MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Advances in Robotic Technology
ISSN: 2997-6197

Heart Disease Prediction Using Multiple Machine Learning Algorithms

Pradip M Paithane* and Atharva J

Head of Department, Artificial Intelligence and Data Science, VPKBIET College, India

***Corresponding author:** Pradip M Paithane, Artificial Intelligence and Data Science, VPKBIET College, Baramati, Maharashtra, India, Email: paithanepradip@gmail.com

Review Article

Volume 2 Issue 1

Received Date: May 10, 2024

Published Date: May 31, 2024

DOI: 10.23880/art-16000114

Abstract

Heart disease is a significant health concern globally, and the ability to predict and diagnose it accurately is crucial for effective treatment and prevention strategies. Machine learning algorithms have shown promise in enhancing the prediction of heart disease by analysing complex medical data. So in this paper, we have analysed and compared different machine learning algorithms like Logistic Regression, SVM and Naive Bayes(Gaussian Naive Bayes) for the prediction of heart disease. In proposed work the data used consist of different medical attributes like age, heart rate, chest pain type, restingBP, max heart rate, etc. To increase the accuracy of the models I used cross validation technique (Kfold). The Support vector machine received highest accuracy as compared to other approaches.

Keywords: Machine Learning; Logistic Regression; Naive Bayes; SVM; Data Preprocessing; Healthcare Analytics; Heart Disease

Abbreviations: CVDs: Cardiovascular Disease; ML: Machine Learning; SVM: Support Vector Machine; EDA: Exploratory Data Analysis.

Introduction

Cardiovascular disease (CVDs) are the disease or disorders related to the heart or the blood vessels. Globally, cardiovascular diseases (CVDs) constitute the primary cause of death, taking millions of lives annually [1]. Effective cardiac disease prediction can greatly improve public health by facilitating early detection and preventive care. Within the medical area, machine learning (ML), a subset of artificial intelligence, has become a powerful tool, opening up new possibilities for the diagnosis and prediction of a wide range of diseases, including heart issues. This work aims to predict whether the person is likely to get diagnosed with heart disease based on his or her past medical history and analyzing various medical attributes related to heart [2].

The work mainly focuses on predicting of heart disease based majorly three machine learning algorithms namely: (1) Logistic regression (2) SVM (3) Naive Bayes [3]. These all machine learning techniques fall under the supervised learning. By employing these algorithms for heart disease prediction ,we aim to develop a predictive model that not only identifies individuals at risk of heart disease but also helps in preventing and mitigating these risks at early stage which can lead to improved patients outcomes and reduced medical costs [4].

The dataset that is used in the project includes various attributes like age, gender, chest pain type, max heart rate, cholesterol level, fasting sugar, etc which helps in predicting whether the person is likely to be diagnosed with cardiovascular heart diseases. The dataset consist of total twelve attributes based on which the three machine learning algorithm logistic regression, SVM, naïve bayes are trained [5].

Literature Review

Heart Disease Prediction using Machine Learning Algorithms: Authors: Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath [6]. In this paper the author have prepared a system to predict heart disease based on various machine learning algorithms like logistic regression, KNN and Random forest classifier.

An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naive Bayes Algorithm: Authors: Rucha Shinde, Sandhya Arjun, Priyanka Patil, Prof. Jaishree Waghmare [7]. In this paper authors have proposed a heart disease prediction system using naïve bayes and k-means clustering algorithms. For grouping the attributes k-means clustering and for prediction naïve bayes is used.

Effective Heart Disease Prediction Using Machine Learning Techniques: Authors: Chintan M. Bhatt, Parth Patel, Tarang Ghetia, and Luigi Mazzeo [8]. In this paper the authors have proposed a model to predict cardiovascular diseases. They have applied the model on a dataset with around 70,000 instances (Table 1).

Title	Author	Novel Approach
Heart disease prediction using machine learning algorithms	Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath	Logistic regression, KNN, Random Forest Classifier
An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm	Rucha Shinde, Sandhya Arjun, Priyanka Patil ,Prof. Jaishree Waghmare	Naïve Bayes, K-means Clustering
Effective Heart Disease Prediction Using Machine Learning Techniques	Chintan M. Bhatt, Parth Patel, Tarang Ghetia ,Luigi Mazzeo	MLP, Random forest, Decision Tree, XGBoost

Table 1: Related Work on Heart Disease.

Methodology

This consists of comparison of different machine learning techniques to predict heart disease. The machine learning algorithms that are used in this paper are Logistic regression, SVM, Naïve Bayes(Gaussian naïve bayes) [8]. The methodology that is been used includes step- 1 data collection, than in the second step EDA(exploratory data analysis) to gain initial insights from the data. The third step is data pre-processing in which different methods are applied to the data to transform the raw data into the required and desired format. It includes outlier removal, scaling the data, encoding the categorical data, etc [9]. After that splitting the data into training and testing sets, with a split size of 80:20. Than in the next step training the different models on the training data. After training the models predict the values for the testing data and then calculating the accuracy of the different models in Figure 1.

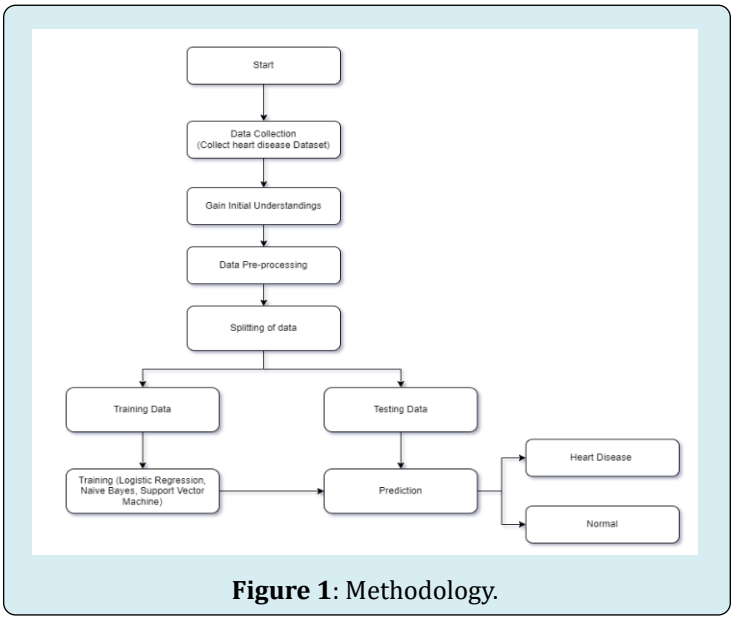

Figure 1: Methodology.

In our work we also implemented cross validation to improve the accuracy of the models. Cross validation is a technique used in machine learning for evaluating the model performance. It helps in ensuring the generalization of predictive models to unseen data. It involves partitioning the data into different sets and averaging the results obtain from different set of partitions. There are different types of cross validation like stratified k-fold cross validation, k-fold cross validation, leave one out cross validation and more [10]. In are work we have used k-fold cross validation. In k-fold cross validation the dataset is divided into k folds and each fold is used as validation set ones and the accuracy for each iteration is measured and the final accuracy is the average of all the k iterations [11].

Logistic Regression

The logistic regression is based on the sigmoid function. It is a supervised machine learning algorithm which is used for classification problems. It uses the logistic (or sigmoid) function to transform a linear combination of input features into a probability value ranging between 0 and 1.
The logistic(sigmoid) function is as followed:

$$F(x)=1/1+e^{-x} \quad (1)$$

Naive Bayes

The naive bayes classifier is based on the bayes theorem. The bayes theorem can be stated as:

$$P(H|X) = P(X|H)\ P(H)\ /P(X) \quad (2)$$

Gaussian naïve bayes is a classification technique used in machine learning which is based on probabilistic approach and gaussian distribution. Gaussian or normal distribution is a type of continuous probability distribution for a real-valued random variable [12].

Support Vector Machine

Support vector machine(SVM) is a supervised machine learning algorithm used for classification and regression tasks. SVM aims to find out the best hyperplane that separate the data points of different classes. SVM can also handle non-linearly separable data by using the method called kernel tricks in Table 1 [13].

Attributes

Attribute	Description
Age	Age of the patient in years
Sex	Gender of the patient
ChestPainType	Chest pain type
RestingBP	Resting Blood pressure
Cholestrol	Serum cholesterol
FastingBS	Fasting blood sugar
RestingECG	Resting electrocardiogram results
MaxHR	Max heart rate
ExcerciseAngina	Exercise induced angina
Oldpeak	ST numeric value measured in depression
ST_Slope	The slope of the peak exercise ST segment
HeartDisease	Target Variable

Table 2: Dataset Attributes.

In this research the proposed heart disease prediction system used three different ML algorithms 1) Logistic regression 2) Naïve bayes 3) SVM. We also utilized k-fold cross validation to evaluate the model performance. Comparing the three models SVM showed highest accuracy around 86% outperforming logistic regression 83% and Gaussian naïve bayes 0.79%. Model accuracies of models with cross as shown in Figures 2,3.

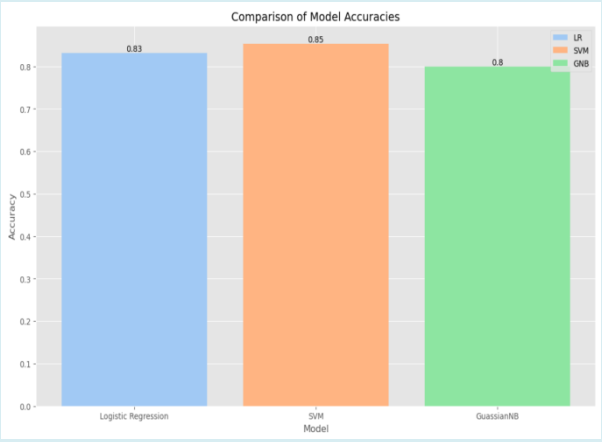

Figure 2: Model comparison without k-fold.

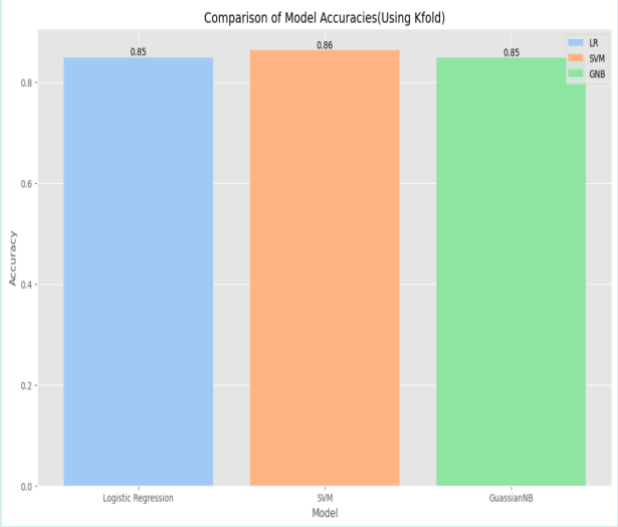

Figure 3: Model comparison using k-fold.

For this work the system that is been used is i7- 13gen hx processor with 16 ram and 512 gb of ssd and a 4050 rtx graphic card. The dataset consisted of 12 different attributes and 918 instances [14]. The data was split into train and test based on 80:20 ratio.

Conclusion

In this paper three different ML algorithms namely logistic regression, SVM, naïve bayes are employed for heart disease prediction. The study showed that SVM outperformed logistic regression and gaussian naïve bayes with an accuracy of around 86%. The SVM algorithm showed an average accuracy around 86% across all the k-fold. In this research article, aimed in developing and comparing different ML algorithms for prediction of heart diseases. The findings shows the importance of ML in prediction of heart disease which will help in early detection and mitigation of heart related risks improving patients outcomes and reducing medical costs. Future scope of the works lies in implementing more complex models for prediction of heart disease and utilizing real time data to monitor patients at risk.

References

1. Mukundrao P, Kakarwal SN (2022) Automatic pancreas segmentation using a novel modified semantic deep learning bottom-up approach. International Journal of Intelligent Systems and Applications in Engineering 10(1): 98-104.

2. Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications 17(8): 43-48

3. Arumugam K, Naved M, Priyanka PS , Orlando LC, Antonio H, et al. (2023) Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings 80 (2023): 3682-3685.

4. Shah, Devansh, Samir P, Santosh KB (2020) Heart disease prediction using machine learning techniques. SN Computer Science 1(6): 345.

5. Harshit J (2021) Heart disease prediction using machine learning algorithms. IOP conference series: materials science and engineering.

6. Rucha MS, Arjun PP, Jaishree W (2015) An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies 6(1): 637-639.

7. Chintan M, Parth P, Tarang G, Pier LM (2023) Effective heart disease prediction using machine learning techniques. Algorithms 16(2): 88.

8. Pradip M. Paithane, Sarita JW, Sangeeta K (2023) Optimization of route distance using k-NN algorithm for on-demand food delivery. System research and information technologies 1: 85-101.

9. Sarita JW, Pradip M. Paithane, Pradip MP, Patil SN (2021) Applications of Fuzzy Logic in Assessment of Groundwater Quality Index from Jafrabad Taluka of Marathawada Region of Maharashtra State: A GIS Based Approach. International Conference on Hybrid Intelligent Systems. Springer International Publishing, pp: 354-364.

10. Pradip M. Paithane (2022) Yoga posture detection using machine learning. Artificial Intelligence in Information and Communication Technologies, Healthcare and Education. Chapman and Hall/CRC, pp: 27-33.

11. Pradip M. Paithane (2023) Random forest algorithm use for crop recommendation. ITEGAM-JETIA 9(43): 34-41.

12. Pradip M. Paithane, Sarita JW (2022) Automatic Quality Control Scrutiny of Sugar Crystal using K-Means Clustering Algorithm Image Processing." American Scientific Research Journal for Engineering, Technology 9(12): 2395-0056.

13. Pradip M. Paithane, Sarita JW (2023) Novel modified kernel Fuzzy C-Means algorithm used for cotton leaf spot detection. System research and information technologies, pp: 85-99.

14. Mohan, Sk, Chandrasegar T, Gautam S (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE access 7: 81542-81554.