Predicting Heart Attacks with Machine Learning Algorithms

V Karthick, Associate Professor
Department of CSE
Rajalakshmi Engineering College
Chennai, India
vkarthick86@gmail.com

Nandha Krishna, UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701518@rajalakshmi.edu.in

Vishal B, UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701312@rajalakshmi.edu.in

Sri Prasath S,UG Student
Department of CSE
Rajalakshmi Engineering College Chennai,India
210701261@rajalakshmi.edu.in

ABSTRACT:

The heart, a vital organ in humans, plays a crucial role in purifying and circulating blood throughout the body. Heart attacks, a leading cause of death worldwide, manifest through symptoms like chest pain, rapid heartbeat, and breathing difficulties. This study presents a comprehensive examination of heart attacks and contemporary prediction techniques. It provides an overview of different machine learning methodologies employed for heart attack prediction, including Decision Tree, Logistic Regression, Support Vector Machines (SVM), Naive Bayes, Random Forest, K-Nearest Neighbours (KNN), and XGBoost Classifier. The comparison of these algorithms is conducted based on their respective features, offering insights into their effectiveness. The purpose of this study is to support ongoing efforts to create precise and effective models of prediction for cardiac diagnosis and prevention.

KEYTERMS: Random Forest, XGBoost, Naive Bayes, Decision Tree, KNN, Logistic Regression, SVM, Machine Learning, and predictive Cardiac Attacks.

1.Introduction:

The heart, being a vital organ, demands meticulous care to ensure overall well-being. Given its association with

well-being. Given its association with various diseases, including heart attacks, predictive studies in this domain are crucial. Currently, a significant number of individuals succumb to heart attacks, often diagnosed at advanced stages due to the lack of precise predictive tools.
Healthcare industries grapple with the challenge of early heart attack prediction owing to the complexity and variability of health data. Researchers endeavour to develop prototypes capable of early detection, encountering both benefits and limitations in their approaches. Machine learning systems, leveraging data processing and utilization, offer promising avenues for predictive analytics. By harnessing natural constraints like cholesterol levels, blood pressure, age, and gender, various algorithms such as Decision Trees, Random Forest, Naive Bayes, K-Nearest Neighbours, Support Vector Machines (SVM), and Logistic Regression, and XGBoost are compared for their predictive accuracy.

This research attains approximately 80% accuracy on the testing set during evaluation, although the translation of data into practical use remains time-consuming. To address the challenge of accuracy and efficiency, the Random Forest method emerges as a preferred choice, facilitating more precise outcomes in less time.

2.LITERATURE REVIEW:

Numerous studies conducted in medical research focus on developing heart attack prediction systems utilizing a range of algorithms for machine learning.

[1]This review paper discusses various machine learning algorithms used for predicting heart disease, including logistic regression, decision trees, support vector machines, neural networks, and ensemble methods. It highlights the strengths and weaknesses of each approach and the importance of feature selection and data preprocessing in enhancing model performance by Dilsizian, S. E., Siegel, E.

[2]This study compares several machine learning algorithms, such as Naive Bayes, decision trees, and k-nearest neighbors (KNN), in predicting heart disease. The results show that decision trees and KNN provide high accuracy, demonstrating the potential of these algorithms in clinical settings by Asri, H., Mousannif, H., Al Moatassime, H., Noel, T.

[3]The paper presents a hybrid model combining multiple machine learning algorithms to improve prediction accuracy. The study emphasizes the use of feature engineering and hybrid approaches, integrating methods like Random Forest and Gradient Boosting, to acHaq, A. U., Li, J. P., Memon, M. H., Nazir, S., Sun, R.hieve.

[4]This review explores the application of deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for predicting cardiovascular risk. By Raghupathi, W., Raghupathi, V.

[5] This research applies machine learning techniques such as logistic regression, support vector machines (SVM), and artificial neural networks (ANN) to predict heart disease by Yadav, S. S., Jadhav, S. M., support vector machines, and decision trees.

[7]This comprehensive survey provides an overview of various machine learning algorithms used for heart disease prediction, including ensemble methods like bagging and boosting. The study also discusses the role of big data and the importance of data quality in the predictive modeling process by Al-Madi, N.

[8]This paper investigates the use of genetic algorithms for feature selection to enhance the performance of machine learning models in heart disease prediction by Sudhakar, K.

[6]This case study focuses on the use of explainable AI techniques such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) to interpret the predictions of machine learning models for heart disease by Ribeiro, M

.[9] The review analyzes studies that utilize logistic regression, decision trees, random forests, neural networks, and deep learning models. It also explores the role of feature engineering and the impact of various features on the predictive performance of models by Dey,A.

[10]This survey provides a detailed overview of various machine learning and data mining techniques applied to heart disease prediction.It covers a wide range of

algorithms including Decision Trees, Support Vector Machines (SVM), Neural Networks, K-Nearest Neighbors (KNN), and ensemble methods by Khan, M.

3.PROPOSED SYSTEM:

The primary objective of the proposed system is to employ diverse machine learning methodologies for predicting heart attacks.

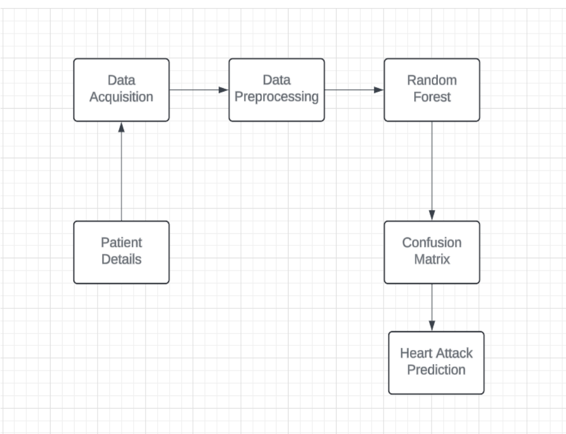


Fig 1. Methodology of proposed system.

3.1Data Acquisition:

Acquiring data involves recording real-world physical scenarios and transforming them into numerical values suitable for computer processing.

3.2Data Pre-Processing:

One of the most important steps in developing a machine learning model is data preprocessing, which includes obtaining and preparing raw data for analysis. However, in many projects, access to clean and prepared data is not readily available. Consequently, before any data manipulation can occur, it is imperative to clean the data by removing unwanted entries. This ensures that everyone can utilize preprocessing services effectively.

3.4 Model Stacking:

Model stacking involves combining multiple regression and classification models into a two-layer estimator system. Baseline models predict outcomes on test datasets, which are then used as input for meta classifiers in the second layer.

3.5 Logistic Regression:

According to the variables that are independent in a dataset, logistic regression is a statistical model for predictive evaluation and categorization that calculates the probability of an event happening. The logit function is used to convert odds and predict possibilities that range from zero to one.

The logistic function takes the form:

p(x)={1/1+e^{-(x-mu)/s}

where s represents a scale parameter and μ denotes a location parameter

3.5K-NearestNeighbour Classifier:

Using the supervised learning technique K-Nearest Neighbour (KNN), new data is categorized according to how similar it is to preexisting data points. To calculate the separations between data points, it applies the Euclidean distance formula.

The interval between the data points is determined using the Euclidean distance formula.

$$\text{A and B} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

3.6 Decision Tree:

A decision tree is a graphical model that shows potential outcomes depending on a number of different parameters. In order to create effective machine learning models, raw data must be cleaned and prepared. This process is known as data preliminary processing.

$$\text{Entropy: } H(S) = -\sum^{n} pi\,(S) * log2pi\,(S)$$

3.7 Naive Bayes:

Based on Bayes' theorem, Naive Bayes is a straightforward but effective classification technique. It's frequently employed for classification duties in a variety of industries.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

3.8 Support Vector Machine (SVM):

SVM seeks to locate the best hyperplane in space of multiple dimensions for efficiently classifying data points.

3.9 Random Forest:

Using ensemble learning principles, random forest is a supervised learning technique that may be applied to both regression and classification tasks. It increases prediction accuracy by producing data subsets for decision trees.

The suggested method predicts if a person is having a heart attack based on a dataset that includes parameters including the type of chest discomfort, resting plasma pressure, serum cholesterol, fasting plasma sugar, and ECG readings. Every characteristic adds to the thorough examination of heart function, facilitating precise forecasting.

.

3.10 XGBOOST:

Importantly, exang (0 for No, 1 for Yes) represents angina generated by exercise and is used in heart attack models for prediction.

One very flexible and very accurate gradient boosting technique is XGBoost. It prioritizes performance and quickness over computing capacity for boosted tree algorithms.

The ST depression brought on by exercise as opposed to rest is indicated by the old peak. This feature offers important information about the degree of heart stress during exercise.

Performance Evaluation: One of the most crucial aspects of the procedure for machine learning is the assessment of performance, which must be carried out meticulously. Three primary subtasks are evaluated: performance measurement, data resampling, and outcomes data with statistical significance

Cardiac Attack Forecast

After the previously mentioned processes are finished, the project generates predictions based on input data, which allows it to anticipate the accuracy score for a certain dataset and decide whether or not a patient needs to be confirmed to have a heart attack.

4.RESULT ANALYSIS

This project's main goal is to estimate the probability that a person will have a heart attack and then offer practical advice based on the estimate. In this situation, the Random Forest algorithm has proven to have the ability to attain high accuracy rates is 90.16.

AGE OF PATIENT		50	72	55	32
CP OF PATIENT		3	4	1	3
TRESTBPS OF PATIENT		110	162	153	171
CHOL OF PATIENT		222	217	277	261
FBS OF PATIENT		1	0	1	0
THALACH OF PATIENT		159	187	154	111
EXANG OF PATIENT		1	0	1	0
OLD PEAK OF PATIENT		0.4	1.2	2.1	1.5
THAL OF PATIENT		2	1	2	1
TARGET VALUE		1	0	1	0

Table 1. Table of data

The data provided in Table-1 offers sufficient information for predicting the likelihood of a person experiencing a heart attack. Each feature within the dataset reflects various aspects of cardiac function.

For example, 'cp' indicates the kind of chest pain, which can be classified into four categories: Asymptomatic, Non-anginal pain, uncharacteristic angina, and Characteristic angina. These characteristics, listed in Table-1, are important markers for estimating the likelihood that a heart attack may occur and offer insightful information on cardiac functions.

• trestbps - This parameter indicates the level of plasma pressure during the relaxation phase.

• chol - Serum cholesterol level measured in milligrams per decilitre (mg/dL).

• fbs - Fasting plasma sugar levels, where a value greater than 120mg/dL is denoted as 1, otherwise as 0.

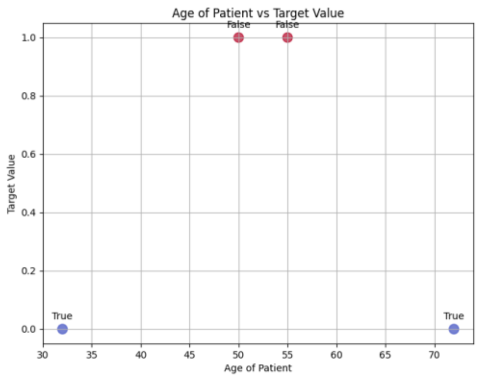
• restingecg - Results obtained from an electrocardiogram conducted while the patient is at rest.

• exang - Indicates whether angina was induced by exercise, where 0 signifies "No" and 1 signifies "Yes".

• old peak - Represents the ST depression induced by exercise in comparison to the resting state.

AGE OF PATIENT	50	72	55	32
CP OF PATIENT	3	4	1	3
TRESTBPS OF PATIENT	110	162	153	171
CHOL OF PATIENT	222	217	277	261

FBS OF PATIENT	1	0	1	0
RESTECG OF PATIENT	178	163	174	114
THALACH OF PATIENT	159	187	154	111
EXANG OF PATIENT	1	0	1	0
OLD PEAK OF PATIENT	0.4	1.2	2.1	1.5
SLOPE VALUE	1	64	43	69
THAL OF PATIENT	2	1	2	1
TARGET VALUE	1	0	1	0
RESULT OF HEART ATTTACK	False	True	False	True



Algorithms	**Accuracy**
Logistic Regression	90.5
Naïve Bayes	80.2
SVM	87.3
KNN	67.21
Decision Tree	81.97
Random Forest	90.16
XGBoost	78.69

Table-2: Table of data with results.

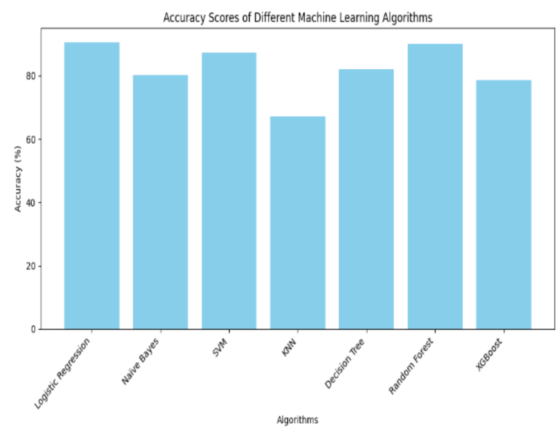


Table-3: Accuracy Results.

5.CONCLUSION

One particularly strong collaborative educational framework that may be used for classification as well as regression tasks is the Random Forest method. To produce a final prediction, it generates many decision trees and aggregates their outputs. This method allows the algorithm to quickly reach a high prediction accuracy, especially in the initial phases. Utilizing this technology for medical records, especially those connected to heart health, has enormous promise for promoting early identification of heart-related problems and, ultimately, saving lives.

The prediction of heart attacks is a major difficulty in the modern world. By entering pertinent parameters from their medical reports, people can use these programs to predict the risk of a heart attack in situations where they do not have instant access to medical personnel. Equipped with this knowledge, individuals can decide whether to seek medical attention, potentially reducing the severity of the crisis.

6.FUTURE SCOPE

In the future, this platform holds potential for augmentation through the integration of new features. For instance, upon the prediction of a heart attack, a feature could be implemented to dispatch notifications to all family members of the patient. Simultaneously, pertinent information would be relayed to the nearest hospital for immediate attention. Moreover, an additional functionality could facilitate online consultations between physicians, enhancing collaborative medical discussions.

It's crucial to emphasize that machine learning (ML) technologies are used for more than only predicting and analysing heart attacks. Additionally, ML algorithms play a key role in the improvement of several medical sectors, including radiology, bioinformatics, and medical imaging analysis.

7.REFERENCES

[1] Dilsizian, S. E., & Siegel - E.Machine Learning Algorithms for Predicting Heart Attack , June 2020.

[2] Nazir .H - Hybrid Machine Learning for predicting heart attack,August 2020.

[3] Raghupathi.W, Deep Learning Models for Predicting Cardiovascular Risk,November 2020.

[4] Yadav .S - Machine Learning Models for Heart Disease Prediction,March 2021.

[5] Asri, H., Mousannif, H.,Moatassime, H., Noel, T. - Study comparing machine learning algorithms for predicting heart disease,May 2021.

[6] Ribeiro, M. - Case study on explainable AI techniques for interpreting machine learning models in heart disease prediction,August 2021.

[7] Madi, N. - Comprehensive survey on machine learning algorithms for heart disease prediction, including ensemble methods,September 2021.

[8] Sudhakar, K. - Paper investigating the use of genetic algorithms for feature selection in heart disease prediction,March 2022.

[9] Zhang, S. (2024). Enhancing heart disease prediction through ensemble learning and feature engineering: A comparative analysis,December 2023.

[10] Dey, A. - Review analyzing various machine learning models and feature engineering for heart disease prediction, June 2023.