



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 6)

Available online at: www.ijariit.com

The classification scheme for the heart disease prediction

Saloni Kapoor

salonikapoor31@gmail.com

Chandigarh University, Ajitgarh, Punjab

Ashwinder Tanwar

ashwinder.cse@cumail.in

Chandigarh University, Ajitgarh, Punjab

ABSTRACT

With the enormous enhancement of diseases in medical and the other communities of healthcare, it is extremely important to have an analysis of the heart diseases at the early stages. Since nowadays it is very important to detect the diseases and lessen the death of patients at early stages. Every person has different values of cholesterol, blood pressure and many more that are linked with heart disease prediction. But it has scientifically proven that the normal person blood pressure is counted to be 120/90 along with this the pulse rate and the cholesterol value is 72. In this paper, the various “machine learning algorithms” are explained that include Support Vector Machine, Decision tree, neural network and many more are explained so that complete description can be provided. Along with this, the entire description of the heart disease has been provided that depicts the need for the topic to be selected. There are some of the issues present in the Data Mining algorithm that are also described in the paper. The ultimate aim is to improve efficiency in different parameters by describing the classification approach for detecting heart disease. The parameters on which the prediction can be done are the age, serum cholesterol, gender, blood pressure, pulse rate. The accuracy and the efficiency in the prediction can be increased only if the number of attributes is more. For the classification of heart disease, the most efficient algorithm is the Support Vector Machine algorithm since it will not only reduce the prediction time but will also improve the efficiency of the algorithm.

Keywords— Data mining, Machine learning algorithms, Heart disease prediction

1. INTRODUCTION

1.1 Data mining

Different amount of information is accumulated in files, databases, and many other appliances, as the confidential and the private data should not be stored anywhere else. So, it is very important to discover and propose a system in which all the data and information [1] can be stored safely and securely. Sometimes it becomes very difficult for the users to extract and use only relevant data from this large data. So, in order to overcome this situation, Data Mining is used. Data Mining is a method of selecting, choosing and extracting only those facts which are useful and relevant for that particular instant of time. It allows the user to access their data anytime and from anywhere. Data mining is commonly known as Knowledge Discovery in Databases (KDD), which is the non-trivial extraction of implicit, previously stored and most popularly used data from the stored data.

There are a large amount of relevant and irrelevant data being stored on the databases and many other areas. It has given rise to the term Data Mining, which can be further useful in the decision making process. It is the phenomenon of extracting useful and important data from a large amount of data being stored almost everywhere on the internet. It mostly deals with the already stored and gathered data for any other purpose instead of data mining purpose. This shows that the main objective of data mining is to collect and choose the relevant data from the previously stored data. There are enormous steps of data sets being used in data mining.

It consists of a repetitive sequence of steps:

1. Data Cleaning: It removes unwanted noises and inconsistent data.
2. Data Integration: It combines multiple data sources.
3. Data Selection: in this step, the data which is relevant is retrieved and extracted from the databases.
4. Data Transformation: in this, the data is converted into its most suitable form which can be done by performing summary or aggregation operations.
5. Data mining: It is the most important step, in which very advanced methods are used for the extraction of data.
6. Knowledge Presentation: In this step, visualization and representation techniques are used to show the mined data to the user.

Data mining is the logical mechanism used to search important and relevant data from a large amount of data stored. This technique is used to find the patterns that were already used and known. When these patterns are found they are further used to [2] make decisions for the betterment of the industries and business.

Three steps involved are:

- Exploration: It is the first step in data mining, in which the data is cleaned and converted into another format, important variables are used and the nature of data based problems are resolved.
- Pattern Identification: When the data is explored, refined and resolved for some specific purpose then a pattern is formed for the identification. In this, a pattern is identified and selected which make the best prediction.
- Deployment: In this step, different patterns are deployed to have the desired outcome.

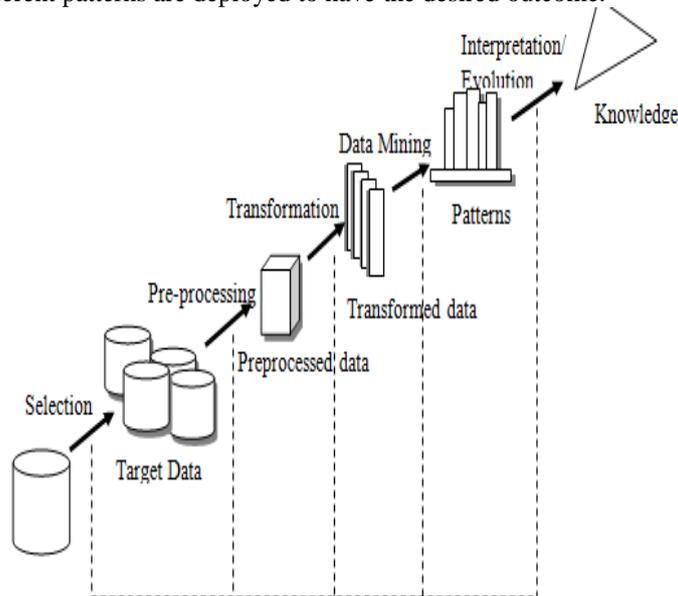


Fig. 1: Data Mining is the core of knowledge discovery datasets process

1.2 Heart diseases prediction

Heart diseases are one of the major causes of death nowadays. Smoking, consumption of alcohol in large quantity, cholesterol, and pulse rate are the reason for heart diseases. The heart is the operating system of the human body, if it will not function properly then it will directly affect the functioning of the other body parts. Some of the major factors lead to heart diseases are family history, high blood pressure, high rate of cholesterol, age, poor diet and many more. The stretching of blood vessels will increase the blood pressure which will further cause the cardiac rest. Smoking is one of the major causes of heart diseases; almost 40% of the population is dying because of this. Because it limits the supply of oxygen in our body and prevents the proper flow of blood and tightens the blood vessels. Different types of data mining techniques are employed for the prediction of data mining such as Naïve Bayes, KNN algorithm, Decision tree, Neural Network. In KNN algorithm K user-defined value is used to find the values of factors that lead to heart diseases. A decision tree is used to deploy a classified report on the heart suffering patients. The naïve Bayes is employed to predict the probability of the heart diseases. Last but not least, the neural networks are used to minimize the errors occurred at the time of prediction. By using all these techniques, the records are classified as well [3] as maintained regularly. The activity of every patient is properly checked, if there is any change, and then the level of risk is informed to the patients. With the help of all these classifiers, the doctors are able to predict the heart diseases at the very initial stage.

2. CLASSIFIERS USED IN HEART DISEASES PREDICTION

There are various classifiers used for the prediction of heart diseases like a decision tree, SVM, neural networks, naïve bayes, and KNN. These classifiers are employed and used so that the risk of heart diseases can be predicted at its early stage. This will help the doctor as well to the patient to have the proper treatment on time. This will help to diagnose the diseases at its early stages. These classifiers are briefly explained as follow:

2.1 Classification by decision tree induction

Decision tree frames the classification in the form of a tree. It divides the datasets into smaller and smaller subsets while on the other hand decision tree is incrementally developed. The final structure is the tree along with the decision nodes and leaf nodes. A decision node has two or more than two branches; leaf denotes the classification or decision. The root node is the top node which is the best predictor. It can handle both the numerical as well as categorical data.

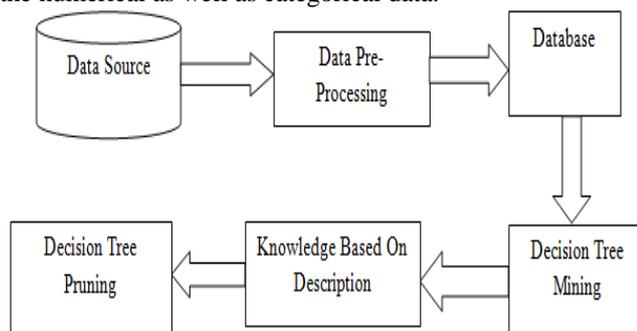


Fig. 2: Decision Tree

It is denoted by the supervised learning method as well as the statistical method used for the classification. It allows capturing uncertain conditions about the model in a fundamental way by determining possibilities of the outcomes. It is useful in solving diagnostic [4] and predictive issues. It is used to solve practical learning algorithms in which prior knowledge and observed data are combined. It underlines the prospective used for the evaluation of the learning algorithms. It evaluates the explicit probabilities and its robust noise in the input data.

2.2 K-Nearest neighbor

The classifier that utilizes training samples in order to perform learning in it is known as K-nearest neighbor classifier. A point within an n-dimensional space is represented by each sample. Further, within this n-dimensional pattern space, all the training samples are stored. The pattern space for k-training samples which are nearest to the unknown sample is identified by the k-nearest neighbor classifier in the case when an unknown sample is given. In terms of Euclidean distance, the closeness is defined here. Equal weight is assigned to each attribute by the nearest neighbor classifiers. In order to perform prediction, these classifiers can also be used due to which for given unknown sample, a real-valued prediction can be provided by this approach.

2.3 Neural Networks

According to the biological nervous system which includes multiple correlated processing constituents known as neurons, in this, a gradient approach is used by the neural network. This classifier enhances the working of learner network; the rules are extracted by the trained networks. The neurons which are arranged in an organized manner are used to perform this classifier to solve some specific prediction issues. In order to reduce the chances of errors, the structure is modified and the size is reduced. As per the information present inside and outside the network at the time of learning, the weight is adjusted in such a way that the rates of errors are very less.

2.4 Support Vector Machine (SVM)

It is an effective classifier which is used [5] to perform regression, classification as well as general pattern recognition and hence it is called SVM classifier. It highly generalized the performance without including any prior knowledge even in the case when the input space is very high; this classifier is one of the best classifiers amongst all. It is used for binary classification and for solving multiclass issues.

3. LITERATURE REVIEW

Min Chen (2017) introduced [11] a technique called data mining, which is the mechanism of discovering a new set of information from huge data. It is used to analyze a large volume of data and the patterns were extracted to convert the irrelevant information into useful information. This collected information is fed into several classifiers like KNN, naïve Bayes, SVM and decision trees, each of them performs some specific tasks. These techniques are used to predict heart diseases at their early stage. It shows very effective performance in order to achieve the correct and perfect diagnose for the heart-related diseases. There are certain advantages of this approach such as the diseases can be predicted at their very initial stages and can be diagnosed correctly and properly on time. Therefore, the researcher concluded that this method is very useful in preventing heart-related issues.

Priyanka N B.E., et. al (2017) proposed [6] two methods of data mining that is naïve Bayes and decision tree for the early prediction of severe diseases like a brain tumor, cancer, and heart attack. The results and accuracy of both the introduced methods are compared and then decides which one is better and most widely used. Due to deficient resources available in medical and health region, it was quite difficult to predict the severe diseases on time. Appropriate and suitable technologies should be used to get better and effective results. It is very important to provide good facilities for the hospitals and the patients. Data mining is specially designed to overcome this issue. Therefore, the study of the respective paper concludes that both the methods were compared and results that both are suitable and best at different fields, also provides more accurate and efficient result for heart diseases detection.

Monira Islam et.al (2017) proposed a non-invasive technique to detect the heart rate from Photoplethysmography (PPG) signal. [7] This proposed technique plays a very major role in the diagnosis of heart-related disease. A suitable heart diseases detection mechanism is used with the help of PPG and extracted from human facial videos. This can overcome the expensive ECG machine for heart-related disease detection. The FFT can detect the heart rate and re-check with the conventional heart rate using ECG machines. The contactless approach provides the solution and identifies the heart-related disease it could be diagnosed by the doctor. Therefore, the result of this approach concludes that it ensures the safety by diagnosing the diseases by its early symptoms. From the study and the experimental results it also concludes that the green channel predicts more accurate heart rate than any other. It has been found 97.7% accurate results from green channel.

Tülay Karayölan et.al (2017) proposed backpropagation algorithm for the prediction of heart diseases with the help of artificial neural networks [8]. It has some clinical features in which neural networks are used as input and is trained along with this proposed backpropagation technique. It can predict the heart-related diseases with an accuracy of 95%. The already proposed methods were not sufficient for the early prediction of heart diseases. The advancement of technology will lead to the use of machine learning techniques for the prediction of cardiac diseases in their initial stages. Therefore, the researcher draws the conclusion that the proposed approach has almost 100% accuracy in prediction heart-related diseases at their early stages. It gives better results in comparison to the other techniques.

Tahira Mahboob et.al (2017) introduced various learning practices which assist the detection of innumerable the heart disease [9]. There are certain techniques were used like Hidden Markov Models, Support Vector Machine, Computational intelligent classifier, and data mining and so on. As the treatment of the cardiac disease is very expensive and unaffordable to any normal

individual so, these types of advanced technology are developed to overcome this problem. These techniques are also useful in early stage predictions. It avoids any other future sufferings by making slight changes in daily routine. Hence, the author concludes that the predicted approach has several advantages and is very useful.

Procheta Nag et.al (2017) proposed [10] a very effective technique which is very useful in the prediction of heart diseases at the initial stage. The researcher has developed a prototype called Acute Myocardial Infarction (AMI). Heart attack having various symptoms like chest pain, breathing problem, palpitation, vomiting, and continuous sweating. Therefore, the researcher draws the conclusion that the advancement of computer technology in medical and health region provides useful aids and people are becoming more dependent on these technologies. The results of data mining are very beneficial and are used for better assistance to many physicians a lot of data is related to diseases.

4. CONCLUSION

The relevant information is fetched from the various dataset using data mining techniques. The similar and dissimilar data is clustered after calculating a similarity between different input dataset. Different approaches have been explained and from all of them, Support Vector Machine (SVM) is used. As it is used to classify both similar and dissimilar data type in which central point is calculated by calculating an arithmetic mean of the dataset. This technique is the best technique since it provides efficient results among all other classifiers. According to the type of input dataset a clustered data is classified using a decision tree classifier scheme. In this research work, the SVM classifier is used for classification of various diseases and this helps in predicting the heart disease of a patient.

5. REFERENCES

- [1] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", IJERT, Vol 1, Issue 8, 2012.
- [2] Syed Umar Amin, Kavita Agarwal, Rizwan Beg, "Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors", IEEE, 2013.
- [3] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", IJCA, Volume 47– No.10, June 2012.
- [4] P. Bhandari, S. Yadav, S. Mote, D.Rankhambe, "Predictive System for Medical Diagnosis with Expertise Analysis", IJESC, Vol. 6, pp. 4652-4656, 2016.
- [5] Nishara Banu, Gomathy, "Disease Forecasting System using Data Mining Methods", IEEE Transaction on Intelligent Computing Applications, 2014.
- [6] Priyanka N B.E., and Dr.Pushpa RaviKumar B.E., M.Tech., Ph.D., "Usage of Data mining techniques in predicting the Heart diseases – Naïve Bayes & Decision tree," 2017, IEEE
- [7] Monira Islam, Md.Ashikuzzaman, Tahmida Tabassum and Md. Salah Uddin Yusuf, "A Non-Invasive Technique of Early Heart Diseases Prediction from Photoplethysmography Signal," 2017, IEEE
- [8] Tülay Karayölan, Özkan Köllüç, "Prediction of Heart Disease Using Neural Network," 2017, IEEE
- [9] Tahira Mahboob, Rida Irfan, Bazelah Ghaffar, "Evaluating Ensemble Prediction of Coronary Heart Disease using Receiver Operating Characteristics," 2017, IEEE
- [10] Procheta Nag, Saikat Mondal, Foysal Ahmed, Arun More, and M.Raihan, "A Simple Acute Myocardial Infarction (Heart Attack) Prediction System Using Clinical Data and Data Mining Techniques," 2017, IEEE
- [11] Min Chen, Yixue Hao, Kai Hwang, Fellow, "Disease Prediction by machine learning over big data from healthy communities," 2017, IEEE, vol.15,2017, pp-215-22.