VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

IBM Project
Team ID: PNT2022TMID39416

TEAM LEAD
S.RAJASRI– 510119104018

TEAM MEMBERS
N.ARCHANA - 510119104002
S.DHENISHA - 510119104004
M.VISHNUPRIYA - 510119104021

OF
BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE
&
ENGINEERING

1.A Literature Review on Heart Disease Prediction Based on Data Mining Algorithms

Waveform analysis, time-frequency analysis, Neuro Fuzzy RBF ANN and Total Least Square-based Prony modeling algorithms are some of the techniques used to identify heart disease in the literature. However, in a study by Marshall et al (Marshall et al 1991), classification accuracy was not good with this technique (up to 79%) and the range of improvements to select the appropriate model was still sufficient. They also demonstrated the efficiency of neural networks in diagnosing heart attacks (acute myocardial infarction) by comparing multiple neural network classifiers, the multilayer perceptron and the Boltzmann perceptron classifier. Most of these approaches relate to diagnosis, not to the understanding of fundamental knowledge. (Celia et al. 2000) concluded that attribute selection can be performed prior to genetic programming to find a high level of understandable knowledge. Following this study, Hongmei Yan developed a multiplayer perceptron-based decision support systemWaveform analysis, time-frequency analysis, Neuro Fuzzy RBF ANN and Total Least Square-based Prony modeling algorithms are some of the techniques used to identify heart disease in the literature. However, in a study by Marshall et al (Marshall et al 1991), classification accuracy was not good with this technique (up to 79%) and the range of improvements to select the appropriate model was still sufficient. They also demonstrated the efficiencyofneuralnetworks in diagnosing heart attacks (acutemyocardial infarction) by comparing multiple neural network classifiers, the multilayer perceptron and the Boltzmann perceptron classifier. Most of these approaches relate to diagnosis, not to the understanding of fundamental

2.A literature survey of predicting heart disease

Heart disease defines a range of conditions that affect human heart. The name "heart disease" is often used commonly with the name "cardiovascular disease". Heart disease is a term that allow to a large number of medical circumstances related to heart. These medical circumstances characterize the irregular health condition that directly affects the heart and all its parts. Heart disease generally allows to some conditions that involve narrowed or blocked blood vessels which can lead to a heart attack, stroke or chest pain. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease .There are various types of cardiovascular disease.The most similar types are heart failure (HF) and Coronary Artery Disease (CAD). The main root cause of heart failure (HF) is occur due to the blockade or narrowing down of coronary arteries. Coronary arteries also supply blood to the heart. Data mining is a non trivial extraction of implicit, previously unknown potential useful information called as knowledge from the medical data using complex algorithms. Big data (BD) can be referred as huge record of information set. Big Data and Data Mining are two various things. The task carried out by these two methods are similar focusing on collecting the huge amount of data, handling them and preparing report on the data by taking out the information which is knowledgeable. Data Mining is basically an activity of observing the patterns in the data which is relevant and with particular information by using Big Data. The useful patterns with hidden patterns, unknown correlation.

3.Heart Disease Prediction using Exploratory Data Analysis

A study in 2016 found that human beings are collectively generated data more than ten exabytes from various sources.Exploratory Data Analysis is a method to analyze data using advanced techniques to expose hidden structure, enhances the insight into a given dataset, identifies the anomalies and builds parsimonious models to test the underlying assumptions. Exploratory Data Analysis (EDA) is classified into Graphical or non-graphical and Univariate or multivariate Univariate data consider one data column at a time while multivariate method considers more than two variables while analyzing. The diagnostic methods of diseases are of two types namely, Invasive and Non-invasive Invasive diagnostic method includes incise procedures in which instruments are used to cut the skin, mucus membrane and connective tissues. In contrast, non-invasive methods are used to diagnose diseases without opening the skin. Some of the machine learning algorithms based on non-invasive methods are Support Vector Machine (SVM), K- means clustering, K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Naive Bayes, Logistic Regression and rough set .Predicting and diagnosing heart disease is the biggest challenge in the medical industry and it is based on factors like physical examination, symptoms and signs of the patient [1-3]. Factors which influence heart diseases are cholesterol level of the body, smoking habit, and obesity, family history of diseases, blood pressure and working environment. Machine learning algorithms play a vital and accurate role in predicting heart disease. The advancement of technologies allows machine language to pair with big data tools to handle unstructured and exponentially growing data. In the paper, K means clustering method is proposed in big data environment and the visualization is made with the tableau dashboard.

4.Use of electronic health data for disease prediction

This paper presents a systematic literature review with respect to application of data science to heart failure (HF) datasets with the intention of generating both a synthesis of relevant findings and a critical evaluation of approaches, applicability and accuracy in order to inform future work within this field. This paper has a particular intention to consider ways in which the low uptake of techniques within clinical practice could be resolved. Literature searches were performed on Scopus (2014-2021), ProQuest and Ovid MEDLINE databases (2014-2021). Search terms included 'heart failure' or 'cardiomyopathy' , 'data analytics', 'data mining' or 'data science'. 81 out of 1688 articles were included in the review. The majority of studies were retrospective cohort studies. The median size of the patient cohort across all studies was 1944 (min 46, max 93260). The largest patient samples were used in readmission prediction models with the median sample size of 5676 (min. 380, max. 93260). Machine learning methods focused on common HF problems: detection of HF from available dataset, prediction of hospital readmission following index hospitalization, mortality prediction, classification and clustering of HF cohorts into subgroups with distinctive features and response to HF treatment. The most common ML methods used were logistic regression, decision trees, random forest and support vector machines. Information on validation of models was scarce. Based on the authors' affiliations, there was a median 3:1 ratio between IT specialists and clinicians. Over half of studies were co-authored by a collaboration of medical and IT specialists. Approximately 25% of papers were authored solely by IT specialists who did not seek clinical input in data interpretation. The application of ML to datasets, in particular clustering methods, enabled the development of classification models assisting in testing the outcomes of patients with HF. There is, however, a tendency to over-claim the potential usefulness of ML models for clinical practice.

5.Improving heart disease prediction using feature selection approaches

Heart Disease is the disorder of heart and blood veins. It is very difficult for medical practitioners and doctors to predict accurate about heart disease diagnosis. Data science is one of the more important things in early prediction and solves large data problems now days. This research paper describes the prediction of heart disease in medical field by using data science. As many researches done research related to that problem but the accuracy of prediction is still needed to be improved. So, this research focuses on feature selection techniques and algorithms where multiple heart disease datasets are used for experimentation analysis and to show the accuracy improvement. By using the Rapid miner as tool; Decision Tree, Logistic Regression, Logistic Regression SVM, Naïve Bayes and Random Forest; algorithms are used as feature selection techniques and improvement is shown in the results by showing the accuracy.

REFERENCES:

[1] Apruv Patel,kunjan D. khatri,Smit Kiri,Kathan Patel Research Scholars,Department of computer engineering,Vishwakarma Government Engineering College Ahmedabad, India.

[2] Bo Jin ,Chao Che, Zhen Liu, Shulong Zhang, Xiaomeng Yin And Xiaopeng Wei, "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling" ,IEEE Access 2018.

[3] R.IndrakumariaT.PoongodibSoumya RanjanJenaca,cAssistant Professor, School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P., IndiabAssociate Professor,School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P., India.

[4] IEEE/ACM transactions on computational biology and bioinformatics 18 (2), 745-758, 2019

[5] 2019 16th international bhurban conference on applied science and technology (IBCAST), 619-623, 2019