

A Survey on Heart Disease Prediction using Hybrid Technique in Data Mining

Ankita Pimputkar¹, Prof. J. S. Dhobi²

¹Student (Master of Engineering), Computer Engineering, Government Engineering College, Gandhinagar, Gujarat, India

²Associate Professor, Computer Department, Government Engineering College, Gandhinagar, Gujarat, India

Abstract

Data Mining techniques have been widely used in These methods have been extremely successful in planning clinical emotionally supportive networks as a result of their capacity to find concealed examples and connections in medicinal information. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Heart disease prediction is treated as a most complicated task in the field of medical science. Coronary illness forecast is dealt with as a most confused errand in the field of therapeutic science. The massive amounts of data generated for prediction of heart disease which is too difficult and baggy to be processed and analyzed by conventional methods. Data mining provides the methodology and technology to transform these data into useful information for decision making. Use of data mining algorithms will result in quick prediction of disease with high accuracy.

Keywords: Data mining, Heart Disease, Coronary Disease(CVD), Data Mining Techniques

1. Introduction

1.1 Data Mining:

Due to a wide availability of huge amount of data and a need to convert this available huge amount of data to useful information necessitates the use of data mining techniques. Data Mining and KDD have become popular in recent years. The popularity of data mining and KDD shouldn't be a surprise since the size of the data collections that are available are far too large to be examined manually and even the methods for automatic data analysis based on classical statistics and machine learning often face problems when processing large, dynamic data collections consisting of complex objects.

Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific and medical research. The widening gap between data and information calls for a systematic development of data mining tools that will turn data tombs into "golden nuggets" of knowledge (Han and Kamber).

Data mining refers to extracting or "mining" knowledge from large amount of data. The notation of finding useful patterns in data have been given a variety of names including data mining, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing but recently the terms data mining and KDD are dominating in the Management Information Science (MIS) communities and database fields. Data mining is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

Health Informatics is a rapidly growing field that is concerned with evolving Computer Science and Information Technology to medical and health data. Medical Data Mining is a domain of challenge which involves a lot of misdiagnosis and uncertainty. A general framework proposed for medical data mining is shown in fig⁸.

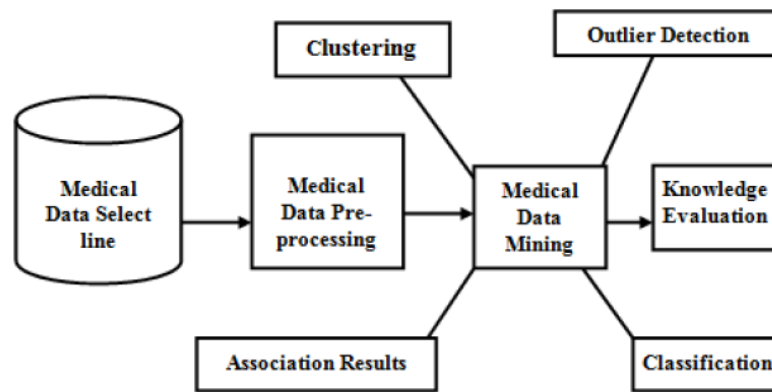


Fig -1 Framework of Medical Data Mining

1.2 Data Mining Tasks:

- As Han and Kamber stated data mining tasks are used to specify the kind of patterns to be found in data mining process. In general, data mining tasks can be classified into two categories: **Predictive** and **Descriptive**.
- A Predictive model makes a prediction about values of data using known results found from different data and its goal is to Identify strong links between variables of a data table (columns) while the, examples include providing a diagnosis for a medical patient on the basis of a set of test results, estimating the probability that customers will buy product A given a list of other products they have purchased, or predicting the value of something six months from now, given current and past values of the index .
- Descriptive model identifies patterns or relationships in data. It simply summarizes data in convenient ways or in ways that will lead to increased understanding of the way things work. For example, a physician might be interested in discovering the influence of climate among malaria patients by grouping patients in different climate zones, or may be a physician might want to cluster contraceptive users on their education background.

1. Association Rule Discovery:

- One of the best known data mining techniques. A pattern is discovered based on a relationship of a particular item on other items in same transaction. It tell us about the relationship between attributes used for analysis.

2. Classification:

- Classification analysis is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection.
- Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects.

3. Prediction:

- Prediction has attracted considerable attention given the potential implications of successful forecasting in a business context.
- There are two major types of predictions: one can either try to predict some unavailable data values or pending trends, or predict a class label for some data. Regression Analysis is generally used for prediction. Prediction can also be used for identification of distribution trends based on available data.

4. Clustering:

- Similar to classification, clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes.
- Clustering is also called unsupervised classification, because the classification is not dictated by given class labels.
- There are many clustering approaches all based on the principle of maximizing the similarity between objects in a same class (intra-class similarity) and minimizing the similarity between objects of different classes (inter-class similarity).

1.3 Application of Data mining in Health Care

- Treatment effectiveness
- Healthcare management
- Customer relationship management
- Fraud and abuse
- Medical Device Industry
- Hospital Management
- Disease Prediction

2. Heart Disease

- The heart is important organ or part of our body. Life is itself dependent on efficient working of heart. If operation of heart is not proper, it will affect the other body parts of human such as brain, kidney etc. Heart is hollow muscular organ that pumps blood through the body. The heart, blood, and blood vessels make up the circulatory system, which is responsible for distributing oxygen and nutrients to the body and carrying away carbon dioxide and other waste products.
- It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it. It is also known as a Cardiovascular Disease.
- CVD refers to a class of disease that involve a heart or blood vessels. It is commonly related to atherosclerosis, a process whereby fatty deposits ("plaques") form in your arteries, causing them to narrow and possibly block completely.
- When atherosclerosis affects the major arteries in the body it can cause a heart attack, stroke or peripheral arterial disease.
- Heart Attack: Heart attack occurs when a blood clot blocks the blood vessel that supplies blood to the heart muscle itself. Angina (temporary chest pain or discomfort) is often a warning sign of an impending heart attack.
- Stroke: Stroke occurs when blood flow to the brain is interrupted.
- As the leading cause of death in the world, heart disease, according to WHO, accounts for 3.8 million and 3.4 million deaths in males and females, respectively.

Symptoms

Symptoms of Heart Disease are as follows:

- Discomfort, pressure, heaviness, or pain in the chest, arm, or below the breastbone.
- Discomfort radiating to the back, jaw, throat, or arm.
- Fullness, indigestion, or choking feeling (may feel like heartburn).
- Sweating, nausea, vomiting, or dizziness.
- Extreme weakness, anxiety, or shortness of breath.
- Rapid or irregular heartbeats

Risk factors for Heart Disease:

- Age
- Angina
- Blood cholesterol levels
- Diabetes
- Diet
- Genes
- Hypertension
- Obesity
- Physical Inactivity
- Smoking
- Work

3. Related Work

3.1 Literature Survey:

1. Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors¹.

- In this paper, proposed approach focused at developing an intelligent data mining system based on genetic algorithm optimized neural network for the prediction of heart disease based on risk factors categories. It first determine the number of inputs, layer and hidden neurons of the neural network and then it uses the back propagation algorithm to train the networks using the weights optimized by GA.

2. Classification of Heart Disease using K-Nearest Neighbor and Genetic Algorithm².

- In this paper, proposed approach combines KNN with GA for effective classification. Proposed approach consists of two parts: 1. evaluating attributes using genetic search. 2. Building classifier and measuring accuracy of classifier. In this approach accuracy of heart disease data is improved 5% over classification algorithm without GA.

3. Diagnosing of Heart Diseases using Average K-Nearest Neighbor Algorithm of Data mining³.

- In this paper, to get improved classification accuracy and efficiency a new approach called AKNN is proposed. To remove the drawbacks and to make the KNN a faster algorithm AKNN is proposed here. In case of AKNN, super sample is created for each class which is the average of every training sample in that particular class. When the test samples are given the AKNN searches sample data and find the closest to the input. The closest neighbor is identified by measuring distance between the neighbors. The proposed approach gives higher efficiency and reduces complexity based on attribute reduction.

4. Classification and Prediction of Heart Disease Risk using Data Mining Techniques of Support Vector Machine and Artificial Neural Network⁴

- In this paper, the proposed approach incorporates the classes of heart disease utilizing Support Vector Machine and ANN. This paper proposes the architecture which includes Pre-processing, Preparing, Training and Testing with individual models, Evaluation of results and Prediction of heart disease risk. The proposed approach test and evaluate the projected system model in terms of specificity, precision, accuracy and sensitivity.

5. Heart Disease Prediction using Naïve Bayes⁵

- In this paper, the proposed approach uses Naïve Bayes Classifier. It proposed a web application that enables users to induce instant steerage on their cardiopathy through an intelligent system online. The application is fed with numerous details and also the heart disease related to those details. The application permits user to share their heart connected problems. It then processes user specific details to see for numerous health problem that might be related to it.

6. Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace smoothing Technique⁶

- This paper approach uses classification algorithm namely Naïve Bayes algorithm and a smoothing technique Laplace smoothing. It uses details such as age, gender, fasting blood sugar, cholesterol, blood pressure etc. to predict the heart disease. Smoothing technique is used to make an approximating function that attempts to capture important patterns in the data while avoiding noise or other fine scale structures or rapid phenomena. If we end up with a probability value of zero for some value of attribute it will return a zero probability. To avoid this problems Laplace smoothing technique is used known as Laplacian correlation or Laplacian estimator.

7. Heart Disease Prediction System using Data Mining Method⁷

- This paper proposes a heart disease prediction system based on three different data mining techniques. 1. Naïve Bayes, 2. J48 Decision tree, 3. Random Forest. The proposed approach predict the heart disease by using medical profiles and also the performance is compared by calculation of confusion matrix. Overall system provides high performance and accuracy.

3.2 Comparative Study of Existing System:

Table 3.1 Comparative Study of Existing System

No.	Title	Techniques	Strong Points	Weak Points
1.	Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors.	GA,NN	Predicts the risk of heart disease with an accuracy of 89%	Time for optimization high, very slow convergence
2.	Classification of Heart Disease using K-Nearest Neighbor and Genetic Algorithm.	KNN, GA	Accuracy of heart disease dataset improved 5% over classification algorithm without GA.	Low efficiency, Dependency on the selection of good values of k, not good to find local optima, identifying suitable fitness function is difficult
3.	Diagnosing of Heart Diseases using Average K-Nearest Neighbor Algorithm of Data mining.	AKNN	To remove voting drawback of KNN and makes KNN faster AKNN is used.	KNN takes more time to train data, Does not work if data is noisy.
4.	Classification and Prediction of Heart Disease Risk using Data Mining Techniques of Support Vector Machine and Artificial Neural Network	SVM,ANN, BPNN	Accuracy obtained using SVM and ANN is 84.7% and 81.8% respectively	Needs training, Neural Networks cannot be retrained.
5.	Heart Disease Prediction using Naïve Bayes.	Naïve Bayes	Proposes web application that enables user to induce instant steerage on their cordiopathy through an intelligent system online	Class conditional independence, loss of accuracy, Probability value zero problem, very strong assumption on the shape of data distribution.
6.	Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace smoothing Technique.	Naïve Bayes, Laplace Smoothing Technique	Provides 86% accuracy, probability value zero problem is solved	dependencies cannot be modelled
7.	Heart Disease Prediction System using Data Mining Method	Naïve Bayes, J48, Random Forest	Reducing error occurrence	Loss of accuracy, dependencies cannot be modelled by Naïve Bayes

4. Problem Statement

Data mining technique have been widely used in clinical decision support system for prediction and diagnosis of various disease with good accuracy. As huge amount of information is produced in a medical fields yet this information is not properly utilized. The health care system is data rich however knowledge poor. There is an absence of successful analysis methods to find connection and pattern in health care data. For this reason data mining techniques can be utilized. According to literature survey I found certain limitation for data analysis like accuracy, speed, error rate etc.

5. Conclusion

Heart disease is the leading cause of death for the two men and ladies. Know the notice signs and side effects of a heart assault with the goal that you can act quick on the off chance that you or somebody you know may show at least a bit of kindness assault. The odds of survival are more prominent when crisis treatment starts rapidly. This paper principally concentrates on the investigation of different methodologies of heart assault sickness forecast explore papers are broke down and considered. The prediction of existing frameworks can be enhanced, so In future, new algorithms and methods are to be produced which defeat the disadvantages of the current system.

7. References:

- [1] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg , “Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors” , IEEE Conference on Information and Communication Technologies, IEEE, 2013.
- [2] M. Akhil jabbar, B.L Deekshatulu, Priti Chandra, “Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm”, International Conference on Computational Intelligence: Modeling Techniques and Applications, 2013.
- [3] C. Kalaiselvi, PhD, “Diagnosis of Heart Disease Using K-Nearest Neighbor Algorithm of Data Mining”, IEEE, 2016.
- [4] Mrs S. Radhimeenakshi, “Classification and Prediction of Heart Disease Risk using Data Mining techniques of Support Vector Machine and Artificial Neural Network ”, IEEE, 2016
- [5] Garima Singh, Kiran Bagwe, Shivani Shanbhag, Shraddha Singh , Sulochana Devi, “Heart Disease Prediction using Naïve Bayes”, International Research Journal of Engineering and Technology (IRJET), March-2017.
- [6] Vincy Cherian, Bindu M. S., “ Heart Disease Prediction using Naïve Bayes Algorithm and Laplace Smoothing Technique”, International Journal of Computer Science Trends and Technology (IJCST), Mar-Apr 2017.
- [7] Keerthana T. K., “ Heart Disease Prediction System using Data Mining Method”, International Journal of Engineering Trends and Technology”, May 2017.
- [8] Dr. B. Umadevi, M. Snehapriya, “ A Survey on Prediction of Heart Disease Using Data Mining Techniques”, , International Journal of Science and Research, April 2017.
- [9] Ankita Dewan, Meghna Sharma, “Prediction of Heart Disease using Hybrid Technique in Data Mining Classification”, IEEE,2015.
- [10] Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, ELSEVIER.