International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Email:editorijoer@gmail.com http://www.ijoer.in

Vol.4., Issue.2., 2016
(Mar-Apr)

REVIEW ARTICLE

ISSN: 2321-7758

A HYBRID APPROACH FOR FAST AND ACCURATE DETECTION OF HEART DISEASES USING DATA MINING ALGORITHMS- A SURVEY

PRASADGOUDA PATIL[1], SOUMYA PATIL [2]

[1,2]Department of Computer Science and Engineering, Secab Institute of Engineering & Technology, Vijayapur, Karnataka, Nauraspur, Bagalkot road ,Vijayapur, Karnataka,India

ABSTRACT

Cardiovascular disease is the leading global cause of death, accounting for 17.3 million deaths per year, a number that is expected to grow to more than 23.6 million by 2030. Cardiovascular disease is a term used to describe a variety of heart diseases, illnesses, and events that impact the heart and circulatory system. There is a huge amount of 'knowledge-enriched data' in hospitals, which needs to be processed in order to extract useful information from it. This data is very useful in making valuable medical decisions. Data mining technique in the history of medical data found with enormous investigations found that the prediction of heart disease is very important in medical science. The main aim of this research is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. Decision tree algorithm (ID3 algorithm) is used for classifying the user information for making the decision based on the symptoms and also used to choose the test attributes by using information gain measurements on each tree node. Fuzzy Cluster Analysis Algorithm is used to provide an optimally selected the features and remove those unrelated and compress similar class features. The proposed algorithms will predict the diseases with more accuracy. The goal of classification is to accurately predict the target class for each case in the data.

Keywords— Cardiovascular disease; data mining; fuzzy logic; weka tool; decision tree; naive bayes; classification via clustering.

I. INTRODUCTION

Today medical field have come a long way to treat patients with various kinds of diseases. Among this the most threatening one is the Heart disease/CVD which cannot be observed with a naked eye and comes instantly when its limitations are reached. CVDs are the number 1 cause of death globally: more people die annually from CVDs than from any other cause. An estimated 17.5 million people died from CVDs in 2012, representing 31% of all global deaths. Of these deaths, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke. Over three quarters of CVD deaths take place in low- and middle-income countries. Out of the 16 million deaths under the age of 70 due to non

communicable diseases, 82% are in low and middle income countries and 37% are caused by CVDs. People with cardiovascular disease or who are at high cardiovascular risk need early detection and management using medicines as appropriate treatment. Bad clinical decisions would cause death of a patient which cannot be afforded by any hospital. Many hospitals use hospital information systems to manage their healthcare or patient data. Hence, more adequate systems for diagnosis of cardiovascular disease need to be developed. There is a bulk of hidden information in this data that is not yet explored which give rise to an important query of how to make useful information out of the data. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. In today's era, data mining has its successful application in various fields including healthcare. On the other hand, fuzzy logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. Our work attempts to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record using hybrid feature selection algorithm which is the Combination of CFS and Filter Subset Evaluation gives better accuracy for classification.

II. RELATED WORK

Researchers have been applying different data mining techniques such as naïve bayes, decision tree and clustering algorithm. Nidhi Bhatla and Kiran Jyoti[1] proposed a work to reduce the number of attributes used in heart disease diagnosis that will automatically reduce the number of tests which are required to be taken by a patient. The author illustrated that Decision Tree and Naive Bayes using fuzzy logic has outplayed over other data mining techniques. Syed et.al[2] developed a technique for prediction of heart disease using major risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. The technique involves two most successful data mining tools, neural networks and genetic algorithms. Manmuna Fatima[3] has done a research to mine the historical unstructured data of heart patients and to extract significant features and patterns where K-Means clustering technique is applied to find out clusters in data which are further used to extract hidden patterns related to heart patients. V Krishnaiah[4] developed a prototype to remove uncertainty of unstructured data that contains different attributes, an attempt was made by introducing fuzziness in the measured data. It was found that Fuzzy K-NN classifier suits well as compared with other classifiers of parametric techniques. Collins et.al[5] undertook a work to investigate the effectiveness of using auto associative neural networks and optimization algorithms in missing data prediction and classification tasks. The results reveal GA, SA and PSO to be more efficient when compared to RF in terms of predicting the forest area to be affected by fire. P.K. Anooj (2012) [9] developed Clinical Decision Support System for heart disease using weighted Fuzzy Rules. E.P. Ephzibah et al (2012) [10] framed Fuzzy Rules for Heart Disease diagnosis using 6 attributes. M. Anbarasi et al (2010) [11] developed an Enhanced Prediction System for heart disease with feature subset selection using Genetic Algorithm. Moreover, three classifiers Decision Tree, Naive Bayes and Classification via Clustering have been used and Decision Tree performed with good prediction probability of 99.2%. B.Patil et al (2009) [12] used Artificial Neural Network for developing heart disease prediction system. Carlos (2006) [13] compared Association Rules and Decision Trees for disease prediction.

III. DATAMINNING TECHNIQUES

Data mining techniques are used to explore, analyze and extract medical data using complex algorithms in order to discover unknown patterns. Classification is a technique that predicts categorical class labels. It classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data. Classification is a two – step process consisting of Model Construction and Model Usage. Model Construction is defined as a process of

describing a set of predetermined classes whereas Model Usage is helpful for classifying future or unknown objects. There are mainly three classifiers Decision Tree, Classification via Clustering and Naive Bayes to diagnose presence of heart disease in patients.

Decision tree classifier- Decision Tree Classifier is a simple and widely used classification technique. It applies a strait forward idea to solve the classification problem. Decision Tree Classifier poses a series of carefully crafted questions about the attributes of the test record. Each time time it receive an answer, a follow-up question is asked until a conclusion about the calss label of the record is reached. The decision tree classifiers organized a series of test questions and conditions in a tree structure.

Naive Bayes classifier-In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations.

Classification via clustering- Clustering is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data. Classification is the task of generalizing known structure to apply to new data. Hence, classification is performed based on clustering.

IV. METHODOLOGY

In our research studies of heart disease features and pattern extraction the proposed system aims to efficiently diagnose the presence of heart disease in an individual. A fuzzy set is a collection of distinct elements with a varying degree of relevance or membership. Fig. 1 illustrates the overview of our proposed system.

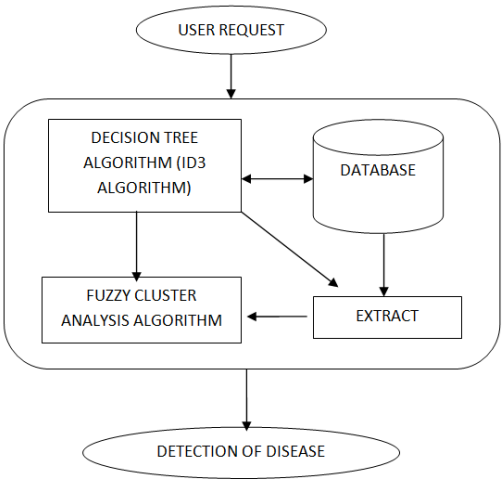

Fig. 1 Overview of proposed system

A fuzzy set is a generalization of the concept of a set whose characteristic function takes only binary values. A fuzzy inference model can be created using the properties of fuzzy set. The proposed system uses hybrid feature selection algorithm which is the Combination of CFS and Filter Subset Evaluation gives better accuracy for classification. Using the hybrid feature selector, the data is applied to the classification algorithm in which Naïve bayes gives higher Accuracy comparing to the other classifiers. The goal of classification is to accurately predict the target class for each case in the data. The hybrid feature selection method reduced the data is given to the classification and clustering algorithms to analyze which is best fit for heart disease prediction. Classifiers have been trained to classify the medical data into four classes viz. “Normal”, “Low Risk”, “Medium Risk” or “High Risk”. Based on these risk factors and attributes undertaken, the disease is predicted using naive bayes algorithm. Among many classification algorithm naive bayes algorithm accuracy is higher.

V. CONCLUSIONS

The overall objective of the research work is to predict the heart disease patients with more accuracy which is useful for health care information systems with reduced number of attributes. Decision Tree Classifier using fuzzy logic is used for detection of heart disease and compress the attributes for test. We can build an intelligent system which could predict the disease using risk

factors hence saving cost and time to undergo medical tests and checkups and ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases. This method will prove to be very efficient and helpful in saving a patients' life from heart disease.

REFERENCES

[1] Chaitrali S. Dangare and Sulabha S. Apte , "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" , *International Journal of Computer Applications (0975¿¿¿888)* , vol. 47 , no. 10 , 2012

[2] K. Srinivas, G. Raghavendra Rao and A. Govardhan , "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques" , *The 5th International Conference on Computer Science & Education*

[3] Niti Guru, Anil Dahiya and Navin Rajpal , "Decision Support System for Heart Disease Diagnosis Using Neural Network" , *Delhi Business Review* , vol. 8 , no. 1 , 2007

[4] Yanwei Xing, Jie Wang and Zhihong Zhao , "Combination data mining methods with new medical data to predicting outcome of" Coronary Heart Disease" , *International Conference on Convergence Information Technology*

[5] T. John Peter and K. Somasundaram , "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)*

[6] Jiawei Han, Micheline Kamber and Jian , *Pei-Data Mining: Concepts and Techniques* , 2011 Sellappan Palaniappan and Rafiah Awang , "Intelligent Heart Disease Prediction System Using a Data Mining Techniques" , *IJCSNS International Journal of Computer Science and Network Security* , vol. 8 , no. 8 , 2008

[7] S.H Ishtake and S.A. Sanap , "Intelligent Heart Disease Prediction System Using Data Mining Techniques" , *International J. of Healthcare & Biomedical Research* , vol. 1 , no. 3 , pp.94 -101 , 2013

[8] R. chitra and v. seenivasagam , "review of heart disease prediction system using data mining and hybrid intelligent techniques" , *ICTACT journal on soft computing* , vol. 03 , no. 04 , 2013

[9] P .K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules"; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40.

[10] E.P.Ephzibah, Dr. V. Sundarapandian, "Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis"; International Journal of Fuzzy Logic Systems (IJFLS) Vol.2, No.1, February 2012.

[11] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm"; International Journal of Engineering Science and Technology, Vol. 2(10), 2010.

[12] 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, ISSN 1450-216X Vol.31 No.4, 2009.

[13] Carlos Ordonez, "Comparing association rules and decision trees for disease prediction"; ACM, 2006.