

Heart Disease Prediction using Machine learning and Data Mining Technique

Anish Xavier

Department of Information Tech Engineering
University of Mumbai
Vashi, India

Safa Sadat

Department of Information Tech Engineering
University of Mumbai
Vashi, India

Sachin Chakalal

Department of Information Tech Engineering
University of Mumbai
Vashi, India

Abstract — Nowadays Heart disease is considered one of the major causes in today's world. It cannot be easily predicted by the medical doctors as it is a difficult task that demands expertise and higher knowledge for prediction. There is a lot of data available within the healthcare systems on the internet. However, there is a lack of effective analysis tools to discover hidden relationships and patterns in data. An automated system in medical diagnosis would enhance medical efficiency and reduce costs. This web application intends to predict the occurrence of a disease based on data gathered from medical research particularly in Heart Disease. The goal is to extract the hidden patterns by applying data mining techniques on the dataset, which are noteworthy to heart diseases and to predict the presence of heart disease in patients where the presence is valued on a scale. The prediction of heart disease requires a huge size of data which is too complex and massive to process and analyze by conventional techniques. Our objective is to find out the suitable machine learning technique that is computationally efficient as well as accurate for the prediction of heart disease.

Keywords: Data mining, Heart Disease Prediction, Classification algorithms, machine learning.

I. INTRODUCTION

The highest mortality of both India and abroad is mainly because of heart disease. According to the World Health Organization (WHO), heart-related diseases are responsible for taking 17.7 million lives every year, 31% of all global deaths. Hence, this is a vital time to check this death rate by identifying the disease correctly in the initial stage. We can use data mining technologies to discover knowledge from the datasets. The discovered knowledge can be used by healthcare administrators to improve the quality of service. The discovered knowledge can also be used by medical practitioners to reduce the number of adverse drug effects, to suggest less expensive therapeutically equivalent alternatives. Anticipating patient's future behaviour on the given history is one of the important applications of data mining techniques that can be used in healthcare management. A major challenge facing healthcare organizations (hospitals, medical centres) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering effective treatments. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost

of clinical tests. They can achieve these results by employing appropriate computer-based information

and/or decision support systems. Healthcare data is massive It includes patient data, resource management data, and transformed data. Healthcare organizations must have the ability to analyse data. Treatment records of millions of patients can be stored, and computerized and data mining techniques may help in answering several important and critical questions related to health care. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcomes. This suggestion is promising as data modelling and analysis tools, e.g., data mining, have the potential to generate acknowledge-rich environment that can help to significantly improve the quality of clinical decisions.

II. LITERATURE SURVEY

There is bulk of information available on technological advances for heart disease prediction System. This includes introduction of a new classification approach for the classification of heart disease, which uses Artificial Neural Network and feature subset selection. Feature subset selection reduces the number of attributes. Pre-processing is through using Principal Component Analysis (PCA). Results demonstrate that the proposed approach indicates enhanced accuracy over traditional classification techniques.

Some of the existing systems are:

- 1) Data Mining Classification Methods in Cardiovascular Disease Prediction (Author: Dr. K. Usha Rani ,et.al)
- 2) Offline voice assistant app for android based devices Heart diseases dataset using neural network (Author: Milan Kumari,et.al)
- 3) Review of Heart disease prediction system using data mining and hybrid intelligent techniques (Author: Kiran Jyoti,et.al)

II. HEART DISEASE

As we know the heart is considered a major organ of our body. If the operation of a heart is not proper, it will affect the other body parts of a human such as a brain, kidney, etc. It is more like a pump, which flows blood in and out throughout the body. If the circulation of blood in the body is inefficient the organs like the brain suffers and if the heart stops working altogether, the end occurs within minutes. Life is completely dependent on the efficient working of the heart. The term Heart disease refers to the disease of heart & the blood vessel system within it. There are several factors which increase the chances of Heart disease:

- Hypertension
- Physical inactivity
- Poor diet
- High Blood pressure
- High blood cholesterol
- Obesity
- Cholesterol
- Resting ECG
- Thal
- Peak exercise ST segment
- Diagnosis of heart disease
- Angina

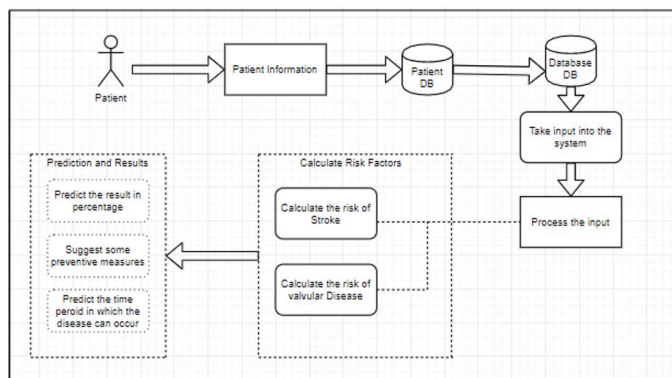


Fig: 1 Block Diagram

III. METHODOLOGY

For the implementation of the project, we used different algorithms i.e. Naïve Bayes and genetic algorithm. Heart disease prediction is a web-based machine learning application, trained by a UCI dataset. The user inputs its specific medical details to get the prediction of heart disease for that user. The algorithm will calculate the probability of the presence of heart disease. The result will be displayed on the webpage itself. Thus, minimizing the cost and time required to predict the disease. The format of data plays a crucial part in this application. At the time of uploading the user data application will check its proper file format and if it's not as per need then an error dialog box will be prompted.

1) Website: The system will consist of a website, where users will register themselves for getting the report of the

health of their heart in terms of predictive analysis about their heart disease. Users will have to fill a form initially for registration. Then the user will get redirected to the profile page where they will have to complete their profile by filling all the information related to their heart. After submitting the health information, the patient will be able to have look at the report where they will be knowing the status or risk of their heart in terms of percentage. If the user will have risk greater than 60% then the user will be redirected to another form where he will have to enter additional symptoms so that system will give a prediction about the category of heart disease from two most common categories i.e. CAD (Coronary Artery Disease) and Valvular disease

2) Database: The server will be using a MySQL database. The system 's database consists of the following tables.

Users table – This table will consist of all the user information which includes the user's name, e-mail id, phone number, address, etc.

Medical history table – This table will consist of all the health-related information of users which is related to heart that includes attributes such as age, gender, resting blood pressure, cholesterol, fasting blood sugar, old peak, etc. Data set We take the data set in this paper with 700 records and 14 attributes collected from the online dataset repository of archive.ics.edu/ml/datasets. The dataset parameters are listed in Table

IV. CONCLUSION

In this project, various suggestions and classification methods are executed on the heart datasets to predict the heart diseases. Classification algorithms are used to predict a small set of relations between attributes in the databases to build a correct classifier. The main contribution of the present study to attain high calculation accuracy for early diagnoses of heart diseases. we have evaluated the popular and effective heart disease prediction methods from the literature survey and finally, select the most effective algorithms of Naïve Bayes and Logistic regression for their performance analysis on the heart disease prediction. A skilled system is developed for the end-user to check the risk of heart diseases based on assumed parameters and the best associative classification method. The experimental results show that a large number of the rules support the better determines of heart diseases that even support the heart professional in their diagnosis in decisions.

FUTURE SCOPE

As future work, considering the more attributes as input data can expand this study. Furthermore, the work could be done on early detection of heart disease by processing family's historical data. There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system.

REFERENCES

- [1] S.. Ishtake and S. . Sanap, "“ Intelligent Heart Disease Prediction System Using Data Mining Techniques ”," International Journal of healthcare & biomedical Research, vol. 1, no. 3, pp. 94–101, 2013.
- [2] V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," Caribbean Journal of Science and Technology, vol. 1, pp. 208–217, 2013.

- [3] D. S. Chaitrali and A. S. Sulabha, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks," International Journal of Computer Engineering & Technology (IJCET), vol. 3, no. 3, pp. 30–40, 2012.
- [4] R. Rao, "SURVEY ON PREDICTION OF HEART DISEASE USING DATA MINING TECHNIQUES," International Journal of Data Mining & Knowledge Management Process (IJDMP), vol. 1, no. 3, pp. 14–34, 2011.
- [5] S. Vijayarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey," International Journal of Computer Applications & Information Technology, vol. II, no. I, pp. 17–21, 2013.
- [6] T. J. Peter and K. Somasundaram, "AN EMPIRICAL STUDY ON PREDICTION OF HEART DISEASE USING CLASSIFICATION DATA MINING TECHNIQUES," 2012.
- [7] Nidhi Bhatla Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques," International Journal of Engineering Research & Technology (IJERT), 2012.
- [8] K. Sudhakar, "Study of Heart Disease Prediction using Data Mining," vol. 4, no. 1, pp. 1157–1160, 2014.
- [9] R. Chitra and V. Seenivasagam, "REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES," Journal on Soft Computing (ICTACT), vol. 3, no. 4, pp. 605–609, 2013.
- [10] N. A. Sundar, P. P. Latha, and M. R. Chandra, "PERFORMANCE ANALYSIS OF CLASSIFICATION DATA MINING TECHNIQUES OVER HEART DISEASE DATA BASE," International Journal of Engineering Science & Advanced Technology, vol. 2, no. 3, pp. 470–478, 2012.