A Fuzzy Expert System for Heart Disease Diagnosis

Ali.Adeli, Mehdi.Neshat

Abstract— The aim of this study is to design a Fuzzy Expert System for heart disease diagnosis. The designed system based on the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation data base. The system has 13 input fields and one output field. Input fields are chest pain type, blood pressure, cholesterol, resting blood sugar, maximum heart rate, resting electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan, sex and age. The output field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4 (distinguish presence (values 1, 2, 3, 4)). This system uses Mamdani inference method. The results obtained from designed system are compared with the data in upon database and observed results of designed system are correct in 94%. The system designed in Matlab software. The system can be viewed as an alternative for existing methods to distinguish of heart disease presence.

Index Terms— Heart Disease, Fuzzy Expert System, Fuzzy Logic, Medical Distinguish

I. INTRODUCTION

Nowadays the use of computer technology in the fields of medicine area diagnosis, treatment of illnesses and patient pursuit has highly increased. Despite the fact that these fields, in which the computers are used, have very high complexity and uncertainty and the use of intelligent systems such as fuzzy logic, artificial neural network and genetic algorithm have been developed [1].

In domain of heart disease risk, smoke, cholesterol, blood pressure, diabetes, sex and age are main risk factors that affect on heart disease risk [4]. Because of the many and uncertain risk factors in the heart disease risks, sometimes heart disease diagnosis is hard for experts. In the other word, there exists no strict boundary between what is Healthy and what is diseased, thus distinguish is uncertain and vague [2]. Having so many factors to analyze to diagnose the heart disease of a patient makes the physician's job difficult. So, experts require an accurate tool that considering these risk factors and show certain result in uncertain term.

Motivated by the need of such an important tool, in this study, we designed an expert system to diagnose the heart disease. The designed expert system based on Fuzzy Logic. This fuzzy expert system that deals with diagnosis has been

A.Adeli.S.Department of computer Engineering University Islamic Azad of shirvan branch, Iran.(student) Phone: +989153846831 ,Fax:+985842251884 ,E-mail: adeli_ali405@yahoo.com. implemented and experimental results showed that this system did quite better than non-expert urologist and about 94 % as a well as the expert did.

Remaining of the paper organized as follows. We will show previous research concerning heart disease diagnosis in next section. In section 3, we introduce dataset for this system. In section 4, we introduce method of designing that includes fuzzy expert system designing in section 4.1, fuzzy rule base in section 4.2 and in section 4.3, we show fuzzification and defuzzification. In section 5, we show results.

II. PREVIOUS RESEARCH

As for other clinical diagnosis problems, classification systems have been used for heart disease diagnosis problem, too. When the studies in the literature related with this classification application are examined, it can be seen that a great variety of methods were used which reached high classification accuracies using the dataset taken from UCI machine learning repository. Among these, [4] ToolDiag, RA obtained 50.00% classification accuracy by using IB1- 4 algorithm. [4] WEKA, RA obtained a classification accuracy of 58.50% using InductH algorithm while ToolDiag, RA reached to 60.00% with RBF algorithm. [4] Again, WEKA, RA applied FOIL algorithm to the problem and obtained a classification accuracy of 64.00%. [4] MLP+BP algorithm that was used by ToolDiag, RA reached to 65.60%. [4] The classification accuracies obtained with T2, 1R, IB1c and K* which were applied by WEKA, RA are 68.10%, 71.40%, 74.00% and 76.70%, respectively. [4] Robert Detrano used regression algorithm and obtained 77.0% logistic classification accuracy.

Novruz Allahverdi & Serhat Torun & Ismail Saritas proposed a fuzzy expert system to determination of coronary heart disease risk (CHD) of patient for the next ten-years, in 2007[1]. The designed system gives user the ratio of the risk and may recommend using one of three results: (1) normal live, (2) diet, (3) drug treatment. The result of this fuzzy expert system in 79% as a well as the expert did.Moreover, Cheung utilized C4.5, Naive Bayes, BNND and BNNF algorithms and reached the classification accuracies 81.11%, 81.48%, 81.11% and 80.96%, respectively [4].

III. DATASET

Designed system based on the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation dataset [3]. This dataset is part of the collection of databases at the University of California, Irvine collected by David Aha. The purpose of

M.Neshat. Department of Computer Engineering, Faculty University Islamic Azad of shirvan branch, Iran Phone : + 9805118526851 ,Fax:+989157005321,E-mail : neshat_mehdi @ yahoo . com.

this dataset is to diagnose the presence or absence of heart disease given the results of various medical tests carried out on a patient. This database contains 76 attributes and 303 examples of patient, but all published experiments refer to using a subset of 14 of them and we have just used 12 attributes. This system uses 11 attributes for input and 1 attribute for result.

Input fields (attributes) are chest pain type, blood pressure, cholesterol, resting blood sugar, resting maximum heart rate, sex, electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan and age. The output field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 4 (distinguish presence (values 1, 2, 3 and 4)); increasing value shows increasing heart disease risk.

In this study, we use low density lipoprotein (LDL) cholesterol. About the blood, we use systolic blood pressure.

In this dataset, fields divide to some sections and each section has a value. For instance, in this dataset, sex has two section (1=male and 0=female), chest pain has 4 section (1=typical angina, 2=atypical angina, 3=non-angina pain and 4=asymptomatic), resting blood sugar has 2 section (0=false and 1=true) and it is true when FBS>120, ECG has 3 section (0=normal, 1=having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) and 2=Hypertrophy (showing probable or definite left ventricular hypertrophy by Estes' criteria)), exercise has 2 section (3=normal, 6=fixed defect and 7=revisable defect). This system uses these sections exactly and for dividing of the other fields we use expert's idea that will be introduced in next section.

IV. METHOD

In this section, we show the fuzzy expert system designing, membership functions, fuzzy rule base, fuzzification and defuzzification.

A. FUZZY EXPERT SYSTEM DESIGNING

The most important application of fuzzy system (fuzzy logic) is in uncertain issues. When a problem has dynamic behavior, fuzzy logic is a suitable tool that deals with this problem. First step of fuzzy expert system designing is determination of input and output variables. There are 11 input variables and 1 output variable. After that, we must design membership functions (MF) of all variables. These membership functions determine the membership of objects to fuzzy sets.

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.

Input variable are:

1. Chest pain: This input variable supports 4 chest pain types. We have defined a value in this system for each chest pain type that we use these values for system testing. Each chest pain type is a fuzzy set. In this field, fuzzy sets do not have overlap and sets define in form of crisp because the patient has just one chest pain type on time. Chest pain types with their values have shown as follow.

1=typical angina 2=atypical angina 3=non-anginal pain 4=asymptomatic

2. Blood Pressure: Different values of blood pressure change the result easily. In this field, we use systolic blood pressure. This input variable has divided to 4 fuzzy sets. Fuzzy sets are "Low", "Medium", "High" and "Very high". Membership functions of "Low" and "Very high" sets are trapezoidal and membership functions of "medium" and "high" sets are triangular. We have defined fuzzy membership expressions for blood pressure input field (Eq. (1)). These fuzzy sets will be shown in Table 1. Membership functions of blood pressure field will be shown in FIG.1.



FIG. 1. Membership functions of Systolic Blood Pressure

$$\mu_{low}(x) = \begin{cases} 1 & x < 111 \\ 134 - x/23 & 111 \le x < 134 \end{cases}$$
$$\mu_{medium}(x) = \begin{cases} x - 127/21 & 127 \le x < 139 \\ 1 & x = 139 \\ 153 - x/21 & 139 \le x < 153 \end{cases}$$
$$\mu_{high}(x) = \begin{cases} x - 142/21 & 127 \le x < 157 \\ 122 & 127 \le x < 157 \\ 142 & 127 \le x < 157 \\ 122 & 127 \le x < 157 \\ 122 & 127 \le x < 172 \end{bmatrix}$$
$$\mu_{veryhigh}(x) = \begin{cases} x - 154/21 & 154 \le x < 171 \\ 122 & 127 \le x < 171 \\ 122 & 127 \le x < 171 \\ 122 & 127 \le x < 171 \end{bmatrix}$$

(1)

3. Cholesterol: Cholesterol has salient affect on the result and can change it easily. For this input field, we use the value of low density lipoprotein (LDL) cholesterol. Cholesterol field has 4 fuzzy sets (Low, Medium, High and Very high). These fuzzy sets have been shown in Table 2. Membership functions of "Low" & "Very high" sets are trapezoidal and membership functions of "Medium" & "High" sets are triangular. Membership functions of cholesterol field will be shown in FIG.2. Eq. (2) shows these membership function expressions of cholesterol.



FIG. 2. Membership functions of Cholesterol



(2)

4. Blood Sugar (Diabetes): Blood sugar field is one of the most important factors in this system that changes the result. This input field has just one fuzzy set. In this system, we have defined that if the amount value of blood sugar is higher than 120 (>120) then man has blood sugar. FIG. 3 shows the membership function of blood sugar. Membership function of this fuzzy set is trapezoidal. We will see fuzzy membership expressions for blood sugar field in Eq. (3).



FIG. 3. Membership function of Blood Sugar

$$u_{veryhigh}(x) = \begin{cases} (x - 105) / & 105 \le x < 120 \\ 66 & \\ 1 & x \ge 120 \end{cases}$$

(3)

5. Resting Electrocardiography (ECG): In this field, we have 3 fuzzy sets (Normal, ST_T abnormal & Hypertrophy). Membership functions of "Normal" & "Hypertrophy" fuzzy sets are trapezoidal and membership Function of "ST_T abnormal" fuzzy set is triangular. We see these membership functions in FIG. 4. In Table 3, we show fuzzy sets with their values. In this Table, in "RANGE" column, we have defined a value for each fuzzy set in left side of each interval and we use just these values for system testing.





ST_T wave abnormality = T wave inversions and/or ST elevation or depression of > 0.05 mV.

Hypertrophy = showing probable or definite left ventricular hypertrophy by Estes' criteria.

6. Maximum Heart Rate: The value of this input field is maximum heart rate of man in 24 hours. By increasing of age in man, maximum of heart rate in 24 hours decreases. In this field, we have 3 linguist variables (fuzzy sets) (Low, Medium and High). In Table 4, we have defined these fuzzy sets. Membership functions of "Low" & "High" fuzzy sets are trapezoidal and membership function of "Medium" fuzzy set is triangular that will be shown in FIG. 5. Eq. (4) shows fuzzy membership function expressions.



FIG. 5. Membership function of Maximum Heart Rate



(4)

7. Exercise: This input field has just 2 values (0, 1) and one fuzzy set (true). If doctor determines exercise test for patient, value 1 will enter in system, otherwise, value 0 will enter in it.

8. Old Peak: This input field means ST depression induced by exercise relative to rest. Old peak field has 3 fuzzy sets (Low, Risk and Terrible). These fuzzy sets have been shown in Table 5 with their ranges. Membership functions of "Low" and "Terrible" fuzzy sets are trapezoidal and membership function of "Risk" fuzzy set is triangular that have been shown in FIG. 6. We have defined fuzzy membership function expressions in Eq. (5).



9. Thallium Scan: This input field consists 3 fuzzy sets: Normal, Reversible Defect and Fixed Defect. For each fuzzy set we have defined a value that we use them for system testing. These fuzzy sets with their values will be shown in Table 6.

TABLE 6. CLASSIFICATION OF THALLIUM SCAN

INPUT FIELD	RANGE	FUZZY SETS	
Thallium scan	3	Normal	
	6	Fixed Defect	
	7	Reversible Defect	

10. Sex: This input field just has 2 values (0, 1) and sets (Female, Male). Value 0 means that patient is male and value 1 means that patient is female.

11. Age: This input field divides to 4 fuzzy sets (Young, Mild, Old, Very old). These fuzzy sets with their ranges will be shown in Table 7. Membership functions of "Young" & "Very old" are trapezoidal and membership functions of "Mild" & "Old" are triangular. The membership function expressions have been shown in Eq. (6).



FIG. 7. Membership functions of Age



Output Variable:

(6)

The "goal" field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 4. By increasing of integer value, heart disease risk increases in patient. In this system, we have considered a different output variable, which divides to 5 fuzzy sets (Healthy, Sick (s1), Sick (s2), Sick (s3), sick (s4)). Table 8 shows these fuzzy sets with their ranges. Membership functions of "Healthy" & "Sick (s4)" fuzzy sets are trapezoidal and membership functions of "Sick (s1)", "Sick (s2)" and "Sick (s3)" are triangular. These membership functions will be shown in Fig. 8.





B. FUZZY RULE BASE

Rule base is the main part in fuzzy inference system and quality of results in a fuzzy system depends on the fuzzy rules. This system includes 44 rules. Antecedent part of all rules has one section. This system designed with another rule bases (64 rules, 15 rules, 10 rules and 5 rules) and results showed in 44 rules system are best in comparison with results of the other rule bases. In the other hand, results with 44 rules tend to the expert's idea and laboratory results. The rules have been shown in FIG.9.

1. If (chest, pain is applied) then (result is healthy) (1)
2 If (chest, pain is atangina) then (result is sick(s1)) (1)
3. If (chest_pain is nonangina) then (result is sick(s2)) (1)
4. If (chest_pain is asymptomatic) then (result is sick(s3)) (1)
5 If (chest_pain is asymptomatic) then (result is sick(s4)) (1)
6 If (sex is female) then (result is sick(s1)) (1)
7 If (sex is male) then (result is sick(s2)) (1)
8 If (blood, pressure is low) then (result is healthy) (1)
9 If (blood_pressure is medium) then (result is sick(s1)) (1)
10. If (blood, pressure is bigh) then (result is sick(s2)) (1)
11. If (blood_procedure is high) then (result is sick(s3)) (1)
12 If (blood_pressure is verybigh) then (result is sick(s4)) (1)
13. If (cholesteral is low) then (result is healthy) (1)
14. If (cholesteral is medium) then (result is sick(s1)) (1)
15. If (cholesteral is high) then (result is sick(s2)) (1)
16. If (cholesteral is high) then (result is sick(s2)) (1)
17. If (cholesteral is verybigh) then (result is sick(s4)) (1)
18. If (blood, sugar is true) then (result is sick(s2)) (1)
19. If (ECG is normal) then (result is healthy) (1)
If (ECG is normal) then (result is sick(s1)) (1)
 If (ECG is ST-Tabnormal) then (result is sick(s2)) (1)
If (ECG is hypertrophy) then (result is sick(s3)) (1)
 If (ECG is hypertrophy) then (result is sick(s4)) (1)
If (max_heart_rate is low) then (result is healthy) (1)
25. If (max_heart_rate is medium) then (result is sick(s1)) (1)
 If (max_heart_rate is medium) then (result is sick(s2)) (1)
 If (max_heart_rate is high) then (result is sick(s3)) (1)
 If (max_heart_rate is high) then (result is sick(s4)) (1)
29. If (exercise is true) then (result is sick(s2)) (1)
 If (oldpeak is low) then (result is healthy) (1)
31. If (oldpeak is low) then (result is sick(s1)) (1)
 If (oldpeak is terrible) then (result is sick(s2)) (1)
 If (oldpeak is terrible) then (result is sick(s3)) (1)
34. If (oldpeak is risk) then (result is sick(s4)) (1)
35. If (thal is normal) then (result is healthy) (1)
36. If (thal is normal) then (result is sick(s1)) (1)

37. If (that is fixdef) then (result is sick(s2)) (1)
38. If (that is rev.def) then (result is sick(s3)) (1)
39. If (that is rev.def) then (result is sick(s4)) (1)
40. If (age is young) then (result is healthy) (1)
41. If (age is nild) then (result is sick(s1)) (1)
42. If (age is old) then (result is sick(s2)) (1)
43. If (age is old) then (result is sick(s3)) (1)
44. If (age is veryold) then (result is sick(s4)) (1)
FIG. 9. Rule Base of System



D. FUZZIFICATION AND DEFUZZIFICATION

Designed system uses inference mechanism Mamdoni approach. In this system, we don't have any logical combination of inputs with AND/OR because antecedent part of all rules has one section. We have defined a validity degree (k) for each rule that will be shown as follow.

k1= chest pain(x)
...
k12= blood pressure (x)
...
k24= max_heart_rate (x)
...
k44= age(x)

For aggregation of rules, maximum of validity degree are calculated as follow.

K=max (k1, k,..., k44)

For defuzzification process, designed system uses "Centroid" method.

We have shown surface viewer of some fields as follow.







FIG. 11. Surface Viewer of Blood Sugar and Cholesterol



FIG. 12. Surface Viewer of Sex and Blood Pressure

V. SYSTEM TESTING

We have tested the designed system with following values for each field and graphical result is shown in FIG.13.

TABLE 9. SYSTEM TESTING								
Chest Pain	BP	LDL-C	Blood Sugar	ECG	Maximum Heart Rate			
4	117	230	130	0	160			
(asymptomatic)	(low)	(high)	(true)	(normal)	(high)			

Exercise	Old Peak	Thallium	Sex	Age	Result of System
1 (true)	1.4 (low)	7 (reversible)	1 (male)	60 (very old)	2 (Sick (s2))



FIG. 13. Result of the Patient in Designed System



FIG. 14. Membership functions of test

Patient with follow situation have been tested by this system. The results in V.A. Medical Center, Long Each and Cleveland Clinic Foundation data base and this system are same[3].

VI. CONCLUSION

Fuzzy Expert System for Heart Disease Diagnosis designed with follow membership functions, input variables, output variables and rule base. Designed system has been tested with expert-doctor. Designing of this system with fuzzy base in comparison with classic designed improves results. Results have been shown from this system in compression with past time system are logical and more efficient. This system simulates the manner of expert-doctor. This system is designed in way that patient can use it himself. This fuzzy expert system that deals with diagnosis has been implemented. Experimental results showed that this system did quite better than non-expert urologist and about 94 % as a well as the expert did.

REFERENCES

- [1] 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
- [2] M.Nikravesh & Janusz & Lotfi A.Zadeh, Foring New Frontier: Fuzzy Pioneer I, Springer 2007
- [3] Robert Detrano & M.D & PhD, V.A. Medical Center, Long Each and Cleveland Clinic Foundation. Available: www.archive.ics.uci.edu/ml/datasets/Heart+Disease
- [4] Kemal Polata,* & Salih Güne,sa & Sülayman Tosunb, Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing, ELSEVIER, PATTERN RECOGNATION, 2007.
- [5] K. Polat & S.Sahan, H. Kodaz & S. Güne, s, A new classification method to diagnosis heart disease: supervised artificial immune system (AIRS), in: Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN), 2005.
- [6] S. Sahan & H. Kodaz & S. Günes & K. Polat, A new classifier based on attribute weighted artificial immune system, Lecture Notes in Computer Science, vol. 3280, ISSN 0302-9743, 2004, pp. 11–20.
- [7] K. Polat & S. Sahan & S. Güne, s, A new method to medical diagnosis: artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia, Expert Systems with Applications 31 (2) (2006) 264–269.