



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 6)

Available online at: www.ijariit.com

Review on prediction of heart disease using data mining

Siddharth Mundra

siddharthmundra1@gmail.com

Manjara Charitable Trust's Rajiv Gandhi Institute of
Technology, Mumbai, Maharashtra

Kiran Manjrekar

manjrekarkiran7@gmail.com

Manjara Charitable Trust's Rajiv Gandhi Institute of
Technology, Mumbai, Maharashtra

Nimit Lalwani

nimit.lalwani@gmail.com

Manjara Charitable Trust's Rajiv Gandhi Institute of
Technology, Mumbai, Maharashtra

Nilesh Rathod

nilesh.rathod@mctrigit.ac.in

Manjara Charitable Trust's Rajiv Gandhi Institute of
Technology, Mumbai, Maharashtra

ABSTRACT

The heart is the next major organ comparing to the brain which has more priority in the Human body. Prediction of occurrences of heart diseases in the medical field is significant work. Data analytics is useful for prediction from more information and it helps medical centers to predict various diseases. A huge amount of patient-related data is maintained monthly. The stored data can be useful for the source of predicting the occurrence of future disease. Some of the data mining and machine learning techniques are used to predict heart disease, such as Artificial Neural Network (ANN), Decision tree, K-Nearest Neighbor(KNN), Naïve Bayes and Support Vector Machine (SVM). This paper provides an insight into the existing algorithm and it gives an overall summary of the existing work.

Keywords— Cardiovascular disease, Data mining, Classification, Naive Bayes

1. INTRODUCTION

Heart disease is one of the prevalent diseases that can lead to reducing the lifespan of human beings nowadays. Each yearly 18.5 million people are dying due to heart disease. Life is dependent on component functioning of heart because the heart is a necessary part of our body. Heart disease is a disease that affects the function of the heart. An estimate of a person's risk for heart disease is important for many aspects of health promotion and clinical medicine. A risk prediction model may be obtained through a multivariate regression analysis of a longitudinal study. Due to digital technologies are rapidly growing, healthcare centers store huge amount of data in their database that is very complex and challenging to analysis. Data mining techniques and machine learning algorithms play vital roles in the analysis of different data in medical centers. The techniques and algorithms can be directly used on a data set for creating some models or to draw vital conclusions, and inferences from the data set.

Data Mining is the study of substantial data sets to extricate the hidden and formerly unidentified patterns, relationships, and knowledge that are hard to investigate with conventional measurements. Data mining methods are the consequence of a long method of research and product improvement. Data Mining is alienated into two assignments such as Predictive Tasks and Descriptive Tasks. Predictive Tasks forecast the estimation of an explicit attribute based on other attributes. Categorization, Regression and Deviation Deduction come under Predictive Tasks. Descriptive Tasks design that outlines the connection between the data. Clustering, Association Rule Mining, and Sequential Pattern Discovery are future under Descriptive Tasks. Data Mining comprises of few steps from raw data collection to some form of new knowledge. The iterative process comprises of following stages like Data Integration, Data cleaning, Data transformation, Data Selection, Data Mining, Knowledge Representation, and Pattern Evaluation.

Common attributes used for heart disease are Age, Sex, Fasting Blood Pressure, Chest Pain type, Resting ECG(test that measures the electrical activity of the heart),Number of major vessels colored by fluoroscope, Fasting blood sugar, Exang(exercise included angina), ST depression (finding on an electrocardiogram, trace in the ST segment is abnormally low below the baseline),Threshold Pressure (high blood pressure), Serum Cholesterol(determine the risk for developing heart disease), Thalach(maximum heart rate achieved) ,painless (chest pain location substernal=1, otherwise=0)),, smoke, Hypertension, Food habits, weight, height and obesity.

2. LITERATURE SURVEY

Nahar, Jesmin, et al., [1], examines the sick and healthy aspects that contribute to cardiopathy for males and females. Association rule mining, a machine data technique, is employed to find these parts and conjointly the UCI Cleveland dataset, organic information, is taken into account aboard the three rule generation algorithms –Predictive Apriori, Apriori, and Tertius.

Sudha [2] this paper investigates the classification tree ways in the data processing. The target of this paper is to explore the investigational results of the performance of varied categorization techniques for a cardiopathy dataset. The classification tree algorithms used and experimented during this work area unit Random Forest, call Stump and LMT Tree algorithmic program.

Chaurasia, Vikas, and Saurabh Pal [3] this analysis paper proposes to deliver a survey of existing ways of information discovery in databases victimization data processing techniques which is able to be valuable for medical practitioners to yield undefeated call. The foremost goal of this work is to forecast is to determine the existence of cardiopathy with faded variety of attributes.

[4] Have done piece victimisation, one partial clustering algorithmic program (K-Means) and one hierarchical clustering algorithmic program (agglomerative). K-means algorithmic program has higher effectiveness and quality and converges fast once production with massive information sets. Hierarchical bunch constructs a hierarchy of clusters by either often merging 2 smaller clusters into a bigger one or splitting a bigger cluster into smaller ones. Using WEKA, they calculate the performance of k-means and graded bunch algorithmic program on the premise of accuracy and runtime.

Research paper [5] intends to provide a survey of current techniques of knowledge discovery in databases victimisation processing techniques that are in use in today's medical analysis notably in heart disease prediction. It is discovered that decision tree was in shut competition with Bayesian, whereas algorithms like KNN, Neural Networks, etc. The reduction of actual data size by genetics additional improved the potency.

Jayami Patel et al, [6] urged cardiopathy prediction using data mining and machine learning algorithmic programs. The goal of this study is to extract hidden patterns by applying processing techniques. The best algorithmic program J48 supported UCI information has the most effective accuracy rate compared to LMT.

Ashwini Shetty et al, [7] suggested developing the prediction system which is able to diagnose heart disease from patient's medical dataset. 13 risk factors of input attributes have taken into thought to form the system. After the analysis of knowledge from the dataset, information cleansing and information integration were performed.

Sharan Monica.L [8] planned analysis of cardiovascular disease. This paper projected processing techniques to predict diseases. It will provide the survey of current techniques to extract data from dataset and it will be useful for attention practitioners. The performance is also obtained on the basis of time taken to develop the decision tree for the system.

Shabana Asmi P et al [9], Authors have worked to form the system a lot of accurately predict the presence of cardiovascular disease with adscititious attributes of the disease using association rules.

S. Seema et al,[11] focuses on algorithms that can predict chronic disease by mining the datasets using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure better performance at an accurate rate. From this analysis, SVM gives highest accuracy rate.

R. Ansari et al [12] reviewed data mining on heart disease prediction and diagnosis. This study tries to support healthcare people to early analyze cardiovascular disease and evaluate related risk factors. At last the main cardiovascular disease diagnosis indices were recognized using expert's beliefs. Then, data mining techniques were applied to cardiovascular related dataset. Finally, cardiovascular illness attributes were identified and a model was developed based on extracted rules and classifiers. Visual Studio was used for the coding of algorithm.

3. CONCLUSION AND FUTURE WORK

Heart disease is one of the leading causes of death worldwide and also the early prediction of cardiopathy is vital. The data mining-assisted cardiopathy prediction system helps the doctor as a tool for cardiopathy diagnosis. From the analysis it's ended that, data mining plays a significant role in cardiopathy classification. In the above paper we have studied various algorithms that can be used for classification of heart disease databases, also we have seen different techniques that can be used for classification and the accuracy obtained by them. There are many possible improvements that could be explored to improve accuracy of this heart disease prediction system. Due to time limitation, the following research/work need to be performed for the future and would like to make use of testing different data mining techniques and different decision tree and to explore different rules such as association rule, logistic regression, and clustering algorithms.

4. REFERENCES

- [1] Nahar, Jesmin, et al., "Association rule mining to detect factors which contribute to heart disease in males and females", Expert Systems with Applications, Vol. 40, No. 4, pp. 1086-1093, 2013.
- [2] Vijayarani, S., and S. Sudha, "An efficient classification tree technique for heart disease prediction", International Conference on Research Trends in Computer Technologies (ICRTCT-2013) Proceedings published in International Journal of Computer Applications (IJCA)(0975-8887). Vol. 201, 2013
- [3] Chaurasia, Vikas, and Saurabh Pal, "Data mining approach to detect heart diseases", 2014.

- [4] Nidhi Singh and Divakar Singh, "Performance Evaluation of K-Means and Hierarchical Clustering in Terms of Accuracy and Running Time", Ph.D. Dissertation, Department of Computer Science and Engineering, Barkatullah University Institute of Technology, 2012.
- [5] Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.
- [6] Jayami Patel, Prof. Tejal Upadhyay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.
- [7] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277- 281
- [8] Sharan Monica.L, Sathees Kumar.B, "Analysis of cardiovascular Disease Prediction using Data Mining Techniques", International Journal of Modern Computer Science, vol.4, 1 February 2016, pp.55-58.
- [9] Asmi, Shabana P., and S. Justin Samuel. "An analysis and accuracy prediction of heart disease with association rule and other
- [10] Nilesh Rathod, Seema Shah, Kavita Shirsat,"An Interactive Online Training & Placement System", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 12, December-2013
- [11] Dr.S.Seema Shedole, Kumari Deepika, "Predictive analytics to prevent and control chronic disease", <https://www.researchgate.net/publication/316530782>, January 2016.
- [12] R. Ansari, P. Azimi, and M. R. Taghva, "Heart Disease Diagnosis Using Data Mining Techniques," 2017.