IJCSMS International Journal of Computer Science & Management Studies, Vol. 12, Issue 03, September 2012
ISSN (Online): 2231 –5268
www.ijcsms.com

An Intelligent Fuzzy Based Scheme To Analyze Heart Diseases

Neeru Anand[1], Dr. Rajendar Chhillar[2]

[1]M. Tech. Student, DCSA, MDU, Rohtak, Haryana, India
anandneeru2@gmail.com

[2]Professor, DCSA, MDU, Rohtak, Haryana, India
chhillar02@gmail.com

Abstract

The main objective of this research work is to define some fuzzy based ruleset to identify the patient disease based on person medical related information. This information includes the person age, calestrol level, heart beat and blood pressure. In this work these four characters of a patient are studied and presented in the form of fuzzy rule. As the fuzzy rule is applied on these characteristics collectively and based on it the decision about the heart disease is taken. Any of the medical diagnostic application cannot return true results; there are always some chances of perception while dealing with medical area. Even though they obtained results shows that effective use of application for the initial prediction of heart disease based on user characteristics.

Keywords: *Fuzzy Rule, Data Mining, Medical Diagnostic System, Fuzzy Logic, Physical Characteristics.*

I. Introduction

The influence of data mining on the quality of Health Care cannot be understated. All Health Care organizations retain detailed and comprehensive records of patient data. Trends and patterns identified in these records can positively impact the quality of Health Care. The huge amounts of patient data, makes identification of these trends an arduous task. However data mining applications, built for this purpose, can make this very simple and produce efficient results.

There have been several cases, where application of data mining techniques, have helped resolve a problem in the health industry. For instance, data mining on pneumonia patient records in a hospital, showed that patients who were administered medication immediately on arrival responded better than patients who were not administered medication on arrival. In order to arrive at this conclusion the data mining application, used several inputs, such as the tests and other information of the patients who showed better medication results. Various relations were drawn between the inputs. One of these was the relation between the results and the time taken to administer medication after arrival. It was found that, shorter the time, better the result.

There were several other key issues that were addressed at this time. The data mining tests proved that several tests, which were largely extraneous, were conducted on the patients. These led to a delay in the administration of medication and thereby affected the recovery of the patient. To overcome this, a standardized plan was created to treat pneumonia patients. The identification of these associations between inputs and finding the resultant best outcome was possible only because of data mining techniques.

A) Mining the Data

Health care organizations store huge amounts of data in the form of patient databases. Trends in these databases can be identified using data mining practices, which sort and model the data in order to arrive at a conclusion. The data mining applications present the data in the form of data marts. This allows end users to choose the specific sets of data, which they want to be analyzed. The data in these data marts can then be presented using a graphical user interface, arranging the data into columns and rows.

In the Health care industry, however, the lack of standard clinical vocabulary has hindered the process of data mining to a certain extent. For example a simple term such as 'hypertension' can be expressed in various ways in health care. This could lead to unnecessary problems, during the process of data mining. The

increase in the use of standardized terms will reduce the percentage of errors in the data mining process.

Cleaning the data before it can be mined is also an important step in the data mining process. In many Health care organizations, the mode of preparing patient reports can lead to a good deal of confusion. For instance, in a certain hospital, a report was prepared, before and after a patient went in for an X-ray check. This could be construed as two different reports, when analyzing the data and produce erroneous results. Further in certain organizations, in order to reduce the number of reports, a patients' record contains only the name of the attending physician and not the names of other physicians consulted or tests performed at a later stage, leading to erroneous predictions.

The data mining effort thus requires the wholehearted participation of all health care personal to produce comprehensive and correct reports, which can be mined. Further, the number of input variables for the data mining application has to be determined correctly. The number of inputs should not be so large, that it produces not be limited to such an extent, that they produce biased results. Co-operation between the physicians and analysts is also recommended, since some of the results might be more easily understood by the health care personal.

II Literature Survey

In year 2000, Shusaku Tsumoto performed a work," Problems with Mining Medical Data". Thus, it is highly expected that data mining methods will find interesting patterns from databases as reuse of stored data and be important for medical research and practice because human beings cannot deal with such a huge amount of data. In this paper, we focus on the characteristics of medical data and discuss how data miners deal with medical data. In year 2004, Y. Alp Aslandogan performed a work," Evidence Combination in Medical Data Mining". We combine the beliefs of three classifiers: k-Nearest Neighbor (kNN), Naïve Bayesian and Decision Tree. Dempster's rule of combination combines three beliefs to arrive at one final decision. Our experiments with k-fold cross validation show that the nature of the data set has a bigger impact on some classifiers than others and the classification based on combined belief shows better overall accuracy than any individual classifier. We compare the performance of Dempster's combination (with differentiation-based uncertainty assignment) with those of performance-based linear and majority vote combination models. We study the circumstances under which the evidence combination approach improves classification. In year 2006, Wong Kok Seng performed a work," Collaborative Support for Medical Data Mining in Telemedicine". This paper will discussed an idea on how to overcome above mentioned issues and proposed a solution that can be served as the platform for future medical data sharing in telemedicine. The successful development of the working prototype will greatly enhance the functionality of existing data sharing in the hospital. At the same time, the tools and algorithms designed in this idea will helps to solve some ofthe data mining challenges. In year 2008, Hai Wang performed a work," Medical Knowledge Acquisition through Data Mining". Data mining has been widely considered as an effective tool for knowledge discovery. This paper discusses the important role of medical experts for medical data mining, and presents a model of medical knowledge acquisition through data mining. In year 2010, Zhao Yongyi performed a work," Intelligent Data Mining for Economic Prediction and Analysis". This paper describes importance that the application of economic data in the data mining algorithm and its application, which combines with the current economic data of national macro-economic indicators, presents the data warehouse model structure and its implementation characteristics, and uses SQL Server 2005 data warehouse and data mining solutions on economic data for the application of data mining solution, system architecture, algorithms implementation, and finally discusses the application of data mining algorithms development trends and key technologies in the economic field. In year 2010, Shiguo wang performed a work," A Comprehensive Survey of Data Mining-based Accounting-Fraud Detection Research". Bayesian network, and stack variables etc. Regression Analysis is widely used on hiding data. Generally the detecting effect and accuracy of NN are superior to regression model. General conclusion is that model detecting is better than auditor detecting rate without assisting. There is a need to introduce other algorithms of no-tag data mining. Owing to the small size of fraud samples, some literature reached conclusion

based on training samples and may overestimated the effect of model.

In year 2010, Mahnoosh Kholghi performed a work," Classification and Evaluation of Data Mining Techniques for Data Stream Requirements". Generally, two main challenges are to design fast mining methods for data streams and the need to promptly detect changing concepts and data distribution because of highly dynamic nature of data streams. The goal of this article is to analyze and classify the application of diverse data mining techniques in different challenges of data stream mining. For this goal, this article tries to categorize and analyze related researches for better understanding and to reach a framework that can map data mining techniques to data stream mining challenges and requirements.

III Proposed Scheme

A medical diagnostic system cannot be implemented without an expert personal. In this present we have defined a medical application to identify the person heart disease or the chances of heart disease. For this heart disease analysis we have used one of the major soft computing techniques called fuzzy logic. The fuzzy logic is one of the intelligent schemes to deal with uncertainty. The fuzzy logic takes the intelligent decision based on the characteristics of current population set. The fuzzy rule is implemented on the physical characteristics of a person. The basic flow the proposed work is shown below

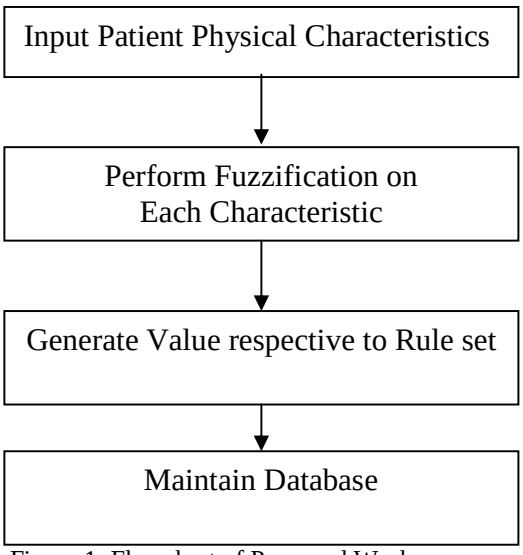
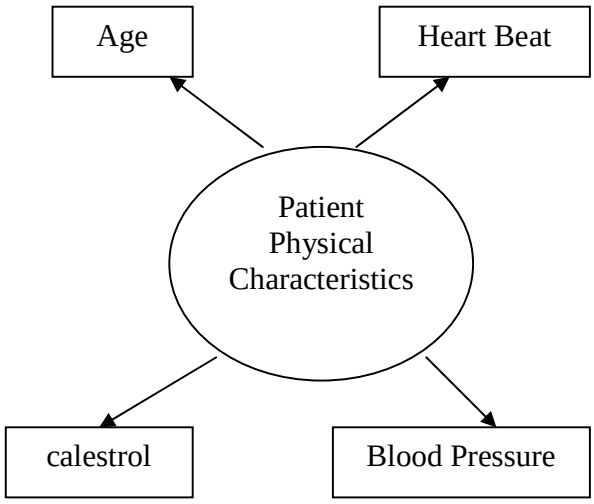


Figure 1: Flowchart of Proposed Work

The basic characteristics included in this proposed work are-



At first, we will describe the input variables with their membership functions. In second step, we introduce the output variable with its membership functions. In next section, we'll show the rules of system.

A. Blood Pressure

Blood Pressure is the major factor that can predict the chances of heart disease. Generally a heart patient has the higher blood pressure. According to the fuzzy rule set we have divided the patient blood pressure in four major categories called "Very High", "High", "Normal" and "Low". If the blood pressure is greater than 172, it is considered very high. If the blood pressure is between 154 and 172 it is considered high blood pressure. If it is between 127 and 154 then it is considered medium otherwise it is considered "Low". The fuzzy generation of this ruleset is shown as under in figure 3.

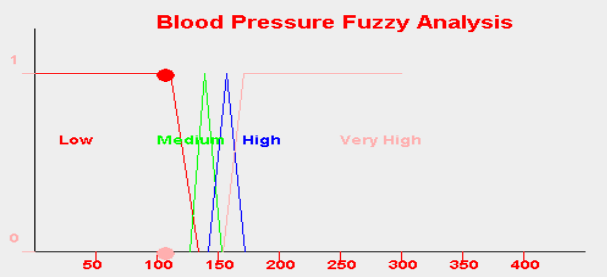


Figure 3: Blood Pressure Analysis

B. Cholesterol

Calestrol is the another major factor that can increase the chances of heart disease in a patient. Higher the Calestrol value more chances of heart disease. According to the study we have divided the calestrol level in a patient in four major categories called "Very High", "High", "Normal" and "Low". If the Calestrol Level is greater than 307, it is considered very high. If Calestrol level is between 217 and 307 it is considered high blood pressure. If it is between 188 and 250 then it is considered medium otherwise it is considered "Low". The fuzzy generation of this ruleset is shown as under in figure 4.

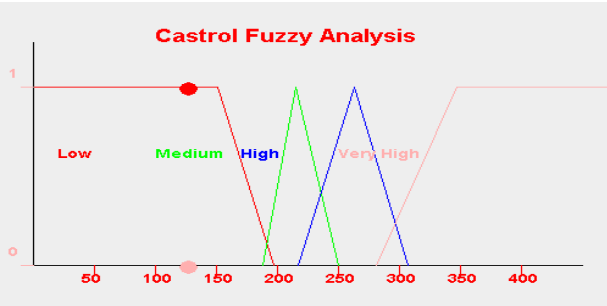


Figure 4: Calestrol Level Analysis

C. Heart Beat

Heart beat is the major factor that can increase the chances of heart disease in a patient. Higher the Heart Beat value more chances of heart disease. According to the study we have divided the Heart Beat level in a patient in three major categories called "High", "Normal" and "Low". If the Heart Beat Level is greater than 194, it is considered high. If Heart Beat level is between 141 and 194 it is considered Average Heart Beat. otherwise it is considered "Low". The fuzzy generation of this ruleset is shown as under in figure 5.

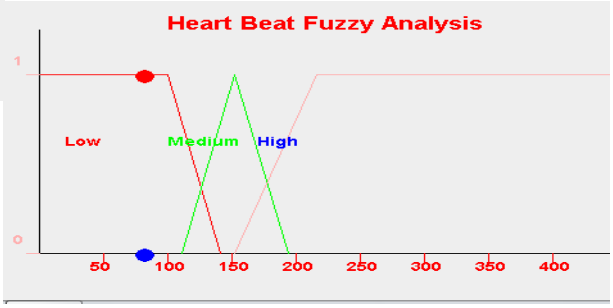


Figure 5: Heart Beat Fuzzy Analysis

D. Age

Age is the another major factor that can increase the chances of heart disease in a patient. Higher the Heart Beat value more chances of heart disease. According to the study we have divided the Heart Beat level in a patient in three major categories called "High", "Normal" and "Low". If the Heart Beat Level is greater than 194, it is considered high. If Heart Beat level is between 141 and 194 it is considered Average Heart Beat. Otherwise it is considered "Low". The fuzzy generation of this ruleset is shown as under in figure 6.

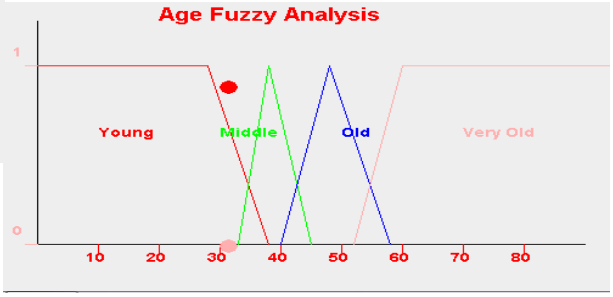


Figure 6: Age Level Analysis

References

[1].Rafael S. Parpinelli, Heitor S. Lopes, Member, IEEE, and Alex A. Freitas.
[2].Predictive data mining in clinical medicine: a focus on selected methods and Applications Riccardo Bellazzi, Fulvia, Ferrazzi and Lucia Sacchi.
[3].Predictive data mining in clinical medicine: Current issues and guidelines by Riccardo Bellazzia,, Blaz Zupanb
[4].DATA MINING FRAMEWORK by Hemambika Payyappillil, College of Engineering and Mineral Resources at West Virginia University.
[5].FUZZY LOGIC IN CLINICAL DECISION SUPPORT SYSTEM by Jim Warren, Gleb Beliakov and Berend van der Zwaag.
[6].Russell, S. and Norvig, P., Artificial Intelligence: A modern approach, pp. 23, Prentice-Hall International, 1995.
[7]. Buchanan, B. and Shortliffe, E., Rule-Based Expert Systems, Addison-Wesley, Reading, Massachusetts, 1984.
[8].Waterman, D.A., A Guide to Expert Systems, Reading, MA: Addison-Wisley, 1986.
[9].Kolodner, J. L., Case-Based Reasoning, California: Morgan Kaufman Publishers, 1993.
[10]. Phan, T. and G. Chen, Some Applications of Logic in Rule-Based Expert Systems, Expert Systems, vol.19, No.4, pp.208-223, 2002.
[11]. Clancey, W. J. and Shortliffe, E. H. (ed.)., NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching. In: Readings in Medical Artificial Intelligence: The First Decade, Addison-Wesley, pp.361-381, 1984.
[12]. Zadeh, L. A., sets, Information and Control, 8, pp.338-353, 1995.
[13]. Leung R.W.K, Lau H.C.W., and Kwong C.K., on a responsive replenishment system: a logic approach, Expert Systems, vol. 20, pp. 20-32, 2003.