

A REVIEW ON DATA MINING FOR HEART DISEASE PREDICTION

Gunsai Pooja Dineshgar

Department of Computer Engineering,
HJD institute Technology and Research Kera,
Kutch-Gujarat

Mrs. Lolita Singh

Assistant Professor of Computer Department
HJD institute Technology and Research Kera
Kutch-Gujarat

ABSTRACT:- This paper intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques. The Healthcare industry is generally “information rich”, but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. Advanced data mining techniques are used to discover knowledge in database and for medical research, particularly in Heart disease prediction. The system uses medical terms such as sex, blood pressure, cholesterol etc like attributes to predict the likelihood of patient getting a Heart disease. The performance of these techniques is compared, based on accuracy. The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHGPS) using data mining modeling technique, namely, Clustering. it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. The data mining Clustering techniques, clusters Method are used partitioning method K-mean and K-medoid algorithm are analyzed on Heart disease database.

Keywords- Data Mining, Heart disease Prediction, clustering, K-Mean, K-Medoid.

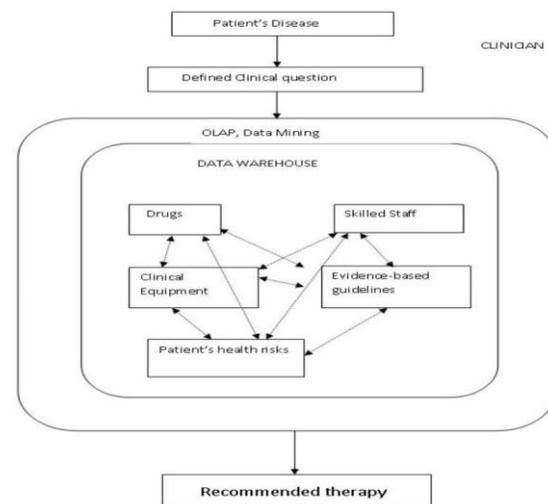
1. INTRODUCTION

Data mining refers to extracting or “mining” knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named “knowledge mining from data,” which is unfortunately somewhat long. “Knowledge mining,” a

shorter term, may not reflect the emphasis on mining from large amounts of data.[10].

Data Mining in Healthcare

The growth of medical databases is very high. This rapid growth is the main motivation for researchers to mine useful information from these medical databases. Data mining techniques play an important role in finding patterns and extracting knowledge to provide better patient care and effective diagnostic capabilities. Data mining provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with traditional statistical methods. These techniques and methods are based on statistical techniques, visualization, machine learning, Fig.1 depicts how data mining play an important role in modern clinical practice [12]



2. Heart Disease Prediction (HDP)

Heart disease prediction plays a vital role in data mining because in worldwide most of the death occur in heart

diseases. Medical diagnosis plays an important role and yet complicated task that needs to be executed efficiently and accurately. To reduce cost for achieving clinical tests an appropriate computer based information and decision support should be needed. Comparative studies of various techniques are available for an accurate implementation and good efficient for automated systems [17]. There are different types of heart diseases. They are discussed as follows:

Coronary Artery Disease- When the combination of fatty material, calcium and scar tissue (plaque) builds up in the arteries that supply the heart with blood through this, the disease should develops. Through these arteries called the coronary arteries, the heart muscle (myocardium) gets the oxygen and other nutrients it needs to pump blood.

Coronary artery disease is America's No.1 killer, affecting more than 13 million Americans. *Enlarged Heart*

Cardiomegaly- A heart condition that causes the heart to become larger than normal as a result of heart disease.

Cardiomegaly is most often linked to high blood pressure, but it can also occur as a result of other heart conditions, such as congestive heart failure, and other non-cardiac causes such as long-term anemia.

Heart Attack- A heart attack is the death of, or damage to, part of the heart muscle because the supply of blood to the heart muscle is severely reduced or stopped.

Heart Valve Disease- Valvular heart disease refers to several disorders and diseases of the heart valves, which are the tissue flaps that regulate the flow of blood through the chambers of the heart.

Congenital Heart Disease- Congenital heart disease refers to a problem with the heart's structure and function due to abnormal heart development before birth. Congenital means present at birth [18].

Heart disease is caused by disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. The major causes of cardiovascular disease are tobacco use, physical inactivity, an unhealthy diet and

harmful use of alcohol. These are the three causes of heart diseases (1) chest pain (2) stroke and (3) heart attack. To prevent and identification of these diseases different techniques of data mining is used through this easily find out heart related diseases and this is the aim of this research studies.

Many studies have been done that have focus on prediction of heart disease. Researchers have applied different data mining techniques for predicting heart disease and achieved different probabilities for different prediction models. The performances of the algorithms are compared with each other based on Sensitivity, Specificity, Accuracy, True Positive rate and False Positive rate.

2.1 Benefits of Developments of Disease Prediction

1. Prevention and Diagnosis:
2. Workout of Treatment Plan:
3. Reduction of Cost of Patient Management:

Attributes and a class attribute. Both test and training data are used for performance analysis. In a trained network, the test data is given as the input. the output of the net is calculated. In future this work will be enhanced by applying Clustering techniques using partitioning based method [19].

Attribute are:-

- 1) Age
- 2) Gender
- 3) Smoking
- 4) Heart rate
- 5) Cholesterol
- 6) Blood pressure
- 7) Blood sugar

And introduce two new attribute in my research that are

- 1). L.V.F
- 2). C.C.F

3. DATA MINING TECHNIQUES

3.1. Clustering:-

Clustering and classification are both fundamental tasks in Data Mining. Classification is used mostly as a supervised learning method, clustering for unsupervised learning (some clustering models are for both). The goal of clustering is descriptive, that of classification is predictive. Clustering groups data instances into subsets in such a manner that similar instances are grouped together, while different instances belong to different groups [13].

Type of clustering Method

1. Partitioning Methods
2. Hierarchical Methods
3. Density-Based Methods
4. Grid-Based Methods

Partitioning Method

Partitioning methods relocate instances by moving them from one cluster to another, starting from an initial partitioning. To achieve global optimality in partitioned-based clustering. Iterative optimization partitioning algorithms are subdivided into k-medoids and k-means methods [13].

3.1.1 K-Medoid Algorithm

Two early versions of k-medoid methods are the algorithm PAM (Partitioning Around Medoids) and the algorithm CLARA (Clustering LARge Applications). PAM is iterative optimization that combines relocation of points between perspective clusters with re-nominating the points as potential medoids. CLARANS (Clustering Large Applications based upon RANdomized Search) in the context of clustering in spatial databases. CLARANS uses random search to generate neighbors by starting with an arbitrary node and randomly checking max neighbor neighbors.

- 1 Select k representative objects arbitrarily
- 2 For each pair of non-selected object h and selected object i , calculate the total swapping cost TC_{ih}

- 3 For each pair of i and h ,
 - If $TC_{ih} < 0$, i is replaced by h
 - Then assign each non-selected object to the most similar representative object
- 4 repeat steps 2-3 until there is no change.

3.1.2 K-Means Algorithm

The k-means algorithm takes the input parameter, k , and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the inter cluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or center of gravity.

Algorithm: k-means. The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

1. Partition objects into k nonempty subsets
2. Compute seed points as the centroids of the clusters of the current partition. The centroid is the center (mean point) of the cluster.
3. Assign each object to the cluster with the nearest seed point.
4. Go back to Step 2, stop when no more new assignment.

3. Literature Survey

1. Durairaj M, Revathi V [1] have used MLP Back Propagation Nerual Network is proposed to predict the Heart Disease. the accuracy of the different training functions of Multilayer Perceptron Algorithm were compared to select the best taining function. The TRAINBR algorithm gives more prediction accuracy than other algorithms. The experimental results show the encouraging results that the MLP with proper training algorithm can be an effective tool to predict Heart Disease with improved accuracy.

2. AKASH JARAD1, ROHIT KATKAR2, ABDUL REHAMAN SHAIKH3, ANUP SALVE [2] have used 14 attribute to predict heart disease. Many more input attributes can be taken but our goal is to predict with less number of attributes and faster efficiency to predict the risk of having Classification method to predict heart disease K-nn, Decision List Algorithm ,Naïve Bayes , performance is based on accuracy and the time taken to build the model. Naïve Bayes algorithm gives 52.23% of accurate result.

3. Ms. Shinde Swati B.[3] have used the heart disease using data mining techniques such as Naive Bayesian and KNN algorithm etc. used Text Mining for mining the data which is not structured. can also integrate data mining and text mining. This system generates and extracts hidden information i.e. patterns and relationships between different attributes from the historical heart related database. accuracy are respectively 86% and 76% .

4. Dr. G. Rasitha Banu MCA., M.Phil., Ph.D.1, J.H.BOUSAL JAMALA MCA [7] have used new unsupervised classification system is adopted for heart attack prediction at the early stage using the patient's medical record. FCM clustering algorithm for finding the risk of heart attack of a patient using the profiles collected from the patients. data sets obtained from 270 patients, shows that the classifier has achieved better accuracy than most of the existing algorithms. The performance of the proposed FCM is proved to be a well known approach in terms of accuracy.

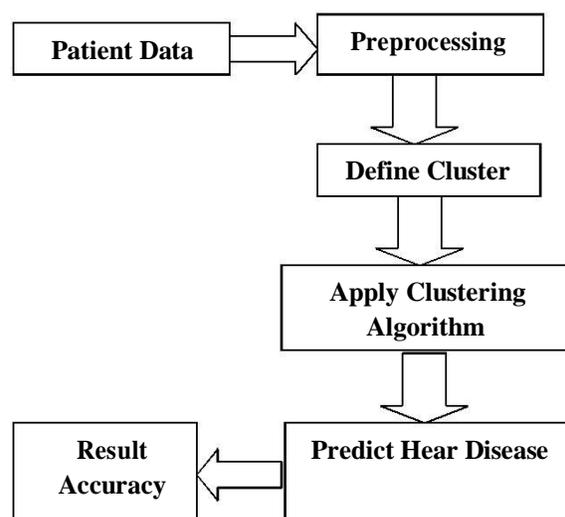
5. Miss. Chaitrali S. Dangare1, Dr. Mrs. Sulabha S. Apte2 [5] have presented Heart disease prediction system (HDPS) using data mining and artificial neural network (ANN) techniques. Find accuracy using 13 and 15 attribute. From the ANN, a multilayer perceptron neural network along with back propagation algorithm is used to develop the system. The experimental result shows that using neural networks the system predicts Heart disease with nearly 100% accuracy.

6 T.Georgeena. S.Thomas ,Siddhesh.S.Budhkar, Siddhesh.K.Cheulkar, Akshay.B.Choudhary, Rohan Singh [9] have used The decision-tree algorithm is one of the most effective and efficient classification methods available. The data will judge the efficiency and correction rate of the algorithm. They have used platform asp.net.

4. Proposed Method

From the Literature Survey we analyzed that in previous work have no used K-medoid algorithm. In our research work we will used K-medoid algorithm and compare with K-mean algorithm. The main objective of this research is to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. To develop this system, medical terms such as sex, blood pressure, and cholesterol like input attributes are used. To get more appropriate results. The data mining clustering techniques partitioning Method algorithm are used. The objective of this research is to create an intelligent & cost effective system which will overcome the limitations of existing system and improve its performance.

Flow of Method:-



Conclusion:-

The main objective of this research is to develop a Model of Heart Disease Prediction System using data mining modeling technique, namely, Clustering using partitioning method. HDP system can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for Analysis heart disease and thus assist healthcare practitioners to make clinical decisions. HDP System which compare two techniques of data mining and analysis accuracy using heart disease attribute. Comparison techniques is clustering like K-mean and K-medoid find accuracy of both techniques.

Reference:-

- [1] Durairaj M, Revathi V,” Prediction Of Heart Disease Using Back Propagation MLP Algorithm”, ISSUE 08, AUGUST 2015.
- [2] AKASH JARADI, ROHIT KATKAR², ABDUL REHAMAN SHAIKH³, ANUP SALVE⁴, “INTELLIGENT HEART DISEASE PREDICTION SYSTEM WITH MONGODB”, Volume 4, Issue 1, January-February 2015.
- [3] Ms. Shinde Swati B. ,” Decision Support System on Prediction of Heart Disease Using Data Mining Techniques”, Volume 3, Issue 2, March-April, 2015
- [4]R. Tamilarasi*, Dr. R. Porkodi, “A Study and Analysis of Disease Prediction Techniques in Data Mining for Healthcare”, March 2015.
- [5] Miss. Chaitrali S. Dangare¹, Dr. Mrs. Sulabha S. Apte², “A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS”, Volume 3, Issue 3, October - December (2012).
- [6] #1Neeru Pathania and Ritika*²,” Implementation of Fuzzy Controller for Diagonse of Patient Heart Disease”, Vol. 2 Issue 4, April 2015.
- [7] Dr. G. Rasitha Banu MCA., M.Phil., Ph.D.1, J.H.BOUSAL JAMALA MCA, ” HEART ATTACK PREDICTION USING DATAMINING TECHNIQUE”, Volume 02, Issue 05, [May – 2015]
- [8] Hlaudi Daniel Masethe, Mosima Anna Masethe,” Prediction of Heart Disease using lassification Algorithms”, WCECS 2014, 22-24 October, 2014.
- [9] T.Georgeena.S. Thomas, Siddhesh.S. Budhkar, Siddhesh.K. Cheulkar, Akshay.B.Choudhary, Rohan Singh,” Heart Disease Diagnosis System Using Apriori Algorithm”, Volume 5, Issue 2, February 2015.
- [12] Dharmendra, Teena Verma, Uuneet Kumar and Parita Jain,” Evidence based Guideline System”, November 2012.
- [13] Lior Rokach, oded Maimon CLUSTERING METHODS Department of Industrial Engineering Tel-Aviv University.
- [17] G. Purusothaman* and P. Krishnakumari, “A Survey of Data Mining Techniques on Risk Prediction: Heart Disease” , Vol 8(12), 58385, June 2015.
- [18] S.Vijayarani S.Sudha, “A Study of Heart Disease Prediction in Data Mining”, Vol. 2, No.5, October 2012.
- [19] Aieman Quadir Siddique, Md. Saddam Hossain , ” Predicting Heart-disease from Medical Data by Applying Naïve Bayes and Apriori Algorithm”, Volume 4, Issue 10, October-2013