

<https://doi.org/10.51470/IJNS.2024.01.04.7>

International Journal of Nature Science (IJNS)



HEART DISEASE PREDICTION AND DIAGNOSING USING DATA MINING AND MACHINE LEARNING TECHNIQUES

Dr.T.Ramaprabha¹, Mrs R.Auxiliaanitha mary²

Associate Professor¹, Dept of Information Technology,

Nehru Arts and Science College Coimbatore, nascramaprabhacs@nehrucolleges.com

Research Scholar², Dept of Information Technology, Nehru Arts and Science College Coimbatore, auxiliaanithamary@gmail.com

Corresponding Author:

nascramaprabhacs@nehrucolleges.com

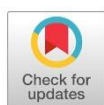
Article History

Volume:1, Issue:4, 2024

Received: 27th September, 2024

Accepted: 8th October 2024

Published: 20th December, 2024.



Abstract: Heart is the next major organ comparing to the brain which has more priority in the Human body. It circulates blood throughout the body, supplying it to all organs. Prediction of incidence of cardiac disorders in the medical field is significant effort. Heart-related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but all through the whole world. So, there is a need for a reliable, accurate, and feasible system to diagnose and predict such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Data analytics is useful for predicting diseases based on more information, and it assists medical centres in doing so. Some data mining and machine learning techniques, such as Navie bayes, Random Forest, and Support Vector Machine (SVM) and xGboost algorithm, are used to predict heart diseases. The diagnosis and prediction of heart disease has become an increasingly challenging endeavour for doctors and hospitals both in India and abroad. A speedy and efficient detection tool must be established in order to prevent the enormous number of deaths caused by cardiac ailments. The main objective of this research paper is to predict the heart disease of a patient using machine learning algorithms.

Keywords: Machine Learning, Supervised learning, Support vector machine, Random Forest, Navie bayes, xGboost

Authors citation: T.Ramaprabha and R.Auxiliaanitha Mary., Heart Disease Prediction and Diagnosing Using Data Mining And Machine Learning Techniques.Int.J.Nat.Sci. Vol.1(4). 2024.Pp:7-12. <https://doi.org/10.51470/IJNS.2024.01.04.7>

1. INTRODUCTION

Heart attacks are the leading cause of death among all fatal diseases. Several surveys on heart illnesses are conducted by medical professionals in order to collect information about heart patients, their symptoms, and the progression of their disease. Heart attacks are a prevalent and potentially dangerous occurrence these days. Some symptoms hinted at what was to come. Medical science has made outstanding use of technological advances to improve healthcare quality. These technological advancements have paved the way for accurate sickness detection and prognosis. When it comes to forecasting cardiac ailments, machine learning may be a terrific alternative for you to achieve a high level of accuracy. As a result, three algorithms will be put in place. It is a combination of the Support Vector Machine, Random Forest, Navie Bayes, and xGboost algorithms. Furthermore, these four strategies produce substantially faster and more consistent results. People presently live in luxury all around the world, and they labour like machines to amass wealth and renown. Because of their stressful schedules, people forget to take care of their health.

As a result, their diet and way of life have shifted. Young individuals develop blood pressure, diabetes, and a range of other disorders as a result of strain and stress in their lives. All of these factors play a role in the development of heart disease.

Many studies have been undertaken in an attempt to identify the most relevant risk factors for heart disease and to accurately estimate the overall risk. Heart disease is also referred to as a silent killer because it causes mortality without evident signs. Early detection of cardiac disease is critical in making decisions about lifestyle adjustments in high-risk patients, reducing consequences.

1.1 Description

The major goal of this work is to target all conceivable combinations of attributes against various methods. The technique that works the best to forecast heart disease at an early stage is then identified among all the techniques. Implementing four algorithms, including Support Vector Machine, Random Forest, Navie Bayes, and xGboost, would make it easier to identify and classify the disease. To categorise and train the model, a dataset is used. Following model training, the most accurate and successful algorithm was employed to forecast the disease.

1.2 Problem statement

The most difficult element of cardiovascular disease is detecting it. There exist technologies for forecasting heart disease, but they are either too expensive or ineffective for quantifying the risk of heart disease in humans. It has been demonstrated that early detection of cardiac problems reduces mortality and overall repercussions. However, it is not possible to properly monitor patients every day in all circumstances, and 24 hour consultation with a doctor is not available because it requires more intellect, time and expertise. In today's world, we may utilise various machine learning algorithms to evaluate data for hidden patterns because we have a lot of data. Hidden patterns in medical data can be used to diagnose illnesses.

1.3 Proposed Approach

Current methods for predicting cardiovascular risk miss many patients who would benefit from preventative care, while others receive unneeded intervention. Machine learning has the potential to increase accuracy by taking advantage of complicated connections between risk factors. We investigated if machine learning can help forecast cardiovascular risk.

2. LITERATURE SURVEY

The summary of the literature review can be seen in Table 1. Several approaches have been performed on this popular dataset, but the accuracy obtained by all the approaches is more with time computations. Mihir Patel et al., [1], The Heart Disease Prediction System predicts the disease using the 13 attributes and can be extremely useful during a critical period. When three classification algorithms, Logistic Regression, Random Forest, and Decision Tree, were tested, Logistic Regression had the greatest accuracy of 92%. In an emergency, the ability to foresee quickly is advantageous.

Baban.U. Rindhe et al., [2], to predict heart illness utilising machine learning approaches for heart disease categorization provides superior results for Using the model dataset, the Support Vector Classifier scored 84.0%, the Neural Network scored 83.5%, and the Random Forest Classifier scored 80.0%. ApurbRajdhan and et al., [3], Using the UCI machine learning repository dataset, this study analyses the accuracy score of Decision Tree, Logistic Regression, Random Forest, and Naive Bayes algorithms for predicting heart disease. According to the findings of this study, the Random Forest algorithm is the most efficient algorithm for predicting cardiac disease, with an accuracy score of 90.16%.

Senthil Kumar Mohan et al., [4], Effective Prediction of Heart Disease Using Hybrid Machine Learning Techniques in which the strategy's goal is to uncover important components by utilising Machine Learning, resulting in improved accuracy in the prediction of cardiovascular disease. The expectation model is made up of numerous combinations of features and a few well-known arrangement algorithms. We achieve an improved exhibition level with a precision level of 88% using the prediction model for heart disease with hybrid random forest with a linear model (HRFLM). They were also educated on various data mining approaches and expectation techniques, such as KNN, LR, SVM, NN, and Vote, which have recently gained popularity in distinguishing and predicting heart disease.

Anshila Raina et al., [5], This study used machine learning techniques such as backward elimination algorithm, logistic regression, and REFCV on a dataset available publically on the Kaggle website, with the findings further analysed using confusion matrix and cross validation. used to predict cardiac disease The model Logistic Regression has an accuracy of 85%. Lakshmana Rao et al., [6], Machine learning bases reasoning on previously recorded data. Machine Learning Techniques for Cardiovascular Disease Prediction in which the factors that contribute to heart disease are difficult to differentiate. The concept of CHD disease is perplexing; hence, the disease must be approached with caution. Furthermore, from the standpoint of medicinal science, data mining is used for discovering various types of metabolic machine learning a technique that causes the framework to gain from previous information testing, models without being expressly customised. Early detection may have an influence on the heart or cause sudden death.

Avinash Golande and et al., [7], The research looks into various ML algorithms that can be used to classify cardiac disease. The accuracy of Decision Tree, KNN, and K-Means algorithms that can be utilised for classification was investigated. This study concluded that Decision Tree had the highest accuracy and that it may be made more efficient by combining multiple strategies and parameter tuning.

Devara Sandhya and et al., [8], The methodologies employed in the study are Random Forest and Logistic Regression; they discovered that Random Forest had higher accuracy than Logistic Regression. Our goal is to improve the Random Forest's performance by removing unneeded and irrelevant features from the dataset and providing better results. Singh Yeshvendra K. et al., [9], Investigate supervised machine learning algorithms like Random Forest, Support Vector

Machine, Logistic Regression, Linear Regression, and Decision Tree approaches. They used 297 tuples in data pre-processing, dividing the data 70% for training and 30% for testing. The first algorithm used is Linear Regression, which has an accuracy of 83.82% when employing a sigmoid function. In supervised machine learning, the classification algorithm Support Vector Machine provides 83.83%.

Rohit Bharti et al., [10], In this study, various machine learning methods and deep learning are used to compare and analyse the outcomes of the UCI Machine Learning Heart Disease dataset. The dataset is made up of 14 primary attributes that will be used in the research. Various outcomes are obtained and validated using an accuracy and confusion matrix. 94.2% accuracy using a deep learning approach.

3. COMPARISON ANALYSIS

From the comparison, we get to know that Logistic Regression is preferable as it gives better classification and higher prediction results for the detection of heart diseases.

S.N O	AUTHOR	YEAR	FINDINGS
1.	Mihir Patel [1]	2022	Logistic Regression, Random Forest and Decision Tree in which LogisticRegression has given highest accuracy of 92%.
2.	Baban.U. Rindhe [2]	2021	Support Vector Classifier: 84.0 % Neural Network: 83.5 % Random Forest Classifier: 80.0 %
3.	ApurbRajdhan [3]	2020	Using the UCI machine learning repository dataset, a study evaluates the accuracy score of Decision Tree, Logistic Regression, Random Forest, and Naive Bayes algorithms for predicting heart disease. The Random Forest algorithm is the most effective method for predicting cardiac disease, with an accuracy score of 90.16%.
4.	Senthilkumar Mohan [4]	2019	1) Decisiontree 2)LanguageModel 3)SupportVector HRFLM- 88% Accuracy
5.	Anshila Raina [5]	2020	Machine Learning, Logistic regression- 85% accuracy
6.	A.Lakshmanarao [6]	2019	For random Oversampling, SVMgiven thebestaccuracy of82.30%.ForSyntheticMinorityOversampling,Random Forestgiven thebestaccuracy of91.3%ForAdaptivesyntheticsampling,RandomForest(90.3%Accuracy)

7.	Avinash Golande[7]	2019	A study on Decision Tree, KNN and K-Means algorithms. Decision tree provide high accuracy comparing with KNN and K-Mean.
8.	DevaraSandhya[8]	2022	The approaches applied are Random Forest and Logistic Regression, with Random Forest outperforming Logistic Regression in terms of accuracy.
9.	Singh Yeshvendra K.,[9]	2018	Linear Regression-70% Support Vector Machine-83.3%
10	Rohit Bharti[10]	2021	Logistic regression 83.3% K neighbors 84.8 % SVM 83.2 % Random forest 80.3 % Decision tree 82.3 % DL 94.2%

TABLE 1. A Comparative Study Of Various Machine Learning Algorithms In Literature Review

4.CONCLUSION

In this paper, It is essential to create a system that can forecast heart diseases precisely and effectively given the rise in mortality brought on by sudden heart disorders. The goal of the study was to discover the best effective machine learning algorithm for detecting cardiac problems. Using the output of several methods employed in various studies, this study evaluates the accuracy score of the Support vector machine, Random Forest, Navie Bayes, and xGboost algorithms for predicting heart disease. Targeting all potential combinations of the qualities against various algorithms is the main goal. The strategy that works the best to forecast heart disease at an early stage is to detect and deliver superior results, assisting health professionals in predicting heart disease effectively and efficiently.

REFERENCES

- [1] Mihir Patel, Rahul Patange, Chaitanya Patil, Anuradha Kapoor, "Predicting Heart Disease Using Machine Learning Algorithms.", International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 04, 2022.
- [2] Baban.U. Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagare, Manisha Darade, "Heart Disease Prediction Using Machine Learning", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 5, Issue: 1, 2021.
- [3] ApurbRajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi, "Heart Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Vol. 9 Issue 04, April-2020.
- [4] Senthilkumar Mohan, ChannrasegarThirumalai, Gautam, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", Digital Object Identifier, IEEE Access, 2019, VOLUME 7.
- [5] ANSHILA RAINA, B. MAHALAKSHMI, and DHRUV JAIN, "HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS", International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 03, 2022.
- [6] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar, "Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific and Technology, Volume 8, Issue: 11, 2019.
- [7] Avinash Golande, Praveen Kumar T, "Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology & Engineering, Volume 8, 2019.
- [8] Devara Sandhya, Dr.Kamalraj, "Heart Disease Prediction using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 03, 2022.
- [9]. Singh Yeshvendra K, Nikhil Sinha and Sanjay k Singh, "Heart Disease Prediction System Using RandomForest", International Conference on Advances in Computing and Data Sciences, Singapore, 2016. Springer.
- [10] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande and Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Hindawi Computational Intelligence and Neuroscience, 2021, Article ID 8387680, <https://doi.org/10.1155/2021/8387680>.