International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Unveiling Cardiac Insights: Evaluating Machine Learning Algorithms for Accurate Heart Disease Forecasting

Simhadati Prasanna[1], Ramesh Chappa *[2]

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: Heart diseases, also referred to as cardiovascular diseases, include a diverse range of conditions that have an emotional control on the functioning of the heart. Heart failure is currently one of the world's top reasons of death. Developing a robust prediction system specific to this condition is imperative to address this challenge. Machine learning, a widely employed tool in data science, traditionally forecasts an output based on input data. In the current context, machine learning is applied to medical records to make clinical predictions regarding individual patients' illnesses. The machine learning algorithms discern patterns within the provided input data and leverage this knowledge to predict the presence of diseases using real-world data. To indicate if patients have cardiac disease, this study investigates the use of several machine learning (ML) models, such as Logistic Regression, naive Bayes, decision tree models, Random forest models, XGBoost, Support Vector Machines (SVM), and Artificial Neural Networks (ANN). We perform extensive data preprocessing and exploratory data analysis (EDA) to ensure accurate and pertinent data. We use a dataset with 303 entries and 14 attributes, including age, sex, type of CP, resting BP serum cholesterol, blood sugar during fasting, the highest heart rate achieved, exercise-induced angina, old peak, slope, number of major blood vessels colored by fluoroscopy, and thalassemia. The outcomes specify that machine learning models, predominantly Random Forest, demonstrate significant potential in predicting heart disease and exhibit superior predictive performance of 95.08% accuracy.

Keywords: *Cardiovascular Disease, Clinical Prediction, Machine Learning Models, Predictive Modelling*

1. Introduction

Since heart disease nevertheless remains one of the biggest killers in the world, early detection and prevention are crucial. According to WHO reports, almost 18 million deaths are linked to coronary artery disease cardiovascular disease (CVD) is responsible for 1 in 3 global deaths, yet most impulsive heart disease and stroke cases are escapable. In the year 2010, the rate of CVD was estimated at US$863 billion, and this figure is projected to increase by 22% to US$1,044 billion by 20302. Surprisingly, nations with low to middle incomes account for 80% of deaths from CVD.[1] This staggering number emphasizes the urgent need for increased awareness, research, and preventive measures to combat this devastating condition.[2] Traditional diagnostic methods, while effective, can be time-consuming and may not always identify individuals at risk before symptoms manifest. This is where machine learning (ML) comes into play, proposing a transformative approach to predicting heart disease with increased accuracy and efficiency.

An estimated 12 million fatalities globally occur from heart disease each year, according to the World Health Organization. It takes a life in the US every 34 seconds and is a major cause of death in many developing nations. Similarly, it is the prime reason for death in India, highlighting the significant threat that heart disease poses to adult lives globally. [3] We can reduce the death rate by predicting the chance of heart disease. However, this is not easy with the diagnosis procedure. In addition, these methods are costly too. Therefore, intending to create a prediction system to address the problem, we examine various machine-learning algorithms in this research.

Cardiovascular disease (CVD) is the most substantial health concern in Asia and a global issue. The increase in coronary heart disease with age is primarily linked to high cholesterol levels and blood pressure. It is estimated that up to two-thirds of cardiovascular disorders in the Asia-Pacific region are due to hypertension, highlighting the crucial need for blood pressure reduction measures.

Despite advancements in the availability of safe and effective prevention methods worldwide, cardiovascular disease (CVD) remains the prominent cause of premature death, responsible for six million deaths annually globally in 2019, with fifty-eight million occurring in Asia. Given Asia's vast population, diverse ethnic and cultural diversity, social status, and healthcare systems, the region faces numerous challenges in preventing and treating CVD. Addressing the CVD epidemic in Asia requires the development of cost-effective policies, strategies, and interventions, which in turn necessitates timely information on the costs and alternatives of CVD medical specialization in the region. Therefore, this analysis aimed to outline the

[1] *Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali – 532201, India*
ORCID ID: 0009-0005-1448-4039
[2] *Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali – 532201, India*
ORCID ID: 0000-0001-9829-0146
** Corresponding Author Email: chappa.ramesh01@gmail.com*

key features of the CVD pandemic and emphasize the barriers to CVD prevention and treatment in Asian countries.

Heart disease comprises various conditions, such as coronary artery disease, arrhythmias, and failure of the heart, each with its own set of risk factors. Cardiovascular diseases (CVDs) constitute a staggering 26% of all deaths in India. This emphasizes the urgent need for attention and action to address this critical health issue. Some of the risk factors related to CVDs among adults in India are [1]

- Approximately 15% of the residents use tobacco. [1]

- At 4.3 litres per person, the average pure alcohol consumption reveals an important aspect of our society's habits. [1]

- Just over 21% of the population has hypertension, increasing the risk of heart attack, heart failure, kidney disease, or stroke. [1]

Traditional menace assessment tools, such as the Framingham Risk Score, primarily rely on statistical methods and a limited number of risk factors. However, these methods may not capture the complex, nonlinear relationships between multiple risk factors and heart disease outcomes.

The fig. below shows the latest prevalence of heart disease globally. [2]

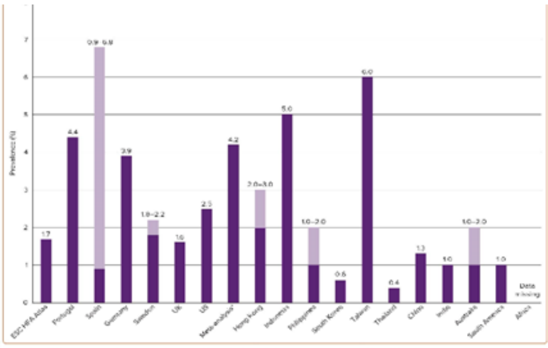


Fig 1. heart failure rate country wise statistics

Most of the data collected from the World Health Federation, states that the deaths caused by CVDs have increased by 6.5 % when compared to 1990 and this may worsen the situation. The graph below shows how the death rate increases yearly as per the World Health Report.

The regions with the greatest age-standardized rates of CVD death in both 1990 and 2019 were Central Asia, Eastern Europe, and Central Europe for both men and women. In 1990, there were 670.2 and 467.2 deaths per 100,000 people; in 2019, the figures were 524.1 and 345.7 deaths per 100,000 people. Furthermore, with 376.7 and 339.8 fatalities per 100,000 persons for men and women, respectively, the Middle East and North Africa area had the second-highest proportions for both sexual categories in 2019.

In the majority of regions, ischemic heart disease stands as the foremost cause of mortality due to cardiovascular disease (CVD) for both males and females. It's important to note, however, that in the Sub-Saharan Africa region, stroke takes precedence as the leading cause of CVD mortality among women. Similarly, in South Asia, stroke emerges as the primary cause for both males and females. As a regional average, stroke ranks as the second most prominent reason for CVD mortality across various regions. [3]

Males are more susceptible than females to high blood pressure, diabetes, alcoholism, tobacco use, physical inactivity, and other cardiovascular disease (CVD) risk factors. Stoutness is the sole risk element that is higher in women. [3]

Machine learning models, on the other hand, excel at analyzing large datasets with numerous features, uncovering hidden patterns that may not be apparent through conventional analysis.

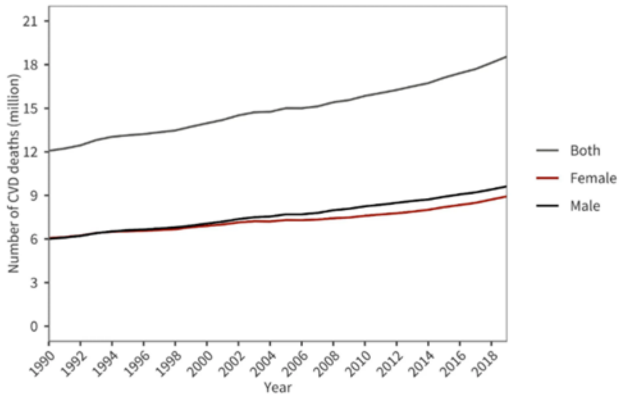


Fig 2. Death rate among adults due to heart failure

Healthcare practitioners can create prediction models that take into

account a variety of parameters by utilizing machine learning techniques. These variables may include demographic data, medical history, lifestyle factors, and even genetic information.

1.1 Machine Learning models :

When predicting heart disease using machine learning, it is essential to carefully choose the appropriate model. Considerations such as dataset size, the complexity of feature relationships, and the need for model interpretability are essential. Typically, a blend of models and techniques is utilized to ensure the most precise predictions. With the lavishness of medical data available, machine learning can support us in determining valuable patterns and information. Machine learning is typically used for disease estimation in the medical sector, but it has a wide range of applications.[5] Supervised learning models are trained on labeled data in which the desired result (target element) is known. Employing this training data, the model acquires the capability to associate the input features with the output. [4]

2. Related Work

After an intense study of data collected from different scholars, we categorized some outcomes. In this section, we elaborate on some research findings. Harshit [8] investigated many machine-learning models, including KNN, Random Forest Classifier, and Logistic Regression, to predict cardiac disease. The results showed an average accuracy rate of 87.5%. Bhanu Prakash [9] examined genetic algorithms and radial basis functions to predict coronary illness. He analyzed the dataset with 14 attributes resulting in 85.40% accuracy. Ghada [10] discussed different machine learning algorithms such as ANN, Bagging- QSVC, SVM, QNN, and VQC performed analysis using the UCI ml repository's Cleveland dataset with 303 instances and got an average of 90% accuracy. Ashok Kumar[27] addressed the logistic regression method and

some more machine learning technologies, resulting in the highest classification accuracy of 85% and the sensitivity of 89 and 81% respectively. The summary of the Literature study using different machine learning models is detailed in the table below.

Table 1. Literature Study Report

S.No	***Author***	***Data source***	***Methodologies***	***Findings***
1	Latha Parthiban[111]	Cleveland t dataset	CANFIS, Genetic Optimization	MSE is 0.000842
2	N. Aditya Sundar[12]	Cleveland database	WAC Naïve Bayes	Accuracy: 84%
3	Tanawat Tantimongcolwat[13]	MCG recordings	BNN(back propagation neural networks) DK-SOM	BNN: sensitivity 89.7%, specificity 54.5% accuracy 74.5%
4	Jyoti Soni[14]	Cleveland dataset	Naïve Bayes, Decision Tree, ANN	Accuracy NB: 86.53% DT: 89% ANN: 85.53%
5	Purushottam [15]	Dataset UCI	SVM RBF MLP PART	Accuracy 86.7%
6	SENTHILKUMAR MOHAN[16]	Cleveland dataset	Naïve Bayes LR, DL, DT, RF, SVM	Accuracy: 88%
7	Resul Das[17]	Dataset Cleveland	Neural Networks Ensemble-based Methodology	accuracy: 89.01% Sensitivity:80.95% Specificity: 0.95%
8	S Sarah[18]	Cleveland Dataset	NB, SVM, KNN, Logistic Regression, DT, and RF	LR accuracy: 85.25%
9	Davide Chicco[19]	299 records dataset Hospital Faisalabad	Random forests, Decision tree, Gradient Boosting, Linear Regression, ANN, Naïve Bayes, SVM, KNN	RF Accuracy:74%
10	Rati Goel [20]	Kaggle dataset	Logistic Regression, KNN,SVM, NaïveBayes, DT, Random Forest	SVM accuracy: 86%

3. Data PreProcessing

Data preprocessing is a fundamental step in constructing effective ML model, especially in the perspective of predicting heart disease entails converting raw data into an organized, structured manner suited for analysis and modeling. The dataset lacks any null values; however, there is a substantial need for properly handling numerous outliers. Additionally, the dataset exhibits improper distribution.

3.1 Dataset Description

We obtained this dataset from Kaggle (The cardiovascular system Disease dataset). There are 303 items in the dataset overall, 165 of them are male patients and 138 are female patients of various ages. Of these, 45.54% of patients have cardiac issues, while 54.46% do not. When a patient has cardiovascular disease, the target class "target" equals 1, when the patient is healthy, it equals 0. Let us provide a detailed description of our dataset in the table below.

Table 2. List of Features

S.NO	FEATURE	DATATYPE
1	Age	Int
2	Sex	Int
3	CP	Int
4	Trestbps	Int
5	Chol	Int
6	Fbs	Int
7	Restecg	Int
8	Thalch	Int
9	Exang	Int
10	Oldpeak	float
11	Slope	Int
12	Ca	Int
13	Thal	Int
14	Target	float

Let us describe the target value in detail based on the given dataset: The dataset for heart disease prediction comprises several medical attributes critical for diagnosing the condition. It includes demographic information such as Age and Sex, where Sex is encoded as 1 for males and 0 for females. Clinical features are well-represented with CP (chest pain type), categorized into four types ranging from typical angina to asymptomatic pain, Trestbps (resting blood pressure), and Chol (serum cholesterol level in mg/dl).). Whether fasting blood sugar levels are higher than 120 mg/dl is indicated by the Fbs characteristic. Restecg is a three-valued term that represents the outcomes of resting electrocardiograms (0, 1, 2). Exercise-induced angina is indicated by Exang, while Thalch records the highest heart rate reached. The Oldpeak feature calculates the difference between exercise-induced ST depression and rest. The number of main vessels (0–3) colored by fluoroscopy is counted by Ca, while the slope characterizes the slope of the peak workout ST segment. Finally, Thal evaluates the outcomes of the thallium stress test by assigning values to normalcy, fixed defect, and reversible defect. The goal variable is a binary indicator that shows whether cardiac disease is present (1) or not (0). This dataset offers an extensive feature set for developing cardiac prediction algorithms. The mean value indicates that approximately 54.46% of the entries have a value of 1. The first quartile value, denoted as 0.000000, suggests that 25% of the entries are 0. The median value, denoted as 1.000000, shows that close to 50% of the entries have a value of 1. The maximum value in this feature is 1, as expected for a binary feature.

3.2 Establishing column correlation for feature selection:

Understanding the relationships among features (columns) in the dataset is crucial in machine learning projects. One way to examine these relationships is by checking the correlation between columns. The direction and intensity of a linear relationship between two variables are ascertained using correlation. Correlation helps in Feature Selection, Data Understanding, and Dimensionality Reduction along with Model Performance. Strong relationships are indicated by a correlation coefficient near 1 or -1, whereas weak relationships are indicated by a number near 0. The target variable has a perfect correlation with itself (1.000). Among the features, exercise-induced angina (exang) shows a moderate positive correlation of 0.437 with heart disease, indicating that its presence is moderately associated with heart disease. Chest pain type (cp) also has a notable correlation of 0.434, suggesting that specific types of chest pain are linked to heart disease. Other significant correlations include oldpeak (0.431), maximum heart rate achieved (thalach) (0.422), and the number of major vessels colored by fluoroscopy (ca) (0.392). The slope of the peak exercise ST segment (slope) and thallium stress test results (thal) have moderate correlations of 0.346 and 0.344, respectively. Sex (0.281) and age (0.225) show weaker, yet relevant, correlations. Resting blood pressure (trestbps) and resting electrocardiographic results (restecg) have lower correlations (0.145 and 0.137, respectively), while serum cholesterol level (chol) and fasting blood sugar (fbs) have the weakest correlations (0.085 and 0.028), indicating that these features are less predictive of heart failure in this dataset.

3.3 Exploratory Data Analysis:

EDA is a crucial phase in a Machine learning model. Before creating models, it helps guarantee that the data is thoroughly comprehended and prepared. The first step towards EDA would be analyzing our target value. The graph below represents the ratio of patients suffering from heart disease and those without it.

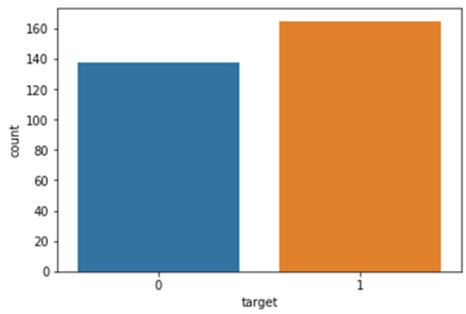


Fig 3. Analysing “target” feature

After a keen analysis, we conclude that the dataset contains 54.46%(165 patients) suffering

from heart disease whereas the remaining 45.45%(138 pati ents) do not.

Now, let us analyse the different features that have a major correlation with the target value.

The analysis with different features is presented below, and the impact is represented in graphical format.

3.3.1 Analysing the sex feature:

When analyzing the sex feature in a heart disease dataset, it is important to examine its distribution, its relationship with the target variable (heart disease presence or absence), and i ts interaction with other features. Here we visualize the feat ure using a bar plot.

Through the analysis of the 'sex' feature, we have discovere d that cardiac diseases are more common in women than in men.

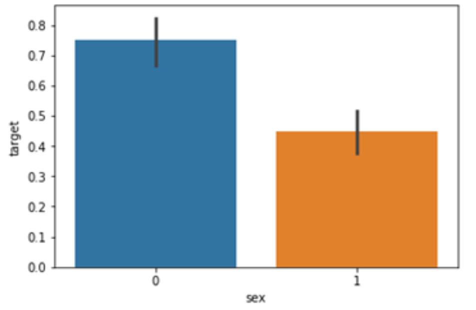


Fig 4. Analysing “sex” feature

3.3.2 Analyzing the 'Chest Pain Type' feature:

The cp (chest pain type) feature plays a weighty role in analyzing heart disease, as different types of chest pain are associated with varying risks and underlying conditions related to heart disease.

CP features have values that range between 0 and 3, as mentioned.

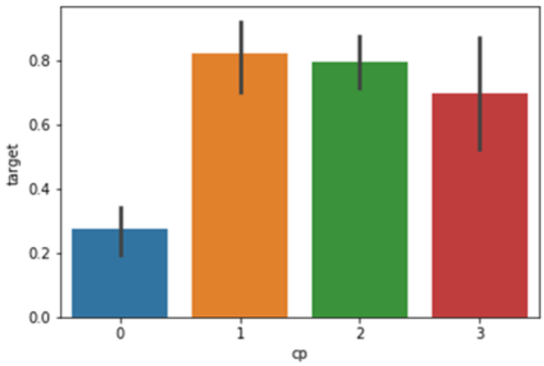


Fig 5. Analysing “CP” feature

We find that those who score '0' for chest discomfort, which is indicative of normal angina, have a significantly lower risk of cardiac issues

3.3.3 Analyzing FBS feature:

By examining the rapport between fasting blood sugar levels and the presence of heart disease, you can identify if elevated FBS levels are connected with a higher risk of heart disease.

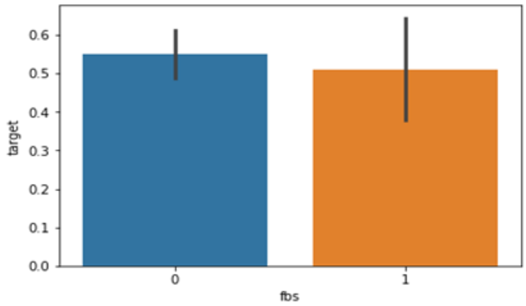


Fig 6. Analysing “fbs” feature

As the FBS feature has the least correlation with the target value, we can observe that it has no significant impact.

3.3.4 Analysing the 'exang' feature

Examine how exercise-induced angina (exang) correlates with the occurance of heart disease (target).

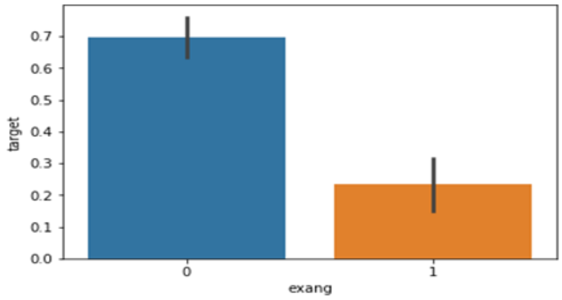


Fig 7. Analysing “exang” feature

Heart issues are far less common in those with exang=1.

3.3.5 Analysing the Slope feature

Here, we investigate patterns and correlations between the slope of the ST segment and other clinical variables.

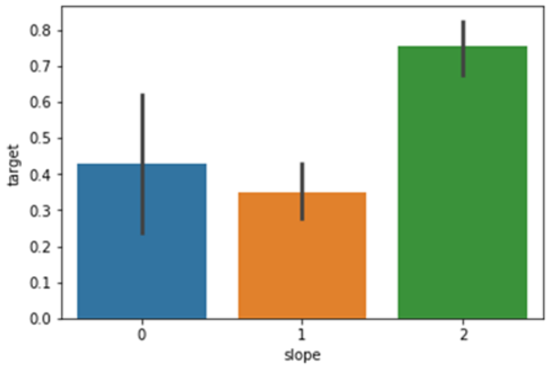


Fig 8. Analysing the “Slope” feature

We find that compared to Slopes 0 and 1, Slope 2 considerably worsens heart discomfort.

3.3.6 Analysing the 'ca' feature

Here we examine how the number of major vessels colored by fluoroscopy (ca) correlates with the existence of heart disease (target). The ‘ca’ value ranges between 0 to 4 whereas the detailed description is shown in the above table 2.

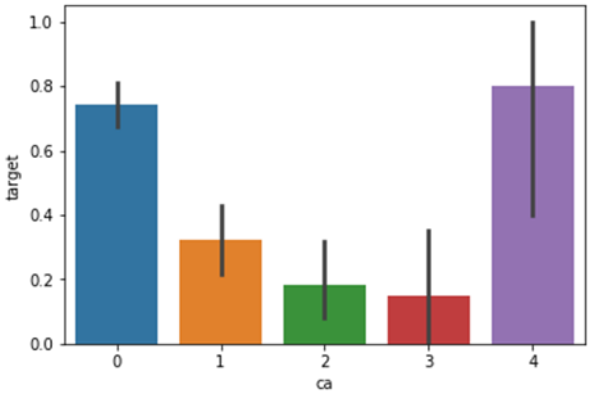


Fig 9. Analysing “ca” feature

This graphical depiction demonstrates the high number of heart disease patients in the ca=4 population.

3.3.7 Analysing the ‘thal’ feature

The scattering of the 'thal' feature in the dataset provides insights into the prevalence of different types of thalassemia. A distribution plot representing the 'thal' feature is included below for reference.

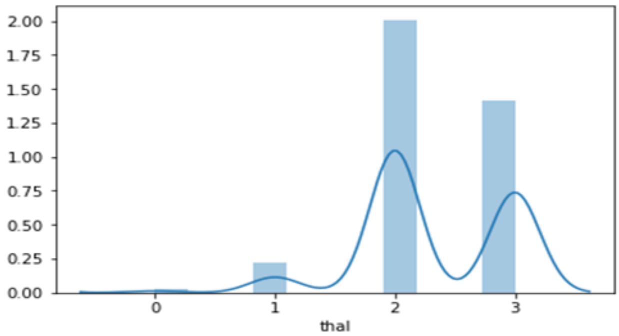


Fig 10. Barplot for “thal” feature

3.4 Modelling

The primary dataset is used to create a 20% training dataset (242 values) and a 20% testing dataset (61 values). Next, the training dataset is used to train the model, and the testing dataset is used to assess the model's performance. We use a variety of classifiers to thoroughly evaluate their performance on the clustered dataset, including the random forest classifier, XGBoost, and decision tree classifier.[21]

3.5 Proposed Methodologies

We are in the process of developing an end-to-end analytical model to envisage the probability of CVD. Our approach involves using a total of 303 entries with 14 attributes to diagnose and assess the risk of cardiovascular issues in patients. We are aiming to create a robust and scalable solution that effectively addresses real-world scenarios in data science research

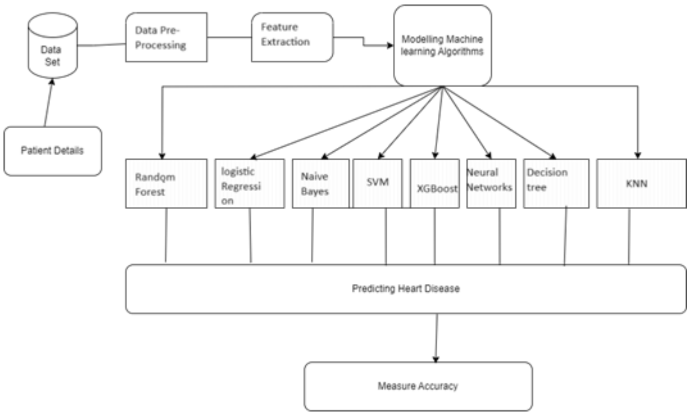


fig 11. Architecture for proposed model

a. Logistic Regression

Logistic regression estimates the likelihood of an event based on given data. It deals with binary data (1, 0), 1 for the occurrence of an event, and 0 for non-occurrence.[22] [Logistic regression can be used to simulate the chance that a certain patient has heart disease based on different predictor variables in the context of heart disease prediction

$$\mathbf{P(y = 1 \mid X)} = 1/\mathbf{1 + e - (\beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn)} \quad (1)$$

Formula (1) evaluates the probability that target variable y is 1 given the feature vector X.

The accuracy score of 85.25% in this study was attained by the use of logistic regression.

b. Naïve Bayes

Using Bayes' Theorem as its foundation, Naïve Bayes is a simple yet powerful categorization algorithm. It makes the assumption that predictors are independent of one another, which means that characteristics or attributes should not be connected. Even if there is a dependency, all these features or attributes independently contribute to the probability, hence the name "naïve."[23] The key idea is to use Bayes' theorem to compute the posterior probability of each class given a set of features, and then to predict the class with the highest posterior probability.

$$\mathbf{P(M \mid N) = P(N \mid M) \cdot P(M)}/\mathrm{P(N)} \qquad (2)$$

Formula (2) results in the posterior probability, which is the probability of class M (e.g., presence of heart disease) give n the features N (e.g., medical attributes).

Bayes' Theorem allows us to revise our initial beliefs (prior s) in light of new evidence (likelihood), providing a mathe matical framework for updating the probability of a hypoth esis based on observed data.

While the Naïve Bayes algorithm attained an accuracy of 8 5.25%, there is room for improvement considering the Ran dom Forest algorithm's impressive accuracy score of 95.08 %.

c. Support Vector Machine

Support Vector Machines (SVM) are a powerful set of supe rvised learning algorithms used for classification, regressio n, and outlier detection tasks. SVMs are effective in high-d imensional spaces and are particularly useful when the num ber of dimensions exceeds the number of samples. They als o can handle non-linear classification. By focusing on these critical data points, SVM achieves robust generalization to new, unseen data. In the current research, the accuracy achi eved using Linear SVM is 81.97%.

d. K Nearest Neighbors

The K-nearest neighbor algorithm is a powerful supervised algorithm used to predict a class in a sample. It confidently classifies results by using the majority of k-NN categories. The primary objective is to accurately predict the number o f individuals who will undergo heart disease in the upcomin g year using the K-Nearest Neighbor (KNN) method. This e ffective algorithm works based on the nearby distance betw een entities to accurately classify the data. [24] I understan d that the accuracy of the model using KNN was 67.21%.

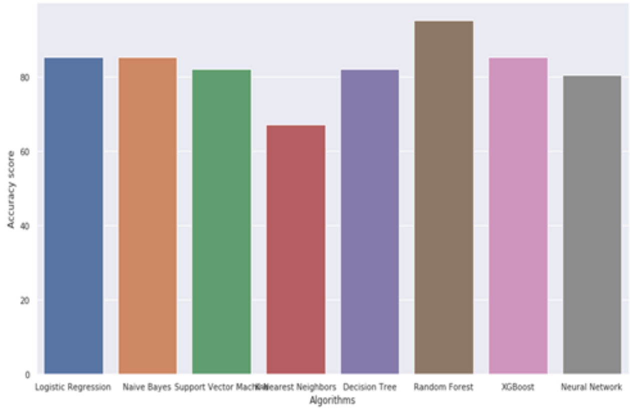
e. Decision Tree

Machine learning tasks including regression and classificati on are performed with decision trees. In essence, these algo rithms create a tree-like structure by repeatedly segmenting the data into subsets according to the most significant attrib ute at each node. Decision trees are particularly useful in th e prediction of heart disease because of their interpretabilit y and adaptability when handling both categorical and num erical data. The Decision Tree approach yielded an accurac y score of 81.97%.

f. Random Forest

Random Forest is an ensemble learning algorithm widely u sed for classification and regression tasks. It operates by co nstructing multiple decision trees during training and mergi ng their results to improve accuracy and control overfitting. Each tree in the forest is built using a random subset of the training data and features, which introduces diversity amon g the trees. When a prediction is needed, each tree in the fo rest provides its output, and the algorithm combines these r esults—through majority voting for classification or averag ing for regression—to produce a final prediction. This proc ess enhances the model's robustness and generalization abil ity, making Random Forest particularly effective for large d atasets with numerous features and for handling missing va lues and noisy data. The Random Forest method was origin ally proposed by Tin Kam Ho of Bell Labs in 1995, and its popularity continues to grow due to its robustness and vers atility in handling complex classification problems.[25]

Algorithm for Random Forest:



Step 1: Bootstrap Sampling: Creates different training sets by randomly sampling with replacement.

Step 2: Train Decision Tree: Build a decision tree for each bootstrap sample by:

- *Randomly choosing a subset of features at each node.*
- *Finding the best split for the selected features.*
- *Recursively splitting nodes until stopping conditions are met.*

Step 3: Make Predictions: Aggregate predictions from all trees for a new data sample using majority voting (classification) or averaging (regression).

effectiveness in addressing classification and regression problems due to its high performance and scalability. It employs an ensemble technique, sequentially building models that correct errors made by the preceding ones. With built-in regularization terms to prevent overfitting and the capability to handle missing values internally, XGBoost proves to be a powerful tool for heart disease prediction, proposing high accuracy and robustness through advanced boosting techniques and regularization. Notably, in a recent study, XGBoost exhibited an impressive accuracy score of 85.25%.

f. Neural Networks

A potent machine learning method that draws motivation from the composition and operations of the human brain is neural networks. They work especially well with intricate data patterns and linkages. This research study establishes a decision support system for predicting a patient's heart disease. A database of previous heart disease instances is used to make the forecast. Thirteen input attributes and medical terms like blood pressure, cholesterol, and sex are used by the algorithm. [26] On the dataset, the neural network model yielded an accuracy of 80.33%.

4. Evaluating Results

In this study, a dataset of 13 features is used to create prediction models, and the accuracy of the modeling approaches is determined. All experiments were carried out on a Core I5 with a 2.4GHz CPU and 16GB RAM. An extensive range of machine learning and ensemble learning methodologies thoroughly validate the suggested method. The best output was determined by evaluating methodologies such as RF, K Nearest Neighbors, SVM, XGBoost, Neural Networks, Logistic Regression, and Decision Tree. After training and testing machine learning algorithms, we got 95.02% accuracy with Random Forest. This algorithm outperforms others for this problem. Other machine learning algorithms such as Logistic Regression, Naïve Bayes, and XGBoost, performed well, with average accuracy rates of 85.25%. By comparing all the techniques mentioned above, we conclude that the Random Forest algorithm worked better on the dataset. The Fig below provides a pictorial illustration of the accuracy of the various comparing models and the proposed technique.

5. Conclusion and Future Scope

Machine learning (ML)--based heart disease diagnosis has significant potential for early identification and intervention, which could save countless lives. The use of machine learning models in the healthcare industry makes it possible to analyze large, intricate data sets and find trends and risk factors related to heart disease. Through the utilization of diverse algorithms, ranging from logistic regression to more sophisticated approaches like neural networks and ensemble methods, we can create resilient predictive models that support medical professionals in making well-informed judgments. To optimize their impact and reliability, it is imperative to guarantee that these models are interpretable, verified across a range of populations, and seamlessly incorporated into clinical processes. We can make far more predictions by gathering comprehensive datasets from electronic health records (EHRs), patient histories, and medical examinations. Intending to keep the model current and accurate throughout time, we update it frequently with fresh data. Work together with epidemiologists, data scientists, and cardiovascular specialists to continuously improve the model and add the most recent findings from studies. using medical images (e.g., echocardiograms, ECG) to improve the predictive ability of the model. To improve the study, we can look into other variables including dietary patterns, stress levels, and socioeconomic circumstances. Enhancement can be achieved by giving healthcare professionals continual education and training so they can apply the model efficiently. Provide patient education materials that highlight the advantages and restrictions of the model to build their confidence and involvement. The cardiac disease prediction model has the potential to improve patient outcomes and advance preventive cardiology by expanding its accuracy, personalization, and impact by tackling three future scope areas. By addressing these future scope areas, the heart disease prediction model can evolve to become more accurate, personalized, and impactful in improving patient outcomes and advancing preventive cardiology. The utilization of supervised machine learning techniques has facilitated the detection of intricate connections yet latent patterns among diverse risk indicators, culminating in the development of increasingly resilient and dependable prediction models. Furthermore, the development of unsupervised learning techniques, like as deep learning, has opened up new possibilities for the detection and prognosis of cardiovascular illness. [28] Even while research is moving in an optimistic direction right now, there is still a lot of room for improvement.

Author contributions

Prasanna S: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation, Field study

Ramesh Ch.: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Institute for Health Metrics and Evaluation (IHME). The Global Burden of Diseases, Injuries and Risk Factors Study 2010 (GBD 2010). Generating Evidence, Guiding Policy Report, World Economic Forum, The Global Economic Burden of Non-communicable Diseases, Harvard School of Public Health, 2011.

[2] Cardiac Failure Review 2023; 9:e11. https://doi.org/10.15420/cfr.2023.05 .

[3] https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf

[4] 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

[5] J. Soni et.al.,“Intelligent and effective heart disease prediction system using weighted associative classifiers,” International Journal on Computer Science and Engineering, vol. 3, no. 6, pp. 2385–2392, 2011

[6] Aljanabi, Maryam & Qutqut, Mahmoud & Hijjawi, Mohammad. (2018). Machine Learning Classification Techniques for Heart Disease Prediction: A Review. International Journal of Engineering and Technology. 7. 5373-5379. 10.14419/ijet.v7i4.28646.

[7] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 2017, pp. 204-207, doi: 10.1109/ISCC.2017.8024530.

[8] Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072" heart disease prediction using machine learning algorithms" DOI 10.1088/1757-899X/1022/1/012072

[9] Doppala, B.P., Bhattacharyya, D., Chakkravarthy, M. et al. A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset. Distrib Parallel Databases 41, 1–20 (2023). https://doi.org/10.1007/s10619-021-07329-y

[10] Ghada Abdulsalam , Souham Meshoul, and Hadil Shaiba , "Explainable Heart Disease Prediction Using Ensemble-Quantum Machine Learning Approach", DOI: 10.32604/iasc.2023.032262

[11] Latha Parthiban and R.Subramanian "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm"

[12] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra,performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.

[13] Tantimongcolwat T, Naenna T, Isarankura-Na-Ayudhya C, Embrechts MJ, Prachayasittikul V. Identification of ischemic heart disease via machine learning analysis on magnetocardiograms. Comput Biol Med. 2008 Jul;38(7):817-25. doi: 10.1016/j.compbiomed.2008.04.009. Epub 2008 Jun 11. PMID: 18550044.

[14] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction"

[15] Purushottam et al. / Procedia Computer Science 85 (2016) 962 – 969," Efficient Heart Disease Prediction System", doi: 10.1016/j.procs.2016.05.288

[16] S.Mohan et al.: Effective Heart Disease Prediction Using Hybrid ML Technique, doi: 10.1109/ACCESS.2019.2923707

[17] Resul Das , Ibrahim Turkoglu , Abdulkadir Sengur "Effective diagnosis of heart disease through neural networks ensembles" https://doi.org/10.1016/j.eswa.2008.09.013

[18] Sarah, S., Gourisaria, M. ., Khare, S., Das, H. (2022). Heart Disease Prediction Using Core Machine Learning Techniques—A Comparative Study. In: Tiwari, S., Trivedi, M.C., Kolhe, M.L., Mishra, K., Singh, B.K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 318. Springer, Singapore. https://doi.org/10.1007/978-981-16-5689-7_22

[19] D. Chicco, G. JurmanMachine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone BMC Med. Inf. Decis. Making, 20 (1) (2020), pp. 1-16

[20] Goel, Rati, Heart Disease Prediction Using Various Algorithms of Machine Learning (July 12, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, http://dx.doi.org/10.2139/ssrn.3884968

[21] Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms 2023, 16, 88. https://doi.org/10.3390/a16020088

[22] Hidayatullah Arghandabi , Parvaneh Shams, A Comparative Study of Machine Learning Algorithms for the Prediction of Heart Disease.

[23] Digvijay Kumar Int J Sci Res CSE & IT, September-October-2020; 6 (5) : 46-54 Cardiovascular Disease Prediction Using Machine Learning doi: https://doi.org/10.32628/CSEIT20659

[24] Dimas Aryo Anggoro , Naqshauliza Devi Kurnia, Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease, https://doi.org/10.30534/ijeter/2020/32852020

[25] M.A.Jabbar, B.L.Deekshatulu, & Priti Chandra. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. Journal of Network and Innovative Computing, 4, 10. https://cspub-jnic.org/index.php/jnic/article/view/118

[26] Dangare, Chaitrali and Apte, Sulabha, A Data Mining Approach for Prediction of Heart Disease Using Neural Networks (November 14, 2012). International Journal of Computer Engineering and Technology (IJCET),Volume-3,Issue-3,October-December-2012.,Available at SSRN: https://ssrn.com/abstract=2175569

[27] Dwivedi, A.K. Performance evaluation of different

machine learning techniques for prediction of heart disease. Neural Comput & Applic 29, 685–693 (2018). https://doi.org/10.1007/s00521-016-2604-1

[28] Johnson, Soto, Glicksberg, Shameer, Miotto, Ali, Ashley, DudleyJournal of the American College of Cardiology, 2018Volume 71, Issue 23, Pages 2668-2679.