DECIPHERING CARDIAC HEALTH: A DEEP DIVE INTO MACHINE LEARNING FOR PREDICTIVE HEART DISEASE MODELING

Simhadati Prasanna [1] and Dr. Chappa Ramesh [2*]

[1] PG Student, Department of CSE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[2] Professor, Department of CSE, Aditya Institute of Technology and Management, Tekkali, AP, India.

DOI: 10.5281/zenodo.11634855

Abstract

Heart disease is a grave concern that affects millions across the globe. Preventive healthcare relies on timely exposure and accurate prediction of heart disease This research paper examines the practice of machine learning algorithms to envisage heart disease, leveraging diverse datasets and advanced analytical techniques. The study begins with a thorough review of existing literature on ML-based heart disease estimation models, highlighting the strengths and confines of various algorithms. Subsequently, a comprehensive dataset comprising demographic, clinical, and lifestyle factors is employed for model development. The research focuses on popular ML algorithms such as Random forests, KNN, SVM, and Neural Networks, comparing their performance in terms of sensitivity, specificity, and overall accurate results published will vary by the datasets, sizes, and many other features [1] Cardiovascular diseases (CVDs) refer to a group of medical conditions that have a control on the heart and blood vessels. It's important to be aware of the conditions that can affect our hearts and overall health. Among these are coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other related ailments. According to WHO, CVDs are the major reason for death around the world, claiming an estimated 17.9 million lives each year.

Keywords: Cardio Vascular Disease, Neural Networks, ML-Based Prediction.

1. INTRODUCTION

The major problem observed in today's world is heart disease. The heart stops functioning when it finds a lack of blood circulation. This may lead to a severe heart attack. According to WHO reports, nearly 18 million deaths are seen due to coronary artery disease [2]. The unprecedented growth in health data availability and advancements in ML methodologies present a unique opportunity to develop robust predictive models that can leverage diverse patient information, including clinical records, lifestyle factors, and genetic markers. Within the realm of heart disease, a multitude of distinct conditions manifest, each presenting unique challenges and implications for cardiovascular health. Among these, coronary artery disease (CAD) emerges prominently, categorized by the gradual evolution of plaque within the coronary arteries. This process diminishes blood flow to the heart muscle, setting the stage for critical events such as heart attacks. Concurrently, heart failure, marked by the heart's compromised pumping ability, underscores the varied nature of cardiovascular disorders and their impact on overall health. Machine learning algorithms offer a unique advantage in heart disease prediction as they can scrutinize large amounts of patient data, identify patterns, and create truthful predictions.[3]

By analyzing several crucial threat factors, comprising blood pressure, insulin levels, cholesterol, pulse rate, and body mass index, algorithms can accurately predict an individual's susceptibility to heart disease. This is a powerful tool that can be used to identify potential risks early on, allowing for preventative measures and potentially life-saving interventions. By analyzing these factors, machine learning algorithms can create tailored risk prediction models that can assist in initial detection and intervention strategies for heart disease. Machine learning algorithms learn and adapt from new data, improving accuracy and performance over time. Businesses can use them to streamline operations, understand customers better, and make more informed decisions.[4]

It will then describe the methodology applied here, which includes the selection of machine learning algorithms and the collection as well as data preprocessing. This document outlines the findings of our study on the predictive accuracy of different machine learning algorithms in relation to heart disease. The results of this study may be of interest to a wider audience as they provide valuable insights into the performance of these algorithms. Our findings underscore the significance of careful selection and proper evaluation of machine learning algorithms. Furthermore, this paper will discuss the limitations and potential challenges that arise while using machine learning algorithms in heart disease prediction and future research directions in this field.

1.1 Overview of Machine Learning Practices:

Machine learning models improve healthcare by analyzing data to identify patterns, leading to better diagnoses and treatment plans. [3] These techniques involve the use of algorithms that facilitate the acquisition of knowledge by computers via data analysis, thus enabling them to draw inferences without explicit programming. Machine learning algorithms are advanced tools that enable us to analyze and understand complex data sets, making informed decisions and predictions based on that data easier. By constantly learning and adapting to new information, these models can provide valuable insights and help us optimize our processes and strategies for greater success. They are broadly categorized into three main types, each with a unique learning approach. Supervised learning algorithms are machine learning algorithms that work on labeled data. In other words, each example in the dataset is associated with a known outcome or label. The algorithm learns to recognize patterns in the input data and associate them with the corresponding labels. These patterns can be used to create prophecies on new, unlabeled data. For instance, a supervised learning algorithm can be used to identify handwritten digits by presenting many examples of handwritten digits, along with their corresponding labels. The algorithm uses these examples to learn how to recognize patterns in the data and associate them with the correct digit label.

Unsupervised learning algorithms are powerful tools that enable us to extract insights from data without the need for predefined labels or outcomes. These algorithms are adept at identifying patterns and structures within the data, thereby paving the way for discoveries and insights. For example, we can use unsupervised learning algorithms to group customers based on their purchase behavior, thereby understanding their preferences and needs more deeply. This knowledge can then be leveraged to create more personalized and effective marketing strategies.

Reinforcement learning algorithms function by engaging with an environment and obtaining feedback through rewards or punishments. The algorithm aims to acquire a strategy that will optimize a cumulative reward signal over a period. For example, a reinforcement learning algorithm can train an agent to play a video game by incentivizing it to obtain high scores and penalizing it for losing lives. The algorithm

ISSN 1462 2815

becomes more proficient at the game over time by learning to take actions that generate greater rewards and avoid actions that result in lesser rewards.

1.2 Machine Learning Techniques in Medical Study

Machine learning practices have great potential in medical diagnostics, particularly heart disease prediction. [5] Machine learning algorithms can detect hidden patterns in patient data that may not be obvious to human clinicians. These algorithms can deliberate a wide range of features and variables, like demographic data, medical thesis, symptoms, and test results, to make accurate prophecies about an individual's risk aimed at heart disease. Additionally, Machine learning algorithms can update their models with new data. Allowing for improved accuracy and adaptability in predicting heart disease [6]. One of the key advantages of using machine learning in heart disease prediction is its ability to handle complex and high-dimensional data. Traditional statistical models may struggle to capture all the intricate relationships between various risk factors and their impact on heart disease.

Machine learning algorithms, on the other hand, can effectively handle this complexity and consider multiple variables simultaneously. This enables more comprehensive and accurate predictions, leading to initial revealing and intervention for individuals at the possibility of heart disease. Furthermore, machine learning algorithms can also handle missing data and outliers more effectively than traditional statistical methods. Various machine learning algorithms can be utilized to predict heart disease. Some frequently used machine learning algorithms for heart disease extrapolation include a k-nearest neighbor, decision tree, linear regression, support vector machine, logistic regression, random forest, and neural networks. These algorithms are trained on labeled data sets that consist of features related to the individual's demographics, medical history, lifestyle factors, and diagnostic test results. The data collected is processed by the algorithms to identify patterns and correlations. These patterns are then leveraged to categorize individuals into different groups based on factors such as the existence or nonexistence of heart disease. By comparing the accuracy of different algorithms, researchers can determine which algorithm performs best in predicting heart disease.

1.3 Heart Disease Causes And Symptoms:

Heart disease is a wide-ranging term that involves several conditions affecting the structures or functions of the heart. There are several medical conditions related to the heart, including coronary artery disease, heart failure, arrhythmias, and valvular heart diseases. The causes of heart disease can vary, but some common risk factors include high BP, high cholesterol levels, diabetes, obesity, smoking, family history of heart disease, and an inactive lifestyle. The symptoms of heart disease can also fluctuate based on the specific condition but recognizing symptoms such as chest pain, shortness of breath, fatigue, dizziness, irregular heartbeat, and swelling is crucial for effective management. Prompt identification can make all the difference. Machine learning algorithms have the prospective to renovate heart disease prediction by analyzing large volumes of patient records and identifying patterns that may not be evident to clinicians. By leveraging machine learning algorithms, healthcare professionals can make more accurate predictions about a person's risk for heart disease. These predictions can help inform personalized treatment plans, lifestyle modifications, and certain preventive measures to help reduce the rate of heart disease. Moreover, the use of machine learning in heart disease estimation allows for continuous learning and improvement. By continuously analyzing new data and updating the predictive models, machine learning algorithms can become accustomed to new patterns and advance their accuracy over time.

Sectional Division: The present paper is systematized into five different sections. Section 2 delivers an in-depth analysis of the literature review on the prophecy of heart disease. Section 3 outlines the suggested methodology of the paper. Section 4 presents the experimental setup for the proposed system, which comprehends empirical data, data pre-processing model setting, parameter setting, and paralleled methodologies along with the performance measurement. Section 5 probes into the results and analysis, while section 6 offers a conclusive summary of the paper.

2. LITERATURE STUDY

After a keen survey of data collected from different researchers, we elaborate on some findings from the literature on heart disease prediction. Using a heart disease dataset, Kalla[6] performed chi-square tests along with linear regression analysis to foretell heart disease based on symptoms, including chest pain. Logistic regression was used to build a prediction model with an accuracy of 85.12%. Dun, E Wang, et. al[7] have investigated several machine learning methodologies like logistic regression, Random forest, and KNN for analyzing heart disease with hypermeter tuning and feature selection, resulting in maximum accuracy of 78%. Roohallah Alizadehsania, Molou. and Abdarb[8] have discussed a comprehensive review of coronary artery disease diagnosis. Here, a total of 67 datasets have been considered that are collected from different countries of 3 continents. A lot of machine learning algorithms were used for LAD stenosis, and the highest accuracy was 86.10%. Huru Hasanova and Muhammad Tufail [9] examined several physiognomies supervised to analyze some early stages of heart disease. The dataset was collected from the UCI repository and included 303 instances and 76 attributes. Only 14 key attributes are used here. The use of SCA_WKNN leads to a 13.05% increase in accuracy compared to using WKNN. After comparing various models, the Random forest algorithm performed well among them. Literature studies on heart disease prediction using several advanced machine learning techniques are represented in Table 1 below.

Table 1: Literature study on heart disease prediction

S. No	Dataset	Method	Performance	Evolution Factor	References
1	Random clinical data from Kaggle data source	K-Nearest Neighbours, Naïve Bayes, SSA-NN	Accuracy-86.7%, Sensitivity- 80%	Precision, recall, f-measures	[10]
2	Random clinical data	K-Nearest Neighbour, Random Forest, Support vector machine	Accuracy- 0.963 (Random Forest)	Accuracy, Sensitivity, Specificity	[11]
3	Dataset from UCI Repository	Logistic regression and Robust method	Logistic Regression - 86.2%,(accuracy)	Accuracy, Sensitivity	[12]
4	Random Clinical Data	SVM, KNN	SVM- 0.775(accuracy), 0.650(sensitivity), 0.930(Specificity),	Accuracy, Sensitivity, Specificity	[13]

			KNN-0.605(accuracy), 0.590(Sensitivity), 0.620(Specificity)		
5	Fingertip'svideo dataset	Modified Plant Optimization Algorithm, Logistic Regression, Naïve Bayes, XGB and ANN	Random forest-88.82%(accuracy)	Accuracy, Precision, f1 score	[14]
6	Random clinical data	KNN, Decision tree, Naïve Bayes, Neural Networks	Accuracy- 52.7%, Random forest	Accuracy, Average	[15]
7	NHANES data from 1999–2000 to 2015-2016	LASSO	Recall value-0.77 Specificity-0.81 Test accuracy-0.82	Recall, specificity, Test Accuracy	[16]
8	Cleveland database with 76 attributes and 302 instances	Naïve Bayes, Logistic Regression, Random Forest, SVM	PCA-82.4% (specificity), precision (86.4%)	Accuracy, Specificity, F1 Score, Sensitivity, False Omission rate	[17]
9	'3,126heart recordings data	CNN, Gaussian SVM, Coarse Gaussian Support Vector Machine	Accuracy-75.7%	Accuracy	[19]

3. PROPOSED METHOD

Regression techniques have broad applications in medical research as they can assess connections, forecast results, and manage the impact of confounding variables. Logistic regression, for example, is an effective method to examine the possibility of a binary conclusion based on the independent variables. By utilizing elements of linear regression found in the logit scale, logistic regression detects the most powerful linear combination of variables that is likely to detect the observed result. Key concerns when using logistic regression involve the selection of independent variables based on established theory, previous research findings, clinical insights, and univariate statistical evaluations while also considering potential perplexing factors that need to be addressed.[18-20]

In logistic regression, by applying the sigmoid function, a linear combination of input features can be transformed into a value that falls between 0 and 1. Logistic Regression employs a more intricate cost function, which can be expressed as the 'Sigmoid function', also referred to as the 'logistic function'.The sigmoid function is given by. [21]

$$\boldsymbol{\sigma(z)=1/1} + e^{-x}$$

"Here, 'z' refers to the output obtained by multiplying the input features with their respective weights and then adding them together in a linear fashion.".

z is evaluated as shown below

$$z=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n$$

$\beta_0,\beta_1,\ldots,\beta_n$ are the coefficients or weights assigned to each feature ($x_1,x_2,\ldots,x_n$).

To achieve the best results, many data scientists and machine learning practitioners resort to Gradient Descent or other optimization techniques to determine the optimal values for the coefficients that minimize the loss function. The decision boundary is the line (or hyperplane in higher dimensions) that splits the two classes in the feature space After training, the model predicts the probability of a new sample being in class 1.. A threshold is applied (commonly 0.5) to classify the sample into one of the two classes.

When assessing the effectiveness of machine learning models, it is common to employ various measures such as accuracy, recall, F1 score, precision, and ROC-AUC. These measures are used to evaluate the performance of the models in terms of their predictive power and ability to classify data accurately. Evaluating the model:

1. Collect the data for the binary and independent outcome variables.
2. With data pre-processing, we can ensure that the data is reliable by identifying and addressing concerns like missing values and outliers and variable transformations if needed.
3. Split the data into a training set and a testing set to assess the model's performance.
4. Standardize or normalize the independent variables to ensure they have a similar scale.
5. Perform feature selection or extraction techniques to lessen the number of independent variables and improve the model's interpretability and performance.
6. Apply logistic regression by fitting a model that predicts the logit of the outcome variable using the selected independent variables.
7. Obtain the coefficient estimates for each independent variable by maximizing the likelihood function using methods such as maximum likelihood estimation or gradient descent.
8. Evaluate the significance of the coefficient estimates using statistical checks such as the Wald experiment or likelihood ratio test.
9. Interpret the coefficients by assessing their direction and magnitude, indicating how they influence the probability of the binary outcome.
10. Assess the model's goodness of fit using measures such as deviance, AIC, or BIC.

LightGBM is crucial for optimizing memory usage and training data in this research paper.LightGBM, short for Light Gradient Boosting Machine, is an open-source gradient boosting framework that is developed by Microsoft. It is designed for distributed and efficient machine-learning tasks. LightGBM is particularly known for its speed and efficiency in handling large datasets and high-dimensional feature spaces. It utilizes a tree-based learning algorithm, similar to other gradient boosting frameworks like XGBoost and CatBoost, but it introduces some optimizations to improve training speed and reduce memory usage. These optimizations include a histogram-based algorithm for finding the best splits during tree construction and a leaf-wise tree growth strategy. LightGBM supports various types of tasks, including classification, regression, and ranking. It's widely used in academia and industry for various machine-learning tasks due to its performance and scalability.

Steps involved while performing LightGBM for data Preprocessing:

1. Prepare your dataset in a format that LightGBM can work with. LightGBM supports various data formats like CSV, numpy arrays, or Pandas Data Frames.
2. Separate your dataset into features (X) and target variables (y).
3. Create and train the LightGBM model.
4. Once the training is completed, evaluate your model's performance using appropriate evaluation metrics.
5. Use the trained model to make predictions on new data.
6. Now calculate the accuracy score of the LightGBM model.

4. EXPERIMENTAL SETUP

4.1 Data collection:

This paper's dataset was taken from the source Kaggle and then processed to deliver the expected model. The processed dataset consists of 14 attributes, which we elaborate on in **Table 2.** The dataset consists of a total of 303 entries, of which 165 are male patients, and 138 are female patients of different ages. Among these, 45.54% of patients have heart problems, and 54.46% do not. The dataset consist**s** of float and integer data types. A detailed view is represented in the table below.

Table 2: Datatype and meaning of each feature of the dataset

Attribute	Data Type	Description
Age	Integer	Describes the age of the person
sex	Integer	Describes the sex (M/F)
cp	Integer	Chest pain
treetops	Integer	The person’s relaxing blood pressure
chol	Integer	Cholesterol levels present
FBS	Integer	Fasting blood sugar level of the person
resting	Integer	Resting electro-cardiographic result
thalach	Integer	Maximum heart rate achieved by the person
exang	Integer	Exercise that includes angina
oldpeak	float	Rate of ST depression caused by activity when compared to rest
slope	Integer	ST/Heart rate slope
thal	Integer	Thalassemia
target	Integer	Risk of heart disease

4.2 Correlation Board

In a machine learning project aimed at predicting heart disease, a correlation table is crucial for understanding the relationships between various features in your dataset. Here’s how you would use and interpret a correlation table.

Fig 1: Correlation heatmap

Although the feature selection process has been completed, there are still outliers present in the data shown in **Fig 2**. Box plots are a valuable tool in feature selection for machine learning projects because they provide a visual summary of a dataset's distribution, central tendency, and variability. The line inside the box represents the median. Data points outside the whiskers are considered outliers and are often plotted as individual points.

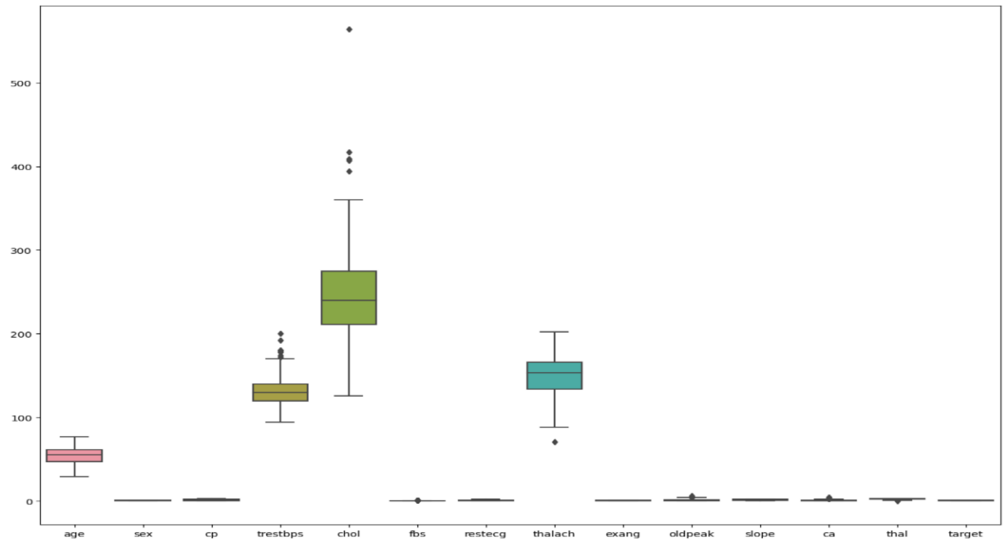


Fig 2

4.3 Data Preprocessing:

Effective data preprocessing can improve machine learning models' effectiveness, accuracy, and interpretability by providing them with high-quality input. Data

ISSN 1462 2815

preprocessing refers to the techniques and procedures applied to raw data to make it suitable for analysis and improve the performance of machine learning algorithms. Ensuring the quality of input data is a critical component of the data analysis process, as it performs an important role in generating precise and meaningful insights, as in any data-driven model.

The method SMOTE is implemented in this experiment to address and resolve the imbalance class problem. This model then introduces the synthetic examples for the minority class, making it less likely that the machine learning model will be predisposed towards the majority class. The SMOTE is a machine learning algorithm that aims to report the problem of class imbalance in datasets. It achieves this by generating synthetic examples for the minority class, thus increasing its representation in the dataset. The SMOTE algorithm works by choosing a random instance from the minority class and finding its k nearest neighbors. It then generates new occurrences along the line segments assembly of these neighbors. This process continues until the preferred balance between the minority and majority classes is accomplished. So, we apply the SMOTE technique to our model before feeding it to classification, which results in better accuracy. As we can observe, the difference in data before SMOTE is shown in Fig. 3. fig.4 represents the target class after SMOTE.

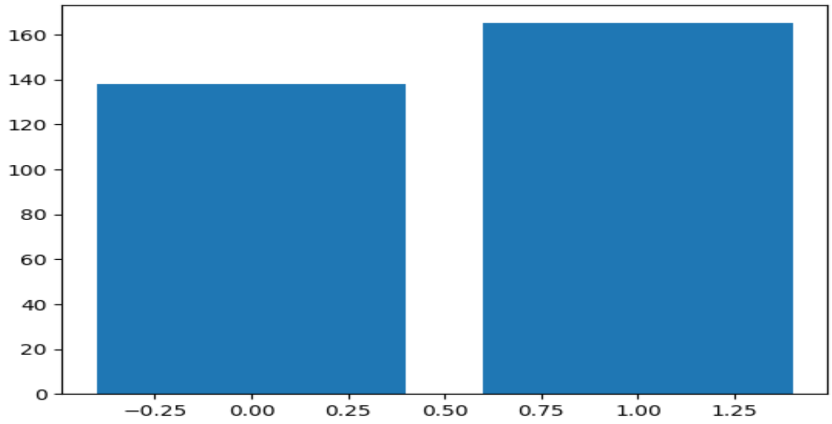


Fig 3: Target class before SMOTE

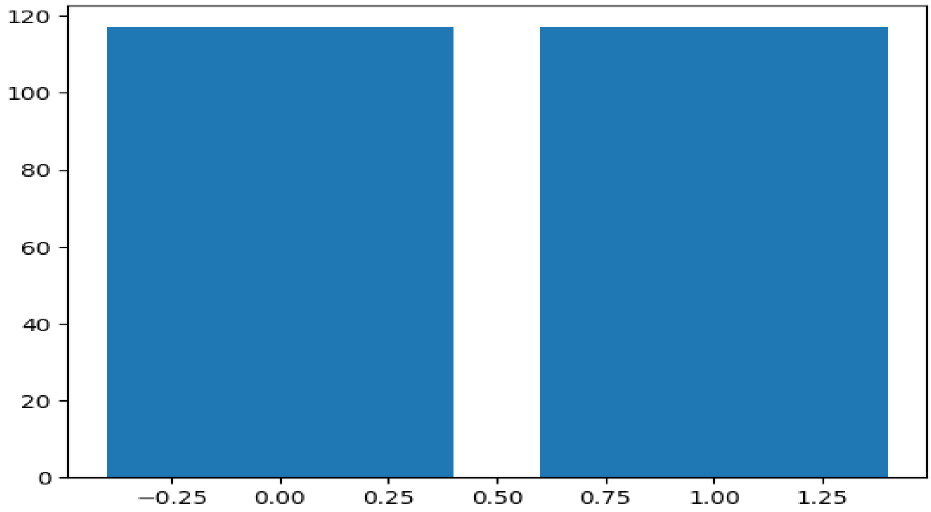


Fig 4: Target class after SMOTE

4.3.1 Feature selection

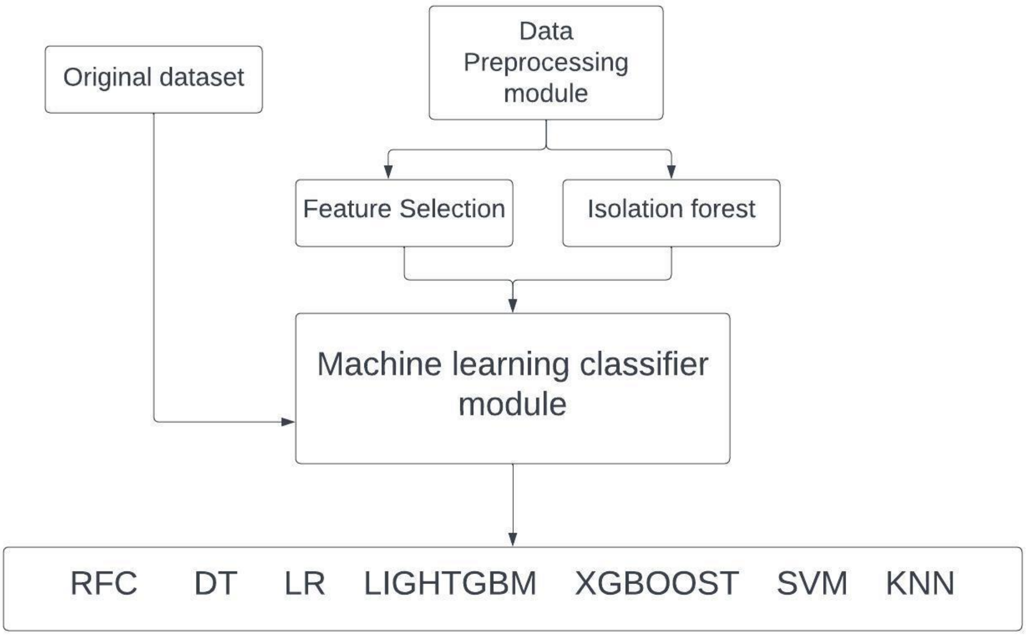
We use the XGBOOST method for effective feature selection in this proposed method. We found this methodology effective when compared to other methods. XGBOOST is an ensemble learning method that chains the predictions of several models to progress overall accuracy and generalization. It uses the gradient of the loss function to update the model in each iteration, making it better at predicting the residuals. This method provides a feature importance score for each feature in the dataset. This score indicates the contribution of each feature to the model's predictions. [21]

4.4 Proposed model using machine learning techniques:

Different classifiers like Decision Tree Classifier were used in the proposed model, and then the most popular ensemble methods, like Random Forest and Logistic Regression, were tested for better accuracy.

In order to effectively manage and analyze complex datasets with a high number of dimensions, the SVM (support vector machine) algorithm was utilized. This approach has proven to be highly effective in handling large volumes of data, making it a popular choice among experts in the field.

The XGBoost method is an effective approach that works with the ensemble and decision tree methods. The schematic diagram is shown below.



Simulation Environment and Hyper Parameters: The present experiment was carried out on a Windows 10 Pro operating system, utilizing an Intel Core i5-6300U CPU with a clock speed of 2.40 GHz. The operating system was 64-bit, and the processor had a speed of approximately 2.5 GHz. Additionally, 8 GB of RAM was installed in the system.

Table 3: Methods and hyperparameters used

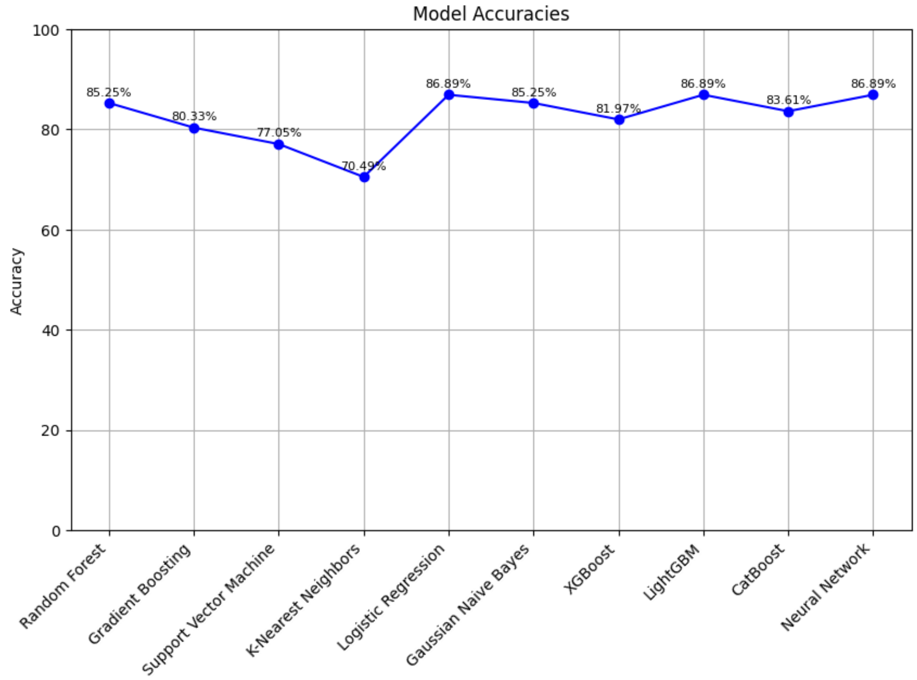
Method	Parameter
Random Forest	'criterion': 'gini', 'max_depth': 10,n_estimators': 200'min_samples_split': 10, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 1
Gradient Boosting	'ccp_alpha': 0.0, 'criterion': 'friedman_mse', 'max_depth': 5, 'n_estimators': 200,
Support Vector Machine	'cache_size': 200, 'decision_function_shape': 'ovr', 'degree': 3, 'gamma': 'scale', 'kernel': 'linear', 'max_iter': -1,
K-Nearest Neighbours	{'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 7, 'p': 2, 'weights': 'distance'}
Logistic Regression	{'C': 1, 'class_weight': None, 'dual': False, 'fit_intercept': True, 'intercept_scaling': 1, 'l1_ratio': None, 'max_iter': 100, 'multi_class': 'auto', 'n_jobs': None, 'penalty': 'l2', 'random_state': None, 'solver': 'lbfgs', 'tol': 0.0001, 'verbose': 0, 'warm_start': False}
XGBoost	{n_estimators', 'max_depth', 'learning_rate', 'subsample'}
Gaussian Naïve Bayes	{'priors': None, 'var_smoothing': 1e-09}

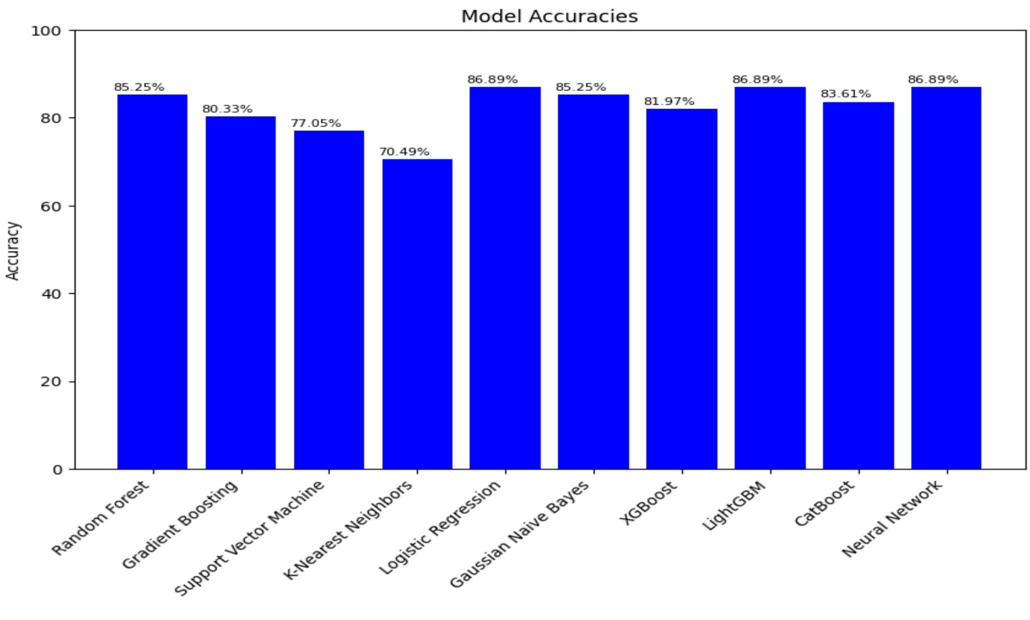
The research paper discusses the application of multiple Python programming frameworks, including sklearn, for performing data pre-processing jobs and implementing classification techniques like machine learning and ensemble learning algorithms. To analyze the data, the study utilized the Numpy and Pandas frameworks, whereas the Matplotlib and Seaborn frameworks were employed to implement functions based on data visualization. The study utilized the Pycm module to manage performance measures in multiclass classification effectively. In addition, the study addressed the class imbalance problem by using the method imblearn with random oversampling and presented diverse methodologies along with their corresponding constraints. The above table displays different methods and their best hyperparameters.

5. RESULT ANALYSIS

This process expounds on utilizing the XGBOOST ensemble technique to calculate the survival of heart failure. The proposed method is thoroughly validated by several ensemble learning methods and machine learning techniques. Upon conducting a comparative analysis of various machine learning practices, including Random Forest, Gradient Boosting, SVM, K-Nearest Neighbours, and Gaussian Naïve Bayes, it was discovered that the logistic regression method yields more accurate results. Based on the aforementioned findings, it is obvious that machine learning algorithms were able to achieve an accuracy rate ranging between 68% and 86%. In particular, Logistic Regression emerged as the most accurate algorithm with a precision rate of 86.89%, followed closely by Gaussian Naïve Bayes with 85.25%, Random Forest with 85.25%, Gradient Boosting with 80.33%, Support Vector Machine with 77.05%, and K-Nearest Neighbors with 70.49%. These outcomes demonstrate the efficiency of the algorithms, as mentioned above, in achieving high accuracy rates. It is noteworthy to mention that the precision rates attained by the machine learning algorithms may differ based on the nature and complexity of the data being analyzed. Therefore, it is recommended that further research be conducted to evaluate the performance of these algorithms across different datasets.

The diagram below shows the accuracy graph attained through the utilization methods.





6. CONCLUSION AND FUTURE SCOPE

Our paper included an exposition of six distinct methods for carrying out comparative analysis, which provided us with promising outcomes. Each method was examined in detail to appraise its effectiveness. The results demonstrated that the methods presented in the paper are effective in conducting comparative analysis. We conclude that machine learning algorithms are more effective for this type of analysis. This finding supports the suggestion made by many researchers in the past that we should use machine learning algorithms when dealing with smaller datasets. Applying data preprocessing improves KNeighbors classifier performance in the ML approach using

13 feature datasets. The duration of computing time has been condensed, which is highly advantageous for model deployment. Furthermore, it has been discovered that standardization of the dataset is necessary in order to prevent the model from becoming overfitted. Failure to standardize the dataset may result in inadequate accuracy while evaluating the model for real-world problems. Real-world data is quite different from the dataset on which the model was trained; hence, standardization is essential to avoid overfitting. Upon analyzing the findings of this study, it is clear that in order to create machine learning-based heart disease diagnosis systems for practical use, it is crucial to conduct experiments that involve real-time patient data and provide clear explanations for the final predictions using interpretable machine learning. A thorough examination of 20 relevant papers has highlighted the pressing need for such an approach. The dataset that we are currently working with is limited in size, which can pose a challenge when using deep learning and machine learning techniques. These methods typically perform better with larger datasets as they can extract more meaningful patterns and relationships. Increasing the dataset size can ominously improve the accuracy and effectiveness of the results generated by these techniques. We can take huge datasets to achieve more promising results, and deep learning can be employed with various other optimizations. Furthermore, the results of the evaluation can be improved by implementing machine learning and optimization techniques. Additionally, exploring various methods of normalizing the data could be beneficial. We can identify the most efficient approach by employing various techniques and conducting a comparative analysis of the resultant outcomes. Furthermore, we can look for ways to integrate machine learning (ML) and deep learning (DL) models trained on heart disease with multimedia to make it easier for both patients and doctors.

References

1) Ahmed, Hager, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A. Ali. "Heart disease identification from patients’ social posts, machine learning solution on Spark." Future Generation Computer Systems 111 (2020): 714-722.

2) Kameswara Rao, B., U. D. Prasan, Mokka Jagannadha Rao, Rajyalaxmi Pedada, and Pemmada Suresh Kumar. "Identification of heart failure in early stages using smote-integrated adaboost framework." In *Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021*, pp. 537-552. Singapore: Springer Nature Singapore, 2022.

3) Anggoro, Dimas Aryo, and Salsa Sasmita Mukti. "Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure." International Journal of Intelligent Engineering & Systems 14, no. 6 (2021).

4) Shorewala, Vardhan. "Early detection of coronary heart disease using ensemble techniques." *Informatics in Medicine Unlocked* 26 (2021): 100655.

5) Prabha, R., G. A. Senthil, A. Lazha, D. VijendraBabu, and Ms D. Roopa. "A novel computational rough set based feature extraction for heart disease analysis." In *I3CAC 2021:* Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC 2021, p. 371. European Alliance for Innovation, 2021.

6) Naveen, Soumyalatha, Sreya Karipody Ravindran, G. Shreya, and S. Nuzhath Ameen. "Effective Heart disease prediction framework using Random Forest and Logistic regression." In *2023 2nd International Conference on Vision Towards Emerging* Trends in Communication and Networking Technologies (ViTECoN), pp. 1-6. IEEE, 2023.

7) Dun, Boyang, Eric Wang, and Sagnik Majumder. "Heart disease diagnosis on medical data using ensemble learning." *Comput. Sci* 1, no. 1 (2016): 1-5.

ISSN 1462 2815

8) Alizadehsani, Roohallah, Moloud Abdar, Mohamad Roshanzamir, Abbas Khosravi, Parham M. Kebria, Fahime Khozeimeh, Saeid Nahavandi, Nizal Sarrafzadegan, and U. Rajendra Acharya. "Machine learning-based coronary artery disease diagnosis: A comprehensive review." Computers in biology and medicine 111 (2019): 103346.

9) Hasanova, Huru, Muhammad Tufail, Ui-Jun Baek, Jee-Tae Park, and Myung-Sup Kim. "A novel blockchain-enabled heart disease prediction mechanism using machine learning." Computers and Electrical Engineering 101 (2022): 108086.

10) Patro, Sibo Prasad, Gouri Sankar Nayak, and Neelamadhab Padhy. "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective." *Informatics in Medicine Unlocked* 26 (2021): 100696.

11) Olsen, Cameron R., Robert J. Mentz, Kevin J. Anstrom, David Page, and Priyesh A. Patel. "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure." *American Heart Journal* 229 (2020): 1-17.

12) Zulkiflee, Nor Fatihah, and Mohd Saifullah Rusiman. "Heart Disease Prediction Using Logistic Regression." *Enhanced Knowledge in Sciences and Technology* 1, no. 2 (2021): 177-184.

13) Squire, Kenneth J., Yong Zhao, Ailing Tan, Kundan Sivashanmugan, Joseph A. Kraai, Gregory L. Rorrer, and Alan X. Wang. "Photonic crystal-enhanced fluorescence imaging immunoassay for cardiovascular disease biomarker screening with machine learning analysis." *Sensors and Actuators B: Chemical* 290 (2019): 118-124.

14) Sharma, Prerna, Krishna Choudhary, Kshitij Gupta, Rahul Chawla, Deepak Gupta, and Arun Sharma. "Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning." *Artificial intelligence in medicine* 102 (2020): 101752.

15) Peirlinck, Mathias, F. Sahli Costabal, K. L. Sack, J. S. Choy, G. S. Kassab, J. M. Guccione, M. De Beule, Patrick Segers, and E. Kuhl. "Using machine learning to characterize heart failure across the scales." *Biomechanics and modeling in mechanobiology* 18, no. 6 (2019): 1987-2001.

16) Dutta, Aniruddha, Tamal Batabyal, Meheli Basu, and Scott T. Acton. "An efficient convolutional neural network for coronary heart disease prediction." *Expert Systems with Applications* 159 (2020): 113408.

17) Rajendran, Rajkamal, and Anitha Karthi. "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers." *Expert Systems with Applications* 207 (2022): 117882.

18) Ali, Farman, Shaker El-Sappagh, SM Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, and Kyung-Sup Kwak. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." *Information Fusion* 63 (2020): 208-222.

19) Kottou, Rafaela, and Veronica Pegkou-Christofi. "Wavelet Based Machine Learning Algorithms to Enhance Cardiovascular Disease Diagnosis." In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1464-1469. IEEE, 2019.

20) Stoltzfus, Jill C. "Logistic regression: a brief primer." *Academic emergency medicine* 18, no. 10 (2011): 1099-1104.

21) Bhatt, Chintan M., Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo. "Effective heart disease prediction using machine learning techniques." *Algorithms* 16, no. 2 (2023): 88.