RESEARCH ARTICLE OPEN ACCESS

Application of Data Mining Techniques for Prognostication of Heart Malady: A Comprehensive Review

Khalid Amin Wani[1], Manjit Kaur[2]
[1]M.Tech Scholar, [2]Assistant Professor,
Department of Computer Science,Lovely Professional University,Punjab

Abstract:

This paper provides an insight on various various techniques, tools andapproaches of data mining for assay of heart maladies.The use of different data mining tools and techniques are showing successful results in various applications one of them being disease diagnosis. Being lethal of all, heart disease is considered as one of the major causes of death throughout the world. Its prognostication requires a high level of expertise and knowledge.Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Data mining makes use of sophisticated mathematical algorithms to segment the data and evaluate the probability of future events. Data mining can be helpful in answering the questions that cannot be addressed through simple query and reporting techniques.

Keywords **— Data mining, Heart Malady.**

I. INTRODUCTION

Heart diseaseremains the greatest cause of deaths for the last two decades. As of late computer technologies built up certain product's to help specialists in making decision regarding heart disease in the beginning. Forecasting the heart disease fundamentally relies upon clinical information of patients. Forecasting of Heart maladies can help specialists for anticipating heart disease current status in view of the clinical information of different patients. Life is,completely subjected to compelling working of the heart. The term Heart disorder implies sickness of heart and vessel system inside it. There are number of components which construct the peril of Heart contamination some of which are:High cholesterol**,** Unhealthy diet**,** Smoking**,** Lack of physical activity, Obesity.

Heart malady is a wide term that consolidates distinctive sorts of afflictions impacting assorted fragments of the heart. A few sorts of Heart maladies are:- Coronary illness, Angina pectoris Congestive heart disappointment Cardiomyopathy, Innate Coronary illness, Arrhythmia, Myocarditis.

II. DATA MINING AND SOME TECHNIQUES

It is defined as obtaining the valuable information from the large amount of data. It can be also defined as mining ofknowledge from complex data. The various data mining tools and techniques can be utilized to discover unacquainted patterns and trends from an enormous amount of data. The main objective of data mining is to discover the patterns in the available data set with less user effort and inputs available. The main contribution of data mining is in decision

making and auguring market trends. Some data mining techniques are:

A. ASSOCIATION

It is the best known and very much inquired about technique for information mining. It is additionally called connection strategy since designs which are found from the dataset depend on the connection between the things.

B. CLASSIFICATION

Classification is a mining utility which causes us to dole out the things in a gathering to target classes. Principle reason for the classification is to anticipate precisely the objective class for each case in the dataset.

C. CLUSTERING.

Unlike classification and prediction, which examine class-named data objects, clustering dissects data objects without a known class name.

D. PREDICTION.

It is an information mining procedure which finds connection amongstdependent and independent variables.

III. DATA MINING TOOLS

There are various tools used for data mining purpose. These are WEKA, TANAGRA, MATLAB and .NET FRAMEWORK.

*WEKA:*It is a data mining tool which was developed in New Zealand by the University of Waikato that implements data mining algorithms using JAVA language. WEKA is a collection of machine learning algorithms and their application to the data mining problems. These algorithms are directly applied to the dataset. WEKA supports data file in ARFF format. WEKA is open source software and hence, it is not dependent on any platform. It includes algorithms for data processing, classification, regression, clustering, association and also visualization tools.

*TANAGRA:*It is open source software as researchers can access to the source code and add their own algorithms and compare their performances, if it conforms to the software distribution license. It includes several data mining algorithms from statistical learning, machine learning and data analysis and database area.

*MATLAB:*It is a data mining tool built in high level language. It provides interactive environment for visualization, numerical computation and programming. The built in math functions, language and tool explore various approaches and helps to reach a solution faster than with the spreadsheet of traditional programming languages like C,C++ and JAVA. It analyze data, develop algorithms, and create models and applications.

*.NET FRAMEWORK:*It is a software framework developed by Microsoft which runs primarily on Microsoft windows. It provides secure communication and consistent applications. It provides language interoperability (each language can code written in other languages) across several programming languages.

*RAPID MINER:*Rapid Miner is undeniably the world prominent open source system for data mining. It is available as a separate application for data analysis and as a data mining engine for the integration into own products. Thousands of applications of

Rapid Miner in more than 40 countries give their users a competitive edge.

IV. METHODOLGY USED IN DATA MINING

Data Mining is essential part of Knowledge Discovery Database (KDD). Many people treat Data Mining as a synonym for KDD since it's a key part of KDD process and consists of an iterative sequence of the following steps:

Data Cleaning – Here weremove noise or irrelevant data in our datasets.

Data Integration - Here multiple data sources may be combined.

Data Selection - Where data relevant to the analysis task are retrieved from the database.

Data Transformation - Where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.

Data Mining - An essential process where intelligent methods are applied in order to extract data patterns.

Pattern Evaluation - To identify the truly interesting patterns representing knowledge based on some interestingness measures.

Knowledge Presentation - knowledge representation techniques are used to present the mined knowledge to the user.

V. LITERATURE SURVEY

Dr. Neeraj Bhargava et al. [1] in year 2017 attempted an examination on utilization of mining calculation (basic CART) so as to anticipate the heart assaults and to analyze the best accessible strategy for expectation. The prescient precision controlled by SIMPLE CART (79.90%) calculation recommends that parameters utilized are dependable markers to anticipate the nearness of coronary illness.

Anuragbhatt, et al.[2] used information mining approach to predict and analyze cardiovascular disease. They provided the experimental analysis using the dataset made available online by UCI machine learning repository using the WEKA tool. J48 and Naïve Bayes algorithm were used to perform this analysis. The datasets were investigated utilizing two distinct strategies i.e. first only selected attributes were taken and then all attributes were taken together. Using J48 anprecision of 82.3% with all attributes and 65.64% with selected attributes was achieved while as Naïve Bayes showed the precision of 98.64% using all attributes and 93.2% using selected attributes.

Jagdeepsingh et al.[3] proposed a hybrid model for prediction of heart disease in year 2016. In their study they proposed a technique that can generate classification association rules (CAR'S). Using this method they were able to predict which method gives the best prediction or accuracy in predicting heart disease. Their proposed work achieved an accuracy of 99.19% using ibk (nearest neighbor) with aprior associative algorithm.

KaliaOrphanou andArianna Dagliati[4] used Naïve Bayes classifiers combined with temporal association rules for coronary heart disease diagnosis in year 2016. In their study they compared performance of two classifiers and reached to the conclusion that

periodic classifiers are more accurate than baseline classifiers having accuracy of 71% and 68% respectively.

ParisaNaraei et al.[5] in year 2016 used two different algorithms, "Support Vector Machines" and "Multilayer Perceptron Neural Networks" utilized using WEKA tool to classify a heart disease dataset. The accuracy of classification results of Multilayer Perceptron and Support Vector Machines was found to be 84.48% and 80.52% respectively. In their study 8 attributes were chosen and result of their analysis showed the strength of SVMs in classification of medical data.

T.Santhanam and E. P. Ephzibah[6] proposed a system in year 2015 that will help physicians in earlier prognostication of heart disease. The proposed system has been devised using two computing techniques like genetic algorithms and fuzzy logic. Among all the other classification and prediction models their model provided an accuracy of 86%. With the help of use of genetic algorithms in their study they were able to reduce the number of attributes from thirteen to seven.

K. PrasannaLaxmi and Dr. C.R.K Reddy[7] in year 2015 proposed an efficient technique for heart disease prediction. They proposed a model based on stream associative classification heart disease prediction system. They used SACHDP on various datasets and it showed consistent results when compared with other associative classification techniques. The datasets used for their study were collected from UCI repository which included heart, breast, Pima, hepatitis and lymph disease datasets. As per the results SACHDP outperformed the other traditional associative classification techniques with an average accuracy of 94.94%.

Purushottam et al.[8]in year 2015planned a framework that could effectively mine the guidelines to foresee the hazard level of patients in light of the given parameters about their wellbeing. The experiment was performed on the Cleveland dataset which was collected from UCI repository using 14 attributes. WEKA tool was used for dataset analysis and knowledge extraction based on evolutionary learning (KEEL) was used to find out the classification decision rules. Total correctly classified instances were 86.66%.

M.A.NisharaBanuand B. Gomathy[9] utilized C4.5 algorithm, Maximal frequent item set algorithm (MAFIA) and K-means clustering in the year 2014 using 13 attributes in the dataset and achieved precision of 89%. In their paper they used K-means to cluster relevant data in a database, MAFIA is then applied for finding maximal frequent patterns in heart disease database and C4.5 is used as a training algorithm to classify frequent patterns into different classes.

Ms.Ishtake and Prof. Sanap S.A.[10]Built up a framework for heart malady prediction utilizing decision tree, Neural Network and Naive Bayes procedures utilizing 15 traits in the year 2013.All the three models were able to answer the complex queries each with its own strength with respect to the ease of model interpretation, access to detailed information and accuracy.

M. AkhilJabbar et.al.[11]in year 2013 proposed way to forecast heart malady by a technique called lazy associative classification. The proposed technique was used upon seven data sets collected from UCI repository and one real life dataset of heart patients from Andhra Pradesh. The average accuracy achieved by them was 81.66% using the same technique and in particular on heart diseases the accuracy achieved was 90% when compared with the other Traditional techniques such as Naïve Bayes and J48.

Syed Umar Amin et al.[12] Presented a strategy for forecast of coronary illness utilizing real hazard factors. The system includes two fruitful information mining devices viz. neural systems and genetic algorithms. The framework was produced utilizing Matlab and anticipated the danger of coronary illness with an exactness of 89%.According to results produced they were able to show that genetic algorithm and neural network approach have shown strength in forecasting of heart malady than traditional ANN.

Chaitrali S. Dangare and Sulabha S. Apte[13]in year 2012 analyzed forecasting system for heart malady by making use of three data mining techniques namely Decision tree, Naïve Bayes and Neural Networks. The addition of two more attributes (obesity and smoking) to the previously defined 13 attributes gave them the accuracy of 99.62%, 90.74% and 100% respectively. They concluded that out of these three prediction models Neural Networks predicts the heart disease with highest accuracy.

Yan Zhang et al. [14] in year 2012 made the use of support vector machine method which is based on statistical learning theory to diagnosis of coronary heart disease. Using a total number of 13 attributes the classification accuracy of three kernel functions was determined and they found out that Rbf kernel function has shown the highest classification accuracy of 88.6% followed by linear and then polynomial kernel function

Beenishfida et.al. [15]in year 2011 proposed a classifier ensemble method for an efficient heart disease diagnosis. In this study classification is done through homogenous ensemble and final results obtained are optimized using genetic algorithms. They found out that the proposed method performed best on Cleveland dataset with an accuracy of 98.65%.

Asha Rajkumar and Ms. G.SophiaReena [16] in year 2010 compared the performance of three algorithms viz. Naïve Bayes, Decision tree, K Nearest Neighbor algorithms for heart disease diagnosis and found out that Naïve Bayes has best compact time for processing dataset (609ms) and gives an accuracy of 52.33% as compared to other algorithms.

SellappanPalaniappan and RafiahAwang[17] build an intelligent prediction system for heart disease diagnosis using three data mining models (Naïve Bayes, Decision tree and Neural Networks). Naive Bayes gives the highest probability (95%) with 432 supporting cases, followed closely by Decision Tree (94.93%) with 106 supporting cases and Neural Network (93.54%) with 298 supporting cases. In their study they

found that intelligent heart disease prediction system was able to answer complex 'What If' queries which conventional decision support system cannot.

VI. CONCLUSION

The purpose of our work is to provide a study of different data mining techniques that can be engaged in automated heart disease prediction systems. Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that different techniques are applied in above mentioned papers using different number of attributes. So, different techniques have shown the difference in accuracy to each other. Also applying hybrid data mining techniques has shown promising results in the diagnosis of heart disease. Neural networks also showed the efficient and promising results and thereby can be effectively used in heart disease diagnosis.In this work, we have performed a literature survey on various papers. In future, we are planning to propose an effective disease prediction system to predict the heart disease with better accuracy using different data mining techniques.

VI. REFERNCES.

[1] N. Bhargava and S. Dayma, "An Approach for Classification using Simple CART Algorithm in Weka," pp. 212–216, 2017.

[2] A. Bhatt, S. K. Dubey, and A. K. Bhatt, "Data Mining Approach to Predict and Analyze the Cardiovascular Disease," 2017 Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Advances in Intelligent Systems and Computing 515, DOI 10.1007/978-981-10-3153-3_12.

[3] J. Singh, A. Kamra and H. Singh, "Prediction of heart diseases using associative classification," *2016 5th International Conference on Wireless Networks and Embedded Systems (WECON)*, Rajpura, 2016, pp. 1 -7.

[4] K. Orphanou, A. Dagliati, L. Sacchi, A. Stassopoulou, E. Keravnou, and R. Bellazzi, "Combining Naive Bayes Classifiers with Temporal Association Rules for Coronary Heart Disease Diagnosis," pp. 81–92, 2016.

[5] P. Naraei, V. Street, V. Street, and V. Street, "Application of Multilayer Perceptron Neural Networks and Support Vector Machines in Classification of Healthcare Data," no. December, pp. 848–852, 2016.

[6] T. Santhanam and E. P. Ephzibah, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model," vol. 8, no. May, pp. 797–803, 2015.

[7] K. P. Lakshmi and C. R. K. Reddy, "Fast rule-based heart disease prediction using associative classification mining," *2015 International Conference onComputer, Communication and Control (IC4)*, Indore, 2015, pp. 1-5.

[8] Purushottam, K. Saxena and R. Sharma, "Efficient heart disease prediction system using decision tree," *International Conference on*

Computing, Communication & Automation, Noida, 2015, pp. 72-77.

[9] M. A. N. Banu and A. D. Preprocessing, "2014 International Conference on Intelligent Computing Applications Disease Forecasting System Using Data Mining Methods," pp. 1–4, 2014.

[10] "' Intelligent Heart Disease Prediction System Using Data Mining Techniques ,'" no. April, pp. 94–101, 2013.

[11] M. A. Jabbar, "Heart Disease Prediction using Lazy Associative Classification," pp. 40–46, 2013.

[12] S. U. Amin, K. Agarwal, and R. Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors," no. Ict, pp. 1227–1231, 2013.

[13] C. S. Dangare and M. E. Cse, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," vol. 47, no. 10, pp. 44–48, 2012.

[14] Y. Zhang, F. Liu, D. Li, and X. Zhou, "Studies on application of Support Vector Machine in diagnose of coronary heart disease," 2012.

[15] B. Fida, M. Nazir, N. Naveed, and S. Akram, "Heart Disease Classification Ensemble Optimization Using Genetic Algorithm," pp. 19–24, 2011.

[16] A. Rajkumar and G. S. Reena, Algorithm, vol. 10, no. 10, pp. 38–43, 2010.

[17] S. Palaniappan and R. Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," pp. 108–115, 2008.