ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 6, September 2016

A STUDY OF HEART DISEASE PREDICTION USING PREDICTIVE DATAMINING

Ann mariya joshy
Department of computer science, Sahrdaya College Of Engineering And Technology
Kodakara,Kerala
annmariya1994@gmail.com

Asha mariya joshy
Department of computer science, Sahrdaya College Of Engineering And Technology
Kodakara,Kerala
ashamariya1994@gmail.com

Angel rose joy
Department of computer science, Sahrdaya College Of Engineering And Technology
Kodakara,Kerala
go4angel828@gmail.com

Abstract
A large set of data is stored and processed in day-to-day life. But a less portion of this data is used efficiently. Data mining techniques is used to extract the hidden data from a huge database. Health care industry use data mining techniques for prediction off diseases from the dataset. In this survey paper we have focused on studying and analyzing data mining techniques which are used predicting the heart disease. This paper also gives details about the existing systems and their advantages and disadvantages.

Keywords: Heart disease, Prediction, Association rules, Classification, Clustering, Fuzzy system.

I. INTRODUCTION

Data mining is referred as the process of extracting required data from large repositories. It plays a vital role in medical field. Medical Data mining has high potential in extracting the hidden patterns from the data set. Clinical diagnosis make use of these patterns for prediction. The medical data is heterogeneous and voluminous and distributed widely. The data is gathered and then combined to provide a user oriented approach. Medical diagnosis plays an essential part and yet complicated task that needs to accomplished precisely and effectively. To reduce the cost for achieving clinical test an appropriate computer based information decision support is required. The existing techniques are analyzed and compared to find the efficient system. This paper aim to analyze different techniques used in data mining for the prediction of heart disease. Heart disease is the leading cause of death in India and worldwide [1].Cardio vascular disease is a highly mortal disease with over 17 million deaths globally [2].So early detection and treatment of heart

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 6, September 2016

disease is imperative.

II. HEART DISEASE PREDICTION

In [3], naive bayes is used to classify the data set and proposes excellent classification result . Naive bayes focuses on the rule that presence or absence of disease depends on a feature itself [4].Various learning algorithms are used to train naïve bayes. The data set was collected from UCI machine learning repository. The data set consists of 270 instances and 14 attributes. The main package used in this approach is Weka 6.4.Naive bayes classifier and evaluation is performed using the test set and then results are analyzed.10-fold cross validation is used to divide the data set into train and test data and to record the accuracy of data set. This method eliminates useless and distortive data and indirectly reduces the number of diagnostic tests to be taken by individual. Discretization method and genetic search was used to remove redundant features. This approach claimed highest accuracy compared to the other methods. It proves that One-R with Genetic search for naïve bayes is best suitable for early diagnosis of heart disease[3].

In [5], association rule mining technique is used for predicting heart attack. In this paper, the model proposed is a novel method CBARBSN, for association rule mining based on sequence numbers and clustering the transactional database for predicting heart attack. The important steps of process are, first medical data is transformed into binary and the proposed method is applied to the binary transactional data. The data is collected from Cleveland database[6]. The medical data contains 14 attributes. From the results, the author concluded that the proposed algorithm performs better than the existing ARNBSN ( ) algorithm. The performance of the algorithms is compared based on the execution time.It provides optimum accuracy for clustering.

In [7], the back propagation is used as the classification technique but it fails because it gets stuck in the local minima and have less profit. So genetic algorithm which uses the phenomena of mutation and crossover over various generations was used. The decision tree algorithm used is sJ48 which uses a pruning technique to build a good decision tree. Pruning is a method which tries to eliminate the over fitting data which is not so relevant in making a decision and leads to poor prediction. At last, a tree is build to provide flexibility and accuracy balance.

In [8],the technique being used for analyzing the data repository is neural network .The heart disease database are trained using feed forward neural network model and back propagation learning algorithm which uses variable learning rate and momentum. The input layer consist of 13 neurons which represents the 13 attributes. It is composed of 4 class labels specifically normal person, first stroke, second stroke and end of life. The output layer consists of two neurons to represent these 4 classes. few of the neural networks are made with and without hidden layer. The data file was gathered from Cleveland database.

In [9], the author proposed upgraded K means clustering algorithm for foreseeing coronary heart illness. There are two methodologies utilized for upgrading K-means clustering algorithm. To begin with the author proposed weighted ranking algorithm to conquer the issue of irregular choice of initial centroids. Second the attributes associated with weights concerned by the doctors are taken into account in both ranking and the K-means algorithm rather than allotting unit weight to all the attributes. The heart dataset was gathered from UCI machine learning storehouse [10]. In addition 35 conditions are carried out to allot weights to attributes. From a test the author presumed that the proposed algorithm enhances the consistency and nature of the final clusters. The unique clusters generates in turns of consistency. In future the precision is further improved to

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 6, September 2016

the dataset from the same area of the doctors by allotting weights furthermore moves forward the productivity of the procedure by unique cluster.

In [11], the heart disease is anticipated by the frequent feature selection method. From the utilization of fuzzy measure and relevant nonlinear integral the execution of the algorithm turns out to be great. The feature selection attribute mirrors the significance of non additive of the fuzzy measure. The author predicts the probability of patients getting a heart illness by utilizing medicinal profiles. The proposed methodology was executed in Java. The sample combination of ordinary and danger level heart attack parameters with their values and weight ages are said. The ordinary level of forecast includes the weightage of lesser worth (0.1) and other than 0.1 is considered as higher danger level. At last the proposed algorithm diminishes the computational time and enhances accuracy. The proposed work can be further extended for computerization of coronary illness expectation. Genuine information are gathered from medicinal services associations and compared with optimum accuracy with the available techniques.

In [13], the heart attack symptoms are anticipated using biomedical data mining procedures. The author used data classification which depends on supervised machine learning algorithms. For data classification the Tanagra device is utilized. Using entropy based cross validations and partitioned techniques, the data is assessed and the outcomes are compared. The algorithms utilized as a part of these techniques are K nearest neighbours, K-means and Mean Clustering algorithms(EMC) is the extension of the K-mean algorithm for clustering process which decreases the number of iterations. Therefore the author investigated that the mean clustering algorithm performs well at the point when contrasted with different algorithms. To run the data the time taken is quick and it gives the result of accuracy around 82.90%. Further this work can be upgraded by applying unsupervised machine learning algorithms.

In [14] the coronary artery disease was effectively diagnosed by rotation forest algorithm in order to support clinical decision making process. It uses the Artificial Neural Networks with Levenberg-Marquardt back propogation algorithm of rotation forest ensemble method as base classifiers. The algorithm is implemented in matlab. From an test, the author analyzed the disease by looking at the execution of base classifiers in terms of sensitivity, accuracy, AUC and specificity on two things i) without rotation forest classifier, the most extreme execution of classifiers and ii) with rotation forest algorithm it really enhances the performance of classifiers. Thus it is watched that Levenberg-Marquardt was the best classifier with or without random forest. The precision is enhanced to 91.2% of original classification accuracy which is an improvement of 7%. In future the proposed work might be utilized to create effective master frameworks for the conclusion of heart diseases.

In [15] using data mining techniques the author foresee heart diseases from horoscope of an individual. It discovers the possible outcomes of suffering a individual from heart disease. Horoscope has 12 areas furthermore, every area is called 'house'. Taking into account house the author predicts the heart disease. The analysis is executed in weka. The algorithm utilized as a part of this work are Decision table, Multilayer perception, J48 and LWL. From this, the decision table performed well when contrasted with different algorithms. In future this work can be stretched out by applying clustering algorithms in data mining.

III. CONCLUSION

7% of the Indian population is affected by heart diseases due to either heriditary factors or due to living habits. Heart disease is mainly affected on the people under the age of 65.This paper analyzes different data

ISSN: 2348 9510
International Journal Of Core Engineering & Management (IJCEM)
Volume 3, Issue 6, September 2016

mining techniques which can be used to predict the heart disease. This paper also analyzes how data mining techniques namely classification, clustering, fuzzy system and association rules are applied to the health data sets for predicting heart diseases. Naive bayes is one major algorithm used for classification and it produces highest accuracy compared to any other techniques. The aim of all studies was to reduce the number of attributes for the prediction and reduce the cost of testing. Data mining provides an effective way to solve this.

REFERENCES

[1] M.A Jabbar, B.L Deekshatulu, Priti Chandra," Heart disease classification using nearest neighbor classifier with feature subset selection", Annals Computer science series,Vol XI,Fasc 1,pp 47-54 (2013)

[2] Smith J.S,"Screening for high risk cardio vascular disesase: A challenge for the guidelines, Archives of internal medicine", 170, 40-42(2010)

[3] M.A.Jabbar Associate Professor Aurora's Engineering College, TELANGANA, INDIA , Computational Intelligence Technique for early Diagnosis of Heart Disease, 2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015, Coimbatore, TN, India.

[4] Saba Bashir et.al,"An ensemble based decision support framework for intelligent heart disease diagnosis', I society pp 259-264(2014)

[5] MA.JABBAR, Dr.PRITI CHANDRA, B.L.DEEKSHATULU "Cluster Based Association Rule Mining For Heart Attack Prediction" JTAIT Vol. 32 No.2 October 2011.

[6]."Cleveland heart disease dataset" sci2s.ugr.es/keel/dataset.php?cod=57

[7]. Ankita Dewan ITM University Gurgaon, INDIA " Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification.

[8] Dr. K. Usha Rani "Analysis of Heart Diseases Dataset Using Neural Network Approach" (IJDKP) Vol.1, No.5, September 2011

[9] R. Sumathi, E. Kirubakaran "Enhanced Weighted K-Means Clustering Based Risk Level Prediction for Coronary Heart Disease" European Journal of Scientific Research ISSN 1450- 216X Vol.71 No.4 (2012), pp. 490-500 © Euro Journals Publishing, Inc. 2012

[10]. Blake, C.L., Mertz, and C.J.: "UCI Machine Learning Databases", http://mlearn.ics.uci.edu/databases/heartdisease

[12] S.Sarumathi, N.S.Nithya, "Effective Heart Disease Prediction System Using Frequent Feature Selection Method" International journal of Communications and Engineering Volume 01– No.1, Issue: 01 March 2012

[13] V.V.Jaya Rama krishniah, D.V.Chandra Sekar, Dr.K.Ramchand H Rao, "Predicting the Heart Attack Symptoms using Biomedical Data Mining Techniques" Volume 1, No. 3, May 2012 ISSN – 2278-1080 The International Journal of Computer Science & Applications (TIJCSA)

[14] Milan Kumari, Sunila Godara "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction", IJCST Vol. 2, Issue 2, June 2011

[15] Mr. Pankaj S. Kulkarni, Ms. V. C. Belokar, Dr. S. S. Sane, Prof. N. L. Bhale, "Heart Disease Prediction from Horoscope of a Person Using Data Mining", International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012