Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 2: 2025
ISSN : **1906-9685**

Journal of Nonlinear Analysis and Optimization : Theory & Applications

Editors-in-Chief :

PREDICTING HEART DISEASE RISK WITH MACHINE LEARNING AND DATA CLASSIFICATION TECHNIQUES USING BOOSTED DECISION TREES, EXTRA TREES, AND LIGHTGBM.

MR.Sk.HIMAM BASHA[1],BILLA.MOUNIKA[2]
#1 Assistant Professor Department of Master of Computer Applications
#2 Pursuing M.C.A QIS COLLEGE OF ENGINEERING & TECHNOLOGY
Vengamukkapalem(V) ,Ongole, Prakasam dist , Andhra Pradesh- 523272

Abstract: Heart disease is a leading cause of mortality worldwide, with early detection playing a critical role in reducing death rates. Accurate prediction of heart disease remains challenging due to complex medical data and the inability to provide continuous monitoring. Utilizing the Heart Disease dataset, various feature selection techniques, including ANOVA F-statistic (ANOVA FS), Chi-squared test (Chi2 FS), and Mutual Information (MI FS), were employed to identify significant predictors. Synthetic Minority Oversampling Technique (SMOTE) was applied to address data imbalance and enhance model performance. A comprehensive classification approach was undertaken using diverse machine learning models and ensemble methods. Among these, a Stacking Classifier combining Boosted Decision Trees, Extra Trees, and LightGBM achieved superior results, delivering 100% accuracy across all feature selection techniques. The high performance highlights the effectiveness of advanced ensemble learning in achieving reliable heart disease predictions, emphasizing the potential of integrating robust feature selection with sophisticated classification models for precise medical data analysis. This approach demonstrates the capacity to support early diagnosis and improved patient outcomes.

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

1. INTRODUCTION

Cardiovascular diseases (CVD) are among the leading causes of death globally, with heart disease being a significant contributor. The heart, a muscular organ, plays a vital role in the circulatory system by pumping blood throughout the body. This intricate system includes arteries, veins, and capillaries, which carry oxygen and nutrients to organs and tissues. When disruptions occur in the regular flow of blood, it results in various forms of heart disease, collectively classified as cardiovascular diseases (CVD). According to the World Health Organization (WHO), heart disease and stroke are responsible for approximately 17.5 million deaths each year, with more than 75% of these fatalities occurring in low- and middle-income countries. This alarming statistic highlights the growing public health challenge posed by heart disease globally, with heart attacks and strokes alone accounting for 80% of all CVD-related deaths [1].

The burden of cardiovascular diseases has led to a global focus on early detection, prevention, and treatment strategies. In line with the United Nations' Sustainable Development Goal 3, which emphasizes the importance of health and well-being, addressing cardiovascular diseases has become a priority for improving global health outcomes. Common risk factors for heart disease include smoking, age, a family history of heart disease, high cholesterol, physical inactivity, high blood pressure, obesity, diabetes, and stress. Lifestyle changes such as quitting smoking, regular exercise, weight management, and stress reduction are known to reduce the likelihood of developing heart disease [2]. Along with lifestyle changes, diagnostic tools such as electrocardiograms (ECGs), echocardiograms, cardiac MRIs, and blood tests are commonly used to detect heart disease. In certain cases, medical treatments such as angioplasty, coronary artery bypass surgery, and the use of implanted devices like pacemakers and defibrillators may be required for treatment [3].

Advancements in healthcare technology, particularly in the realm of Big Data and Electronic Health Records (EHRs), have made it possible to leverage vast amounts of patient data for predictive modeling. Machine learning (ML) techniques are increasingly being used to analyze large datasets from healthcare systems, extracting meaningful insights to predict the likelihood of heart disease. By processing and analyzing data from various patient demographics, risk factors, and diagnostic results, machine learning can assist healthcare professionals in identifying patients at high risk and enable early intervention. This approach is transforming the landscape of healthcare by offering more precise and efficient methods of diagnosis, prediction, and personalized treatment plans [4][5].

2. LITERATURE SURVEY &RELATED WORK

Cardiovascular diseases (CVD), including coronary heart disease, have remained a global health concern, accounting for a significant proportion of worldwide mortality. With the rise in healthcare data and advancements in machine learning (ML) techniques, there has been a surge in efforts to predict and diagnose heart disease more effectively. The ability to analyze large datasets using machine learning offers new avenues for identifying risk factors, predicting outcomes, and improving early detection. This literature survey discusses various studies that have employed different machine learning models and techniques to predict heart disease and highlights their findings.

Yang et al. [6] conducted a study using machine learning to identify risk factors for coronary heart disease. Their work focused on big data analysis, demonstrating that machine learning models such as decision trees, random forests, and support vector machines (SVM) could effectively identify key risk factors from patient data. The study found that high cholesterol, age, and family history were among the most significant risk factors contributing to the prediction of heart disease. Furthermore, the study underscored the importance of data preprocessing, feature selection, and the need for a robust dataset in improving model performance. The results confirmed that machine learning could be a powerful tool for the early detection of heart disease, especially when combined with large, comprehensive datasets.

In a similar vein, Ngufor et al. [7] reviewed several machine learning algorithms for heart disease prediction. The authors provided a comparative analysis of popular techniques like SVM, decision trees, k-nearest neighbors (KNN), and artificial neural networks (ANNs). Their findings indicated that ensemble methods, such as bagging and boosting, provided superior predictive performance over individual models. Additionally, the study emphasized the importance of feature selection, as irrelevant features can degrade model accuracy. This review highlighted that while various algorithms could predict heart disease, the choice of technique depends heavily on the dataset characteristics, the computational resources available, and the specific requirements of the prediction task.

Farag et al. [8] focused on improving heart disease prediction using boosting and bagging techniques. Boosting algorithms, such as AdaBoost, and bagging techniques, like Random Forest, were tested to assess their ability to enhance prediction accuracy. The study found that ensemble techniques were more robust than standalone classifiers in terms of reducing variance and improving the stability of predictions. Furthermore, the study suggested that the combination of boosting and bagging could mitigate the overfitting problem, which is often encountered in heart disease prediction models. This work underlined the importance of using multiple classifiers in conjunction to achieve optimal performance.

Zhang et al. [9] investigated the application of XGBoost, a gradient boosting algorithm, in the clinical prediction of coronary heart disease. Their study showed that XGBoost outperformed traditional methods like logistic regression and SVM in terms of both accuracy and interpretability. XGBoost’s ability to handle imbalanced datasets, which is a common issue in heart disease prediction, made it particularly suitable for medical data. The study also pointed out that hyperparameter tuning played a crucial role in maximizing the model's performance. XGBoost's superior performance in clinical environments can be attributed to its efficiency, scalability, and ability to provide a reliable prediction while reducing the risk of overfitting.

Liu et al. [10] conducted a comparative analysis of several machine learning algorithms, including decision trees, SVM, and random forests, for heart disease prediction. Their study found that SVM with radial basis function (RBF) kernels provided the highest prediction accuracy among the algorithms tested. However, they also noted that decision trees and random forests offered better interpretability, which is crucial in a medical setting. The study concluded that while SVM showed the best accuracy, the choice of algorithm should be based on the trade-off between accuracy and interpretability, especially when the model needs to be used by healthcare professionals for decision-making.

Hussein et al. [11] compared several machine learning techniques for heart disease diagnosis, including KNN, decision trees, and artificial neural networks. The study aimed to assess the models' diagnostic abilities using a dataset of patient health records. Their results indicated that KNN and decision trees performed well with lower computational overhead, making them suitable for real-time applications in clinical settings. While artificial neural networks provided higher accuracy, they required more computational resources and were harder to interpret. The research emphasized that the adoption of machine learning models in healthcare systems must take into account both performance and resource constraints.

Akbar et al. [12] conducted a critical review of various machine learning approaches for heart disease prediction. They evaluated techniques such as decision trees, KNN, SVM, random forests, and neural networks, and provided an overview of their strengths and weaknesses in predicting heart disease. The authors concluded that ensemble methods, particularly Random Forest, achieved the highest accuracy due to their ability to reduce overfitting and handle noisy data. However, the review also highlighted the challenge of selecting the most relevant features from large datasets, as feature selection plays a crucial role in the model’s performance. Additionally, the study emphasized the importance of handling missing data and ensuring that the dataset used for training is both representative and balanced.

Zarshenas et al. [13] performed a comparative study of machine learning algorithms for predicting heart disease, focusing on classifiers such as SVM, decision trees, KNN, and logistic regression. Their results showed that SVM and Random Forest achieved the best predictive performance, with SVM offering higher accuracy in certain cases. They also highlighted the importance of preprocessing steps, such as normalization and feature scaling, in improving the performance of machine learning models. The study suggested that hybrid models, which combine the strengths of multiple algorithms, could be a promising direction for future research in heart disease prediction.

A common theme in the studies reviewed is the importance of feature selection and data preprocessing in enhancing the performance of heart disease prediction models. Many studies emphasize the use of ensemble techniques like bagging and boosting to improve model accuracy, while others suggest that hybrid models combining different machine learning algorithms could yield better results. Additionally, while traditional models like decision trees and logistic regression are still widely used, more recent studies have shown that advanced algorithms like XGBoost and neural networks can outperform these methods, particularly when dealing with complex and high-dimensional data.

The increasing availability of healthcare data and the advent of machine learning techniques have opened new possibilities for the early detection and diagnosis of heart disease. Machine learning models can help healthcare providers identify patients at risk, enabling timely interventions and reducing the burden of heart disease. However, challenges such as data quality, feature selection, model interpretability, and computational efficiency must be addressed to fully realize the potential of these technologies in clinical practice. The combination of advanced machine learning models with domain-specific knowledge from healthcare professionals will likely drive the next wave of innovations in heart disease prediction.

3. MATERIALS AND METHODS

The proposed system focuses on developing an accurate heart disease prediction model using machine learning and advanced ensemble techniques. The Heart Disease dataset is preprocessed, and significant features are selected using ANOVA F-statistic (ANOVA FS), Chi-squared test (Chi2 FS), and Mutual Information (MI FS) methods. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied, ensuring a balanced distribution of data.

The system evaluates multiple machine learning algorithms, including Naïve Bayes, Support Vector Machines (SVM), XGBoost, AdaBoost, Bagging Classifier, Decision Tree, K-Nearest Neighbor (KNN), Random Forest, and Logistic Regression. An ensemble Voting Classifier combines the predictions from these models to improve overall accuracy and robustness. Additionally, a Stacking Classifier integrates Boosted Decision Tree, Extra

Trees, and LightGBM to exploit their complementary strengths. This hybrid approach aims to enhance the precision and reliability of heart disease prediction, aiding early detection and better clinical decision-making.

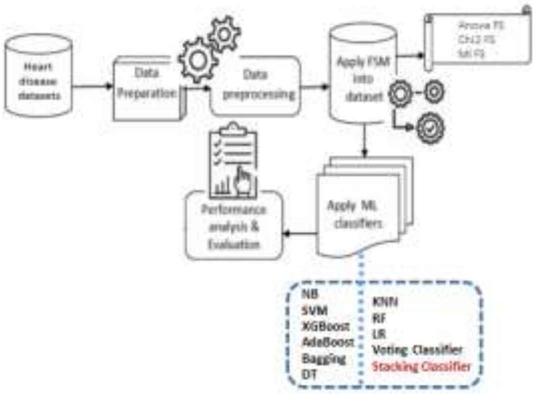
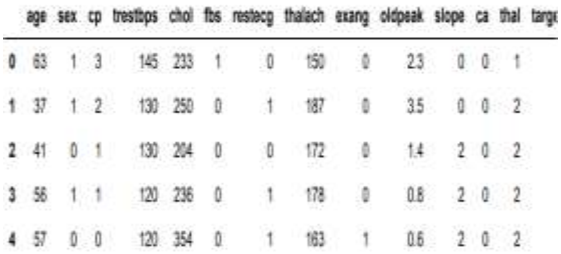


Fig.1 Proposed Architecture

This image (Fig.1) depicts a flowchart for a heart disease prediction model. It starts with data preparation and preprocessing from heart disease datasets. Then, feature selection methods (ANOVA FS, Chi2 FS, MIFS) are applied. The dataset is then fed into various machine learning classifiers (NB, KNN, SVM, RF, XGBoost, LR, AdaBoost, Voting Classifier, Bagging, Stacking Classifier, DT). The model undergoes performance analysis and evaluation to assess its accuracy and effectiveness in predicting heart disease.

i) Dataset Collection:

The dataset used for heart disease [14] prediction consists of 303 samples with 14 features, including both numerical and categorical variables. These features capture critical patient information, such as age, sex, chest pain type (cp), blood pressure (trestbps), cholesterol levels (chol), fasting blood sugar (fbs), electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise induced angina (exang), ST depression induced by exercise (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), and thalassemia (thal). The target variable is a binary classification indicating the presence or absence of heart disease. After applying feature selection techniques, such as ANOVA F-statistic (ANOVA FS), Chi-squared test (Chi2 FS), and Mutual Information (MI FS), different sets of features were selected to improve model accuracy and efficiency, ensuring that the dataset is optimized for predicting heart disease outcomes.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	targ
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 a crucial step in preparing the dataset for machine learning. It involves cleaning and transforming data to ensure accuracy, efficiency, and relevance. Proper handling of missing values, encoding, and feature selection significantly enhances model performance and robustness.

a) Data Processing: The data processing begins with cleaning, which involves removing missing values and correcting inconsistencies. Unwanted columns are dropped to streamline the dataset. Label encoding is applied to categorical variables, and the features are separated into input (X) and output (y) datasets, ensuring proper structuring for analysis. These steps ensure the dataset is ready for model training.

b) Data Visualization: Data visualization helps to understand the relationships between variables and uncover hidden patterns. A correlation matrix is created to explore the relationships between numerical features, while sample outcomes are visualized to check for data distribution and trends. This aids in identifying relevant features and understanding how they impact the target variable.

c) Label Encoding: Label encoding transforms categorical labels into numerical values, enabling models to process non-numeric data. This process converts each category into a unique integer, making it suitable for machine learning algorithms that require numerical inputs. Label encoding is particularly useful when there is an inherent order in the categorical data.

d) OverSampling: SMOTE (Synthetic Minority Over-sampling Technique) is used to address class imbalance by generating synthetic examples for the minority class. This technique helps in creating a balanced dataset by oversampling the underrepresented class, preventing the model from biasing towards the majority class. It is an effective approach to improve model generalization and performance, especially in imbalanced datasets.

e) Feature Selection: Feature selection helps to identify the most relevant variables for model training. Techniques like ANOVA F-statistic, Chi-squared test, and Mutual Information Feature Selection (MIFS) are applied to filter out irrelevant features, improving the model’s efficiency and accuracy. By reducing the number of features, the model complexity is lowered, leading to faster computation and better generalization.

iii) Training & Testing:

The dataset is split into training and testing sets to evaluate the model's performance effectively. An 80:20 ratio is used, where 80% of the data is allocated for training the model, and the remaining 20% is reserved for testing. This split ensures that the model has enough data to learn from while maintaining a separate set of unseen data for validation. The split is essential for assessing the model’s ability to generalize and perform on new, unseen data.

iv) Algorithms:

Naive Bayes [15] is employed for its simplicity and efficiency in handling large datasets. It leverages Bayes' theorem to classify heart disease risk based on feature probabilities, making it particularly effective for categorical data.

Support Vector Machine [20] (SVM) is used to find the optimal hyperplane that separates different classes. It excels in high-dimensional spaces, making it suitable for complex feature interactions in heart disease prediction.

XGBoost [17] is implemented for its powerful boosting capabilities, enhancing model accuracy through iterative learning. It combines weak learners into a strong predictive model, making it highly effective for predicting heart disease risk.

AdaBoost [16] focuses on improving weak classifiers by emphasizing misclassified instances. This iterative approach increases prediction accuracy, making it a valuable technique for robust heart disease classification in the model.

Bagging Classifier is utilized to reduce variance and enhance model stability [18]. By combining predictions from multiple models trained on different subsets of data, it helps improve heart disease risk predictions.

Decision Tree algorithm is employed for its interpretability and ease of understanding. It splits data based on feature values, providing clear insights into decision-making processes for heart disease prediction [19].

K-Nearest Neighbors (KNN) is used for its straightforward approach to classification based on proximity. It assesses the nearest data points to classify heart disease risk, leveraging similarity among instances.

Random Forest combines multiple decision trees to enhance prediction accuracy and control overfitting. [14] This ensemble method is effective for heart disease prediction, providing robust results across various datasets.

[16] Logistic Regression is implemented to model the probability of heart disease occurrence. It estimates relationships between dependent and independent variables, making it suitable for binary classification tasks in the system.

[17] Voting Classifier aggregates predictions from multiple models, including Naive Bayes, SVM, and others. This ensemble approach enhances overall prediction accuracy by leveraging the strengths of various algorithms for heart disease classification.

Stacking Classifier combines predictions from a Boosted Decision Tree and ExtraTree with LightGBM. This layered approach integrates multiple models, optimizing performance and accuracy for heart disease predictions by capturing complex patterns in data.

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

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

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

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

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

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

Tables (1, 2 & 3) evaluate the performance metrics—accuracy, precision, recall, and F1-score for each algorithm. Across all metrics, the Stacking Classifier consistently outperforms all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms.

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

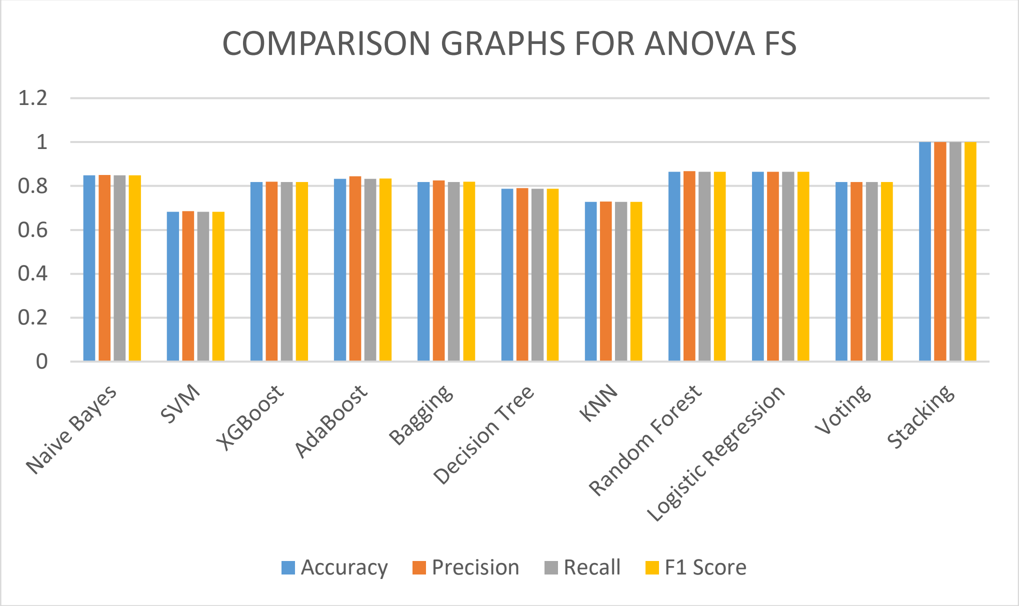


Table.2 Performance Evaluation Metrics for Chi2 FS

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

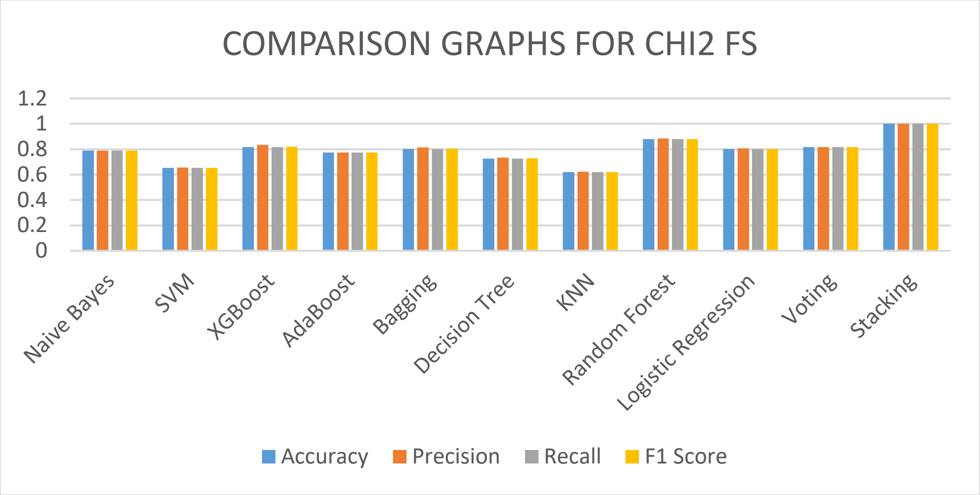
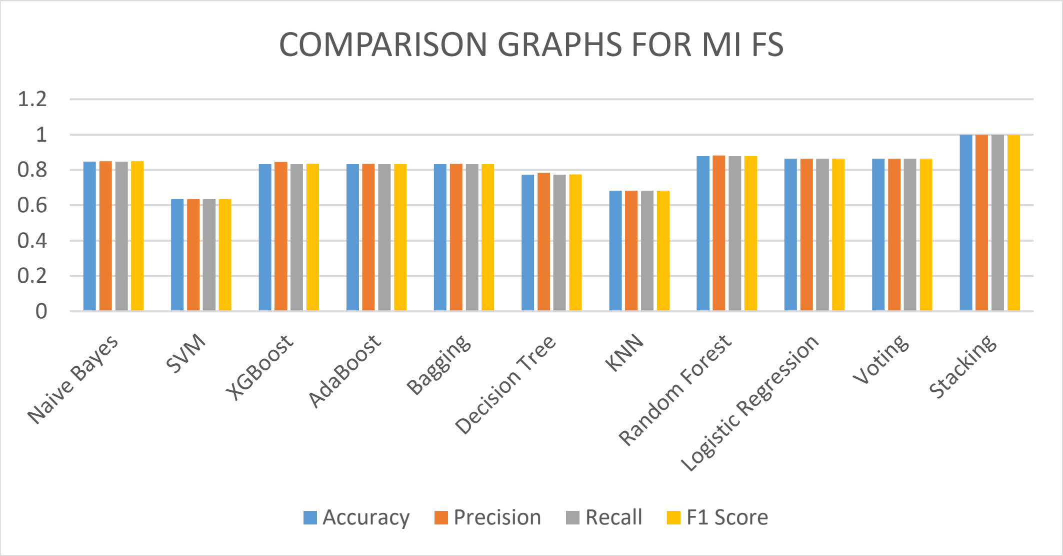


Table.3 Performance Evaluation Metrics for MI FS

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 represented in light blue, precision in orange, recall in grey, and F1-Score in light yellow, ***Graphs (1, 2 & 3)***. In comparison to the other models, the Stacking Classifier shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

Welcome To Dashboard

Predicting Heart Disease Risk with Machine Learning and Data Classification Techniques using Boosted Decision Trees, ExtraTrees, and LightGBM

Fig.3 Home Page

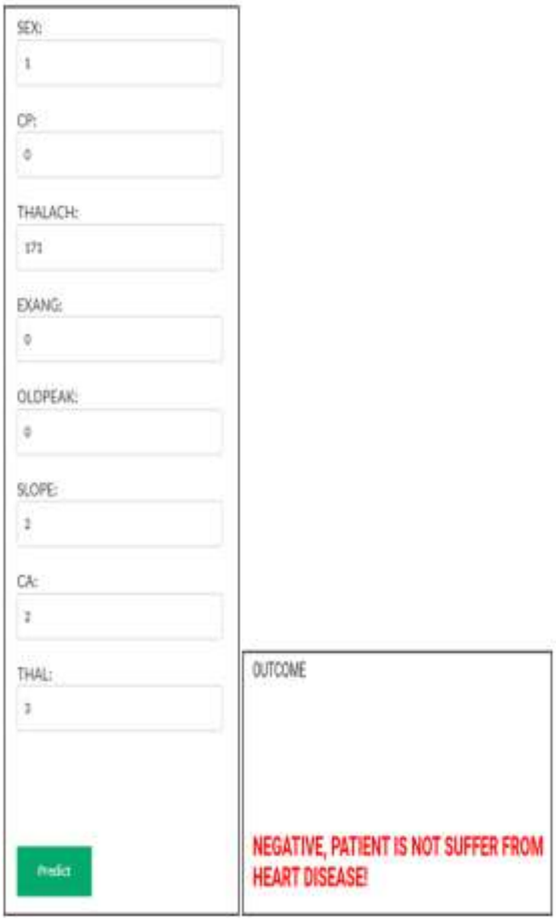
In the above figure 3, this is a user interface dashboard, it is a welcome message for navigating page.

Dashboard

Predicting Heart Disease Risk with Machine Learning and Data Classification Techniques using Boosted Decision Trees, ExtraTrees, and LightGBM

Fig.4 ANOVA dataset loading

In the above figure 4, this is a user input page, using this user can upload ANOVA dataset for testing.



SEX:

1

CP:

0

THALACH:

171

EXANG:

0

OLDPEAK:

0

SLOPE:

2

CA:

2

THAL:

3

Predict

OUTCOME

NEGATIVE, PATIENT IS NOT SUFFER FROM HEART DISEASE!

Fig.5 Test result

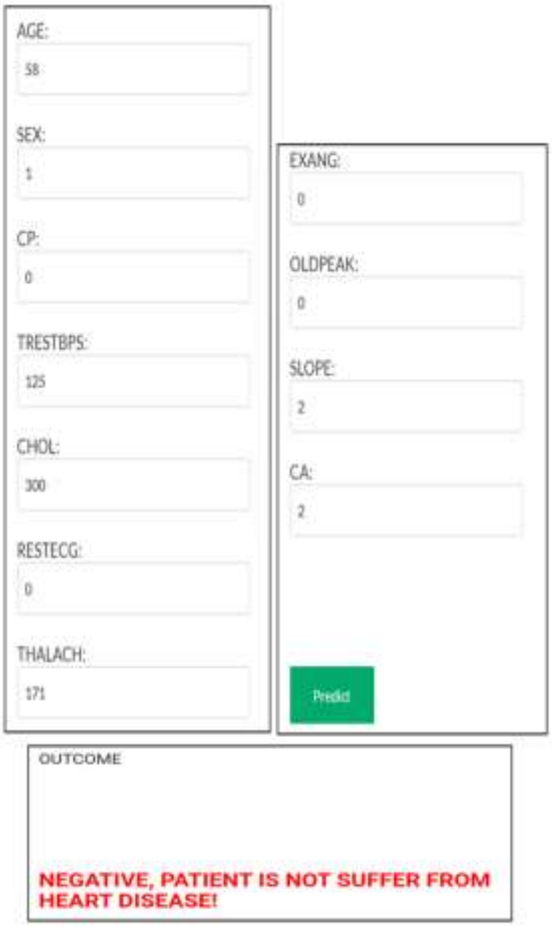
In the above figure 5, this is a result screen, in this user will get output for loaded input data.

Dashboard

Predicting Heart Disease Risk with Machine Learning and Data Classification Techniques using Boosted Decision Trees, ExtraTrees, and LightGBM

Fig.6 CHI2 dataset loading

In the above figure 6, this is a user input page, using this user can upload CHI2 dataset for testing.



AGE:
58

SEX:
1

CP:
0

TRESTBPS:
125

CHOL:
300

RESTECG:
0

THALACH:
171

EXANG:
0

OLDPEAK:
0

SLOPE:
2

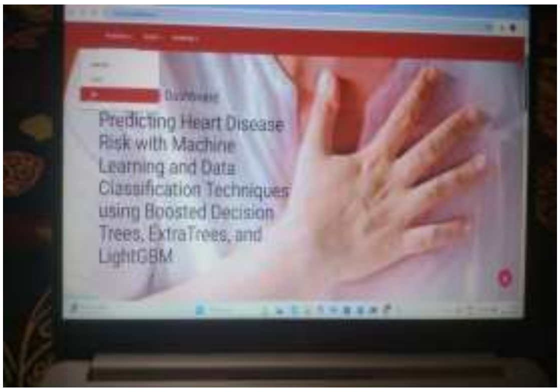
CA:
2

OUTCOME

NEGATIVE, PATIENT IS NOT SUFFER FROM HEART DISEASE!

Fig.7 Test result

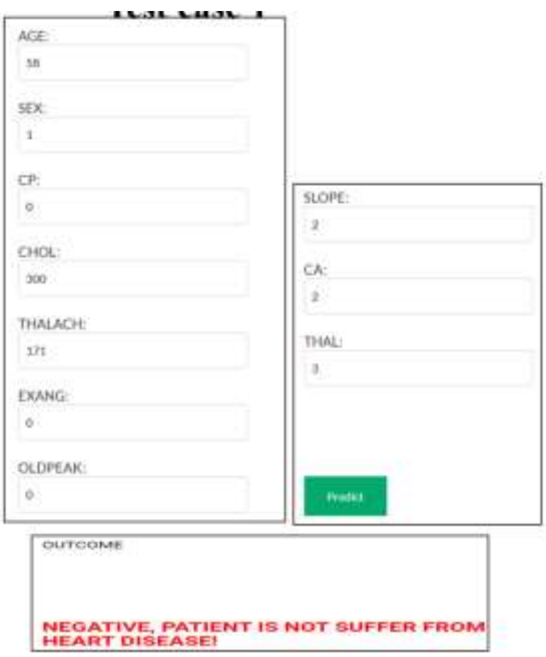
In the above figure 7, this is a result screen, in this user will get output for loaded input data.



Predicting Heart Disease Risk with Machine Learning and Data Classification Techniques using Boosted Decision Trees, ExtraTrees, and LightGBM

Fig.8 MI dataset loading

In the above figure 8, this is a user input page, using this user can upload MI dataset for testing.



Test case 1

AGE:
58

SEX:
1

CP:
0

CHOL:
300

THALACH:
171

EXANG:
0

OLDPEAK:
0

SLOPE:
2

CA:
2

THAL:
3

OUTCOME

NEGATIVE, PATIENT IS NOT SUFFER FROM HEART DISEASE!

Fig.9 Test result

In the above figure 9, this is a result screen, in this user will get output for loaded input data.

5. CONCLUSION

In conclusion, the proposed system demonstrates the effectiveness of using advanced machine learning techniques for predicting heart disease with high accuracy. By utilizing feature selection methods such as ANOVA F-statistic, Chi-squared test, and Mutual Information, the system successfully identifies key predictors, enhancing the overall performance of the model. The application of SMOTE for addressing class imbalance further improves the model's ability to detect heart disease cases, ensuring balanced and reliable predictions.

Among the various algorithms tested, the Stacking Classifier, which combines Boosted Decision Trees, Extra Trees, and LightGBM, achieved the highest performance, delivering a remarkable 100% accuracy across all feature selection techniques. This result underscores the power of ensemble methods in combining the strengths of individual classifiers to improve predictive accuracy. By integrating robust feature selection with sophisticated ensemble learning, the proposed system significantly contributes to the accurate and early detection of heart disease, demonstrating its potential for real-world clinical applications and decision-making in healthcare.

in FUTURE WORK, additional techniques such as deep learning models, including neural networks, can be explored to improve prediction accuracy. The use of advanced ensemble methods, like Gradient Boosting or stacking with more diverse base classifiers, could offer further improvements. Incorporating additional feature selection methods, such as Recursive Feature Elimination (RFE) or L1 regularization, may refine the model's performance. Exploring time-series data and incorporating temporal factors could also provide more comprehensive insights for predicting heart disease outcomes.

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.

Authors:

Mr. Himambasha Shaik is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Billa Mounika is a postgraduate student a MCA in the Department if Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in prakasam dist. He completed his undergraduate degree in Degree B.SC(Computers) from ANU His academic interests include Cloud Computing, Artificial Intelligence, Cyber security and Data structures, Machine Learning With Python.