Open access Journal International Journal of Emerging Trends in Science and Technology

Enhanced Heart Disease Analysis and Prediction System [EHDAPS] Using Data Mining

Authors

Alagugowri. S [1], Dr. T. Christopher[2]

[1]Research Scholar, Department of Computer Science, Government Arts College, Udumalpet – 642 126

Email: *alagugowri7@gmail.com*

[2]Assistant Professor& HOD, Department of Computer Science, Government Arts College

Udumalpet - 642 126

Email: *chris.hodcs@gmail.com*

ABSTRACT

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining techniques are used for variety of applications. Data mining techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in clinical data. One of the most important applications of such systems is in diagnosis of heart disease. The main objective of Enhanced Heart Disease Analysis and Prediction System (EHDAPS) is predicting the heart disease using historical heart database. To develop this system, medical terms such as sex, blood pressure, and cholesterol like seventeen input attributes are used. In this paper association among various attributes which are the causative factors of heart diseases are analyzed. The patient's records are observed before prediction and the factors are grouped as per its severity level. In this system the level of causative factors are categorized using K-Means clustering technique and it distinguishes the risky and non risky factors. Frequent risk factors are mined from the clinical heart database using Apriori algorithm. The risk factors are taken for this study to predict the risk level and find the co-ordination among the factors that helps the medical people to predict the disease with minimum tests and treatments.

Keywords: *Heart disease, Data mining, K-Means Clustering, Apriori algorithm*

1. INTRODUCTION

Knowledge Discovery in Database (KDD) was formalized in 1989, with reference to the general concept of being broad and high level in pursuit of seeking knowledge from data. The term Data mining was then coined, this high level application technique is used to present and analyze data for decision makers. Data mining is one of the steps involved in Knowledge Discovery in Databases. A formal definition of Knowledge discovery in databases is "Data mining is the non trivial extraction of embedded previously unidentified and potentially useful information about data". Data mining is used to find the hidden patterns in data and to predicting the future behaviours. In health care industry the revealed knowledge can be used to reduce the risk of level of the patients who have a heart disease [1].

Disease diagnosis is a complex task which requires much experience and knowledge. Traditional way of diagnosing Heart disease is doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Nowadays, Health care industry contains huge amount of heath care facts, which contains hidden information. This hidden information is useful for making effective decisions. A major challenge facing healthcare organizations (hospitals, clinical centres) is the stipulation of quality services at

reasonably priced costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to terrible consequences which are therefore unacceptable. Healthcare industry must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems [2]. Health care data is enormous. It includes patient centric data, resource management data and transformed data. Health care organizations must have ability to analyze data. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care. The deaths due to heart disease in many countries occur due to work burden, mental pressure and many other problems. On the whole it is found as primary reason behind death in adults. Nowadays, in the world Heart disease is the major cause of deaths. So the prediction is needed to reduce the risk level of a patient.

2. HEART DISEASE

Heart is the foremost part of our body. Life is dependent on efficient functioning of heart. If the functions of heart is not proper, it will distress the other body parts of human such as brain, kidney etc. Heart disease affects the functions of heart. There are number of factors which increases risk of Heart disease [3].

Some of them are listed below:

- Hereditary
- Smoking
- Alcohol Intake
- Bad Cholesterol
- High Blood Pressure
- Over Weight
- Lack of physical exercise

Heart disease is the foremost health problem and it affects a large number of people. Today heart diseases are most common disease in our society. Around 60% of our population is suffering from heart disease because of their modern daily life. According to survey of WHO [4], 17 million total global deaths are due to heart attacks and strokes. In 2008, 17 million people died because of heart disease. WHO estimated by 2030, almost 23.6 million people will die due to heart disease. Predication should be done to reduce risk level of Heart disease [5]. Diagnosis is usually based on symptoms and physical inspection of a patient. Almost all the doctors are predicting heart disease by their knowledge and experience. The diagnosis of disease is a complicated and tedious task in medical field. Predicting Heart disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects [6]. Healthcare industry today generates large amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc. The large amount of data is a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Only human brain power alone is not enough for proper diagnosis. A number of complications will arrive during diagnosis, such as less accurate results, less experience, time dependent performance, knowledge up gradation is difficult etc.

3. LITERATURE SURVEY

An Intelligent Heart Disease Prediction System (IHDPS) was proposed by Sellappan Palaniappan. IHDPS is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees et al.[7]. Each method has its own power to get appropriate results. To build this system hidden patterns and relationship between them is used. It is web-based, user friendly and expandable.

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al.[8]. The dataset contains clinical records with 13 attributes in each record. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data.

Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural

Network, 2009 was proposed by Shantakumar B.Patil, Y.S.Kumaraswamy [9] in which the data warehouse is pre-processed so that it can be made suitable for the mining process. After the pre-processing, the heart disease data warehouse is clustered with the K-means clustering algorithm, which will extract the data applicable to heart attack from the warehouse. As a result the frequent patterns applicable to heart disease are mined with the aid of the MAFIA (Maximal Frequent Item set Algorithm) algorithm from the data extracted. The neural network is strained with the selected significant patterns for the effective prediction of heart attack. Multi-layer Perception Neural Network with Back-propagation as the training algorithm.

An Empirical study on prediction of Heart disease using Classification Data mining techniques [10] was proposed by T.John Peter. The use of pattern recognition and data mining techniques into risk prediction models in the clinical domain of cardiovascular medicine is proposed. The data is to be modelled and classified by using classification data mining technique. Some of the limitations of the conventional medical scoring systems are handled in this research by use of classification models which can implicitly detect complex nonlinear relationships between dependent and independent variables as well as the ability to detect all possible interactions between predictor variables.

4. PROPOSED SYSTEM

4.1 EHDAPS

The Enhanced Heart Disease Analysis and Prediction System [EHDAPS] is an intelligent diagnostic tool that assists healthcare industries to gain advantage. All doctors do not possess expertise in every sub specialty and moreover there is a shortage of resource persons at certain places. Therefore, an automatic medical diagnosis system would probably be exceedingly beneficial by bringing all of them together. It can aid in achieving clinical tests at a reduced cost. The important clinical features, i.e., age, sex, chest pain type, blood pressure, cholesterol, fasting blood sugar, resting ECG, max heart beat rate have an impact on heart functioning. Generally the factors with the test results of exercise induced angina, old peak, slope, number of vessels colored are taken to predict the risk level of heart diseases.

The system proposed an efficient approach for the extraction of significant patterns from the heart disease warehouses for heart disease prediction. The data warehouse is pre-processed to make it appropriate for the mining process. The proposed system that integration of clinical decision support with computer based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modelling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions. The system improves the diagnosis accuracy and to reduce the diagnosis time. It has become a demanding issue to develop an efficient and reliable medical support system to support yet and still increasingly complicated diagnosis decision process. Hence soft computing methods such as data mining have shown great potential to be applied in the development of heart attack prediction system.

This paper analyzes the existing methodologies and it suggests the enhanced predicting technique to the medical practitioners. In this paper seventeen medical attributes are taken for the analysis. The data mining techniques like K-Means clustering, Apriori are used to predict the heart disease severity and also this system analyze the heart patients' data repository to find out the supportive factors of heart disease. Ultimately it provides the awareness about heart disease and causative factors. Association rule mining [ARM] is the one of the most important and well researched techniques of data mining. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction datasets or other data repository. The Apriori algorithm is a great

success in the history of mining association rules. It is by far the most well-known association rule algorithm. On the whole process of the proposed system shown in the Figure 4.1.

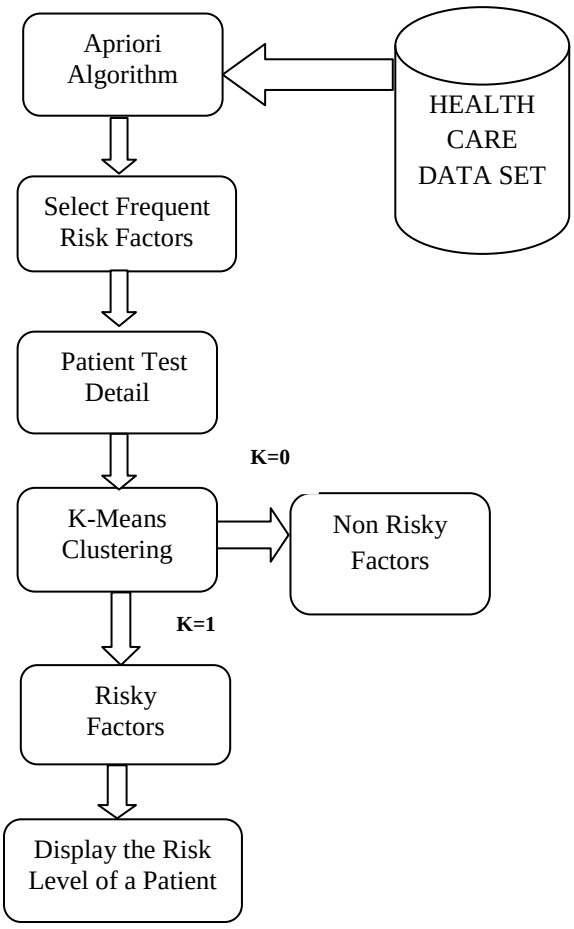

Figure 4.1: Proposed Architecture

4.2 DATA SOURCE

The ARM is used to predict Heart Disease from heart patient Data base. Here the 17 attributes from patients' medical records are taken for this research. Heart Disease Attributes are shown in Table 4.1.

Table 4.1: Heart disease attributes

Parameter	Values
Male and Female	Age < 30 , >30to <50,Age>50 and Age <70,Age>70
Smoking	Never , Past, Current
Overweight	Yes, No
Alcohol Intake	Never , Past, Current
High salt diet	Yes, No
High saturated fat diet	Yes, No
Exercise	Never, Regular
Sedentary Lifestyle/inactivity	Yes, No
Hereditary	Yes, No
Bad cholesterol	Very High >200, High 160 to 200, Normal <160
Blood Pressure	Normal (130/89), Low (< 119/79) High (>200/160)
Blood sugar	High (>120&<400), Normal (>90&<120) Low (<90)
Heart Rate	Low (< 60bpm), Normal (60 to 100) High (>100bpm)
Defect type	Normal, Fixed, Reversible defect
Chest pain type	typical type 1, typical type angina non-angina pain, asymptomatic
Resting electrographic results	Normal having_ST_T, wave abnormal, left ventricular, hypertrophy
number of major vessels colored by fluoroscopy	0-3 values

5. EXPERIMENTAL RESULTS

This system implemented on the .Net platform. The clusters were defined based on the attributes listed in Table 4.1. After clustering using K-Means, risk levels were predicted. EHDAPS, not only predict the risk level and it also analyze the data set to find frequent causative factors. The sample of heart disease risk level prediction (Low, Medium, and High) is as below.

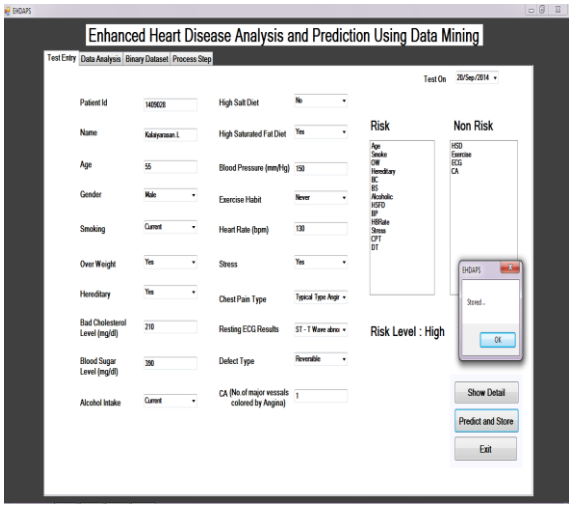

Figure 5.1 Risk level Prediction of a Patient [High Risk]

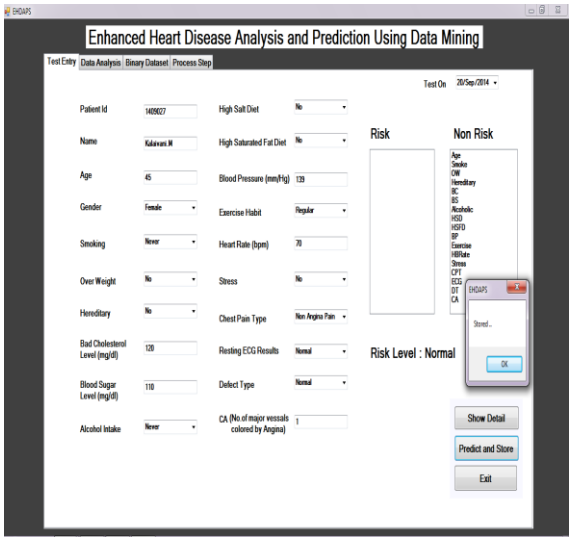

Figure 5.2 Risk level Prediction of a Patient [Normal]

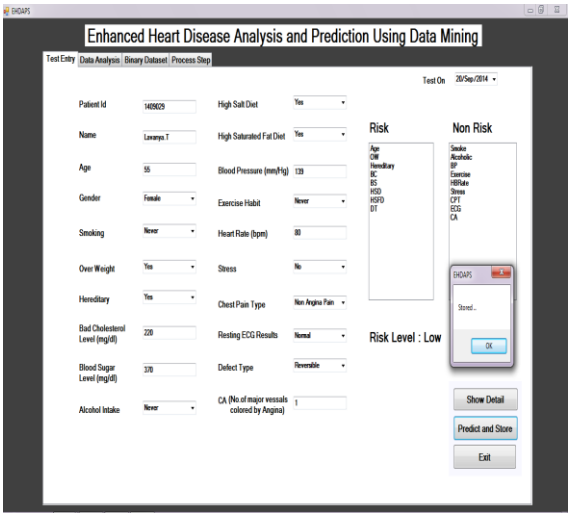

Figure 5.3 Risk level Prediction of a Patient [Low Risk]

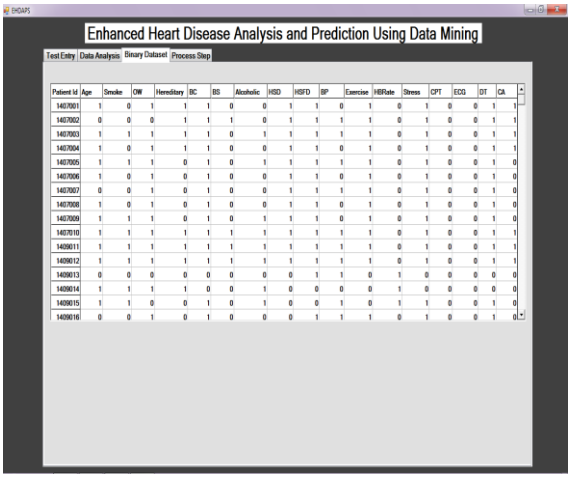

Figure 5.4 Binary Health care Data set

This binary data set illustrates patients' health information in binary mode. Collections of Patient medical details used for transaction databases and set of associations can be represented as binary incidence matrices with columns corresponding to the factors and rows corresponding to the Patients. The matrix entries represent presence (1) or absence (0) of a risk factor in a particular patient.

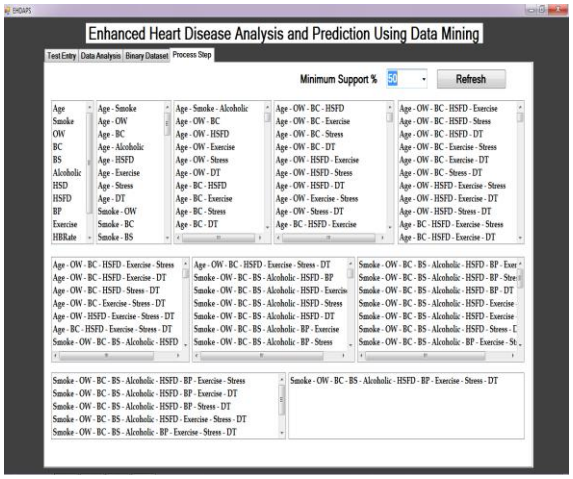

Figure 5.5 Itemset mining[Minimum support 50%]

For a given data set, figure 5.5 shows Apriori algorithm mined risk factors that have minimum of 50% support of all transactions.

6. CONCLUSION AND FUTURE DIRECTIONS

Enhanced Heart Disease Analysis and Prediction System [EHDAPS] provide an intelligent assistance to the persons to predict the risk level of the diseases and also it assists to find out the supportive measures of particular factors.

In future the work can be expanded and enhanced for the automation of various types of disease prediction. It also extended to find various types of diseases with the use of these factors.

REFERENCES

1. K.Srinivas, B.Kavihta Rani, Dr. A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering (IJCSE), Volume 02, February 2010, Page (250-255).
2. Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, "A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining", International Journal of Applied Engineering Research (IJAER), Volume 7, November 2012, Page (1-6).
3. K.Srinivas., G.Raghavendra Rao and A.Govardhan., "Analysis of Attribute Association in Heart Disease Using Data Mining Techniques ", International Journal of Engineering Research and Applications(IJERA), July-August 2012, Page(1680-1683).
4. World Health Organization. The world health report 2010 – Health systems financing: the path to universal coverage, 2010.
5. "Heart disease" from http://wikipedia.org.
6. Nitika, Madan Lal Yadav, "A Fuzzification Approach for Prediction of Heart Disease", International Journal of Engineering Trends and Technology (IJETT), Volume 4, May 2013, Page (2068- 2071).
7. Sellapapan palaniappan., Rafiah Awang., "Intelligent Heart Disease Prediction system using Data mining Techniques", 978-1-4244-1968-5/08/$25.00.,2008 IEEE.
8. Niti Guru,Anil Dahiya, Navin Rajpal., "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Volume8, No. 1, January-June 2007.
9. Shantakumar B.Patil., Y.S. Kumaraswamy., "Intelligent and Effective Heart Attack Prediction System using Data mining and Artificial Neural Network", European Journal of Scientific Research (EJSR),2009, Page(642-656).
10. T.John Peter., "An Empirical study on prediction of Heart disease using Classification Data mining techniques ", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), March 30-31, 2012, Page (514 – 518).
11. K.R. Lakshmi, M.Veera Krishna and S.Prem Kumar., "Performance Comparision of Data Mining Techniques for Predicting of Heart Disease Survivability", International Journal of Scientific and Research Publications, Volume 3,Issue 6, June 2013, Page(1-10).