A Diagnostic Fuzzy Rule-Based System for Congenital Heart Disease

Ersin Kaya, Bulent Oran and Ahmet Arslan

***Abstract*—**In this study, fuzzy rule-based classifier is used for the diagnosis of congenital heart disease. Congenital heart diseases are defined as structural or functional heart disease. Medical data sets were obtained from Pediatric Cardiology Department at Selcuk University, from years 2000 to 2003. Firstly, fuzzy rules were generated by using medical data. Then the weights of fuzzy rules were calculated. Two different reasoning methods as "weighted vote method" and "singles winner method" were used in this study. The results of fuzzy classifiers were compared.

***Keywords*—**Congenital heart disease, Fuzzy rule-based classifiers, Classification

I. Introduction

CONGENITAL heart diseases are defined as structural or functional heart anomaly at birth. Different studies have been reported that 8 per 1000 live births have congenital heart disease. Echocardiography is a very important tool for the diagnosis and following-up of children with congenital heart disease [1]. In the last quarter of 20th centuries, there have been dramatic improvements in echocardiographic equipment, so that it is now possible to obtain direct digital measurements, better color Doppler image resolution and M-mode measurements which are still a reliable method today [2].

In recent years, fuzzy classifier methods are commonly used in medical diagnosis. In the literature, there are a lot of studies for classification of medical data using fuzzy classifiers. In Silvia et al. Mamdani-type fuzzy rule-based classification established for the diagnosis of cerebral cortical malformations. The data of the samples were divided into 4 groups (Localization of the lesion, visual characteristics, Symptoms, Past medical / family history) and membership values and the rule sets are generated based on these groups. As a result, three main groups of diseases (Malformations due to abnormal neuronal proliferation moment, due to an abnormal neuronal migration Malformations, Malformations due to abnormal cortical organization) are classified [3]. In Stavros et al. has studied on data from the UCI database (Pima Indians Diabetes (PID), dermatological diseases (Derm)) with TSK-type fuzzy classifier. The classifier performance has been compared with the results of previous studies in the same dataset [4]. In Ali et al. has studied on nero-fuzzy system which is designed for the diagnosis of breast cancer. The breast cancer data were obtained from the UCI database [5].

In this work a medical data set is classified using a fuzzy rule-based classifier to diagnose 4 different diseases. We established fuzzy rules via Ishibuchi et al. [6]. We used four type weight which introduced by Ishibuchi et al. in 2005[7]. Achievements of the generated classifiers have been determined.

E. Kaya is with Dept of Computer Engineering, Faculty of Engineering and Architecture, Selçuk University, Konya, Turkey 42070 (phone: +90-332-2233333; fax: +90-332-2410064; e-mail: ersinkaya@selcuk.edu.tr).

B. Oran is with Department Pediatric Cardiology, Meram Medicine Faculty, Selcuk University, Konya, Turkey 42070 (e-mail: buloran@yahoo.com.tr).

A. Arslan is with Dept of Computer Engineering, Faculty of Engineering and Architecture, Selçuk University, Konya, Turkey 42070 (e-mail: ahmetarslan@selcu.edu.tr).

II. Medical Dataset

We have conducted this study in the Department of Pediatric Cardiology at Selcuk University, using the data obtain from 2000 to 2003. In this study, pediatric cardiologist made the definition of normal cardiac morphology and four types of congenital heart diseases using physical examination on the patients. Chest roentgenogram and electrocardiogram were obtained for all patients and cardiac diagnosis was confirmed by echocardiographic investigation in Pediatric Cardiology Section. A Hewlett-Packard sonos-1000 system ultrasonic imager with 3,5 MHz transducer was used for echocardiographic assessments. After routine cardiovascular examination, M-mode echocardiography was used. In this study, the data belonging to 4 types of congenital heart disease and normal samples were used.

The medical dataset is composed of 297 measurements and 9 attributes. There are no missing values. The medical dataset includes 8 condition attributes and a decision attribute. The values of condition attributes are real number. The condition attributes are "Patiens Weight (PW)", "Aortic Diameter (AD)", "Left Atrial Diameter (LAD)", "Pulmonary Artery Diameter (PAD)", "Left Ventricular EndDiastolic Diameter (LVEDD)", "Left Ventricular EndSistolic Diameter (LVESD)", "Septal Thickening (ST) and Posterior Wall Thickening (PWT) ". Detailed analyze of the dataset is shown in Table I.

TABLE I
DETAILED ANALYZE OF THE DATASET

Attributes	Max	Min	Median	Mean	SD
PW	5.30	2.40	3.40	3.57	0.73
AD	13.00	7.00	10.30	10.25	1.01
LAD	19.90	7.00	11.70	12.76	3.76
PAD	9.40	6.00	7.80	7.77	0.79
LVEDD	28.20	10.00	16.00	17.19	4.24
LVESD	16.30	4.50	10.00	9.95	2.95
ST	07.00	3.60	4.60	5.09	1.03
PWT	08.70	3.50	4.60	4.87	1.09

Feature, names of the features obtained from biomedical measurements; Max represents the maximum value of the features and Min represents the minimum value of the features; Median represents median value of the features; Mean represents mean value of the features; SD represents standard derivation of the features.

Decision attribute has five different values. The values of decision attribute are “Normal (N)”, “Hypertrophic Cardiomyopathy (HC)”, “Left Atrial Dilatation (LAD)”, “Left Ventricular Hypertrophy (LVH)” and “Septal Hypertrophy (SH)”. The Details of Decision Attribute is shown in Table II.

TABLE II
THE DETAILS OF DECISION ATTRIBUTE

Disease	Number of Samples	Distribution of Samples
Normal	100	33.69%
Hypertrophic Cardiomyopathy	50	16.83%
Left Atrial Dilatation	47	15.82%
Left Ventricular Hypertrophy	50	16.83%
Septal Hypertrophy	50	16.83%

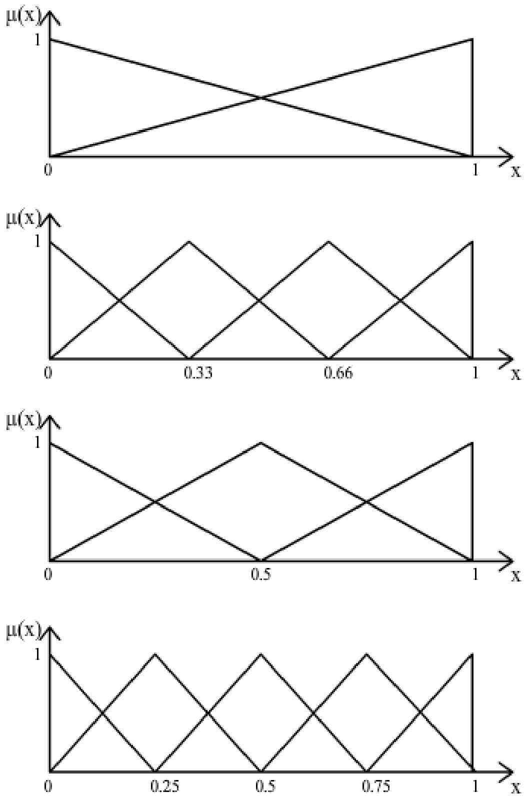

Fig. 1 14 triangular membership functions

III. FUZZY RULE-BASED CLASSIFICATION SYSTEM

Fuzzy rule-based classification system is consists of database, rule base and reasoning method. The database includes fuzzy sets, and linguistic terms. The Rule base composed of fuzzy rules corresponding to the fuzzy subsets. Reasoning method is a mechanism to classifying new samples by using the database and the fuzzy rules [8]. Type of fuzzy rules used in this study is expressed as follows.

Rule R_q: If x_1 is A_{q1} and … and x_n is A_{qn} then Class C_q with CF_q, q=1,2, … ,Q (1)

Where R_q is the label of the qth fuzzy if-then rule, $x=(x_1, x_2, \ldots ,x_n)$ is an n – dimensional pattern vector, A_{qi} is an antecedent fuzzy set, C_q is a consequent class and CF_q is a rule weight. CF_q is a value between 0 and 1. Q is a total number of fuzzy if-then rules in the rule base.

Two different reasoning methods in [9] were used in this study. These are “weighted vote method” and “singles winner method”. We used fuzzy rules generation method of Ishibuchi et al. [6]. According to this method it is assumed that m labeled patterns $x_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ are given from M classes for n-dimension classification problem. To create fuzzy rules, pattern space is separated into fuzzy subsets. We use the 14 triangular membership functions illustrated in Fig.1. For an n- dimension problem, total number of combinations antecedent of fuzzy rules is 14n. Short-length rules can be created using "don’t care"($\mu_{don't\ care}(x) = 1$). The length of the rule (L) is defined by the number of conditions which excluding “don’t care”.

The concepts of confidence and support in association rules are used for determining consequent class of fuzzy rule and selecting fuzzy rules. Fuzzy rule R_q in (1) can be viewed as an association rule $A_q \Rightarrow C_q$, Where A_q is antecedent conditions of fuzzy rule R_q and C_q is consequent class of fuzzy rule R_q. Confidence and Support of fuzzy rule are denoted by c and s respectively. Confidence and Support are defined as they are defined in [11]

$$c(A_q \Rightarrow C_q) = \frac{\sum_{x_p \in Class\, C_q} \mu_{A_q}(x_p)}{\sum_{p=1}^{m} \mu_{A_q}(x_p)} \qquad (2)$$

$$s(A_q \Rightarrow C_q) = \frac{\sum_{x_p \in Class\, C_q} \mu_{A_q}(x_p)}{m} \qquad (3)$$

The confidence is used for finding consequent class of the fuzzy rule. Confidence values of fuzzy rule are calculated for each class. The class which has maximum confidence is determined consequent class for the fuzzy rule. When confidence value of each class is the same, consequent class cannot determine. The fuzzy rule is not generated. Maximum confidence is defined as follows.

$$c(A_q \Rightarrow C_q) = \max\{c(A_q \Rightarrow Class\ h) \mid h = 1, 2, \ldots, M\} \qquad (4)$$

The generated rules are divided into the M groups according to consequent classes. The fuzzy rules are sorted ascending order by using selection criteria for each group. The first N rules are selected from M group. Fuzzy rules are generated by chosen M.N rules.

In Ishibuchi et al. [7], Product of confident value and support value is used as rule selection criterion. In this method, four different weight types were presented. The first, second, third and fourth types of weight are expressed in expression 5, 6, 7, 8 respectively:

$$CF_q^{I} = c\left(A_q \Rightarrow C_q\right) \tag{5}$$

$$CF_q^{II} = c\left(A_q \Rightarrow C_q\right) - \frac{1}{M-1}\sum_{\substack{h=1\\ h\neq C_q}}^{M} c\left(A_q \Rightarrow Class\ h\right) \tag{6}$$

$$CF_q^{III} = c\left(A_q \Rightarrow C_q\right) - \max\left\{ c\left(A_q \Rightarrow Class\ h\right) \middle| h = 1,2,\ldots,M; h \neq C_q \right. \tag{7}$$

$$CF_q^{VI} = c\left(A_q \Rightarrow C_q\right) - \sum_{\substack{h=1\\ h\neq C_q}}^{M} c\left(A_q \Rightarrow Class\ h\right) \tag{8}$$

IV. Experimental Results

Firstly, the datasets are normalized. The features of the datasets are normalized to lie between 0 and 1. The average classification rate is calculated by using leave-one-out (LV1) technique. In LV1 technique, a single pattern is used as the test data. The other patterns are used as the training data. Fuzzy rule-based classifier is designed by using training data and unused data used as test data. This process is repeated so that all patterns are used as a test data.

TABLE III
Simulation result of the medical data by single winner method

Number of Rules	No rule weights	Type 1	Type 2	Type 3	Type 4
5	96,30	96,30	96,30	96,30	96,30
10	95,62	95,96	95,96	95,96	95,96
15	92,26	93,94	94,28	94,28	95,62
20	94,28	96,30	96,63	96,63	**97,98**
25	91,58	96,30	96,63	96,63	**97,98**

TABLE IV
Simulation result of the medical data by weighted vote method

Number of Rules	No rule weights	Type 1	Type 2	Type 3	Type 4
5	96,30	96,30	96,30	96,30	96,30
10	96,30	96,30	96,30	96,30	96,30
15	95,62	95,62	95,62	95,62	96,30
20	97,98	97,98	97,98	97,98	97,64
25	97,64	97,64	97,64	97,67	**98,65**

14 membership functions which illustrated in figure 1 are used for each feature in the medical dataset. Short-length fuzzy rules can be created by using "don't care" membership function. Length of a fuzzy rule is defined by the number of antecedent conditions excluding "don't care" [7, 10]. We generated fuzzy rule of length three or less for the medical dataset.

The consequent classes of obtained fuzzy rules are determined. Fuzzy rules are divided into groups according to the consequent class (M=5). The value of product of confidence and support is used as the rule selection criterion. The fuzzy rule set is generated with N rule selected from each group according to the selection criterion. The fuzzy rule set contains NxM fuzzy rules. In this study, experiments are carried out by giving different values to N such 1,2,3,4 and 5.

Results for the medical dataset are shown Table III and Table IV. The results are presented by using single winner and weighted vote methods in Table III and Table IV, respectively.

V. Conclusion

In this study, the dataset of Congenital Heart Diseases from Department of Pediatric Cardiology at Selcuk University is used. Fuzzy rule-based classifier is used to classify the dataset. The results have shown that weighted vote method generally increased the classification accuracy of Congenital Heart Diseases. In addition, the success of classifier weighted type4 Classifier the type 4 high than others.

References

[1] A. Dragulescu ,LL. Mertens. "Developments in echocardiographic techniques for the evaluation of ventricular function in children," *Arch Cardiovasc Dis*, vol. 103, pp. 603-614, Nov. 2010.

[2] H. Feigenbaum, "Role of M-mode technique in today's echocardiography," *J Am Soc Echocardiogr*, vol. 23, pp. 240-257, Mar. 2010

[3] A. Silvia, R. Richard, K.W. Simon, R.A. Juan, "A fuzzy system for helping medical diagnosis of malformations of cortical development," *Journal of Biomedical Informatics*, vol. 40, pp. 221-235, June 2007.

[4] L. Stavros, M. Ludmil, "Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases" *Artificial Intelligence in Medicine*, vol. 50, pp. 117-126 Oct. 2010.

[5] K. Ali, K. Aytürk, Y. Ugur, "Expert system based on neuro-fuzzy rules for diagnosis breast cancer," *Expert Systems with Applications*, vol. 38, pp. 5716-5726, May. 2011.

[6] H. Ishibuchi, N. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy sets and systems*, vol. 141, pp. 59-88, Jan. 2004.

[7] H. Ishibuchi, N. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Systems*, vol. 13, pp. 428-435, Aug. 2005.

[8] O. Cordon, M.J. del Jesus, F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, pp. 21-45, Jan 1999.

[9] H. Ishibuchi, T. Nakashima, T. Morisawa, "Voting in fuzzy rule-based systems or pattern classification problems," *Fuzzy Sets and Systems*, vol. 103, pp. 223-238, Apr. 1999.

[10] M.Z. Jahromi, M. Taheri, "A proposed method for learning rule weights in fuzzy rule-based classification systems," *Fuzzy Sets and Systems*, vol. 159, pp. 449-459, Feb.2008.

[11] H. Ishibuchi, N. Yamamoto, T. Nakashima, "Fuzzy data mining: effect of fuzzy discretization," in *Proc. 1st IEEE International Conference on Data Mining*, San Jose, CA, USA, 2001 pp. 241-248.