RESEARCH ARTICLE OPEN ACCESS

Decision Support System for Precluding Coronary Heart Disease (CHD) Using Fuzzy Logic

K Cinetha[1], Dr. P. Uma Maheswari[2]
P.G Scholar[1], Professor[2], Department of Computer Science and Engineering,
Info Institute of Engineering Affiliated to Anna University,
Chennai, India

ABSTRACT
Cardiovascular diseases (CVD) remains the biggest cause of deaths worldwide and the Heart Disease Prediction at the early stage is importance. Coronary heart disease (CHD) is the leading cause of death for both men and women and accounts for approximately 600,000 deaths in the United States every year. To design a Decision support System for Precluding Coronary Heart Disease (CHD) risk of patient for the next ten-years for prevention. To assist medical practitioners to diagnose and predict the probable complications well in advance. Identifying the major risk factors of Coronary Heart Disease (CHD) categorizing the risk factors in an order which causes high damages such as high blood cholesterol, diabetes, smoking, poor diet, obesity, hyper tension, stress, etc. Data mining functionalities are used to identify the level of risk factors to help the patients in taking precautionary actions to stretch their life span. Primary prevention is recommended as promoting healthy lifestyle and habits through increased awareness and consciousness, to prevent development of any risk factors using fuzzy logic and decision tree.
Keywords:- Data mining, Heart disease, Coronary Heart Disease, Clustering, Fuzzy Logic, Decision tree

I. INTRODUCTION

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Healthcare industry today generates large amount 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 analysed for knowledge extraction that enables support for cost-savings and decision making. Data mining provides a set of tools and techniques that can be applied to this processed data to discover hidden patterns and also provides healthcare professionals an additional source of knowledge for making decisions.

Coronary heart disease (CHD) can be caused due to risk factors like high blood pressure, high blood cholesterol, tobacco use, obesity, unhealthy diet, physical inactivity, diabetes, advancing age, and inherited disposition. Coronary heart disease (CHD) is the narrowing or blockage of the coronary arteries, usually caused by atherosclerosis. Atherosclerosis (sometimes called “hardening” or “clogging” of the arteries) is the build-up of cholesterol and fatty deposits (called plaques) on the inner walls of the arteries. These plaques can restrict blood flow to the heart muscle by physically clogging the artery or by causing abnormal artery tone and function [21].

Without an adequate blood supply, the heart becomes starved of oxygen and the vital nutrients it needs to work properly. This can cause chest pain called angina. If blood supply to a portion of the heart muscle is cut off entirely, or if the energy demands of the heart become much greater than its blood supply, a heart attack (injury to the heart muscle) may occur. Coronary heart disease (CHD) is the leading cause of death for both men and women and accounts for approximately 600,000 deaths in the United States every year.

It is most commonly equated with atherosclerotic coronary artery disease, but coronary disease can be due to other causes, such as coronary vasospasm, where the stenosis to be caused by spasm of the blood vessels of the heart it is then usually called Prinzmetal's angina. In figure-1, shows Coronary heart disease (CHD) [21] is epidemic in India and one of the major causes of disease-burden and deaths. The leading cause of death worldwide. Previously thought to affect primarily high-income countries, CHD now leads to more death and disability in low- and middle-income countries, such as India, with rates that are increasing disproportionately compared to high-income countries. CHD affects people at younger ages in low- and middle-income countries, compared to high-income countries, thereby having a greater economic impact on low- and middle-income countries. Effective screening, evaluation, and management strategies for CHD are well established in high-income countries, but these strategies have not been fully implemented in India. The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure – a condition that causes around half of all deaths from stroke and heart disease. Heart disease, also known as cardiovascular disease (CVD), encloses a number of conditions that influence the heart – not just heart attacks. Heart disease also includes functional problems of the heart such as heart-valve abnormalities or irregular heart rhythms. These problems can lead to heart failure, arrhythmias and a host of other problems. Effective and efficient automated heart disease prediction systems can be beneficial in healthcare sector for heart disease prediction.

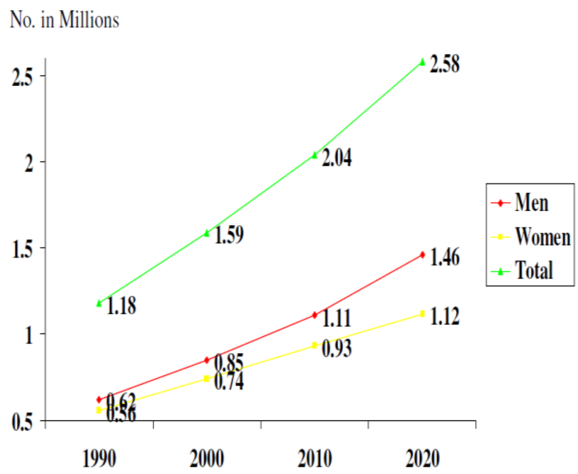

Figure -1 Projections for Coronary Heart Disease (CHD) Mortality in India.

Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. Description focuses on finding human interpretable patterns describing the data. Data Mining has potential applications in several fields, not the least of which is Health Care. The primary purpose of the Data Mining is to help determine trends in patient records to improve Health Care.

II. RELATED WORK

A. Fuzzy logic

Fuzzy logic is a mathematical discipline that we use every day and helps us to reach the structure in which we interpret our own behaviors. Its basis is formed by "true" and "false" values and Fuzzy Set Theory (FST) through which the values in between –"partially true", "partially false"- are determined. FST is a theory that aims to express the uncertainties of life such as "warm" and "cool" which are in between "hot" and "cold" mathematically. And behind these values there is an unclear numerical value. Generally, fuzzy expert systems (FES) are the systems that are based on knowledge or rule. That is, in the basis of a FES lie the fuzzy "if-then" rules. After deciding on designing a fuzzy system the first step to follow is to collect the fuzzy rules of "if-then". These rules are generally collected with the help of an expert [20].

In FES model, the input and output values of the system are crisp values. By fuzzification these crisp input values, its fuzzy membership values and degrees are obtained. These obtained fuzzy values are processed in fuzzy inference mechanism. Here, the fuzzy output values which are also obtained using rule-base are send to the defuzzification unit, and from the this unit the final crisp values are obtained as output. Determination of the next 10-years risk also depends on various items such as weight, smoking, sex, disease history of patients, family genetic factors, blood sedimentation and etc. And there is no formulation for determining the risk according to these items. In this study, a decision support system is designed for determining CHD risk ratio according to next 10-years risk of patient. The system presents alternative results to the doctor instantly, as the data of a patient are inputted to the system.

B. Decision tree

Decision tree include CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and C4.5. These algorithms differ in selection of splits, when to stop a node from splitting, and assignment of class to a non-split node. CART uses Gini index to measure the impurity of a partition or set of training tuples. It can handle high dimensional categorical data. Decision Trees can also handle continuous data (as in regression) but they must be converted to categorical data. The decision tree is built from 1230 training set. We will refer to a row as a data instance. The data set contains three predictor attributes, namely Age, Gender, Intensity of symptoms and one goal attribute, namely disease whose values (to be predicted from symptoms) indicates whether the corresponding patient risk level for the next ten years[19].

C. *Clustering*

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how to decide what constitutes a good clustering? It can be shown that there is no absolute "best" criterion which would be independent of the final aim of the clustering. Consequently, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), in finding "natural clusters" and describe their unknown properties ("natural" data types), in finding useful and suitable groupings ("useful" data classes) or in finding unusual data objects (outlier detection) [19].

III. PROPOSED SYSTEM

The proposed system is to build the Decision Support System for precluding Coronary Heart Disease (CHD) using data mining techniques to identify the level of risk in coronary heart diseases. Coronary heart disease can be caused due to risk factors like high blood pressure, high blood cholesterol, tobacco use, obesity, unhealthy diet, physical inactivity, diabetes, advancing age, and inherited disposition. This system helps the patients in take precautionary actions to stretch their life span and to assist medical practitioners to diagnose and predict the probable complications well in advance. Identifying the major risk factors of CHD categorizing the risk factors in an order which causes high damages such as high blood cholesterol, diabetes, smoking, poor diet, obesity, hyper tension, stress, etc.

Primary prevention is recommended as promoting healthy lifestyle and habits through increased awareness and consciousness, to prevent development of any risk factors. This system which predicts the possibility of heart disease risk of patient for the next ten-years for prevention using clustering algorithm. To assist medical practitioners to diagnose and predict the probable complications well in advance.

A. Fuzzy Methods

The population-based sample used for this report included 2489 men and 2856 women 30 to 74 years old at the time of their Framingham Heart Study examination in 1971 to 1974. Participants attended either the 11th examination of the original Framingham cohort11 or the initial examination of the Framingham Offspring Study.12 Similar research protocols were used in each study, and persons with overt CHD at the baseline examination were excluded. At the 1971–1974 examination, a medical history was taken and a physical examination was performed by a physician. Persons who smoked regularly during the previous 12 months were classified as smokers. Height and weight were measured, and body mass index (kg/m2) was calculated. Two blood pressure determinations were made after the participant had been sitting at least 5 minutes, and the average was used for analyses. Hypertension was categorized according to blood pressure readings by JNC-V definitions10: optimal (systolic ,120 mm Hg and diastolic ,80 mm Hg), normal blood pressure (systolic 120 to 129 mm Hg or diastolic 80 to 84 mm Hg), high normal blood pressure (systolic 130 to 139 mm Hg or diastolic 85 to 89 mm Hg), hypertension stage I (systolic 140 to 159 mm Hg or diastolic 90 to 99 mm Hg), and hypertension stage II–IV (systolic $160 or diastolic $100 mm Hg). When systolic and diastolic pressures fell into different categories, the higher category was selected for the purposes of classification.

Blood pressure categorization was made without regard to the use of antihypertensive medication. Diabetes was considered present if the participant was under treatment with insulin or oral hypoglycemic agents, if casual blood glucose determinations exceeded 150 mg/dL at two clinic visits in the original cohort, or if fasting blood glucose exceeded 140 mg/dL at the initial examination of the Offspring Study participants. Blood was drawn at the baseline examination after an overnight fast, and EDTA plasma was used for all cholesterol and triglyceride measurements. Cholesterol was determined according to the Abell-Kendall technique,13 and HDL-C was measured after precipitation of VLDL and LDL proteins with heparinmagnesium according to the Lipid Research Clinics Program protocol.14 When triglycerides were ,400 mg/dL, the concentration of LDL-C was estimated indirectly by use of the Friedewald formula15; for triglycerides $400 mg/dL, the LDL-C was estimated directly after ultracentrifugation of plasma and measurement of cholesterol in the bottom fraction (plasma density ,1.006).16 Cutoffs for TC (,200, 200 to 239, 240 to 279, and $280 mg/dL), LDL-C (,130, 130 to 159, and $160 mg/dL), HDL-C (,35, 35 to 59, and $60 mg/dL), cigarette smoking, diabetes, and age were considered in this report.

The cholesterol and LDL-C cutoffs are similar to those used for the NCEP ATP II guidelines and were partly dictated by the number of persons with higher levels of TC or LDL-C. For those reasons, we have provided information for cholesterol categories of 240 to 279 and $280 mg/dL and for LDL-C $160 mg/dL. Too few persons had LDL-C $190 mg/dL to provide stable estimates for CHD risk. Study subjects were followed up over a 12-year period for the development of CHD (angina pectoris, recognized and unrecognized myocardial infarction, coronary insufficiency, and coronary heart disease death) according to previously published criteria. “Hard CHD” events included total CHD without angina pectoris.17 Surveillance for CHD consisted of regular examinations at the Framingham Heart Study clinic and review of medical records from outside physician office visits and hospitalizations. Statistical tests included age-adjusted linear regression or logistic regression to test for trends across blood pressure, TC, LDL-C, and HDL-C categories. 18 Age-adjusted Cox proportional hazards regression and its accomp anying c statistic were used to test for the relation between various independent variables and the CHD outcome and to evaluate the discriminatory ability of various prediction models.19,20 The 12-year follow-up was used in the proportional hazards models, and results were adapted to provide 10-year CHD incidence estimates. Separate score sheets were developed for each sex using TC and LDL-C categories. These sheets adapted the results of proportional hazards regressions by use of a system that assigned points for each risk factor based on the value for the corresponding b-coefficient of the regression analyses.

The relative risk, but not the attributable risk, for TC and CHD declines with advancing age.21 Quadratic terms for age were considered in the models for the score sheets. Furthermore, CHD risk is associated with HDL-C in the elderly, 22–24 and interaction terms for TC and age were also considered in the development of the prediction models.22 among women, an age-squared term was found to be significant in the prediction models and was incorporated into the score sheets. Neither age3TC nor age3LDL-C was found to be significant in either sex.

Score sheets for prediction of CHD using TC and LDL-C categorical variables were developed from the b-coefficients of Cox proportional hazards models. The TC range was expanded in 40-mg/dL increments to include $160 mg/dL and $280 mg/dL, the HDL-C range 35 to 59 mg/dL was partitioned to provide three levels for each sex and both optimal and normal blood pressure categories were included. The score sheets provide comparison 10-year absolute risks for persons of the same age and sex for average total CHD, average hard CHD (total CHD without angina pectoris), and low-risk total CHD. Risk factors are shaded, ranging from very low relative risk to very high. Such distinctions are arbitrary but provide a foundation to determine the need for clinical intervention.

IV. SIMULATIONS

Estimates risk for chd over a period of 10 years based on framingham experience in women 30 to 74 years old at baseline. Average risk estimates are based on typical framingham subjects, and estimates of idealized risk are based on optimal blood pressure, tc 160 to 199 mg/dl (or ldl 100 to 129 mg/dl), hdl-c of 55 mg/dl in women, no diabetes, and no smoking. Use of the ldl-c categories is appropriate when fasting ldl-c measurements are available. Pts indicates points.

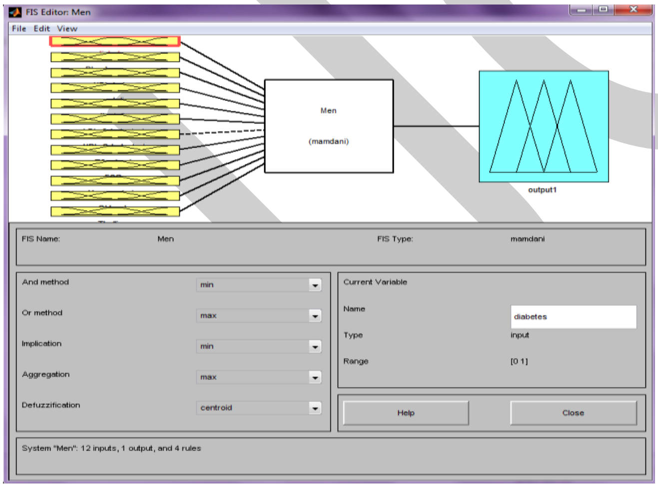

Figure-2 Optimization of fuzzy logic rules using membership function mamdani for Men

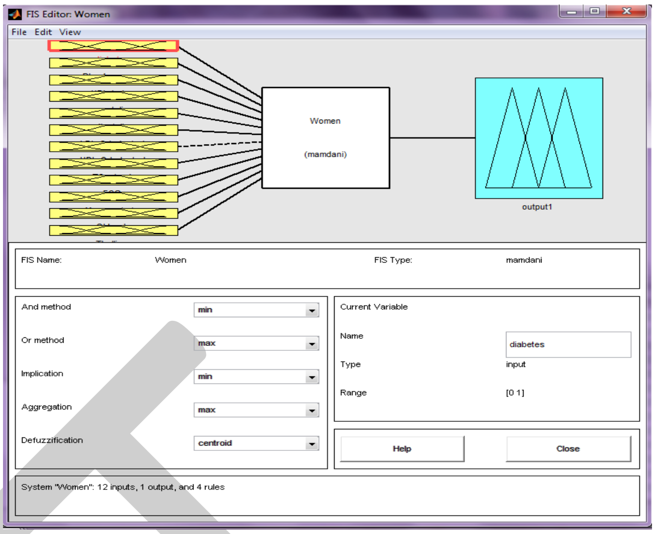

Figure-3 Optimization of fuzzy logic rules using membership function mamdani for Women

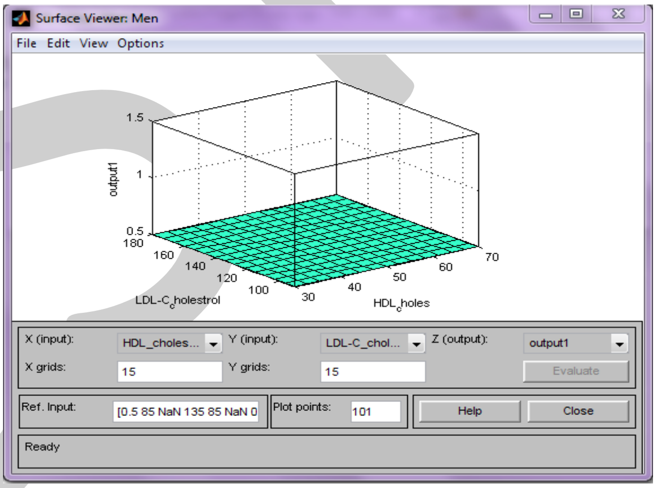

Figure-4 Surface viewer of HDL_C and LDL_C for Men

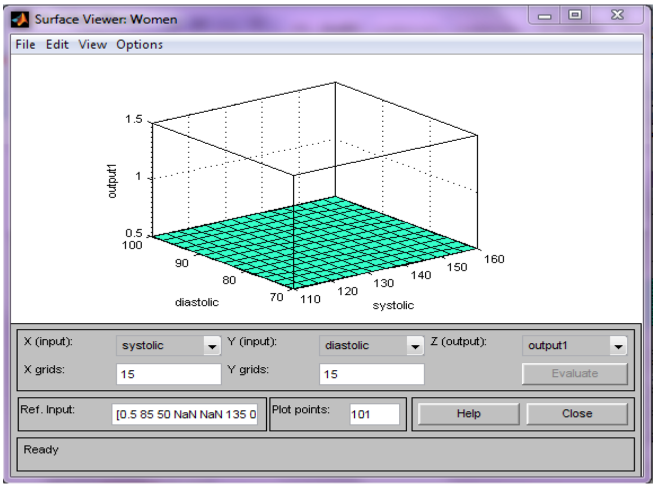

Figure-5 Surface viewer for Systolic and Diastolic for Women

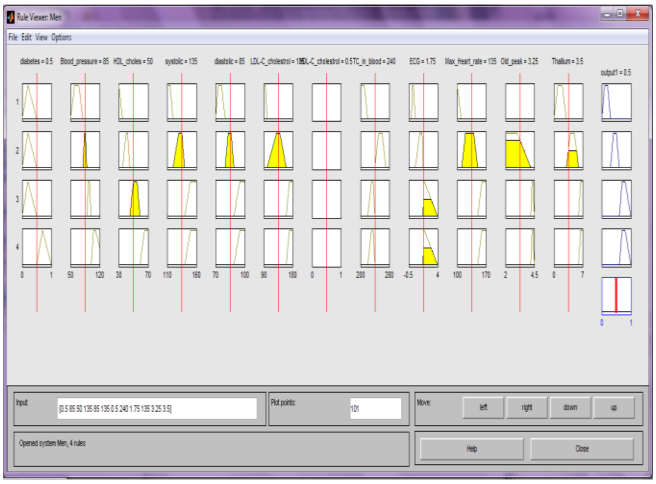

Figure-6 Fuzzy Rule generation of input attributes for Men

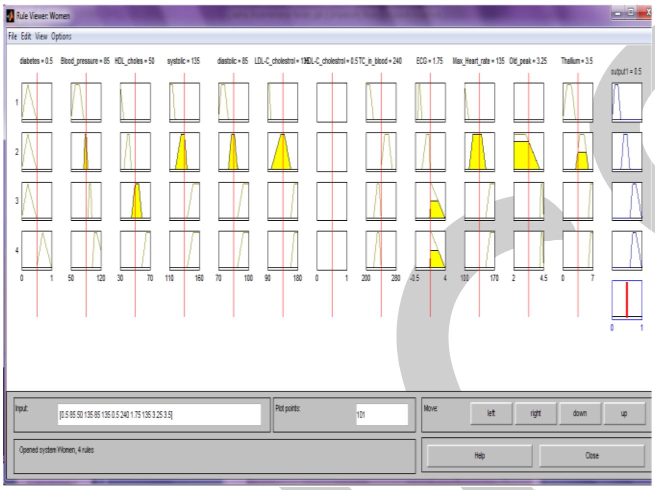

Figure-7 Fuzzy Rule generation of input attributes for Women

A. Performance Analysis

Performance analysis deals with the implementation of Decision tree method in the establishment of fuzzy rules for the diagnosis of coronary heart disease risk (CHD). The method is used to produce the clustered data. Thus, the fuzzy rule is obtained by extracting rules from the cluster using the Least Square Error (LSE). Determination of the best cluster is a rule fuzzy selected at each time of testing performed using variant analysis. The smaller value of variant boundaries for cluster, the more ideal for clustering is. The best ac-curacy of the system of selected rules if it is applied to the TSK inference orde-1method is 97.67%, where the number of rule is 2 which is the result of the clustering process using the reject ratio = 0.8 at radius 0.45 for 1230 training data. Figure-8 shows the result generated by the trace graph is given to the x-graph tool to plot the comparative graph for the estimated accuracy of the system with Fuzzy Logic and neural network. Figure-9 shows the result generated by the trace graph is given to the x-graph tool to plot the comparative graph for the estimated Error rate of the system with Fuzzy Logic and neural network.

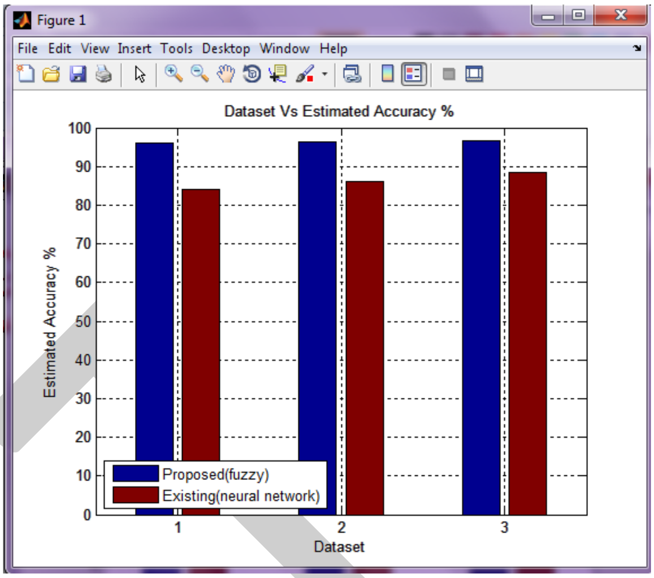

Figure-8 Estimated Accuracy the system using Fuzzy Logic

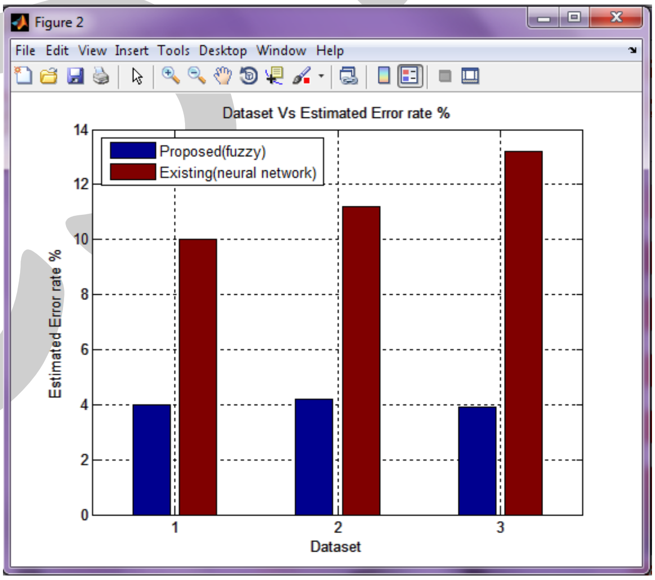

Figure-9 Estimated Error rate of the system

V. CONCLUSION AND FUTURE WORK

This decision support system which predicts the possibility of heart disease risk of patients for next ten-years using fuzzy logic and decision tree which help the patients in taking precautionary actions to stretch their life span. To assist medical practitioners to diagnose and predict the probable complications well in advance. Primary prevention is recommended as promoting healthy lifestyle and habits through increased awareness and consciousness, to prevent development of any risk factors, system which predicts the possibility of heart disease risk for prevention.

Recommended guidelines of blood pressure, total cholesterol, and LDL cholesterol effectively predict CHD risk in a middle-aged white population sample. A simple coronary disease prediction algorithm was developed using categorical variables, which allows physicians to predict multivariate CHD risk in patients without overt CHD with 97.67% estimated accuracy using Fuzzy logic and decision tree with 1230 training dataset.

REFERENCES

[1] *Allahverdi N, Torun S and Saritas I, "Design of a Fuzzy Expert System for Determination of Coronary Heart Disease Risk and Technologies," Computer Systems and Technologies 2006, Veliko Turnovo, 14-15 June 2007.*

[2] *Burch P R J, "Coronary Disease: Risk Factors, Age, and Time," American Heart Journal, Vol. 97, No. 4, 1979, pp. 415-419. doi:10.1016/0002-8703(79)90385-5, 2012.*

[3] *Bahadur Patel, Ashish Kumar Sen D P, Shamsher Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-FuzzyIntegrated Approach Two Level", International Journal Of Engineering And Computer Science ISSN:2319-7242, Page No. 2663-2671, Volume 2 Issue 9 Sept., 2013.*

[4] *Heon Gyu Lee, Ki Yong Noh, and Keun Ho Ryu, "A Data Mining Approach for Coronary Heart Disease Prediction using HRV Features and Carotid Arterial Wall Thickness", International Conference on Bio Medical Engineering and Informatics, 2008.*

[5] *Sulabha S. Apte, "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.*

[6] *Rajalaxmi R R, "A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier", International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C 2012) International Journal of Computer Applications (IJCA), 2012.*

[7] *Srinivas K, Raghavendra Rao R, Govardhan A, "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 Hefei, China. August 24–27, 2010.*

[8] *Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-4, September 2012.*

[9] *Nidhi Bhatla and Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181,Vol. 1 Issue 8, October - 2012.*

[10] *Chitra R and Seenivasagam V, "REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES", ISSN: 2229-6956(ONLINE) ICTACT JOURNAL ON SOFT COMPUTING, JULY 2013, VOLUME: 03, ISSUE: 04, 2013.*

[11] *Yosawin Kangwanariyakul, Chanin Nantasenamat, Tanawut Tantimongcolwat, Thanakorn Naenna, "DATA MINING OF MAGNETOCARDIOGRAMS FOR PREDICTION OF ISCHEMIC HEART DISEASE", EXCLI Journal 2010;9:82-95 – ISSN 1611-2156, published: July 30, 2010.*

[12] *Mai Shouman, Tim Turner, Rob Stocker, "USING DATA MINING TECHNIQUES IN HEART DISEASE DIAGNOSIS AND TREATMENT", 978-1-4673-0483-2-c 2012 IEEE, 2012.*

[13] *Alberto Porta, Giuseppe Baselli, Stefano Guzzetti, Massimo Pagani, Alberto Malliani, and Sergio Cerutti, Senior Member, "Prediction of Short Cardiovascular Variability Signals Based on Conditional Distribution", 2010, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 47, NO. 12, DECEMBER 2000.*

[14] *Tuan D. Pham, Stephen T. C. Wong, "Computational Prediction Models for Early Detection of Risk of Cardiovascular Events Using Mass Spectrometry Data", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 5, SEPTEMBER 2008.*

[15] *Nurul Hikmah Kamaruddin, M.Murugappan, Mohammad Iqbal Omar, "Early Prediction of Cardiovascular Diseases Using ECG signal: Review", IEEE 2012.*

[16] *Chaitrali S. Dangare, Sulabha S. Apte, "A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS", INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY(IJCET), ISSN 0976 – 6367, Volume 3, Issue 3, October - December (2012), pp. 30-40, 2012.*

[17] *Novruz ALLAHVERDI, Serhat TORUN, Ismail SARITAS "DESIGN OF A FUZZY EXPERT SYSTEM FOR DETERMINATION OF CORONARY HEART DISEASE RISK", International Conference on Computer Systems and Technologies - CompSysTech'07, 2007.*

[18] *Rajeswari K, Vaithiyanathan V and Amirtharaj P, "A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique – An Indian Case Study", International Journal of Machine Learning and Computing, Vol. 1, No. 3, August 2011.*

[19] *Persi Pamela I , Gayathri P and Jaisankar N, "A Fuzzy Optimization Technique for the Prediction of Coronary Heart Disease Using Decision Tree", International Journal of Engineering and Technology (IJET), 2013.*

[20] *Rajeswari K, Vaithiyanathan V, "Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.4, April 2011*

[21] *K.Cinetha* and Dr. P. Uma Maheshwari, *"Decision support system for precluding coronary Heart Disease (CHD)"* Volume 3 Issue 2 of IJCSMC. Feb 2014.