid models - which include the best features of several algorithms - they can show a great promise for other studies focused on prediction of heart disease.

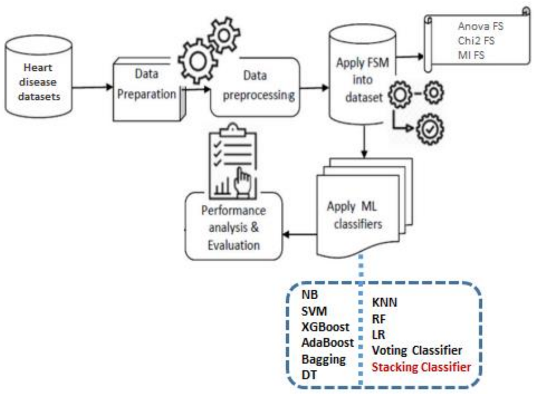
The need to select elements and data preparation in improving the performance of models of cardiac disease prediction is carried out in most of the studies under investigation. While some research suggests that hybrid models include many machine learning algorithms to bring excellent results, several emphasizes the use of file approaches such as bagging and strengthening to increase the accuracy of the model. In addition, although conventional models such as logistics regression and decision -making trees are still widely used, newer studies have shown that, especially in complicated and high -dimensional data, sophisticated algorithms such as XGBOOST and neuron networks can overcome these approaches.

Increasing availability of healthcare data and the development of machine learning methods has created new opportunities for early diagnosis and prevention of heart disease. We help medical specialists to find patients at risk, machine learning models can allow rapid interventions and therefore reduce the burden of heart disease. However, in order to fully meet the promise of these technologies in clinical practice, problems, including data quality, function, model, model and calculation efficiency, must be solved. Advanced machine learning models associated with information specific to domain from medical experts are likely to drive the next generation of developments in the prediction of heart disease.

3. 3. MATERIALS AND METHODS

The aim of the proposed system is to provide accurate machine learning and an advanced model of heart -based prediction. They help to identify remarkable features of pre-processing data set of "ANOVA F-statistic (ANOVA FS), Chi-squared test (Chi2 FS), and Mutual Information (MI FS). The technique of Synthetic Minority Oversampling Technique (SMOTE)" is used to ensure a balanced data distribution, which deals with the class imbalance.

Among several machine learning techniques are systemic analyzes Naive Bayes, "Support Vector Machines (SVM)", XGBOOST, Adaboost, Bag Classifier, Decision Tree, K-Nearest Neighbor, Random Forest and Logistic Regression. Combination of predictions from these models, the file voting classifier increases general accuracy and robustness. To maximize their free strengths, the stacking classifier also combines the strengthened tree of decision -making, extra trees and LightGBM. This hybrid strategy seeks to improve the accuracy and reliability of heart disease, promoting early identification and improved clinical decision -making.



"Fig.1 Proposed Architecture"

This graphic (Fig. 1) illustrates a flowchart for a heart ailment prediction model. The method commences with records training and preprocessing of heart disease datasets. Finally, feature selection techniques "(ANOVA FS, Chi-squared FS, MIFS) are carried out". The dataset is ultimately input into more than one machine learning classifiers (Naive Bayes, k-Nearest acquaintances, aid Vector device, Random forest, XGBoost, Logistic Regression, AdaBoost, vote casting Classifier, Bagging, Stacking Classifier, decision Tree). The model is subjected to performance analysis and evaluation to determine its accuracy and efficacy in predicting cardiac disease.

i) Dataset Collection:

The dataset utilized for heart disease prediction comprises 303 samples with 14 features, encompassing each numerical and categorical elements. these attributes encompass vital patient facts, which include age, intercourse, form of "chest pain (cp), resting blood pressure (trestbps), cholesterol levels (chol), fasting blood sugar (fbs), electrocardiographic findings (restecg), maximum heart rate attained (thalach), exercise-induced angina (exang), ST section despair prompted with the aid of exercise (oldpeak), slope of the peak exercise ST segment (slope), number of important vessels visualized by fluoroscopy (ca), and thalassemia (thal)". The intention variable is a binary classification denoting the existence or non-lifestyles of cardiac disease. Following the implementation of function choice methodologies, including ANOVA F-statistic (ANOVA FS), Chi-squared test (Chi2 FS), and Mutual data (MI FS), numerous feature subsets were identified to decorate model accuracy and efficiency, thereby optimizing the dataset for predicting 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:

Pre-processing is an essential phase inside the practise of the dataset for machine learning. It includes the cleaning and transformation of records to assure precision, efficacy, and pertinence. Effective control of absent values, encoding, and characteristic choice markedly improves model efficacy and resilience.

a) Data Processing: The records processing commences with cleansing, which involves eliminating missing values and rectifying discrepancies. Unnecessary columns are removed to optimize the dataset. Label encoding is utilized for categorical variables, with features divided into input (X) and output (y) datasets to make sure suitable structuring for analysis. Those methods guarantee the dataset is prepared for model training.

b) Data Visualization: data visualization enables the comprehension of relationships amongst variables and famous hid patterns. A correlation matrix is built to look at the correlations among numerical features, whilst sample outcomes are shown to assess data distribution and developments. This facilitates the identity of pertinent trends and elucidates their influence on the target variable.

c) Label Encoding: Label encoding converts categorical labels into number values, facilitating the processing of non-numeric input by using fashions. This process transforms each class into a distinct integer, rendering it appropriate for machine learning methods that need numerical inputs. Label encoding is mainly advantageous when the categorical records possesses a natural hierarchy.

d) OverSampling: SMOTE (synthetic Minority Over-sampling technique) is hired to rectify magnificence imbalance by means of producing synthetic instances for the minority magnificence. This method enables the introduction of a balanced dataset by oversampling the underrepresented class, for this reason preventing the model from displaying bias closer to the majority magnificence. This approach effectively enhances version generalization and performance, particularly in imbalanced datasets.

e) Feature Selection: function selection aids in identifying the maximum pertinent variables for model training. Strategies like as the ANOVA F-statistic, Chi-squared test, and "Mutual information feature selection (MIFS)" are utilized to take away extraneous characteristics, hence enhancing the model's efficiency and accuracy. Lowering the amount of capabilities decreases model complexity, ensuing in expedited computing and more desirable generalization..

iii) Training & Testing:

The dataset is divided into schooling and checking out subsets to evaluate the model's performance accurately. An eighty: 20 ratio is employed, with 80% of the facts detailed for model training and the last 20% allotted for testing. This division guarantees that the model possesses enough records for training while preserving a distinct set of unobserved records for validation purposes. The division is crucial for evaluating the model's potential to generalize and characteristic on novel, unobserved records.

iv) Algorithms:

Naive Bayes [15] is utilized for its straightforwardness and efficacy in managing huge datasets. It utilizes Bayes' theorem to evaluate heart ailment risk by studying characteristic chances, rendering it particularly useful for categorical data.

Support Vector Machine [20] (SVM) is employed to identify the best hyperplane that differentiates distinct lessons. It has superiority in excessive-dimensional spaces, making it appropriate for complex characteristic interactions in coronary heart disease prediction.

XGBoost [17] is hired for its strong boosting capabilities, augmenting model correctness through iterative learning. It amalgamates poor newbies into a strong prediction version, rendering it notably useful for forecasting heart disease risk.

AdaBoost [16] concentrates on enhancing deficient classifiers by prioritizing misclassified occurrences. This iterative technique enhances predictive accuracy, rendering it a tremendous strategy for reliable cardiac ailment classification inside the model.

Bagging Classifier is hired to decrease variation and improve model stability [18]. Integrating forecasts from various models trained on distinct data subsets enhances predictions of heart disease risk.

Decision Tree The algorithm is utilized for its clarity and comprehensibility. It segregates statistics according to function values, yielding explicit insights into decision-making techniques for heart disease prediction [19].

K-Nearest Neighbors (KNN) is employed for its direct methodology in classification predicated on proximity. It evaluates the nearest records points to categorize heart disease risk, utilizing the resemblance among occurrences.

Random Forest integrates many decision trees to improve predictive accuracy and mitigate overfitting. [14] This ensemble technique is efficacious for predicting cardiac disease, yielding dependable consequences throughout diverse datasets.

[16] Logistic Regression is hired to simulate the chance of heart disease manifestation. It assesses the relationships among structured and independent variables, rendering it appropriate for binary class obligations in the system.

[17] Voting Classifier consolidates predictions from various models, consisting of Naive Bayes, support Vector Machines, and others. This ensemble method improves overall predictive accuracy by means of utilizing the strengths of multiple algorithms for heart disease classification.

Stacking Classifier integrates forecasts from a Boosted decision Tree with ExtraTree utilizing LightGBM. This stratified methodology includes various models, enhancing performance and precision in heart disease predictions by way of identifying intricate patterns within the data.

4. 4. RESULTS & DISCUSSION

Accuracy: Test accuracy represents the correct identification between sick patients and healthy ones. To evaluate how accurate an appearance performs one must determine true positive rates and proper negative outcomes for all examined cases. It can be expressed mathematically:

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

Precision: The accuracy shows the proper connection between correctly identified examples which belong to people classified as positive. The accuracy expresses itself through a calculation formula for precision.

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

Recall: All machine learning calculations take into account the capability of models to understand relevant times for a specific elegance. The analysis of precise high quality commentary relative to complete real positivity helps evaluate how well a model detects events in a selected class..

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

F1-Score: The F1 point sum stands as a calculation method to evaluate the accuracy of machine learning models. This measure combines accuracy results while ignoring the complete pattern of the model. Through the accuracy meter analysts can determine which percentage of real predictions a model creates across an entire data set.

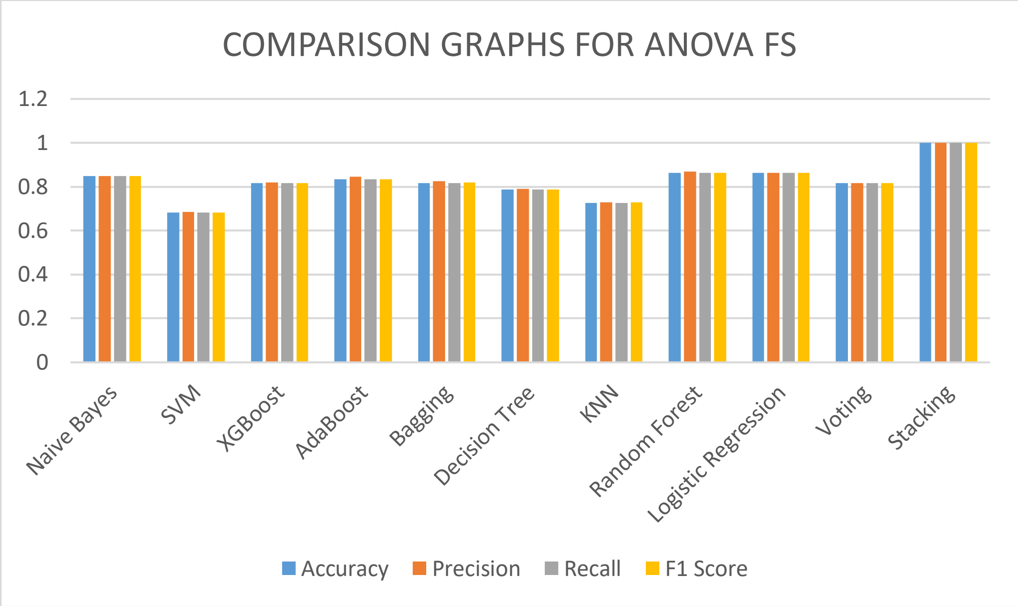
$$F1\,Score = 2 * \frac{Recall\,\mathrm{X}\,Precision}{Recall + Precision} * 100(1)$$

Tables (1, 2 & 3) Measure accuracy alongside accurate results and recall and F1 score for each evaluation method. The Stacking classes execute automatic cross-examinations of every algorithm for all operation stages. All different algorithms in all calculations automatically cross each other through the Stacking classes. The comparison data about alternative strategies can be found in tabular form.

"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**

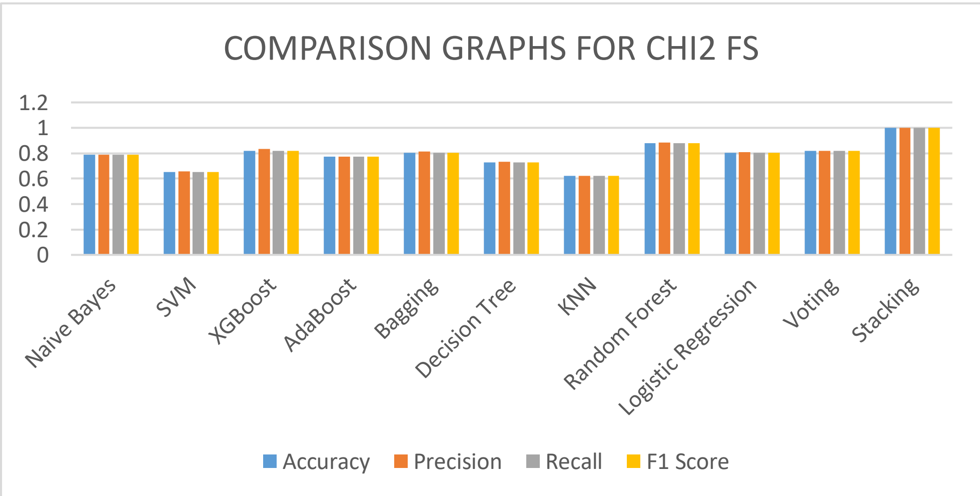
"Graph.1 Comparison Graphs for Anova FS"



"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**

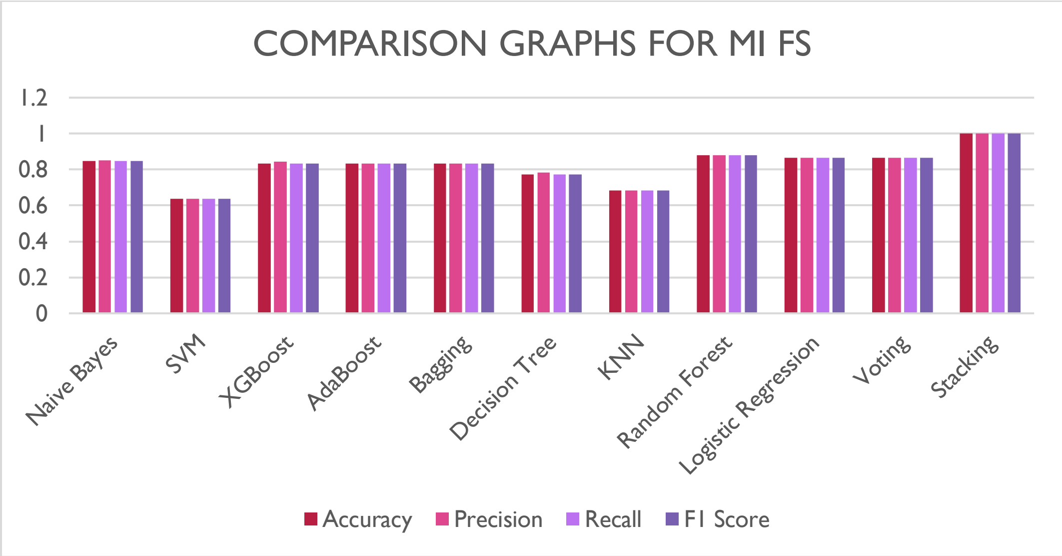
"Graph.2 Comparison Graphs for HHO FS in Chi2 FS'



"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**

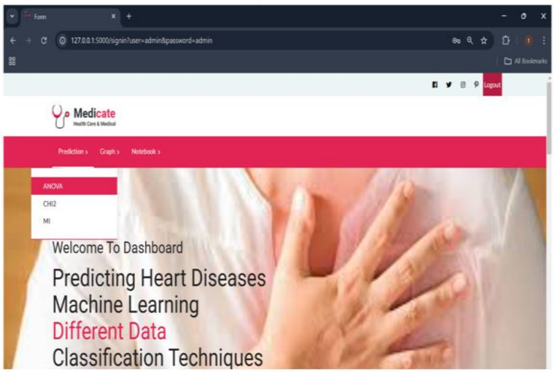
'Graph.3 Comparison Graphs for MI FS"



Accuracy is depicted in light blue, precision in orange, take into account in gray, and F1-score in light yellow, as shown in Graphs 1, 2, and 3. The Stacking Classifier demonstrates superior overall performance relative to the other models, reaching the highest values across all metrics. The aforementioned graphs visually represent these results.

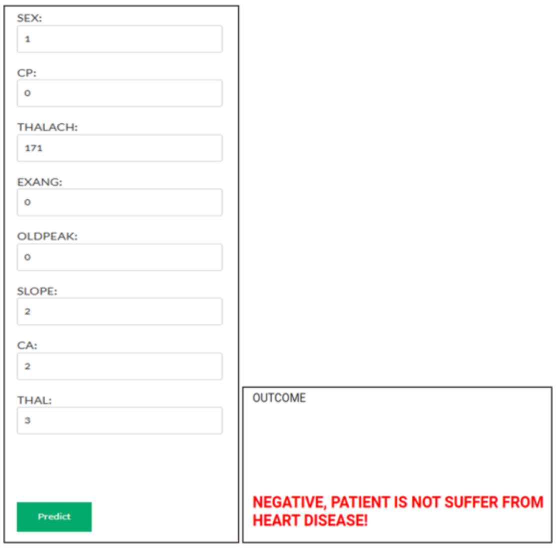
"Fig.3 Home Page"

discern 3 depicts a user interface dashboard featuring a welcome message for page navigation.



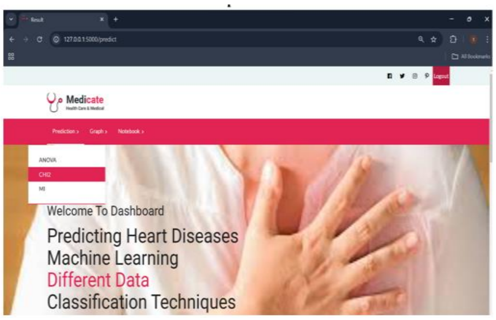
“Fig.4 ANOVA dataset loading”

figure 4 illustrates a user input page that enables the submission of an ANOVA dataset for testing purposes..



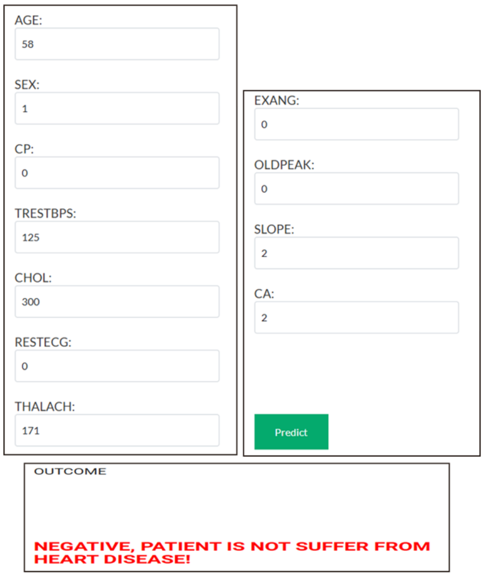
“Fig.5 Test result”

figure 5 illustrates a outcomes display, in which the user gets output for the loaded input data..



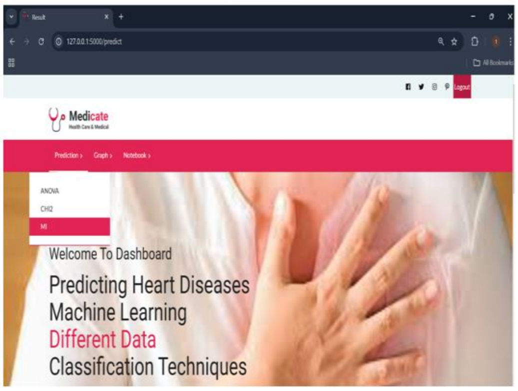
“Fig.6 CHI2 dataset loading”

figure 6 depicts a user input web page that permits the add of the CHI2 dataset for testing purposes.



"Fig.7 Test result"

figure 7 illustrates a effects screen, where the user receives output for the loaded enter data.



"Fig.8 MI dataset loading"

figure 8 depicts a consumer input page that allows the user to submit the MI dataset for testing functions.

"Fig.9 Test result"

figure 9 illustrates a result screen, where the consumer receives output for the loaded input data.

5. 5. CONCLUSION

In end, the proposed system illustrates the efficacy of employing advanced machine learning methodologies for appropriately predicting heart disease. employing feature selection techniques inclusive of the ANOVA F-statistic, Chi-squared test, and Mutual facts, the system efficiently discerns critical predictors, therefore improving the version's common overall performance. The utilization of SMOTE to rectify elegance imbalance complements the model's capacity to perceive heart disease times, hence making sure balanced and dependable predictions.

a number of the algorithms evaluated, the Stacking Classifier, which integrates Boosted decision trees, extra timber, and LightGBM, attained the best performance, achieving an exceptional 100% accuracy across all feature selection strategies. This final results highlights the efficacy of ensemble approaches in amalgamating the strengths of person classifiers to enhance predicted accuracy. The cautioned method enhances the proper and well timed identification of cardiac infection through the integration of advanced function selection and complex ensemble mastering, showcasing its applicability in clinical settings and healthcare decision-making.

In future work, Supplementary methodologies, such as deep learning models and neural networks, may be investigated to enhance predictive accuracy. The implementation of sophisticated ensemble techniques, which include Gradient Boosting or stacking with a broader array of simple classifiers, may additionally yield extra improvements. Integrating supplementary feature selection techniques, such as Recursive feature elimination (RFE) or L1 regularization, may want to enhance the model's efficacy. Investigating time-series data and integrating temporal variables may yield greater thorough insights for forecasting heart disease outcomes.

REFERENCES

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

[2] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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] [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.