Heart Disease Prediction using Machine Learning Techniques

Sharyu U. Kamble, Vaishnavi S. Jawanjal, Pooja P. Velapure, Priya K. Jadhav, Sanjivani S. Kadam

Department of Computer Science and Engineering, SVERI's College of Engineering, Pandharpur, Solapur University, Maharashtra.

Abstract—Diseases related to Heart i.e. Cardiovascular Diseases (CVDs) are the main reason for the number of deaths in the course of the most recent couple of decades and has developed as the most perilous ailment, in India and in the entire world. In this way, there is a need for an accurate, feasible and reliable system to analyze such maladies in time for legitimate treatment.

Machine Learning algorithms and procedures have been implemented to various medical datasets to various medical datasets to investigate of extensive and complex information. Numerous analysts, as of late, have been using several methods to enable the health care industry and the professionals in the diagnosis of heart related diseases.

This paper demonstrates a survey of various models based on such algorithms and techniques and analyze their performance. Models depend on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and ensemble models are discovered extremely prominent among the researchers.

Index Terms—Cardiovascular, datasets, supervised learning algorithms, Support Vector Machines, K-Nearest Neighbour, Naïve Bayes, Decision Trees, Random Forest.

I. INTRODUCTION

Data Mining is a non-minor extraction of certain, beforehand obscure and potentially valuable information about data [1]. In short, it is a procedure of analyzing information from the substitute perspective of view and assembling the knowledge of it [2].

The discovered information can be used for various applications, for example healthcare industry. The Healthcare industry is "data rich", however lamentably not every one of the information are Due, which is required for finding hidden patterns and effective decision making. Data Mining Techniques such as Propelled data mining techniques are utilized to find learning in the database and for medicinal research, especially in the Heart disease prediction.

A major challenge facing the healthcare industry is the nature of the administration. Poor analysis can prompt appalling outcomes which are unacceptable. The datasets are overwhelming for human personalities to fathom, can be effectively investigated utilizing different machine learning techniques.

Accordingly, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately. Now a day’s doctors is embracing many scientific technologies.

Our project’s objective is to foresee the odds of heart disease based on the patient’s dataset and the doctor’s perspective in check-up using machine learning. By utilizing the patient's medical records, a new system is proposed to foresee the chances of heart attack Attributes such as Blood pressure (bp), age, thickness of the artery, etc. are sustained into the dataset and algorithm [3].

II. HISTORY AND BACKGROUND

The motivation to do this problem comes from the World Health Organization estimated. As per the World Health Organization estimation till 2030, practically 23.6 million people will pass on due to Heart illness. Along these lines, to limit the threat, desire for coronary disease should be done. Investigation of coronary sickness has been regularly in perspective on signs, appearances and physical examination of a patient. The most troublesome and complex assignment in medicinal services area is finding of right ailment or right illness. In late patterns investigation on these broad datasets has been able to be fundamental because of monetary weights on medicinal services commercial enterprises.(business endeavors).

Chen et al. Proposed the prediction system for heart disease, i.e. coronary illness using learning vector quantization algorithm. Another study probed on sample database of patients' records. The Neural Network is prepared, tested, trained with 13 input factors such as Age, Blood Pressure, Kumaravel et al. Have proposed automatic diagnosis framework for heart diseases using a neural network system with an accuracy of 63.6–82.9%.

The heart is an important organ of human body part and it is similar to a pump, which circulates blood through the body [4] and If the circulation of blood in the body is inefficient, then vital organs like brain suffer and if heart quits working, demise happens within minutes. Life is totally reliant on the successful working of the heart. The term Heart disease alludes to disease of heart and blood vessel framework inside
Several factors have been demonstrated that increases the chances of Heart disease: Family history, Smoking, Poor diet, High blood pressure (Hyper Tension), High blood cholesterol (Caused by Obesity), Physical inactivity.

True assurance of coronary sickness can’t be conceivable by utilizing just human comprehension. There are stores of parameters that can impact the precise end like less exact outcomes, less experience, time subordinate execution, data up degree and substantially more.

III. DESIGN ISSUES

MATH OR EQUATION

We can compute distance i.e. gap between two points using distance functions as \( d(x,y) \) where \( x, y \) are scenarios composed of features, such that \( x=x_1, \ldots, x_N, y=y_1, \ldots, y_N \). Two functions can be used:

\[
\text{Euclidean function} = d(x_j, x_k) = \sqrt{\sum_i (x_{j,i} - x_{k,i})^2}
\]

\[
\text{Mahanattan function} = d(x_i, x_k) = \sum_i |x_{j,i} - x_{k,i}|
\]

Here we have used Euclidean distance measuring function to find distance between two points, absolute distance measuring is also called as Mahattan distance.

A. KNN Implementation

KNN classifier algorithm stands for k-nearest neighbor algorithm. This classifier is utilized to partition dataset parameters in different clusters based on the distance between points.

In 1951, Hodges et al. introduced a nonparametric technique for pattern classification which is popularly known the K-Nearest Neighbour rule. K-Nearest Neighbour technique is a standout amongst the most rudimentary but very effective classification techniques. It makes no suspicions about the data and is generally be used for classification tasks when there is very less or no prior knowledge about the data distribution. This algorithm involves finding the \( k \) nearest data points in the training set to the data point for which a target value is unavailable and assigning the average value of the found data points to it.

This is the classifier function for KNN. It uses following functions:

a) \( n \_ \text{neighbours} \):
   It hold the value of \( k \) which is an integer value which needs to be passed if required as user defined otherwise its value is 5 by default.

b) \( \text{data.fit()} \): It is used to fit the model in which has two parameters as \( X, Y \) where \( X \) is training data with features and \( Y \) is training data with label.predict().

c) \( \text{accuracy score} \): It is the function which is used to find accuracy of KNN algorithm. Accuracy means the proportion of the correctly predicted data points to all the anticipated data points.

Advantages of KNN are that cost of the learning procedure cost is zero, No presumptions about the attributes of the ideas to learn must be done and complex ideas can be learned by nearby guess utilizing basic strategies, it tends to be considered as baseline method.
Heart diseases are difficult to cure, and it takes away lots of lives every year. Sedentary lifestyle and excessive stress are major reasons that worsen the situation. If the disease is detected early, then its curing becomes easy [5]. There are few treatment methods for a patient once diagnosed with a form of heart disease. Data mining can be of very good help in deciding the line of treatment from such suitable machine learning techniques, datasets and algorithms. Datasets dealing with the same medical issue as Coronary artery disease (CAD) may demonstrate distinctive outcomes while applying the similar machine learning technique. The classification results and accuracy score result the selected important features are based mainly on the efficiency of the medical diagnosis and analysis [6].

### V. CONCLUSION

Heart diseases are difficult to cure, and it takes away lots of lives every year. Sedentary lifestyle and excessive stress are major reasons which worsened the situation. If the disease is detected early, then its curing becomes easy [5]. There are few treatment methods for a patient once diagnosed with a form of heart disease. Data mining can be of very good help in deciding the line of treatment from such suitable machine learning techniques, datasets and algorithms. Datasets dealing with the same medical issue as Coronary artery disease (CAD) may demonstrate distinctive outcomes while applying the similar machine learning technique. The classification results and accuracy score result the selected important features are based mainly on the efficiency of the medical diagnosis and analysis [6].

### REFERENCES

[1] K. Sudhakar, Dr.M. Manimekalai “Study of Heart disease prediction using data mining” ISSN: 2277 128X
[4] Ankita Pimputkar, Jitendra S. Dhobi “A Survey on Heart Disease Prediction using Hybrid Technique in Data Mining” IJARIIE-ISSN(O)-2395-4396
[5] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee “Heart Diagnosis and prediction using machine learning and data mining : Review”


[16] Prediction of Heart Disease using data mining and machine learning algorithms and techniques, Nikhil Kumar Mutlaya, K.V.s. Kaushik DOI: 10.13140/RG.2.2.18185.75362