

A Survey to Predict Cardiovascular Disease Using Data Mining

R Pavithra¹, V Pavithra²

PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India ¹

Asst. Professor, Department of Computer Science, Stella Maris College, Chennai, India ²

ABSTRACT: Cardiovascular diseases are the most common cause of humanity and injury in India. People of all ages are affected particularly the elderly people. They are most prone to emerging heart diseases because of old age. A record of people pretentious by heart disease is of utmost importance. This is essential to predict the group of people commonly affected, age, sex, risk factors, diagnosis and management of heart diseases. A huge number of people are pretentious by heart diseases and hence it is quite difficult to predict accurately. Data mining technique helps to collect file and predict the population affected by heart problems accurately. Data mining is used in various fields. Data mining procedures serve as a key for advancement in the field of medicine. It is very easy to access the information about heart problems and their risk factors including the hidden information for proper diagnosis and treatment. Proper mining methods can save an enormous number of people from mortality due to heart diseases. Various forms of algorithms have been used in data mining.

KEYWORDS: Neural networks, Naïve Bayes, Decision tree, the k-means clustering algorithm, Association rule.

I. INTRODUCTION

Data mining are used for an automatic extraction of the hidden information's from large data set, they are massive. It is a cooperative effort of humans and computers. It is a process that is developed to examine large amounts of data routinely collected which can be used efficiently in the medical organization. In hospitals and medical centers where they generate huge data's which contains a wealth of hidden information, data mining is used to extract the analytic to identify the best way that improves care and reduce costs. Using data mining in the health field is an intelligent diagnostic tool. Data mining also said as Knowledge Discovery in Data (KDD).

1.1 Heartdisease

Cardiovascular diseases are the most common problem faced by the community at present. Almost more than half of the men alone at the age of 60 suffer from some form of heart Diseases and once affected there are explored to lifelong treatment. The common cause of death is due to improperly diagnosed condition and wrong treatment. The most important fact that is essential to control the death rate is proper intervention. This can be achieved by collecting adequate information about the heart diseases.

This procedure known as data mining concert the knowledge we have about heart diseases into data from so that it is easy for studies and researcher. A large amount of data are analyzed in data mining to get an adequate result. Not only known but hidden data is tooanalyzed.

The diagnosis of cardiovascular diseases is relatively a challenge to predict. Thus, an attempt to utilize knowledge about patient's data in form of database is easy. The poor decision may leads to death and hence this computer based information can aid in achieving proper clinical test at a low-cost using data mining techniques. This ensures that the patients are properly diagnosed and treated.

The use of data mining in predicting Heart diseases has found good solutions. This also talks about affected people and to healthy people and people as rich in the community. The ultimate goal is to collect high needed information from row level data's.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

1.2 Major riskfactor

There is the number of danger factor in cardiovascular diseases. Risk factors can either be changeable or unchangeable. Changeable is the one which can be reduced/controlled so that its effect can be reduced to avoid heart diseases. Unchangeable is the one which cannot be controlled/modified.

➤ ChangeableRisk

ATTRIBUTES	DESCRIPTION
Smoking	Damages the blood vessel wall & increases risk of heart diseases.
Blood pressure	Increased blood pressure causes like heart failure and thickening of muscles.
Obesity(overweight)	An abnormal blood lipid (fat) increases risk of coronary heart disease (angina/heart attack).
Diabetes.	High blood sugar which will increase risk of heart diseases
Excessive Alcohol	Consuming alcohol is harmful because it directly damages the heart muscle and cause irregular beating of the heart.
High cholesterol	It increases risk of heart disease due to high (much) cholesterol (fatty substance) in blood.
Diet	An unhealthy diet can lead to high cholesterol and high blood pressure which leads chance in heart disease.

Table-1: Modifiable Risk

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

➤ UnchangeableRisk

ATTRIBUTES	DESCRIPTION
Age	Age increases risk also increases. <ul style="list-style-type: none"> • Men>45 • Women> 55
Sex	Men's have greater risk than women
Genetic	Increases due to family genes i.e., blood relatives

Table-2: Non-Modifiable Risk

1.3 Types of heartdiseases

There are various kinds of heart diseases, which affects people are given below.

TYPES	DESCRIPTION
Angina	Plaque (contains materials such as fat & cholesterol) which restrict oxygen to heart muscles than it needs to function properly.
Coronary artery disease	Plaque which restricts the flow of blood to heart muscles than it needs to function properly.
Arrhythmia	An irregular /non-uniform/ abnormal heartbeat i.e., too slow, too fast.
Heart failure	<i>Congestive heart failure</i> or CHF when part of the heart muscle is so damaged when the heart is not able to pump sufficient blood through the body.
Congenital heart disease	Type of heart disease that causes problems with the heart at birth.

Table-3: Types of cardiovascular diseases

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

II. DATAMINING TECHNIQUES IN CARDIOVASCULARCARE

2.1 NeuralNetworks

Neural Networks (NN) are also known as an Artificial Neural Network (ANN). The neural network is a statistical model. They are used frequently to generate precise grades. It is used to manage when the range of the data is enormous. The neural networks are mainly used to perform data mining tasks. These techniques are used to design a model which is complex, which helps to establish a relationship between input and output. It also helps to find patterns in item set. Feed forward neural network model and back propagation learning algorithm is used frequently for fetching the data that are accurate and sufficient. Feed-forward network architecture includes nodes such as input nodes, hidden nodes, output nodes. It begins with input node, which is determined by finalized data. The determination of hidden node is done through trial and error and finally, the output data is produced. Back- propagation is logical based, it works similar to feed-forward algorithm works. Only that the difference is the flow of data, its helps in calculating the error and transmit it back to the previous layers. With back-propagation algorithm and feed-forward algorithm, the neural networks are well skilled to predict cardiovascular disease. The technique neural networks had exactness 84.9% of fetching data's. The problem with neural network is that it cannot be retained; it is not possible to add editing data with the existing one. Among other techniques this technique has many advantages, high accuracy to get sufficient and perfect data.

2.2 k-means ClusteringAlgorithm

K-mean clustering algorithm is a data mining techniques. It is one of the broadly familiar clustering tools which are used to explore irregularities in the data distribution. It is the simplicity and fast to analyze a huge amount of data, it is a data machine/mining learning algorithm used to cluster. Medical imaging, bio metrics most frequently uses K- means algorithm. It groups the observation into the cluster of related observations. It first clusters observations into K distinct clusters, where k (positive integer) is provided as an input parameter. Then the collected data's are assigned to the group of the centroid adjacent. A centroid is the ordinates that are taken as an average of each coordinate of a sample of the point that allocated the clusters. The purpose of using k-mean clustering is to emphasize the overall squared error function or inter-cluster deviation. Thus, the k-mean clustering algorithm is used in prediction of heart diseases to get accurate and sufficient data. K-mean algorithm is the one which is mainly used in medical applications if a low number of data's exists then the initial grouping will determine the significant cluster. It has the worst runningtime.

2.3 Naïvebayes

The naïve Bayes is a data mining (machine learning) technique is applicable mainly when the dimensional on the input is large. Naive Bayes follows steps which are simple. It is a categorization technique based on Bayes' Theorem. It is used to produce models with the predictive capability and description of patients with cardiovascular diseases. The sympathetic data can be provided in new ways. It shows the probability of each attribute for the state of predictable attributes. That helps in the conversion of data set into a frequency table. Naïve Bayes is used when data is high. The attributes declared are self-sufficient of other attribute and accept more efficient results. Naïve Bayes algorithm is used to get adequacy data in predicting heart diseases patients. It has an accuracy of 72% which is sufficient to identify the physical features of patients who suffer from heart diseases.

2.4 Decisiontree

A decision tree is a tree-like structure that consists of a root node, branches, and leaf nodes. Decision tree in the technique used to build a tree, which is used to apply to the data. It contains internal nodes which indicates attribute, and each branch represents result that has been analyzed. Each leaf node indicates label of class. The tree contains a topmost node which is said to be the root node of the structure. The size of the tree is reduced by removing over fitting data by using Tree pruning technique. This technique gives data that are adequacy and flexible. It has an accuracy of 75.8%. A decision tree is used for feature selection process to get adequacy data are in the prediction of heart disease. Even a little change in input data causes huge changes in the tree. This data mining technique is used in health care

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

organization to predict cardiovascular diseases which produce sufficient and accurate data.

2.5 Associationrule

Association rule is a data mining technique. It is the rule-based method and an analyzing method for data mining. An association rule is used to find regular patterns, association structures among the given item sets/data's from the database. Which can be later used forecast and analyzing. The interaction between correlation items is uttered as association rules. This technique consists majorly two parts, if (predecessor) part and then (result) part. "if" part is the one in which has the data's that are found from the database. "then" part is the one which has the data's that has a combination with the "if" part. Hence, these data mining techniques are used frequently in health care organization. it provides an indication that how data have frequently appeared. The data's which are appeared frequently can be used in prediction of cardiovascular diseases.

III. LITERATURESURVEY

In this paper [5], they use artificial neural networks to predict heart diseases that use hybridization of feed forward and back propagation algorithm. A consideration of single and multilayered neural network models are accomplished for an experiment. The process implemented to speed up learning process at each neuron in all hidden and outputlayer.

In order [7] to get better heart disease prediction result, it produces a model using rule algorithm by calculating the correlation between independent and dependent variables and significantly to improve performance, it applies Jelinek-Mercer smoothing models to Naive Bayesian Classification Algorithm for a classifier results.

The researchers [2] had used certain association rules in data mining for an improved prediction of heart diseases and also concise the association rule. The entire data should be revised instead of an independent sample. The remodeled algorithm with search constraints to minimize the number of association rule and it is validated using train and testapproach.

In this paper [4] various technologies such as Rule set classifiers, Decision Tree Algorithms Neural Network Architecture and Neuro-Fuzzy has been used in this study. Tanagra Data mining tool has been used for per data analysis and warning. The instances in the data set are used to predict the accuracy of heart diseases.

The researchers [3] had put forth MAFIA i.e., Maximal Frequent Item set Algorithm. This had been used to extract data relevant to heart diseases. The pattern not able to heart diseases is chosen based on calculated significant weight age that uses Clustering Using K-Means Algorithm and Frequent Pattern Mining Using Mafia which are used for improving the prediction system.

This paper [1] presented on multiple associative rules for an accurate and efficient classification. A method called predominant correlation used as a fast filter method identifier features which are more relevant.

In paper [8] by hybridizing multiple techniques, a new system evolved showing better results in a diagnosis of heart diseases. The technique operated by training the neural network with the selected pattern for diagnosis of heart diseases, which also followed the generic algorithm for optimization of the network.

This [6] paper used the Naive Bayesian classification data mining technique in the prediction of heart diseases, even complex queries can be handled by this method, which will be used to extract even hidden information. These techniques are used to model interpretation to access the details which are adequacy.

IV. CONCLUSION

A complete elaboration about the role of data mining in a prediction of heart disease has been presented. From this, we come to know that data mining is a milestone in the field of medicine as it plays a major role in effective diagnosis. Not only in medicine, data mining has stepped into all field as it is the easiest and excellent method to collect and maintain sources. Because of its accuracy in adequacy data used in worldwide.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

REFERENCES

- [1]. Li, W., Han, J., & Pei, J. (2001). CMAR: Accurate and efficient classification based on multiple class- association rules. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on (pp. 369-376). IEEE.
- [2]. Ordonez, C. (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334-343.
- [3]. Patil, S. B., & Kumaraswamy, Y. S. (2009). Extraction of significant patterns from heart disease warehouses for heart attack prediction. IJCSNS, 9(2), 228-235.
- [4]. Srinivas, K., Rao, G. R., & Govardhan, A. (2010, August). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In Computer Science and Education (ICCSE), 2010 5th International Conference on (pp. 1344-1349). IEEE.
- [5]. Rani, K. U. (2011). Analysis of heart diseases dataset using neural network approach. arXiv preprint arXiv:1110.2626.
- [6]. Pattekari, S. A., & Parveen, A. (2012). Prediction system for heart disease using Naïve Bayes. International Journal of Advanced Computer and Mathematical Sciences, 3(3), 290-294.
- [7]. Patil, R. R. (2014). Heart disease prediction system using Naive Bayes and Jelinek-mercer smoothing. International Journal of Advanced Research in Computer and Communication Engineering, 3(5), 2278- 1021.
- [8]. Bajaj, P., & Gupta, P. (2014). Review on Heart Disease Diagnosis Based on Data Mining Techniques. International Journal of Science and Research (IJSR), 3(5).
- [9]. Gandhi, M., & Singh, S. N. (2015, February). Predictions in heart disease using techniques of data mining. In Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on (pp. 520-525). IEEE.
- [10]. Patel, A., Gandhi, S., Shetty, S., & Tekwani, B. (2017). Heart Disease Prediction Using Data Mining.