OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

EARLY DETECTION OF HEART DISEASE USING DATA MINING

Bhumika Alte

P.G. Student, Computer Science & Engineering, Everest College of Engineering, Aurangabad, Maharashtra, India.

Abstract: Heart disease is one of the common causes of death in Asia. According to the reports in year 2003-2004 more than 15 million people died and among them around 10 million died because of heart disease. There are certain factors which may cause heart disease; some of them are because of change in life style and addictions. We can list some of them as smoking, alcohol consumption, Obesity, blood pressure and diabetes. Medical database store abundant data in the form of images, charts, numbers and graphs. But all these information cannot be directly used as they are. Some knowledge discovery process must be applied to extract the required features. This extracted information can be used in the process of predicting heart diseases. Death rate because of heart diseases can be reduced, if we could prevent occurrence of it. In this paper, an attempt is made to design a system which will help in predicting the heart disease using data mining techniques and the result can be optimized using optimizer algorithm.

Keywords: disease, knowledge discovery process, data mining techniques, optimizer algorithm

--- ∴∵∴ ---

I INTRODUCTION

Heart disease is commonly seen cause of death in last decades. There are various types of the heart diseases; in which Cardiovascular Disease is one major threat. Heart disease does not show any naked eyes symptoms, depending on which initial treatment can be started. Hence detecting heart disease in early stage is very challenging task, and with this fact the necessity of heart disease detection system arises. For implementing such a system data mining techniques can be used. Data mining techniques are helpful in finding and identifying the relationships in large medical datasets, and so they are popularly known by medical researchers. Data mining is also known as an interdisciplinary field as it helps in exploring Database technology, Machine learning, Neural networks and Information Science along with the Statistical techniques. In this paper we will discuss the data mining technique that can be used in prediction of heart disease. And optimizer algorithm can be used along with the data mining technique to improve the performance of the system.

II OVERVIEW OF HEART AND HEART DISEASE

Function of heart is to circulate the blood all over the body. It works as pump and drive bold flow in circulatory system.

Arteries and veins are two main parts of a heart. Arteries carry fresh oxygenated blood from lungs to the all parts of the body, whereas veins carry deoxygenated blood from all over the body to the lungs.

Inflammation and irritation of the inner lining of the coronary arteries occurs due to the diabetes along with poorly controlled hypertension, and Smroking. Cholesterol gets deposited in the inflamed areas and there plaque formation process gets started, this increases the risk of narrowing of the arteries. If arteries get narrow by 40 to 50 % of their original size the blood flow gets compromise and it results in symptoms of angina.

If this plaque gets ruptured it forms blood clot in the coronary artery. Due to this blood clot oxygen-rich blood could not get delivered beyond that blockage to the heart. Because of no oxygen-rich blood that particular part of heart stops working. This terminology is called as myocardial infarction or heart attack.

There are some observed symptoms of angina. It is also called as pain from the heart. Angina can be defined as a crushing pain or heaviness in the center of the chest with radiation of the pain to the left arm or jaw. Patient can feel shortness of breath or sweating. In some patients indigestion and nausea could be observed whereas some patients gets upper abdominal, shoulder, or back pain. As we go in details of the heart diseases study, it is observed that

OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE &ENGINEERING

heart diseases symptoms vary with different group of people. Like women, diabetic patients and old age people may have different pain description than other ones. Like Some patients may feel weakness, fatigue to perform daily work whereas some patients feel no discomfort at all.

Heart disease symptoms are at the peak over the time as narrowing of coronary artery grows over the time and blood supply to that part of the heart decreases. All these symptoms and heart condition of life taking disease cannot be seen by naked eyes or cannot be detected with just single test. Hence there is necessity of a system which can predict the heart disease depending on the patient's health history. In this paper a heart disease prediction system is planned to implement with the help of data mining techniques.

III LITERATURE SURVEY

There are various data mining algorithms which are used in health care industry. Author Yan used Multilayer Perception approach in 2003 and got 63.6% of accuracy [1]. Author Hara et al used Automatically Defined Groups approach and Immune Multi-agent Neural Network in which they got 67.8%, 82.3% accuracy respectively [2]. Similarly author Osareh in 2010 used PNN,KNN , SVM-RBF, SVM-POLY approaches and received accuracy as 92.86, 94.06, 95.45,92.86%, 95.19 respectively[2] With the time researches adopted a hybrid method for implementing the data mining technique in medical domain. Hybrid methodology says, Instead of using single technology, use of more than one technology with the optimizer gives optimized and more accurate results. Intelligent heart disease prediction system, Intelligent and effective heart attack prediction system, Predictive risk assessment of arthrosclerosis are the successful examples of hybrid implementation of data mining techniques.

IV RESEARCH METHODOLOGY

The research methodology used for this paper is through the study of recent journals and publications in medicine and computer science domain.

V RELATED WORK

1. Ankita Dewan with Meghana Sharma performed a research in 2015 [3]. Their work says that neural network is one of the best classification techniques for predicting and classifying the non-linear data. They have also discussed some algorithms for proposed methodology.

2. Research article by Prakash Mahindrakar and Dr. M. Hanumanthappa[4] is focused on data mining in healthcare domain. Article encloses the points like what is KDD and what is data mining? Which algorithms can be used in the healthcare? Limitations and challenges of data mining algorithms.

3."Predictive data mining for medical diagnosis: An overview of heart disease" by jyoti soni , ujma ansari et al [5] says that in research they have considered around 15 attributes for predicting the heart disease. When data mining techniques are applied it is observed that decision tree and Bayesian classification gives similar accuracy where as other techniques are not performing that well. Secondly genetic algorithm can be used to improve the performance.

4. Heart Disease Prediction System Using Data Mining Technique by Abhishek Taneja [6] . Aim of this research was to design a predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset. Certain data mining algorithms like J48 Classifier, Naïve Bayes and Multilayer Perception are used by Weka 3.6.4 machine learning software. And detail results are shown in confusion matrix.

5. Prediction of Heart Disease using classification Algorithm by Hlaudi Daniel Masethe, Mosima Anna Masethe.[7] This research applies J48, Bayes Net, and Naive Bayes, Simple Cart, and REPTREE algorithm to classify and develop a model to diagnose heart attacks. Data set is consists of the 11 attributes which are needed by health practitioner of predicting the heart disease. Results are shown in confusion matrix in details.
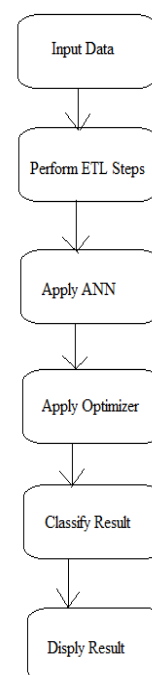
VI PROPOSED SYSTEM

Figure 1.1 System Architecture

OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE &ENGINEERING
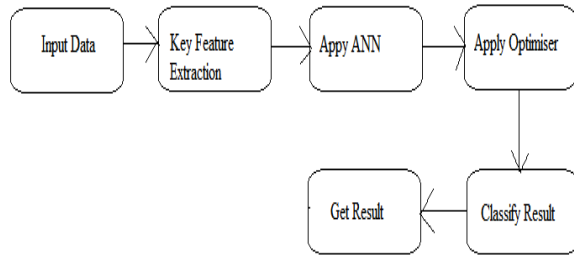

Figure 1.2 DFD Level 1

Figure 1.1 show the system architecture of the proposed system. And DFD level 1 is given in Figure 1.2. Following algorithm can be used in development of the system. 1] Back Propagation 2] Group Search Optimizer 3] Artificial Bee Colony 4] Firefly Algorithm

Back propagation Algorithm: The Back-propagation algorithm is used in multi-layer feed-forward neural network for training the weights in it. Back propagation requires a structure of network with one or more layers. And one layer should be connected to the next layer. A network structure can be defined as one input layer, one hidden layer, and the output layer. The method takes care of adapting the weights to the calculated error in particular input patterns. And then method is applied back to the input layer through the network.

Group search Optimizer: The GSO algorithm can be used to train artificial neural networks. It gives effective results for the three real-world benchmark problems solving. Basically Group search optimizer (GSO), is motivated by animal searching behavior. The GSO is structured on the producer-borrower model, which says that group members search either for producer or for borrower opportunities. This mechanism helps in solving continuous optimization problem.

Artificial Bee Colony: This is meta-heuristic algorithm which is introduced in 2005 by Karaboga. It was motivated by the intelligent foraging behavior of honey bees. The structure consists of three basic blocks: Employed and Unemployed Foraging bees, and Food sources. Both employed and unemployed bee searches for food that is third component, this is the relation between three components. There are different groups of bees in unemployed bee, they are onlooker bees and scout bees. Onlooker bee keep eye on employed bees within colony to select the food source, where as scout bees search for food sources. The work process starts with finding the food by scout bees. Once the food sources are found, employee bees and onlooker bee take the responsibility of carrying the food. When employee bees get exhausted in the process they become a scout bee. And play the role of scout bee to search the food again. In short the exhausted employee changes its role to scout bee and continue the process of food search.

Firefly Algorithm: This algorithm is based on the flashing behavior of the fire flies. Aim of the flash in fire flies is to attract the other fireflies. All the fireflies are unisexual so they attract each other. Brightness is major factor in attraction process. Less brighter one gets attracted towards the brighter one. And process continues. If there is no existence of brighter firefly then than given one, it will start moving randomly. All the above algorithms are planned to utilize in the proposed heart disease predication system.

VII CONCLUSION AND FUTURE SCOPE

As heart disease is listed among the life taking diseases it is necessary to predict it in proper time. Its symptoms cannot be seen by the naked eyes hence considering history of person with some key factors are considered while designing the heart disease prediction system. Technically data mining algorithm are used for the prediction process and the result is optimized with some optimization algorithm. Hybrid or integrated Data Mining technique such as fusion of different classifiers, fusion of clustering with classification or association with clustering or classification etc. can be used for achieving better performance.

In future the heart disease prediction system can be used to design a web based application. Automated heart disease prediction system can be implemented in remote areas like rural regions or country sides, to imitate like human diagnostic expertise for prediction of ailment. System is applicable for assisting the medical practitioners in the process of heart disease prediction

REFERENCES

[1] H. Yan, "Development of a decision support system for heart disease diagnosis using multilayer perceptron", Proceedings of the 2003 International Symposium, vol. 5, (2003), pp. V-709- V-712.

[2] A. Osareh and B. Shadgar, "Machine Learning Techniques to Diagnose Breast Cancer", Health Informatics and Bioinformatics (HIBIT), IEEE, (2010).

[3] Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification By Ankita Diwan And Meghana Sharma , 978-9-3805-4416-8/15/$31.00_c 2015 IEEE

[4] Data Mining In Healthcare: A Survey of Techniques and Algorithms with Its Limitations and Challenges By Prakash Mahindrakar, Dr. M. Hanumanthappa, Prakash Mahindrakar et al Int. Journal of Engineering Research and Applications

OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE &ENGINEERING

www.ijera.com ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.937-941

[5]   Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction By Jyoti Soni , Ujma Ansari, Dipesh Sharma, Sunita Soni, International Journal of Computer Applications (0975 – 8887)
Volume 17– No.8, March 2011

[6] Heart Disease Prediction System Using Data Mining Techniques, By Abhishek Taneja, Oriental Journal Of Computer Science & Technology, ISSN: 0974-6471, December 2013, Vol. 6, No. (4):

[7] Prediction of Heart Disease using Classification Algorithms By Hlaudi Daniel Masethe, Mosima Anna Masethe, Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II WCECS 2014, 22-24 October, 2014, San Francisco, USA

[8] Prediction of Heart Disease Using   a Hybrid Technique in Data Mining Classification by Ankita Dewan and Meghna Sharma

[9] Computation Methods for the   Diagnosis and Prognosis of Heart Disease by Deepthi S and Aswathy Ravikumar

[10] J. Han and M. Kamber, "Data mining: concepts and techniques", 2nd ed.

[11] A. Hara and T. Ichimura, "Data Mining by Soft Computing Methods for the Coronary Heart Disease Database", Fourth International Workshop on Computational Intelligence & Application, IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, (2008) December 10-11.