A STUDY ON EFFECTIVE CLASSIFICATION AND PREDICTION OF HEART DISEASE USING DATA MINING TECHNIQUES

C.UshaNandhini M.Sc., M.Phil.,
Part Time Research Scholar
Assistant Professor in Comp. Applns.,
Vellalar College for Women(Autonomous),
Erode – 638012.
e-Mail: cushanandhini1970@gmail.com
Ph: 9865808812

Dr. P.R. Tamilselvi M.Sc., M.C.A., M.Phil., Ph.D.,
Research Guide
Assistant Professor in Computer Science
Govt. Arts & Science College
Komarapalayam -638183.
e-Mail: selvipr2003@gmail.com

ABSTRACT- The objective of our work is to analyze various data mining tools and techniques in health care domain that can be employed in prediction of heart disease system and their efficient diagnosis.There is a growing need in the health care scenario to store and organize sizeable clinical data, analyze the data, assist the health care professionals in decision making, and develop data mining methodologies to mine hidden patterns and discover new knowledge from clinical data.

Healthcare industry today generates large amounts ofcomplex data about patients, hospitals resources, diseasediagnosis, electronic patient records, medical devices etc.The large amounts of data are a key resource to beprocessed and analyzed for knowledge extraction thatenables support for cost-savings and decision making.Data miningbrings a set of tools and techniques that can be applied to thisprocessed data to discover hidden patternsthat provide healthcare professionals an additional source ofknowledge for making decisions.

Sri Vasavi College, Erode Self-Finance Wing *3rd February 2017*

National Conference on Computer and Communication *NCCC'17*

http://www.srivasavi.ac.in/ | nccc2017@gmail.com

Classification techniques are greatly deployed in several application domains for the purpose of classification, estimation and prediction. In this paper we survey different papers in which one or more algorithms of data mining used for the prediction of heart disease.

Keywords: Data Mining, Heart Disease, Health Care, Classification Techniques

I. Introduction

Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. The patterns must be actionable so that they may be used in an enterprise's decision making process. Data mining is often a complex process and may require a variety of steps before some useful results are obtained. Often data pre-processing including data cleaning may be needed. In some cases, sampling of data and testing of various hypothesis may be required before data mining can start. Data mining has found many applications in the last few years for a number of reasons such as growth in OLTP data, growth in data due to cards, growth in data due to the web, growth in data due to other sources, growth in data storage capacity, decline in the cost of processing, availability of software. A typical data mining process include requirement analysis, data selection and collection, cleaning and preparing data, data mining exploration and validation, implementing , evaluating and monitoring , results visualization.

Cardiovascular Disease

Heart diseases are one of the leading causes of death for men and women. Heart disease, also known as cardiovascular disease, refers to diseases that affect or involve the heart, the blood vessels (which include the arteries, capillaries and the veins) or both. In short, they affect the cardiovascular or circulatory system of the body. The cardiovascular system enables the movement of oxygenated blood from the heart to different parts of the body through arteries and deoxygenated blood back to the heart through veins. There are many types of heart diseases in India and it is necessary for every individual to be aware of them. Coronary artery disease, angina and

valvular heart diseases are some of the heart diseases that are often seen. There are several ways to recognize heart diseases such as heart attacks and quite a few ways to control or prevent them.

Types of cardiovascular disease

1   Coronary heart disease

2   Stroke (cerebrovascular disease)

3   Rheumatic heart disease

4   Deep vein thrombosis and pulmonary embolism

5   Congenital heart disease

Causes of Heart Diseases

The different factors that cause heart diseases are:

• Smoking

• Lack of sleep

• High blood pressure

• High cholesterol levels

• Diabetes

• No or limited physical activity

• Unhealthy eating habits

• Overconsumption of alcohol

• Depression and mental stress

• Pollution

II. Literature Review

In year 2008, LathaParthiban, et. al. [8] performed work, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm". In this paper, a new approach based on coactive neuro-fuzzy inference system (CANFIS) was presented for prediction of heart disease. The proposed CANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the presence of the disease. The performances of the CANFIS model were evaluated in terms of training performances and classification accuracies and the results showed that the proposed CANFIS model has great potential in predicting the heart disease.

In year 2012, Chaitrali S. Dangare, et. al. [2] performed work, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques". This paper has analyzed prediction systems for Heart disease using more number of input attributes. The system uses medical attributes such as sex, blood pressure,

Sri Vasavi College, Erode Self-Finance Wing   *3rd February 2017*
National Conference on Computer and Communication *NCCC'17*
http://www.srivasavi.ac.in/   nccc2017@gmail.com

cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. This research paper added two more attributes i.e. obesity and smoking. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database. The performance of these techniques is compared, based on accuracy.This analysis shows that out of these three classification models Neural Networks predicts Heart disease with highest accuracy.

In year 2013, K.Rajeswari et. al. [10] "Feature Selection for Classification in Medical Data Mining".In this paper they have analyzed the approach of feature selection for classification and also presented a novel approach for the feature selection by using association and correlation mechanism. The aim of this paper is to select the correlated features or attributes of medical dataset so that patient need not to go for many tests and in future it is used for preparing the clinical decision support system which is helpful for decision making of disease prediction in a cheaper way.

Other approach is mentioned in this paper is after removal of some attributes accuracy of classifier is also improved which support our statement of disease prediction in cheaper way by avoiding all unwanted tests for disease prediction. By using association rules and correlation attributes features can be selected. As medical field contains large number of attributes and information so dimensionality reduction is must now. The accuracy of classifiers after removal of attributes is discussed in this paper.

In year 2013 VikasChaurasiaet, et. al. [14] performed work "Early Prediction of Heart Diseases Using Data Mining Techniques". The main objective of this manuscript is to report on a research project where they took advantage of those available technological advancements to develop prediction models for heart disease survivability. They used three popular data mining algorithms CART (Classificationand Regression Tree), ID3 (Iterative Dichotomized 3) and decision table (DT) extracted from a decision tree or rule-based classifier to develop the prediction models using a large dataset. They used 10-fold cross

Sri Vasavi College, Erode Self-Finance Wing   *3rd February 2017*

National Conference on Computer and Communication *NCCC'17*

http://www.srivasavi.ac.in/   nccc2017@gmail.com

validation methods to measure the unbiased estimate. Observation shows that CART performance is havingmore accuracy, when compared with other two classification methods.

In year 2013 ,M.Akhiljabbar et. al. [1] "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm". In this paper they propose a new algorithm which combines KNN with genetic algorithm for effective classification. Genetic algorithms perform global search in complex large and multimodal landscapes and provide optimal solution. Experimental results shows that our algorithm enhance the accuracy in diagnosis of heart disease. In this paper they have presented a novel approach for classifying heart disease. As a way to validate the proposed method, they have tested with emphasis on heart disease on A.P besides other machine learning data sets taken from UCI repository. Experimental results carried out on 7 data sets show that our approach is a competitive method for classification. This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of

mortality in India and in Andhra Pradesh. Identification of major risk factors and developing decision support system, and effective control measures and health education programs will decline in the heart disease mortality.

In year 2014, Mohini D Patilet.et. al. [9] "Effective Classification after Dimension Reduction: A Comparative Study ". The paper discusses in brief about the dimension reduction techniques. It also describes the system developed for dimension reduction and use of the tool WEKA for dimension reduction and preprocessing. Finally a comparative study of the results obtained by the system and WEKA is done. They have presented a comparative study on dimension reduction. Firstly they discussed the concept of dimension reduction, its need and areas of application. Then they focused upon some of the techniques used for reducing dimensions. And this is possible because of the fuzzy rough theory which helps to increase the accuracy and also the use of neural network classifiers which provide a good performance.

In year 2015, [13]S. Udhaya kumara, et.al. "A Novel Neighborhood Rough set Based

Sri Vasavi College, Erode Self-Finance Wing   **3rd February 2017**

National Conference on Computer and Communication *NCCC'17*

http://www.srivasavi.ac.in/   nccc2017@gmail.com

Classification Approach for Medical Diagnosis". In this study, a novel neighborhood rough set classification approach is presented to deal with medical datasets. Fivebenchmarked medical datasets have been used in this research work for studying the impact of proposed work in decisionmaking.Experimentalresultof the proposed classification algorithm is compared with other existing approaches such as rough set,nearest neighbor, support vector machine, Back propagation algorithm and multilayer perceptron to conclude that theproposed approach is cheaper way for disease prediction and decision making. The performance of classification algorithmsmeasured based on various classification accuracy measures.

In year 2015, R. Kavithaet. al. [6] "A Framework for Heart Disease Prediction Using K nearest Neighbor Algorithm". This study proposes a framework to develop a heart disease prediction process using k-nearest neighborwith wrapper filter. Heart disease diagnosis is mostly done with doctor's knowledge and practice. But the cost spentby the patients are more in order to take test in which all the test does not contribute towards

effective diagnosis ofdisease. The feature which contributes towards effective diagnosis istermed as critical feature. In this study proposes a framework to find the subset of critical feature using K nearestneighbor and wrapper filter. This in turn produces a prediction model. Finally they exhibit the ideas of diagnosingheart disease with critical feature.This prediction model helps the experts in efficientdecision making process with fewer attributes todiagnose the heart disease.

In year 2015 A. Shaikhet. al. [12] "Performance Evaluation of Classification Methods for Heart Disease Dataset".This paper evaluates the performance of three basic classifiers such as k-NN, Naïve Bayesian and Decision Tree for medical data, which has been acquired from online machine learning repositories. Decision Tree classifier predicts poorly for the sparse dataset, specially, when dataset comprises of varying attribute values. To evaluate the performance of each different classification method the dataset has been portioned into three different cases. Thus, aim is to evaluate each learning algorithm at different variations related to highly dimensional data. The

experimental results based on the considereddataset represent that k-NN predicted the significantly better results than Naïve Bayesian. Decision Tree produced less accurate results.

In year 2016, S. Kiruthika Devi, et. al [7] "Prediction of Heart Disease using Data Mining Techniques". In this paper data mining classification techniques in the field of health care addressed are namely Decision trees, Naive Bayes, Neural Networks and Support Vector Machines. Hybridizing or combining any of these algorithms helps to make decisions quicker and more precise. Using advanced data mining techniques to excavate valuable information has been considered as an activist approach to improve the quality and accuracy of healthcare service while lowering the healthcare cost and diagnosis time. Using this technique presence of heart disease can be predicted accurately.

III. Conclusion

As the heart disease patients are increasing world-wide each year and huge amounts of data is available for research, researchers are using data mining techniques in the diagnosis of heart disease.

Feature selection really helpful for dimensionality reduction and also for building cost effective model for disease prediction. Analysis presented by different researcher's shows that different data mining techniques and classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that using different techniques and taking different number of attributes we get different accuracies for predicting heart diseases. Taking analysis from some papers published by the researcher's it is shown that techniques such as neural networks give best accuracy in predicting heart diseases thanother techniques.

REFERENCES

1. M.Akhiljabbaret. et. al. "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm" ScienceDirect, Procedia Technology 10 (2013) 85 – 94

2. Chaitrali S. Dangareet. et. al. " Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

3.   Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.