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

[1]Apruv Patel, [2]kunjan D. khatri, [3]Smit Kiri, [4]Kathan Patel

Research Scholars,
Department of computer engineering,
Vishwakarma Government Engineering College Ahmedabad, India

***Abstract*— The medical sector processes vast amounts of data on a regular basis. Handling large data in the old way can affect the results. Advanced data mining techniques are especially used in heart disease prediction to find facts about databases and medical research. Heart disease is the world's largest cause of death. The tremendous amount of data generated for the prediction of heart disease is too difficult and wasteful to process and analyze in the conventional way. Data mining provides methodologies and techniques to transform these mounds into useful information for decision-making. Using data mining algorithms, you can quickly predict disease with high accuracy. In this paper, a single or hybrid combination of data mining algorithms can be used to investigate several papers used in cardiac disease prediction to identify algorithms for future research with high accuracy.**

***Index Terms*— Data mining; Heart disease prediction; Data mining techniques.**

I. INTRODUCTION

Data mining is a technology that is implicit in the available data and has never been previously known, making it accurate in the future. In short, it is the process of analyzing data from different points and gathering knowledge from it. Data mining blends statistical analysis, machine learning, and database technology to extract unknown patterns and relationships from a large database. Learning knowledge can be useful in medical applications, i.e. health care. Health Informatics is a fast-growing field in which evolving computer science and information technology are interested in medical and health data. Medical Data Mining is a challenging area that involves many misconceptions and uncertainties.

Heart disease

The heart is an important organ of all living things that plays an essential role in pumping blood into the remaining organs through the blood vessels of the circulatory system. If the blood circulation in the body is improper, the organs like the brain will suffer, the heart will stop functioning completely and death will occur. Life depends entirely on proper functioning of the heart. The term heart disease refers to diseases of the heart and cardiovascular system. Factors that increase the risk of heart disease:

- Family history
- Age
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

In many cases, diagnosis is generally based on patient's current test report & doctor's experience. Thus the diagnosis is a complex task that requires high skill & much experience.

II. LITERATURE SURVEY

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 system

to support the diagnosis of heart disease (Hongmei Yan 2006). This system is well designed as an input layer of the system with 40 input variables and is categorized into four groups and then encoded using the proposed coding scheme.

Huiyan Wang has pursued a niche in this field by proposing a computerized diagnostic model to promote standardization and popularization of traditional Chinese medicine (TCM) diagnosis (Huiyan Wang 2008). The system works as follows. First, the symptoms are selected by learning the Bayesian network structure from a database of cases that incorporate mutual information theory. In the structure, the Markov blanket of the target variable is selected as the symptom set. Second, the mapping relationship between symptom sets and diagnostic results is based on naive Bayesian classifiers.

The next task in the field was by Huang Jianyong (Hu Jianyong 2004). This study aimed to solve the problem of multi - class classification using binary classifiers. The binary classification tree design is done in such a way that the class group is divided into two different subgroups from one node. The node adopts the class module method to improve the binary classification ability. Partitioning is formulated as an optimization problem, and genetic algorithms are proposed to solve optimization problems. Heart murmurs were an important problem in pediatric cardiovascular disease, and this group had a high incidence of cardiac noise (77-95% as reported), resulting in little congestion due to congenital heart disease. Sanjay (Sanjay 2005) has developed a reliable test device to diagnose murmur in pediatrics.

Carlos Ordonez studied the prediction of heart disease with the help of association rules. They used a simple mapping algorithm. This algorithm treats attributes as numeric or categorical. It is used to convert medical records into transaction formats. Enhanced algorithms are used to minify restricted association rules. The mapping table is prepared and the attribute value is mapped to the item. Decision trees are used for data mining because they automatically divide numeric values . Split points selected in the decision tree are rarely used. Clustering is used to gain an overall understanding of the data.

Deepika studied Pruning-Classification Association Rule (PCAR) to extract association rules. PCAR based on analysis of Apriori algorithm. PCAR covers a set of minimum frequency entries, which Removes irregular items from a set of items. It then sorts the set of items based on the frequency of the set of items and finds a frequent set of items.
Ordonez uses association rules in data mining to obtain better heart disease prediction results. The authors studied the limitations of association rules that mined the entire data set without verification of independent samples. We introduced algorithms that modified the search constraints to reduce the number of association rules and validated them using train and test approaches. They have studied two complementary tasks: predicting absence and predicting the presence of heart disease.

P. Chandra and M. Jabbar used the feature subset selection to create class association rules to detect heart disease. Association rules determine the relationship between attribute values and classification to predict classes in a patient data set. Functional selection, such as genetic search, determines the characteristics that help predict heart disease.
Usha Rani proposed a system for predicting heart disease with a combination of feedforward and backpropagation algorithms with the help of artificial neural networks. Experiments are performed considering single and multi-layer neural network models. Parallel processing is implemented to accelerate the learning process in each neuron of all hidden output layers.

R. Setthukkarasi has developed a new neural fuzzy technique to diagnose the facts of the disease in patient reports. The generalized database is configured for decision-making from a reduced set of attributes, which is the output of the genetic algorithm. Four layered fuzzy neural networks are used for efficient modeling and inference with time dependence under uncertainty.
Chaitrali Dangare has implemented a system for predicting heart disease and has applied three data mining classification techniques: Decision trees, Naive Bayes & Neural Networks. The results show that neural networks are superior to decision trees and Naive Bayes.

M. Akhil jabbar, B.LDeekshatulu, and Priti Chandra propose a new algorithm that combines KNN and Genetic Algorithm for efficient classification. Optimal Solution Perform a global search for complex large and multiple modal data sets to provide genetic algorithms. It is also observed from the results that hybridization with KNN is well performed and provides great accuracy.
Ankita Dewan proposed an efficient genetic algorithm mix using backpropagation for cardiac disease prediction. They concluded that neural networks are the best of all classification techniques for nonlinear data. The BP algorithm is the best classifier of the Artificial Neural Network, a common training method. Where the primary system output is compared to the desired output and the system is adjusted until the difference between the two is minimized. However, there is a drawback to being trapped in local minima.

Shadab Adam Pattekari and Asma Parveen used the Naive Bayesian Classification technique to develop decision support in the heart disease prediction system. The system finds hidden knowledge from past heart disease databases. This is the most effective model for predicting heart disease. The model can respond to complex queries and has its own strength in terms of ease of model interpretation, access to more information and accuracy.

R. Bhuvaneswari et al. use the Naive Bayes classifier for medical use. The authors used two well-known algorithms, the Back Propagation Neural Network (BNN) and the Nave Bayesian (NB) data mining classification, to study the previous experience and to calculate the probability of an object among all objects. Bayesian techniques have been developed for probability concepts. The previous backend is computed by bay rules based on the exact nature of the probability model, and the Naive Bayes classifier is used to study very efficiently in the supervised learning environment.

Nilakshi P. Waghulde and Nilima P. Patil performed cardiac disease data set experiments with multi-layer neural networks in conjunction with back propagation learning algorithms for network training. A weighting algorithm is used to optimize neural network initialization. This study shows the results of the Genetic Neural Network for heart disease prediction, using an optimal neural network architecture that predicts whether a patient is suffering from heart disease, improving accuracy by 98%.

K. Srinivas proposed the application of data mining techniques to the prediction of heart attack. The authors studied classification-based data mining techniques such as rule-based, decision trees, Naive Bayes, and artificial neural networks for a vast amount of medical data. The Tanagra data mining tool was used to examine data analysis, machine learning, and statistical learning algorithms. They used a training data set of 3000 instances with 14 different properties. Instances of the data set represent the results of various test types to predict the accuracy of heart disease.

Sudha et al. studied classification algorithms such as Naive Bayes, Decision tree, and Neural Network to detect stroke disease. Classification techniques such as decision trees, Bayesian classifiers, and backpropagation neural networks were used in this study. Records with irrelevant data were deleted from the data warehouse before the mining process occurred.

D. P. Shukla et. al. performed a task of designing a system to recognize the probability of coronary heart disease. He divided all the parameters into two levels based on the importance of the parameters and gave each step extra weight. Finally, the final decision is made considering both levels. The authors proposed a neuro-fuzzy integration approach at two levels. This error rate is very low and work efficiency is high. The authors concluded that this method could be used to perform an analysis of other diseases.

III. CONCLUSION

This questionnaire investigated a variety of existing technologies that supported data mining to predict heart disease. In this survey, we learned how to apply data mining techniques to predict heart attack. Previously, leaving the system was mostly a design using a single algorithm with poor accuracy. In some researchers' studies we have observed that more accuracy can be achieved through hybridization of two or more algorithms.

REFERENCES

[1]. Ankita Dewan, Meghna Sharma," Prediction of Heart Disease Using a Hybrid Technique inData Mining Classification", 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706.

[2]. R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, R. Boghrati, et al., "Diagnosis of coronary arteries stenosis using data mining," J Med Signals Sens, vol. 2, pp. 153-9, Jul 2012.

[3]. Carlos Ordonez, Edward Omincenski and Levien de Braal ,"Mining Constraint Association Rules to Predict Heart Disease", Proceeding of 2001, IEEE International Conference of Data Mining, IEEE Computer Society, ISBN-0-7695-1119-8, 2001, pp: 433-440

[4]. Deepika. N, "Association Rule for Classification of Heart Attack patients ", IJAEST, Vol 11(2), pp 253-257, 2011

[5]. Usha. K Dr, "Analysis of Heart Disease Dataset using neural network approach", IJDKP, Vol 1(5), Sep 2011.

[6]. R. Setthukkarase, Kannan, "An Intelligent System for mining Temporal rules in Clinical database using Fuzzy neural network",European Journal of Scientific Research, ISSN 1450-216, Vol 70(3), pp 386-395, 2012.

[7]. M Akhil Jabbar, BL Deekshatulu, Priti Chandra," Heart disease classification using nearest neighbor classifier with feature subset selection", Anale. Seria Informatica, 11, 2013

[8]. Chaitrali S Dangare, Sulabha S Apte." Improved study of heart disease prediction system using data mining classification techniques", International Journal of Computer Applications, 47(10):44–48, 2012.

[9]. Shadab Adam Pattekari and Asma Parveen," PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624, Vol 3, Issue 3, 2012, pp 290-294.

[10]. R. Bhuvaneswari and K. Kalaiselvi, "Naïve Bayesian Classification Approach in Healthcare Applications", International Journal of computer Science and Telecommunication", vol. 3, no. 1, pp. 106-112, Jan 2012.