Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 3 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease

By Kantesh Kumar Oad, Xu DeZhi & Pinial Khan Butt

Central South University, China

Abstract - Health care domain systems globally face lots of difficulties because of the high amount of risk factors of heart diseases in peoples (WHO, 2013). To reduce risk, improved knowledge based expert systems played an important role and has a contribution towards the development of the healthcare system for cardiovascular disease. To make use of benefits of knowledge based system, it is necessary for health organizations and users; must need to know the fuzzy rule based expert system's integrity, efficiency, and deployments, which are the open challenges of current fuzzy logic based medical systems. In our proposed system, we have designed a fuzzy rule based expert system and also by using data mining technique we have reduced the total number of attributes. Our system mainly focuses on cardiovascular disease diagnosis, and the dataset taken from UCI (Machine Learning Repository). We explored in the existing work. The majority of the researcher's experimentation was made on 14 attributes out of 76. While, in our system we took advantage of 6 attributes for system design. In the preliminary stage UCI, data participated in suggested system that will get outcomes. The performance of the system matched with Neural Network and J48 Decision Tree Algorithm.

Keywords: fuzzy reasoning, heart disease and diagnose, data mining.

GJCST-C Classification : I.5.1

AFUZZYRULEBASEDAPPROACHTOPREDICTRISKLEVELOFHEARTDISEASE

Strictly as per the compliance and regulations of:

A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease

Kantesh Kumar Oad [α], Xu DeZhi [σ] & Pinial Khan Butt [ρ]

Abstract- Health care domain systems globally face lots of difficulties because of the high amount of risk factors of heart diseases in peoples (WHO, 2013). To reduce risk, improved knowledge based expert systems played an important role and has a contribution towards the development of the healthcare system for cardiovascular disease. To make use of benefits of knowledge based system, it is necessary for health organizations and users; must need to know the fuzzy rule based expert system's integrity, efficiency, and deployments, which are the open challenges of current fuzzy logic based medical systems. In our proposed system, we have designed a fuzzy rule based expert system and also by using data mining technique we have reduced the total number of attributes. Our system mainly focuses on cardiovascular disease diagnosis, and the dataset taken from UCI (Machine Learning Repository). We explored in the existing work. The majority of the researcher's experimentation was made on 14 attributes out of 76. While, in our system we took advantage of 6 attributes for system design. In the preliminary stage UCI, data participated in suggested system that will get outcomes. The performance of the system matched with Neural Network and J48 Decision Tree Algorithm.

Keywords: fuzzy reasoning, heart disease and diagnose, data mining.

I. Introduction

Just recently peoples are stressed over their health and wellness troubles, in the majority of the countries proportion of cardiovascular disease enhancing so quickly and it has become the leading cause / death worldwide [1][2], and it is came to be taken into consideration as a "second epidemic," changing transmittable disease[3][4]. Health domain application is one of one of the most active study area nowadays. Ideal example of health domain application is the detection system for cardiovascular disease based on computer system assisted diagnosis strategies, where the information acquired from numerous other sources and is evaluated based on computer-based application. Before it was very time consuming job to get knowledge from physician and include this knowledge to computer system program by hand into data base medical decision support system and this was totally depending upon clinical experts' concepts which may be subjective.

This trouble has really been resolved using expert systems; get physician, group, knowledge and certain human client details, intelligently. In boosting outcomes at a couple of healthcare companies and strategy internet sites, expert system has actually operated by making needed clinical knowledge quickly readily available to know-how users [5] Taking care of clinical needs, such as making certain specific medical diagnoses, evaluating in a quick manner for avoidable health problem, or avoiding undesirable drug occasions, are the most standard exploitation of Expert System [6]. Expert System could also be possibly lessened costs, progression performance, and reduce client stress. These systems are classified into 2 groups namely (1) Knowledge based and (2) non-knowledge based [7]. The knowledge based system consists of rules (if-then statements). Expert system that is implemented with the assistance of artificial intelligence has the ability to support in a new setting and to learn for instance [8][9]. Given that the concept of computer-based Clinical Decision Support System aroused at first, significant research has actually been made in both academic and practical areas. Many obstacles are longer to impede the effective application of expert systems in scientific environments, among which portrayal and reasoning concerning clinical understanding predominantly under anxiety is the locations that require improved methodologies and strategies [10][11].

In our proposed system, mainly focus on cardiovascular disease diagnosis. We have taken dataset taken from UCI (Machine Learning Repository). UCI database consists of 76 attributes; we investigated in the existing work. Most of the experiments were made by using a subset of 14 from UCI. While, in our system we reduced the number of input attributes that will reduce the number of diagnostic results and we used seven attributes to experiment. From seven attributes, we used six attributes as input, and one attribute for output. Numerical data will enter in into suggested system, and in the last; system will get prediction results. The primary objective of our research is to make and carry out fuzzy rule based system for heart disease people.

Author α: Masters of Engineering in Computer Application Technology. School of Information Science and Engineering, Central South University, Changsha Hunan, China.
e-mail: kanteshoad84@gmail.com
Author σ: Professor, School of Information Science and Engineering, China. e-mail: hunan.xu@mail.csu.edu.cn
Author ρ: Ph.D. (Computer Science) Research Scholar, China.
e-mail: pinial@yahoo.com

II. Related Work

Mehdi.N. Mehdi. Y. (2009) [12] designed a Fuzzy Expert System of diagnosing the hepatitis B intensity rate and making comparisons with Adaptive Neural. M. Neshat, M. Yaghobi, M.B. Naghibi, A. Esmaelzadeh (2008) [13] designed a fuzzy expert system for Diagnosis of liver disorders. P.K. Anooj (2012) [14] developed Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. Persi Pamela I, Gayathri. P, N. Jaisankar (2013) [15] uses a Fuzzy Optimization Technique for the Prediction of Coronary Heart Disease Using Decision Tree. Nidhi Bhatla, Kiran Jyoti (2012) [16] used Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic. Vijay Kumar Magoa, Nitin Bhatia, Ajay Bhatia, Anjali Mago (2013) [17] designed clinical decision support system for dental treatment. Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg (2013) [18] used Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease. Asha Rajkumar, Mrs. G. Sophia Reena (2010) [19] used diagnosis of Heart Disease according to the data mining Algorithm "GJCST Classification J.3. M. Anabarsi. Anupriya *, N. CH. Iyengar (2010) [20] enhanced prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. K. Rajeswari. V. Vaithiyanathan (2011) [21] designed Heart disease diagnosis: an efficient decision support system based on fuzzy logic and genetic algorithm.

III. A Fuzzy Rule Based Approach to Predict Risk Level of Heart Disease

As you can see in figure (1) the process begins with data processing. In a second step, we reduced the number of attributes and then this processed data (symptoms) inserted into fuzzy system using MATLAB programming. After the fuzzy model is successfully developed, the prediction of the symptoms will start and lastly performance based on the result will be analyzed at the end of the development phase.

a) Data Processing

The function of data processing is to draw out the significant information from the raw data collection useful for heart disease prediction, and these data sets should be transformed into the needed style for the level of risk prediction. Due to large amount of data there are chances of errors on it so before processing heart disease datasets, the original raw datasets must be processed;as a result in data processing phase, we cleaned, transformed and analyzed it into the row column format after taking out the unnecessary ones.

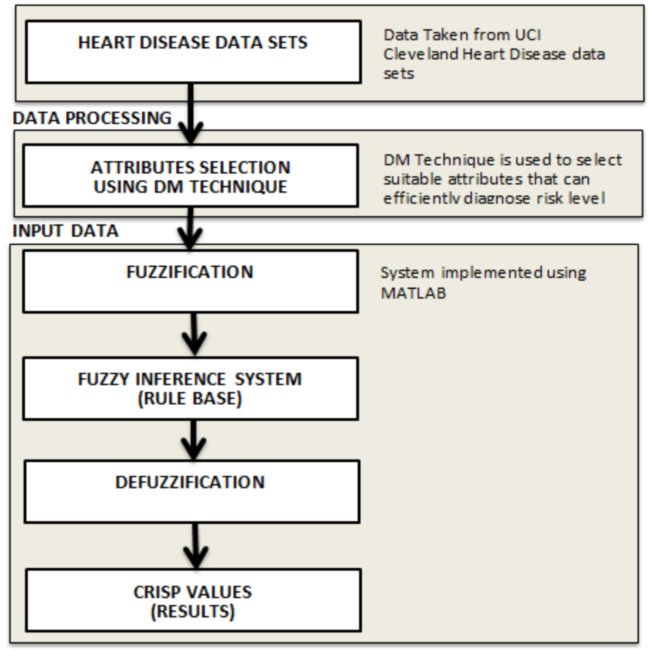

Figure 1 : Process Flow in Constructing Heart Disease Diagnose System

b) Data Formats

For recap or view purpose we should make data in a format so it can be easily reviewed. We have converted data to a .csv (Comma Separated Values) documents format also called comma delimited file. .csv file is a particularly configured text file; which establishments spreadsheet or standard database-style information in a quite easy style, with one record for each line, and each field within that document separated by a comma.

c) Attributes Selection

To minimize disease data sets (feature selection), we have used data reduction technique. Attribute selection lessens the data set size by getting rid of redundant or unnecessary and extracting with a minimized set of characteristics has an extra benefit [22]. Primarily we have actually made use of the data mining tool to get decreased collection of datasets. An attributes selection method contains four steps, named (1) subset generation; (2) subset evolution, (3) stop criterion (4) outcome validation [23] subset generation is a searching technique and we have made use of Best First Search Method using DM tool. Each new subset evaluated and matched with earlier best one according to a certain evolution criterion. It changes the previous finest subset if the new subset turns out to be much better. The process of subset generation and evolution is repeated till a given criterion is pleased.

By using selected method, we have obtained 6 attributes from a total of 14 attributes.

d) System Phases

Fuzzy system used in circumstances where we have a trouble of uncertainty. When the trouble has a dynamic behavior; Fuzzy Rule Base is an appropriate system for handling this issue. In the primary step of fuzzy expert system design; figure out the input and result variables. As it is described in section (4.3) we selected six inputs and one output variable. Then, we have actually made the membership functions for all variables used in the system. Membership function is primarily a visual portrayal of a fuzzy set; and determines the membership degree of objects to fuzzy sets. Rule evolution and defuzzification process described in the next sections.

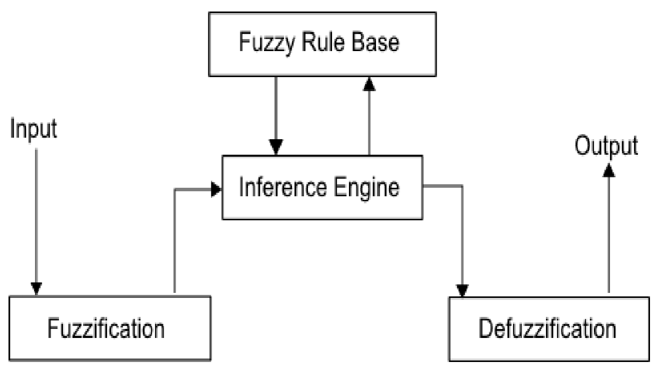

Figure 2 : General Schema of Fuzzy Logic

i. *Fuzzification and Membership Function*

Fuzzification is a process of fuzzifying all inputs and output. Determine the degree to which these inputs and outputs belong to each of the suitable fuzzy sets.

Age:

We have actually separated age into 3 fuzzy sets (Young. Middle and Old) and ranges of these fuzzy sets are determined in Table (1). We have used triangular and trapezoidal Membership function for fuzzy sets.

$\mu young(x) = 1$ when $x \in [1, 29]$
$\mu young(x) = (38-x)/(38-29)$ when $x \in [29, 38]$
$\mu young(x) = 0$ otherwise.

$\mu middle(x) = (x-33)/(39-33)$ when $x \in [33, 39]$
$\mu middle(x) = (45-x)/(45-39)$ when $x \in [39, 45]$
$\mu middle(x) = 0$ otherwise.

$\mu old(x) = 0$ when $x \in [1, 40]$
$\mu old(x) = (x-40)/(60-40)$ when $x \in [40, 60]$
$\mu old(x) = 1$ when $x \in [60, 100]$.

Blood Pressure:

We have actually separated this input fuzzy set into 4 levels called (Low, Medium, High and Very high). Trapezoidal MFs are used for (Low and Very high) and MFs of (medium and high) sets are triangular.

$\mu low(x) = 1$ when $x \in [1, 111]$
$\mu low(x) = (134-x)/(134-111)$ when $x \in [111, 134]$
$\mu low(x) = 0$ otherwise.

$\mu medium(x) = (x-126)/(139-126)$ when $x \in [126, 139]$
$\mu medium(x) = (152-x)/(152-139)$ when $x \in [139, 152]$
$\mu medium(x) = 0$ otherwise.

$\mu high(x) = (x-142)/(157-142)$ when $x \in [142, 157]$
$\mu high(x) = (172-x)/(172-157)$ when $x \in [157, 172]$
$\mu high(x) = 0$ otherwise.

$\mu veryhigh(x) = 0$ when $x \in [1, 154]$
$\mu veryhigh(x) = (x-154)/(172-154)$ when $x \in [154, 172]$
$\mu veryhigh(x) = 1$ when $x \in [172, 300]$.

Cholesterol:

Cholesterol has 3 fuzzy sets (Low, Medium and High). Ranges for these fuzzy sets are determined in Table (1).

$\mu low(x) = 1$ when $x \in [1, 151]$
$\mu low(x) = (197-x)/(197-151)$ when $x \in [151, 197]$
$\mu low(x) = 0$ otherwise.

$\mu medium(x) = (x-188)/(219-188)$ when $x \in [188, 219]$
$\mu medium(x) = (250-x)/(250-219)$ when $x \in [219, 250]$
$\mu medium(x) = 0$ otherwise.

$\mu high(x) = 0$ when $x \in [1, 217]$
$\mu high(x) = (x-217)/(263-217)$ when $x \in [217, 263]$
$\mu high(x) = 1$ when $x \in [263, 500]$.

Heart Rate:

Heart Rate split into 3 fuzzy sets named (Low, Medium and High). Ranges for these fuzzy sets are identified in table (1). MFs of (Low and High) sets are trapezoidal and MF of (Medium) is triangular.

$\mu low(x) = 1$ when $x \in [1, 100]$
$\mu low(x) = (141-x)/(141-100)$ when $x \in [100, 141]$
$\mu low(x) = 0$ otherwise.

$\mu medium(x) = (x-111)/(152-111)$ when $x \in [111, 152]$
$\mu medium(x) = (194-x)/(194-152)$ when $x \in [152, 194]$
$\mu medium(x) = 0$ otherwise.

$\mu high(x) = 0$ when $x \in [1, 152]$
$\mu high(x) = (x-152)/(216-152)$ when $x \in [152, 216]$
$\mu high(x) = 1$ when $x \in [216, 450]$.

Old Peak:

Old Peak divided into 3 fuzzy sets (Low, Risk and Terrible). These fuzzy sets have actually been shown in Table (1) with their ranges.

$\mu low(x) = 1$ when $x \in [0, 1]$
$\mu low(x) = (2-x)/(2-1)$ when $x \in [1, 2]$
$\mu low(x) = 0$ otherwise.

$\mu risk(x) = (x-1.5)/(2.8-1.53)$ when $x \in [1.5, 2.8]$
$\mu risk(x) = (4.2-x)/(4.2-2.8)$ when $x \in [2.8, 4.2]$
$\mu risk(x) = 0$ otherwise.

$\mu terrible(x) = 0$ when $x \in [0, 2.55]$
$\mu terrible(x) = (x-2.55)/(4.2-2.55)$ when $x \in [2.55, 4.2]$
$\mu terrible(x) = 1$ when $x \in [4.2, 6]$.

Thallium Scan:

This input field includes 3 fuzzy sets: (Normal, Fix Defect and Reverse Defect). For each and every fuzzy fuzzy set we have defined a value that we use them for system testing. These fuzzy sets with their values are shown in Table (1).

$\mu normal(x) = (x-1)/(2-1)$ when $x \in [1, 2]$
$\mu normal(x) = (3-x)/(3-2)$ when $x \in [2, 3]$
$\mu normal(x) = 0$ otherwise.

$\mu Fix\ Defect(x) = (x-3)/(4.5-3)$ when $x \in [3, 4.5]$
$\mu Fix\ Defect(x) = (6-x)/(6-4.5)$ when $x \in [4.5, 6]$
$\mu Fix\ Defect(x) = 0$ otherwise.

$\mu Rev\ Defect(x) = (x-6)/(6.5-6)$ when $x \in [6, 6.5]$
$\mu Rev\ Defect(x) = (7-x)/(7-6.5)$ when $x \in [6.5, 7]$
$\mu Rev\ Defect(x) = 0$ otherwise.

Over we have actually selected chosen features, now we have split all inputs into fuzzy sets; we have actually utilized trapezoidal and triangular membership functions in system.

Input Field	Range	Fuzzy Sets
Age	<38 33-45 40>	Young Middle Old
Blood Pressure	< 138 126- 152 142-172 154>	Low Medium High Very High
Cholesterol	< 197 188-250 217>	Low Medium High
Heart rate	< 141 111-194 152>	Low Medium High
Old Peak	<2 1.5 - 4.2 2.5>	Low Risk Terrible
Thallium Scan	3 6 7	Normal Fixed Defect Reversible Defect

Table 1 : Risk factors and Ranges

ii. *Rules Evolution*

In Fuzzy Rule Base System, rules play an important role in the prediction. The rules deliver/provide a sense to linguistic variables and MF (membership function). So we have occupied these fuzzyfied inputs in antecedent part of the rules. In this research, we have actually utilized 19 rules to predict heart disease in the patient.

In our system antecedent part of the rule consist of only single part that will opinion result of antecedent development. Fuzzy logic will govern risk level and this prediction indeed relies on rules that we have made. The made rules we have applied in using Mat Lab R2012a in the rule editor. After then fuzzification, crisp will certainly examine by passing in rule instance.

iii. *Output (Defuzzification)*

In this system, we have one output variable, which divided to 2 fuzzy sets (healthy and sick). For defuzzification procedure, designed system makes use of the Centroid method, determines the area of membership functions within the range of (output) variable.

$$CoA = \frac{\int_{x_{min}}^{x_{max}} f(x).xdx}{\int_{x_{min}}^{x_{max}} f(x)dx}$$

CoA is the center of area/gravity; x is the linguistic variable and x (min) and x (max) signifies the arrays of variables.

IV. System Testing

To compare the performance of our system with Neural Network and J48 Decision Tree, we have divided Cleveland Heart Disease datasets into 2 parts e.g. training data 60% and testing data 40%. And this efficiency/performance is usually matched in term of sensitivity, specificity and accuracy. These terms normally took advantage of diagnostic approaches to enhance analysis results.

$$Sensitivity = TP / (TP + FN)$$
$$Specificity = TN / (FP + TN)$$
$$Accuracy = (TP+TN) / (TP+TN+FP+FN)$$

		Condition		
		Condition Positive	Condition Negative	
Test	Test Positive	TP	FP	+Prediction value = TP / (TP + FP)
	Test Negative	FN	TN	-Prediction value = TN / (FN + TN)
		Sensitivity = TP / (TP + FN)	Specificity = TN / (FP + TN)	

Figure 3 : Illustrates how the Positive Predictive Value, Negative Predictive Value, Sensitivity, and Specificity are Related

Global Journal of Computer Science and Technology (C) Volume XIV Issue III Version I Year 2014

We use specificity to analyse and assess the amount of true positives predicted accurately. Specificity analyses and measure the amount of true negative predicted accurately. Accuracy can be obtained by sum of True Positive and True Negative divided with the total number of instances.

Example: No of healthy (True Positive) and sick (True Negative) peoples predicted correctly.

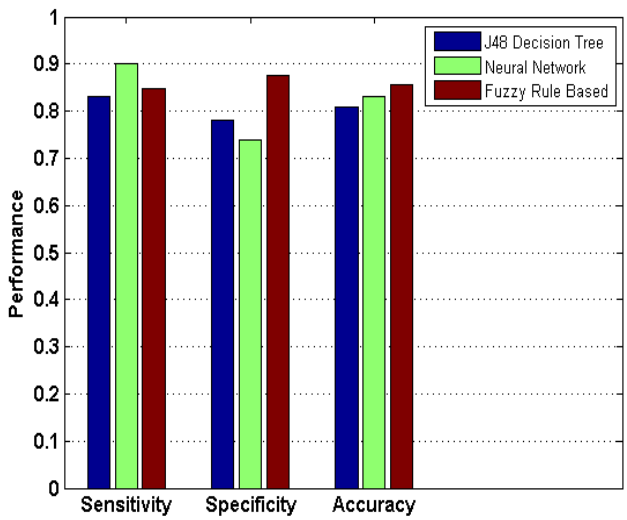

Figure 4 : Training Datasets Performance

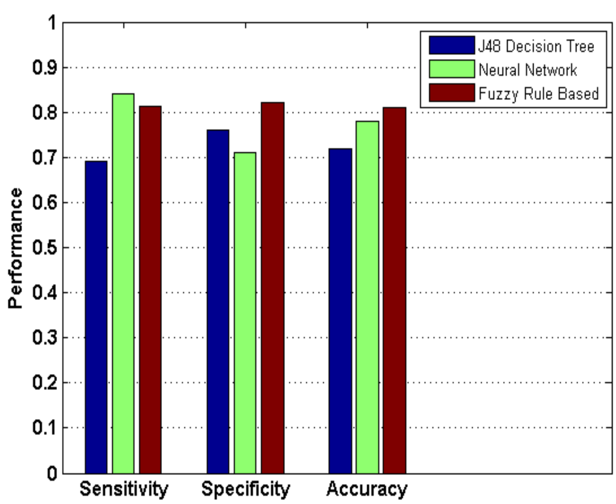

Figure 5 : Testing Datasets Performance

The performance testing has been performed on both training and testing data sets, and we merely used Cleveland heart disease datasets establishes to evaluate our system. And, in last we compared the performance of our system with Neural Network and J48 Decision Tree model using very same data sets. The final result we obtained for training and testing data are shown in figure (5-2a and 5-2b).

V. Conclusion

This proposed system "Fuzzy Rule Based Support System" modelled to predict heart disease intelligently and efficiently, and to replace manual efforts. Experts system can be more proficient and fast so it can be more accurate then manual work. Our system modelled to diagnosis and detecting cardiovascular diseases, the system involves two major phases, one that performs classification and diagnosis, the other one that detects the rate of risks of the respiratory diseases. For this system we have used mamdani inference system. In final this system tested and compared with Neural Network and J48 Decision Tree model to check performance of the system.

VI. Acknowledgement

This work is supported by National Natural Science Foundation of China under grant No.60970096 and 90818004, the Postdoctoral Science Foundation of China under Grant No.20080440988 and the Natural Science Foundation of Hunan Province, China Under Grant No.09JJ4030.

References Références Referencias

1. World Health Organization Report of the Global Atlas http://apps.who.int/globalatlas/.
2. Mendis, S., Puska, P., Norrving, B. (2011). Global Atlas on cardiovascular disease prevention and control. ISBN 978-92-4-156437-3.
3. Gale Nutrition Encyclopaedia (2011). Heartdise-asehttp://www.answers.com/topic/ischaemic-heart-disease (Accessed 25 February 2011).
4. Countries, Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing; Fuster, Board on Global Health; Valentin; Academies, Bridget B. Kelly, editors; Institute of Medicine of the National (2010). Promoting cardiovascular health in the developing world: a critical challenge to achieve global health. Washington, D. C.: National Academies Press. pp. Chapter 2. ISBN 978-0-309-14774-3.
5. Merijohn G. K., Bader J. D., Frantsve-Hawley J., (2008). Clinical decision support chair side tools for evidence-based dental practice. The Journal of Evidence-Based Dental Practice, 8 (3) 2008, pp. 119–132.
6. Garg N.K., Adhikari, McDonaldH. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. Journal of the American Medical Association. PubMed, 293 (10). pp. 1223–1238.
7. AbbasiM.M., KashiyarndiS. (2006). Clinical decision support systems: a discussion on different meth-odologies used in health care.
8. Warren J., Beliakov G., Zwaag B. (2000). Fuzzy logic in clinical practice decision support system. In: Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii. 4–7 January 2000.

9. AndersonJ., Clearing the way for physicians use of clinical information systems Communication of the ACM (1997), pp. 83–90.
10. LinL., HuP.J.H., Sheng O.R.L., (2006). Decision Support Systems. 42.
11. MusenM., ShaharY., ShortliffeE.H., (2001). Clinical decision support systems.
12. Neshat M., Yaghobi M., (2009). Designing a Fuzzy Expert System of Diagnosing the Hepatitis B intensity Rate and comparing it with Adaptive Neural Network Fuzzy System. International Conference in Modelling Health Advances pp.1~6.
13. Neshat M., Yaghobi M., NaghibiM.B., (2008). Fuzzy Expert System Design for Diagnosis of liver disorders. IEEE Computer Society. pp. 252~256.
14. AnoojP.K., (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. "International Journal of Research and Reviews in Computer Science (IJRRCS)" Vol. 3, No. 3. ISSN: 2079-2557.
15. Persi Pamela I., Gayathri. P., JaisankarN., (2013). A Fuzzy Optimization Technique for the Prediction of Coronary Heart Disease Using Decision Tree. International Journal of Engineering and Technology (IJET). Vol 5 No 3.
16. Nidhi Bhatla, Kiran Jyoti (2012). A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic. International Journal of Computer Applications (0975 – 8887) Volume 54– No.17.
17. Vijay Kumar Magoa, Nitin Bhatia, Ajay Bhatia, (2012). Clinical decision support system for dental treatment. Journal of Computational Science. 3 (2012) 254–261.
18. Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg (2013). Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease. International Journal of Advanced Research in Computer Engineering & Technology. (IJARCET) Volume 2, Issue 1.
19. Asha Rajkumar, Mrs. G.Sophia Reena (2010). Diagnosis of Heart Disease Using Data mining Algorithm "GJCST Classification J.3. Global Journal of Computer Science and Technology. P a g e |38 Vol. 10 Issue 10 Ver. 1.0.
20. AnbarasiM., Anupriya*E., IyengarN.CH. (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. International Journal of Engineering Science and Technology. Vol. 2(10). 5370-5376.
21. Rajeswari*K., VaithiyanathanV., (2011). Heart disease diagnosis: an efficient decision support system based on fuzzy logic and genetic algorithm. International Journal of Decision Sciences, Risk and Management, Vol. 3, Nos. 1/2.
22. Han J. and Kamber M., (2006). Data Mining: Concepts and Techniques. Second Edition, Morgan Kaufmann Publishers, San Francisco.
23. DashM. and LiuH., (1997). Feature selection for classification. Intelligent Data Analysis: An International Journal, 1(3):131–156.