Visualizing and Predicting Heart Diseases with an Interactive Dash Board

A popular saying goes that we are living in an "information age". Terabytes of data are produced every day.There is undoubtedly more data available for business use than ever before. In 2022 every person generated 1.7 megabytes of data per second, and internet users create about 2.5 quintillion bytes of data per day. But having data and having insight are two very different things. Unless businesses actually *use* the data, there's no reason to boast in the amount of available data in the world or at your company. It's like saying that you own a lot of golf clubs but not knowing how to swing one. Having data does nothing for a business unless they know how to put their data to use and utilize data to improve their business processes and products. That's where Data Analysis comes in.The health care industry generates a huge amount of data daily. However, most of it is not effectively used. Efficient tools to extract knowledge from these databases for clinical detection of diseases or other purposes are not much prevalent. The aim of this paper is to summarize some of the current research on predicting heart diseases using Data Analysis techniques, analyse the various combinations of mining algorithms used and conclude which technique(s) are effective and efficient. Also, some future directions on prediction systems have been addressed.

INTRODUCTION

The heart is one of the main organs of the human body. It pumps blood trough the blood vessels of the circulatory system. The circulatory system is extremely important because it transports blood, oxygen and other materials to the different organs of the body. Heart plays the most crucial role in circulatory system. If the heart does not function properly then it will lead to serious health conditions including death.

1 . Types of Cardiovascular Diseases

 Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack).There is another heart disease, called coronary heart disease(CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries, the condition is called atherosclerosis. The development of plaque occurs over many years. With the passage of time, this plaque can harden or rupture (break open). Hardened plaque eventually narrows the coronary arteries which in turn reduces the flow of oxygen-rich blood to the heart. If this plaque ruptures, a blood clot can form on its surface. A large blood clot can most of the time completely block blood flow through a coronary artery. Over time, the ruptured plaque also hardens and narrows the coronary arteries. If the stopped blood flow isn't restored quickly, the section of heart muscle begins to die. Without quick treatment, a heart attack can lead to serious health problems and even death. Heart attackis

a common cause of death worldwide. Some of the common symptoms of heart attack [2] are as follows.

1.1. Chest pain
It is the most common symptom of heart attack. If someone has a blocked artery or is having a heart attack, he may feel pain, tightness or pressure in the chest.

1.2. Nausea,Indigestion, Heartburn and Stomach Pain
These are some of the often overlooked symptoms of heart attack. Women tend to show these symptoms more than men.

1.3. Pain in the Arms
The pain often starts in the chest and then moves towards the arms, especially in the left side.

1.4. Feeling Dizzy and Light Headed
Things that lead to the loss of balance.

1.5. Fatigue
Simple chores which begin to set a feeling of tiredness should not be ignored.

1.6. Sweating
Some other cardiovascular diseases which are quite common are stroke, heart failure, hypertensive heart disease, rheumatic heart disease, Cardiomyopathy, Cardiacarrhythmia, Congenital heart disease, Valvular heart disease, Aortic aneurysms, Peripheral artery disease and Venous thrombosis. Heart diseases may develop due to certain abnormalities in the functioning of the circulatory system or may be aggravated by certain lifestyle choices like smoking, certain eating habits, sedentary life and others. If the heart diseases are detected earlier then it can be treated properly and kept under control. Here, early detection is the main key. Being well informed about the whys and wherefores of heart disease will help in prevention summarily.

Prevalence of Cardiovascular Diseases

An estimated 17.5 million deaths occur due to cardiovascular diseases worldwide. More than 75% deaths due to cardiovascular diseases occur in the middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack. India too has a growing number of CVD patients added every year. Currently, the number of heart disease patients in India is more than 30 million. Over two lakh open heart surgeries are performed in India each year. A matter of growing concern is that the number of patients requiring coronary interventions has been rising at 20% to 30% for the past few years.

LITERATURE SURVEY

There are five research papers that explore the computational methods to predict heart diseases. The summaries of them have been presented in a nutshell.

Manpreet Singh et al. [4] proposed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM).They used Canadian Community Health Survey (CCHS) 2012 dataset. Here, twenty significant attributes were used. SEM is used to generate the weight matrix for the FCM model which then predicts a possibility of cardiovascular diseases. A SEM model is defined with correlation between CCC 121(a variable which defines whether the respondent has heart disease) along with 20 attributes. To construct FCM a weight matrix representing the strength of the causal relationship between concepts must be constructed

first. The SEM defined in the previous section is now used as the FCM though they have achieved the required ingredients (i.e. weight matrix, concepts and causality).80% of the data set was used for training the SEM model and the remaining 20% for testing the FCM model. The accuracy obtained by using this model was 74%.

Mrudula Gudadhe et al.[5] presented a decision support system for heart disease classification. Support vector machine (SVM) and artificial neural network (ANN) were the two main methods used in this system. A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease.

Asha Rajkumar et al. [6] worked on diagnosis of heart disease using classification based on supervised machine learning. Tanagra tool is used to classify the data, 10 fold cross validation is used to evaluate the data and the results are compared. Tanagra is a free Data Mining software for academic and research purposes. It suggests several Data Analysis methods from explanatory data analysis, statistical learning, machine learning and database area. The dataset is divided into two parts, 80% data is used for training and 20% for testing. Among the three techniques, Naïve Bayes shows lower error ratio and takes the least amount of time.

Shantakumar B. Patil et al. [7] obtained important patterns from heart disease database for heart attack prediction. Enormous amount of data collected by the healthcare industry is unfortunately not 'mined' properly to find concealed information that can predict heart attack. Here, the authors have proposed MAFIA algorithm (Maximal Frequent Itemset Algorithm) to do so using Java. The data is preprocessed first, and then clustered using k-means algorithm into two clusters and the cluster significant to heart attack is obtained. Then frequent patterns are mined from the item set and significance weightages of the frequent data are calculated. Based on these weightages of the attributes(ex- age, blood pressure, cholesterol and many others), patterns significant to heart attack are chosen. This pattern can be further used to develop heart attack prediction systems.

G Purusothaman et al. [8] have surveyed and compared different classification techniques for heart disease prediction. Instead of applying a single model such as Decision tree, artificial neural network and Naïve Bayes, the authors focus on the working of hybrid models i.e. models which combines more than one classification technique. They have surveyed the works of researchers who studied about the effectiveness of hybrid models. The performances of single models such as Decision tree, artificial neural network and Naïve Bayes are 76%, 85% and 69% respectively. However, hybrid approaches show an accuracy of 96%. Therefore, hybrid models lead to reliable   and promising classifiers for predicting heart diseases with good accuracy.

CONCLUSION

Heart diseases when aggravated spiral way beyond control. Heart diseases are complicated and take away lots of lives every year .When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control. However, it is always advisable to exercise daily and discard unhealthy habits at the earliest. Tobacco consumption and unhealthy diets increase the chances of stroke and heart diseases. Eating at least 5 helpings of fruits and vegetables a day is a good practice. For heart disease patients, it is advisable to restrict the intake of salt to one teaspoon per day.

One of the major drawbacks of these works is that the main focus has been on the application of classification techniques for heart disease prediction, rather than studying various data cleaning and pruning techniques that prepare and make a dataset suitable for mining. It has been observed that a properly cleaned and pruned dataset provides much better accuracy than an unclean one with missing values. Selection of suitable techniques for data cleaning along with proper classification algorithms will

lead to the development of prediction systems that give enhanced accuracy.

In future an intelligent system may be developed that can lead to selection of proper treatment methods for a patient diagnosed with heart disease. A lot of work has been done already in making models that can predict whether a patient is likely to develop heart disease or not. There are several treatment methods for a patient once diagnosed with a particular form of heart disease. Data Analysis can be of very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cardiovascular_disease.

[2] https://www.who.int/health-topics/cardiovascular-diseases/

[3] https://food.ndtv.com/health/world-heart-day-2015-heart-disease-in-india-is-a-growing-concern-ansari-1224160

[4] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago,2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map",IEEE International Conference on Fuzzy Systems (FUZZ),pp. 1377- 1382.

[5] Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010,"Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network",International Conference on Computer and Communication Technology (ICCCT),DOI:10.1109/ICCCT.2010.5640377, 17-19.

[6] Asha Rajkumar, and Mrs G. Sophia Reena, 2010, "Diagnosis of Heart Disease using Data Mining Algorithms",Global Journal of Computer Science and Technology,Vol. 10,Issue 10, pp.38-43, September.