A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning

A. Balaji[1], SD. Gohar Sadika Suhana [2], T. Bhargavi[3], P. Gayathri[4], M. Lourdhu Karthik[5]

[1]Associate Professor, Computer science and Engineering, Tirumala Engineering College, India
[2-5]Students of Computer science and Engineering, Tirumala Engineering College, India

Abstract

Abstract: In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person instantly that might create unexpected consequences for the human life. So diagnosing patients correctly on time is the most challenging task for the medical fraternity. The heart disease treatment is quite high and not affordable by most of the patients particularly in India. The research scope is to develop an early prediction treatment using data mining technologies. Now a day every hospital keeps the periodical medical reports of cardiovascular patients through some hospital management system to manage their healthcare.

The data mining techniques namely decision tree and random forest are used to analyze heart attack dataset where classification of more common symptoms related to heart attack is done using c4.5 decision tree algorithm, alongside, random forest is applied to improve the accuracy of the classification result of heart attack prediction. In this system various data mining technologies are applied to make a proactive approach against failures in early predictions diagnosis of the disease. We proposed an automated system for medical diagnosis that would enhance medical care and reduce cost. Our aim is to provide a ubiquitous service that is both feasible, sustainable and which also make people to assess their risk for heart attack at that point of time or later.

Keywords: Heart attack prediction, ML, Random Forest, Decision Tree, Java, servlet, etc.

1.Introduction

Heart disease is the number one killer according to the World Health Organization(WHO) statistics. Millions of people die every year because of heart disease and a large population of people suffer from heart disease. Prediction of heart disease early plays a crucial role for the treatment. If heart disease could been predicted before, lots of patient deaths would have been prevented and also a more accurate and efficient treatment method could have been provided. A need to develop such a medical diagnosis system arises day by day. The important key points of such medical diagnosis systems are reducing cost and obtaining more accurate rate efficiently. Developing a medical diagnosis system based on machine learning for predication for heart disease provides more accurate diagnosis than the traditional method and reduces the cost of treatment. In this paper, predication of heart disease by machine learning algorithms is proposed to satisfy this need. To build an effective predictive model, large datasets containing relevant patient information are required. This data might include electronic health records, medical imaging, genetic information, and patient-reported outcomes. Identifying the most relevant features (variables) from the dataset is crucial for building accurate models. Various machine learning algorithms can be used for heart disease prediction, including logistic regression, decision trees, random forests, support vector machines, and neural networks.

2. Literature Review

Senthil Kumar Mohan AL, proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which the strategy objective is to find critical includes by applying Machine Learning to bring about improving the exactness in the expectation of cardiovascular malady. We produced an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they are likewise educated about diverse data mining approaches and expectation techniques, such as, KNN, LR, SVM, NN, and Vote have been fairlyfamous of late to distinguish and predict heart disease.

Aditi Gavhane, GouthamiKokkula, Isha Pandya, Prof. Kailas Devadkar (PhD), Prediction of Heart Disease Using Machine Learning, in this paper proposed system they used the neural network algorithm multi-layer perception (MLP) to train and test the dataset. In this algorithm there will be multiple layers, like one for input, second for output and one or more layers are hidden layers between these two input and output layers.There is another identity input called bias. The connection between the nodes can be feed forwarded or feedback based on the requirement.

Mr. Santhana Krishnan.J and Dr.Geetha.S, Prediction of heart disease using machine learning algorithm. This predicts heart disease for Male Patient using Classification Techniques. We produced an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest.

3. Methodology

Traditional approaches for heart disease prediction include fundamental analysis, which looks at a patient's past health data and statistical analysis, which is solely concerned with the number of patients suffering from heart disease and the number of heart patients. The latter is commonly achieved with the help of Genetic Algorithms (GA) or Artificial Neural Networks (ANN's), but these fail to capture correlation the between patient data in the form of long-term temporal dependencies. Another major issue with using simple ANNs for patient data is the phenomenon of exploding / vanishing gradient, where the weights of a large network either become too large or too small , drastically slowing their convergence to the optimal value. This is typically caused by two factors: weights are initialized randomly, and the weights closer to the end of the network also tend to change a lot more than those at the beginning.

We propose machine learning algorithms for effective prediction of heart disease outbreaks in disease-frequent communities. We take input as a collection of heart disease dataset from the Kaggle website and then train and test using a random forest algorithm and then predict based on user input of respective patient. In this system we are implementing an effective heart attack prediction system using Machine Learning algorithm. We can give the input as in CSV file or manual entry to the system. After taking input, the algorithms apply to that input, which is a different machine learning algorithms. After accessing the dataset, the operation is performed and an effective heart attack level is produced.

4. SYSTEM ARCHITECTURE
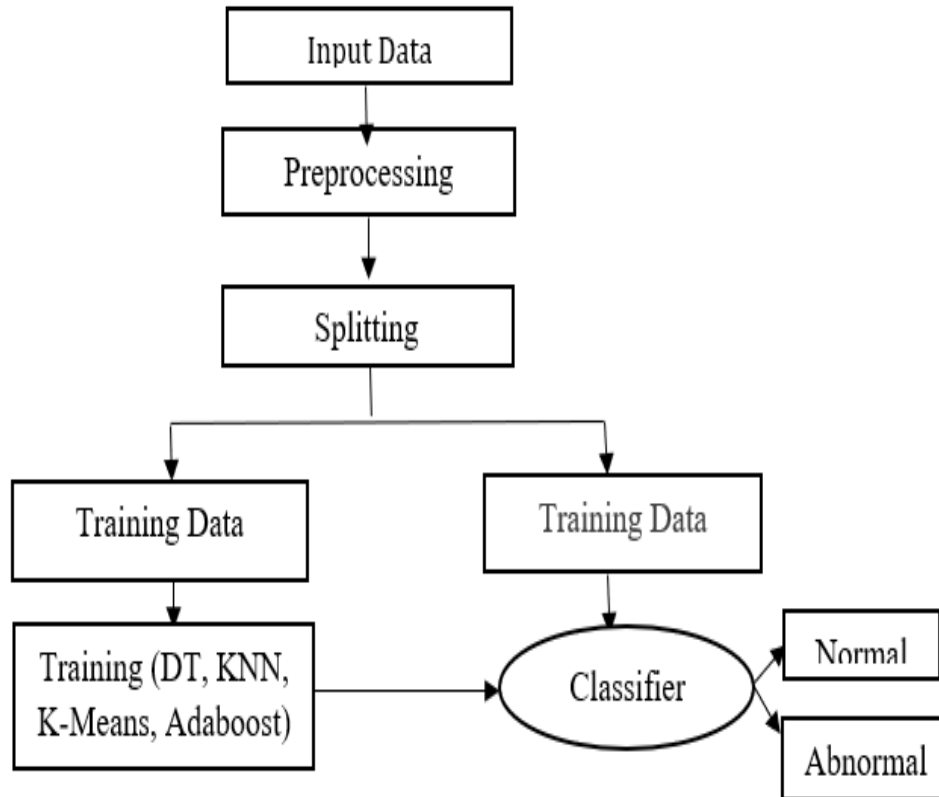
The system will go through various operations such as first input image is given, which will go under preprocessing, segmentation, feature extraction, classification. In preprocessing the unwanted part is removed, segmentation will divide the area of interest into number of parts, feature extraction phase will extract the data and will store it for comparison purpose. After classification phase and comparison finally the heart attack will be predicted. After reviewing various research papers we found that this is the best system architecture for our project. Just the difference between existing system architecture and our proposed system architecture is we used some clinical data instead of data warehouse because data warehouse will be require only when this project will be widely used in medical industry. We simply created easy to use web based application for prediction of heart disease. The proposed system inspired after studying similar systems and the team customized to make it more efficient. The section also explains the working of the proposed architecture in a brief. The performance of the proposed architecture is also explained in the section where it can be considered to be more efficient than that of the existing one.

5. ALGORITHM SELECTION

1. Support vector machine: Support Vector Machine (SVM) is a widely used algorithm in Supervised Learning, primarily used for classification in Machine Learning. Its main goal is to create an optimal hyperplane, dividing an n-dimensional space into distinct classes, allowing easy classification of new data points. SVM selects the most critical points, known as support vectors, to define this hyperplane, giving it its name. This algorithm is primarily used for classification in Machine Learning.

2. K-Nearest neighbor: The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning method developed by Evelyn Fix and Joseph Hodges in 1951. It is essential for pattern recognition, data mining, and intrusion detection. KNN is non-parametric, making it widely disposable in real-life scenarios. It uses prior data (training data) to classify coordinates into groups identified by an attribute, unlike other algorithms like GMM that assume a Gaussian distribution.

3.Navie Bayes: Naive Bayes classifiers are a family of algorithms based on Bayes' Theorem, ensuring feature independence. They are widely used in machine learning due to their simplicity and efficiency. These classifiers are particularly useful in text classification, spam filtering, sentiment detection, and rating classification due to their speed and ease of prediction. Despite their simplistic assumptions, Naïve Bayes classifiers are widely used for their rapid development and prediction capabilities.
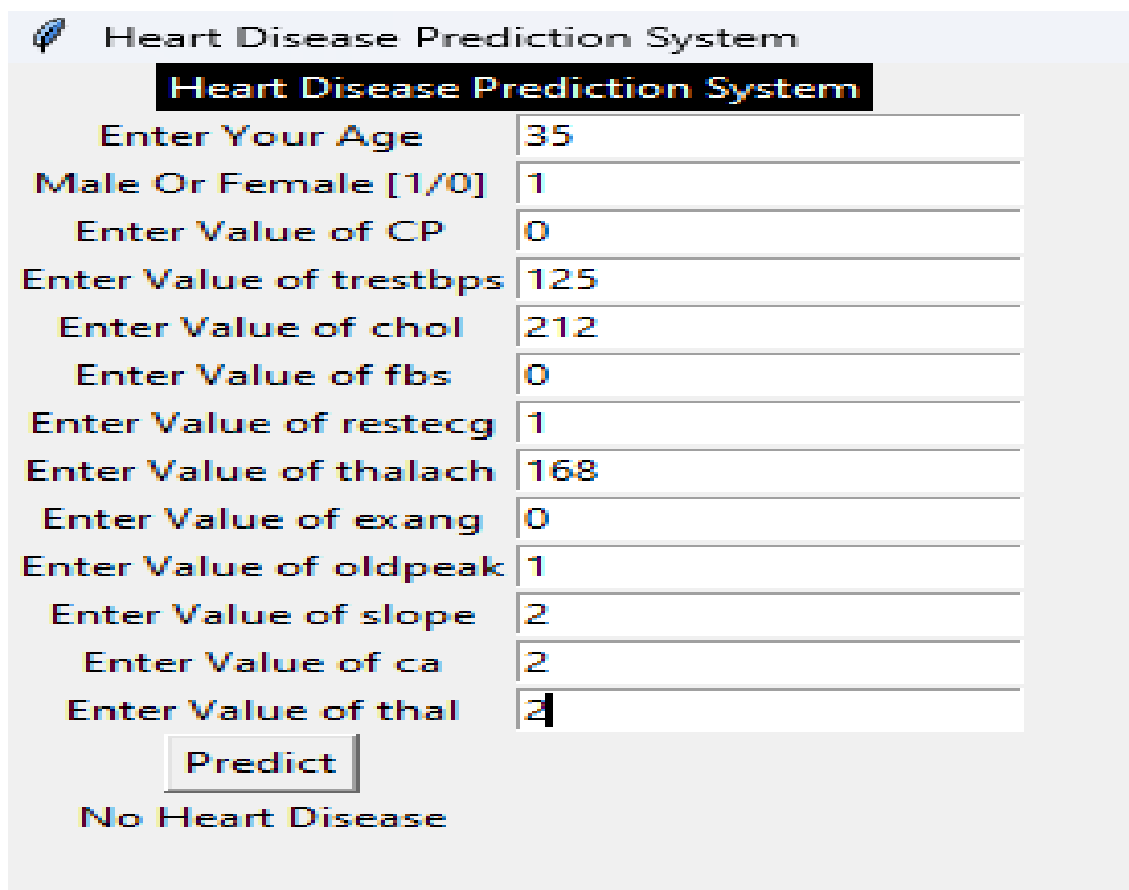
4.Decision tree: Decision Tree is a supervised learning method used for classification and regression tasks, with internal nodes representing dataset features and branches representing decision rules. Decision nodes make decisions, while Leaf nodes display outcomes. The CART algorithm is used to construct a tree-like structure, with decision trees branching off into subtrees based on response (Yes/No). This method provides a pictorial representation of potential solutions to problems.

5.Logistic Regression: Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

6.Random Forest: Random Forest is a well-known machine learning algorithm that falls under the supervised learning category. It can tackle Classification and Regression tasks in the field of ML. This algorithm works by employing ensemble learning, which involves combining several classifiers to effectively address complex problems and enhance model performance. Random Forest is a type of classifier that consists of multiple decision trees built on different subsets of the dataset and combines their predictions to enhance accuracy. Unlike a single decision tree, random forest considers the predictions of all trees and makes a final prediction based on majority vote.

6.RESULT

In Heart Disease Prediction System, we used 13 attributes which is generally common but most important attributes in view of heart attack failure or heart disease. This 13 attributes are age, sex, chest pain type, blood pressure, cholesterol, fasting blood pressure, ecg type, maximum heart rate achieved, exercise induced angina, exercise relative to rest, heart rate slope, cardiac stress and blood disorder.

This is the screenshot of our system which showing that how patient diagnose and how message is display on screen that patient is having negative/positive chances of heart attack.

7.Conclusion

The symptoms of heart disease are familiar for everyone. But no one can predict when it may happen and also the prediction will not give any instance status i.e. normal or abnormal. Our research intension is to notify the patients that having possibility of heart disease or no heart disease. The application of machine learning algorithms shows it better performance results among themselves. This app will be supposed to predict the heart attack from the chest pain at an early stage and guide the person to take treatment early such as to get their ECG done as early as possible and get evaluated from a doctor to make diagnoses of heart attack.

.

8.References

[1]. AvinashGolande, Pavan Kumar T, Heart Disease Prediction Using Effective Machine Learning Techniques, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[2]. T.Nagamani, S.Logeswari, B.Gomathy, Heart Disease Prediction using Data Mining with Mapreduce Algorithm, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3]. Fahd Saleh Alotaibi, Implementation of Machine Learning Model to Predict Heart Failure Disease, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4]. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design And Implementation Heart Disease Prediction Using Naives Bayesian, International Conference on Trends in Electronics and Information(ICOEI 2019).

[5]. Nagaraj M Lutimath,ChethanC,Basavaraj S Pol.,Prediction Of Heart Disease using Machine Learning, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[6]. UCI, Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.

[7]. SayaliAmbekar, Rashmi Phalnikar,Disease Risk Prediction by Using Convolutional Neural Network,2018 Fourth International Conference on Computing Communication Control and Automation.

[8]. Sowmiya, C., and P. Sumitra. "Analytical study of heart disease diagnosis using classification techniques." In Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference, pp. 1-5.IEEE,

[9]. FajrIbrahemAlarsan., and Mamoon Younes Analysis and classification of heart diseases using heartbeat features and machine learning algorithms, Journal Of Big Data,2019;6:81.