HEART DISEASE PREDICTION SYSTEM

Prof. Shinde Reshma U, Miss.Shirisha Agnur, Mis.Bhavna Mushan, Miss.Bhavna Kyatam, Miss.Shweta Mhetre

Assistant professor, Student, Student, Student, Student

Computer science and engineering

NK Orchid College of Engineering and Technology, Solapur, Maharashtra, India

Abstract

Heart disease is one of the most significant causes of mortality in today's world. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. Machine learning has been shown to be effective assisting in making decisions and predictions from the large quantity of data produced by the health care industry. We used popular machine learning algorithms and classifiers performance evaluation metrics such as classification accuracy, correlation coefficient. In this proposed system we develop web application with Flask and integrated ML model in the application. The proposed system can easily identify and classify people with heart disease and plot bar graph for feature importance.

Index Terms – Machine Learning, KNN algorithm,

I. INTRODUCTION

There is no death of records about medical symptoms of patients suffering heart strokes. However, the potential they have to help us foretell similar possibilities in seemingly healthy adults are going unnoticed. For instance: As per the Indian Heart Association, 50% of heart strokes occur under 50 years of age and 25% of all heart strokes occur under 40 years of age in Indians. Urban population is thrice as vulnerable to heart attacks as rural population. Thus, we used two machine learning algorithms to predict the Heart Disease. One is Decision Tree Algorithm. Decision Tree is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data. The results obtained from Decision Trees are easier to read and interpret. The drill through feature to access detailed patient's profiles is only available in Decision Trees. Second is K-nearest neighbor Algorithm. It is most widely used lazy classification algorithm as it reduces misclassification error.

Heart Disease

The heart is important organ of human body part. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it.

A number of factors have been shown that increases the risk of Heart disease:

• Family history

• Smoking

• Poor diet

• High blood pressure

• High blood cholesterol

• Obesity

• Physical inactivity

• Hyper tension

Factors like these are used to analyze the Heart disease. In many cases, diagnosis is generally based on patient's current test results & doctor's experience. Thus, the diagnosis is a complex task that requires much experience & high skill.

Motivation

The motivation for the project was to find the most efficient ML algorithm and create user friendly web site for detection of heart diseases. This study effectively predicts if the patient suffers from heart disease. The health professional or user enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart disease. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance.

Objective
- To effectively predict if the patient suffers from heart disease.
- To provide the best accuracy rate with minimum errors.
- Dynamic platform.
- Simple and easy interface.

Proposed System

The proposed system clearly classifies the data of the patients considering the various parameters and gives the result properly. It also tells if the patient is having the disease or not. It uses two algorithms to increase the accuracy rate. Fig. 1 shows the entire process involved.
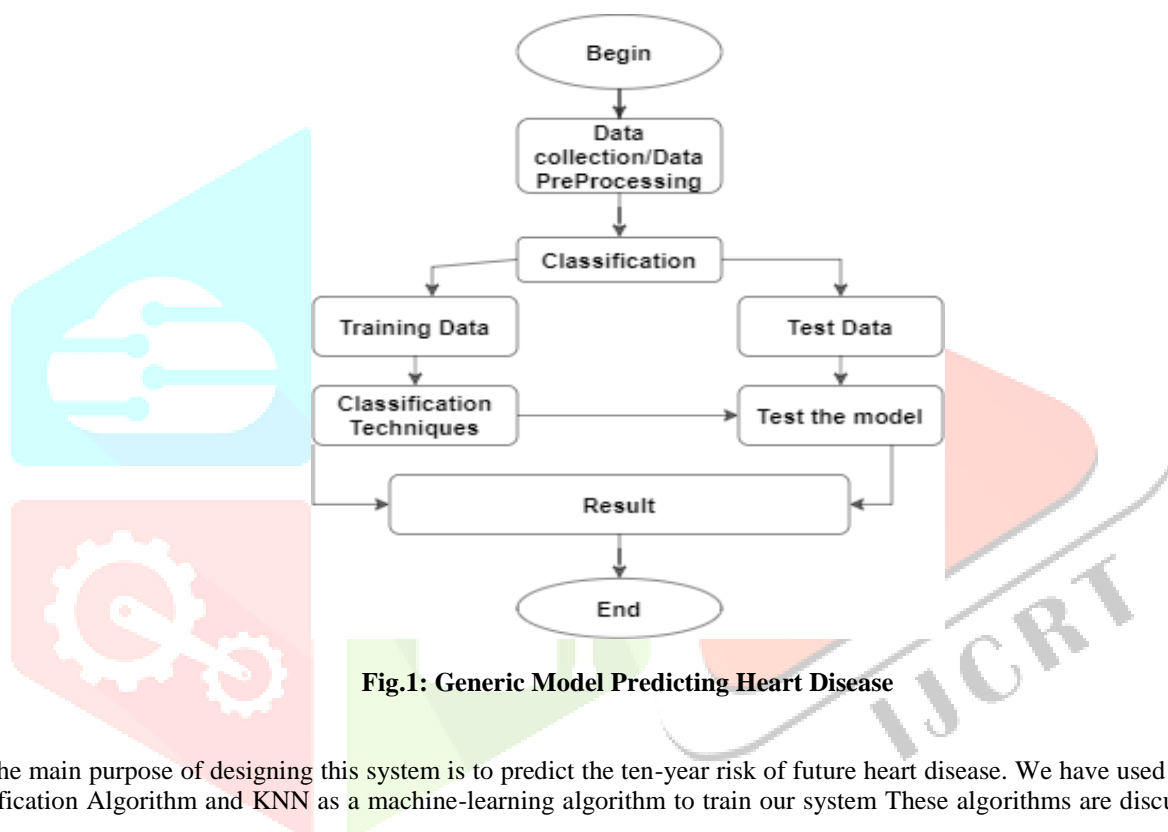
Fig.1: Generic Model Predicting Heart Disease

The main purpose of designing this system is to predict the ten-year risk of future heart disease. We have used Decision Tree Classification Algorithm and KNN as a machine-learning algorithm to train our system These algorithms are discussed below in detail.

Decision Tree Classification Algorithm:

- Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into sub trees.

K-Nearest Neighbor (KNN) Algorithm:

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, and then it classifies that data into a category that is much similar to the new data.

Conclusion:

In heart disease prediction system, we have implemented a GUI based disease prediction system for the patients who want know prior that they have the disease or not. In this system the patient or doctor has to provide their health details to predict the disease. The system predicts the accuracy of disease based on their health checkup data. It provides the accurate results and consumes very less time process the data. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease.

This system can be used by every person who wants to be aware about their health. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible.