2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Direct Higher Order Fuzzy Rule-based Classification System: Application in Mortality Prediction

Abolfazl Doostparast Torshizi
Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
Email: a.doostparast@engineering.ucsb.edu

Linda Petzold
Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
Email: petzold@cs.ucsb.edu

Mitchell Cohen
Department of Surgery
University of California, San Francisco
San Francisco, CA 94110, USA
Email: mitchell.cohen@ucsf.edu

***Abstract*—Trauma is one of the leading causes of death in the U.S. and is ranked third among death causes across all age groups. This paper presents a novel fuzzy rule-based classification approach based on the concept of General Type-2 Fuzzy sets to predict mortality for trauma patients. In this approach each rule in the rule-base has an IF and a THEN part and parameters of the IF part (antecedents) are automatically extracted using powerful general type-2 fuzzy clustering algorithms which enables the model to deal with noisy and/or missing data. To verify efficacy of the proposed model, it has been implemented on several publicly available datasets. Finally, it is used to predict mortality among patients having traumatic injuries based on a large clinical dataset. Accuracy results demonstrate superior capabilities of the proposed approach compared to crisp and fuzzy classification methods in the literature.**

***keywords*- *Trauma*; *mortality prediction*; *fuzzy rule-based classification systems*; *clustering*.**

I. Introduction

More than 192000 deaths from injury take place in the United States every year [1]. Trauma accounts for 30% of all life years lost in the US. It is ranked first in causes of death among age group 1-46, and third across all age groups. The financial cost of trauma exceeds $585 billion a year including both health care costs and loss of productivity [2]. Medical data analysis is an active field of research where one of the important topics is to predict mortality. With regard to mortality in trauma, several research studies have been conducted using various mathematical data analysis techniques [4]. Machine learning and data analysis methods provide the crucial link to precision medicine treatments dynamically after injury.

One of the major difficulties for analysis of clinical data is sparsity. For many reasons, we have a large number of missing values and the available values maybe noisy and occasionally inaccurate. One of the most effective tools in dealing with such datasets having large numbers of missing values is fuzzy logic. Fuzzy Rule Based Classification Systems (FRBCSs) are particularly effective. Some of the successful applications of FRBCSs can be found in [3], [5], [6], [7], [8].

In this paper we present a novel approach in designing FRBCS using General Type-II Fuzzy Sets (GT2 FSs). The proposed approach is easy to implement. Although most of the rule-based fuzzy expert systems rely on predefined antecedent parameters determined by experts, the proposed model is completely data-driven. Our new approach has several advantages over the current rule-based algorithms: 1) there is no need for expert knowledge, so that the parameters of the rule-base can be numerically extracted from the learning data; 2) the number of rules required in the rule-base of the system decreases remarkably which makes the system more agile; 3) there is no need to tune the antecedent parameters.

The model begins with determining the near-optimal number of clusters in each class using a novel Cluster Validity Index (CVI) [17] which is based on General Type-2 Fuzzy Sets (GT2 FSs) and results in robust and flexible data partitions. Based on the optimal cluster numbers determined, the General Type-2 Fuzzy C-Means (GT2FCM) clustering algorithm [14] is then applied to each class, and cluster centers in each dimension are then treated as the mean value of a Gaussian membership function for each rule. The process of completing the rule-base continues by computing the CF values of each rule. Finally, the classification process is accomplished based on the highest firing degree of the obtained rules. The model will be discussed thoroughly in the incoming sections.

The main contributions of this paper can be summarized as follows:

- Direct extraction of fuzzy rules without a need for expert knowledge.
- Automated tuning of rule antecedents based on the results obtained from GT2FCM.
- Reduction of the number of rules in the rule-base through finding the near-optimal number of clusters in each class.

The reminder of the paper is organized as follows: an overview of the basic concepts used in this paper is given in Section 2. In Section 3, the proposed methodology is described. Numerical experiments are provided in Section 4.

II. An overview of preliminary concepts

In this section we present an overview of the basic concepts used in this paper. To do so, first the concept of GT2 FSs is reviewed and its respective representation method being used in this research is introduced. Fuzzy C-Means (FCM) and GT2 FCM clustering algorithms are then briefly reviewed to make the reader familiar with the proposed framework. At the end of this section, we discuss some basics of FRBCSs.

A. General Type-2 Fuzzy Sets

A GT2 FS $\tilde{A}$ is expressed on a universe of discourse using its corresponding T2 membership function ([9][10]):

$$\int_{x\in X}\int_{u\in J_x}\mu_{\tilde{A}}(x,u)/(x,u)dxdu,\quad J_x\subseteq[0,1] \tag{1}$$

where x is the primary variable value, J_x denotes an interval between the lower and the upper membership functions, u denotes the secondary variable, and $\mu_{\tilde{A}}(x,u)$ denotes the secondary membership function. Here, $\int\int$ represents the union over the entire possible values of x, u and $\mu_{\tilde{A}}(x,u)$. Generally, there are three widespread representations for GT2 FSs: the Vertical Slice representation, the Wavy Slice representation, and the α-plane representation [11]. Due to the practicality and good performance of α-planes, we use this representation scheme in this paper.

B. α-Plane Representation of General Type-2 Fuzzy Sets

An α-plane of a GT2 FS $\tilde{A}$ is the union of the entire primary memberships of $\tilde{A}$ whose secondary grades are greater than or equal to α (0$\leq\alpha\leq$1). An α-plane of $\tilde{A}$ is denoted by $\tilde{A}_\alpha$ [12]. In fact, α-planes are Interval Type-2 FSs (IT2 FSs) themselves. Thus any mathematical formulation designed for IT2 FSs can easily be used on each α-plane.

$$\tilde{A}_\alpha=\int_{\forall x\in X}\int_{\forall u\in J_x}((x,u)|f_x(u))dxdu\geq\alpha \tag{2}$$

Consider the secondary membership function $\mu_{\tilde{A}}(x)$. An α-cut on this membership function is denoted by $S_{\tilde{A}}(x|\alpha)$ and can be defined as follows [14]:

$$S_{\tilde{A}}(x|\alpha)=[s_L^{\tilde{A}}(x|\alpha),s_R^{\tilde{A}}(x|\alpha)] \tag{3}$$

Each α-plane is bounded from above by its upper membership function, $\overline{\mu}_{\tilde{A}}(x|\alpha)$, and from the bottom by its lower membership function, $\underline{\mu}_{\tilde{A}}(x|\alpha)$ [13]. The upper and lower membership functions of a plane $\tilde{A}_\alpha$ can be described in terms of α-cuts as follows:

$$\overline{\mu}_{\tilde{A}}(x|\alpha)=\int_{\forall x\in X}s_R^{\tilde{A}}(x|\alpha)dx \tag{4}$$

$$\underline{\mu}_{\tilde{A}}(x|\alpha)=\int_{\forall x\in X}s_L^{\tilde{A}}(x|\alpha)dx \tag{5}$$

C. Fuzzy C-Means (FCM) algorithm

Given a set of data instances $X=[x_1,x_2,...,x_n]$ where $x_i\in\Re^d$, FCM groups X into c clusters by minimizing the following objective function [14]:

$$J(U,V)=\sum_{k=1}^{c}\sum_{i=1}^{n}{u^m}_{ik}||x_i-v_k||^2 \tag{6}$$

where c is the number of clusters, n is the number of data points, m represents the fuzzifier coefficient, and u_{ik} denotes the membership grade of x_i to the cluster C_k with the centroid v_k. This optimization problem has the following constraint:

$$\sum_{k=1}^{c}u_{ik}=1,\quad u_{ik}\geq0 \tag{7}$$

By using Lagrange multipliers, the above optimization problem is unconstrained. Then by setting the unknown parameters to zero, the following alternative iterative equations are obtained to compute the membership of the point i to the cluster k and the center of the cluster k:

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}[\frac{||x_i-v_k||}{||x_j-v_k||}]^{\frac{2}{m-1}}} \tag{8}$$

and

$$v_k=\frac{\sum_{i=1}^{n}{u_{ik}}^m x_i}{\sum_{i=1}^{n}{u_{ik}}^m} \tag{9}$$

D. GT2 Fuzzy C-Means (FCM) algorithm

GT2FCM [14] is an enhancement of the well-known FCM algorithm and employs GT2 FSs in order to capture higher levels of flexibility and accuracy. In order to employ GT2 FSs, GT2FCM utilizes the α-plane representation method. As noted before, α-planes are IT2 FSs themselves. Therefore, IT2 FCM can be implemented on each α-plane and then the obtained results for the entire α-planes can be aggregated through a process called type reduction.

GT2FCM considers the fuzzifier value m as a type-1 fuzzy set. Then it decomposes the fuzzifier into several levels:

$$m=\bigcup_{\alpha\in[0,1]}\alpha/S_m(\alpha) \tag{10}$$

In the above equation, $S_m(\alpha)$ is the α-cut of the fuzzifier m at the level α:

$$S_m(\alpha)=[s_m^L(\alpha),s_m^R(\alpha)] \tag{11}$$

Now, the degree of belonging of a data vector such as X_i to the cluster center V_j can be expressed as a type-1 FS $\tilde{u}_j(x_i)$. Summation of this secondary membership value over the entire patterns gives the GT2 membership value for the cluster center V_j:

$$\tilde{u}_j=\sum_{x_i\in X}\tilde{u}_j(x_i) \tag{12}$$

where

$$\tilde{u}_j(x_i)=\bigcup_{\alpha\in[0,1]}\alpha/S_{\tilde{u}_j}(x_i|\alpha) \tag{13}$$

Finally, the α-plane of the GT2 membership function can be obtained by summing the α-cuts over the entire available patterns. The secondary membership values for each α-level for each pattern are computed as follows:

$$\tilde{u}_j(\alpha)=\sum_{x_i\in X}S_{\tilde{u}_j}(x_i|\alpha) \tag{14}$$

$S_{\tilde{u}_j}(x_i|\alpha)$ has a lower and upper bound of the form $[s_{\tilde{u}_j}^L(x_i|\alpha),s_{\tilde{u}_j}^R(x_i|\alpha)]$. These lower and upper membership values for the pattern x_i at the level α can be computed as follows [14]:

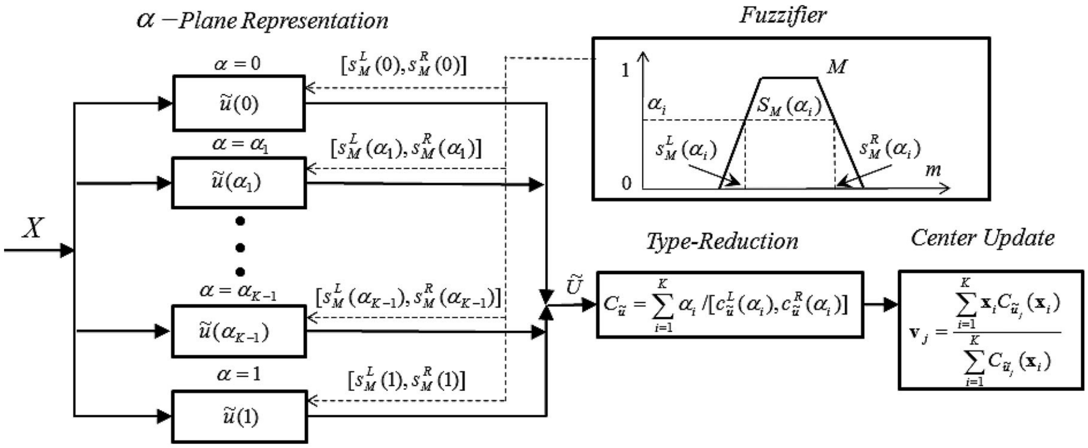

Fig. 1. A schematic view of the GT2FCM algorithm based on iterative computations on each α-plane [14].

$$s_{\tilde{u}_j}^L(x_i|\alpha) = \min\left(\frac{1}{\sum_{l=1}^{c}(d_{ij}/d_{il})^{2/(s_m^L(\alpha)-1)}}, \frac{1}{\sum_{l=1}^{c}(d_{ij}/d_{il})^{2/(s_m^R(\alpha)-1)}}\right) \quad (15)$$

$$s_{\tilde{u}_j}^R(x_i|\alpha) = \max\left(\frac{1}{\sum_{l=1}^{c}(d_{ij}/d_{il})^{2/(s_m^L(\alpha)-1)}}, \frac{1}{\sum_{l=1}^{c}(d_{ij}/d_{il})^{2/(s_m^R(\alpha)-1)}}\right) \quad (16)$$

Using the obtained secondary membership values, left and right centroid values of each α-plane are computed using the Karnik-Mendel (KM) algorithm. The cluster center V_j can be computed as follows [14]:

$$V_j = \frac{\sum_{i=1}^{K} y_i C_{\tilde{u}_j}(y_i)}{\sum_{i=1}^{K} C_{\tilde{u}_j}(y_i)} \quad (17)$$

where K denotes the discretization level of the primary domain and y_i is the position vector of the discretized steps. It is obvious that K is determined by the number of α-planes. As noted before, $C_{\tilde{u}_j}$ is the weighted composition of the interval centroids of each α-plane [14]:

$$C_{\tilde{u}_j} = \bigcup \alpha / \left[c_{\tilde{u}_j}^L(\alpha), c_{\tilde{u}_j}^R(\alpha)\right] \quad (18)$$

Finally, the hard thresholding of the GT2FCM for a pattern x_i is accomplished via the following rule [14]:

$$\text{IF } (\tilde{u}_j(x_i) > \tilde{u}_k(x_i)), \quad k = 1, \ldots, c, k \neq j \quad \text{THEN } x_i \text{ belongs to cluster } j. \quad (19)$$

A schematic view of the GT2FCM algorithm is given in Figure 1.

E. Fuzzy rule-based classification systems

Fuzzy rule-based systems have long been used in control problems as approximation measures of non-linear mappings from input vectors to output values while both of them are non-fuzzy. During the past decade these systems have been employed for classification purposes. Consider a c-class data classification problem. A typical fuzzy IF-THEN rule has the following form [15]:

$$\text{Rule } R_j: \text{If } x_1 \text{ is } A_{j1} \text{ and} \ldots\ x_n \text{ is } A_{jn}, \quad \text{Then class is } C_j \quad j = 1, 2, \ldots, N \quad (20)$$

where $x = (x_1, \ldots, x_n)$ is a n-dimensional vector, A_{ji} is the antecedent linguistic variable for the rule j in the dimension i, C_j is the consequent class and N is the total number of rules. In classic FRBCs, each dimension in the universe of discourse is partitioned into several grids each having a specific linguistic variable in the form of type-1 fuzzy sets. One of the main challenges in dealing with FRBCSs is to tune the antecedent parameters in order to achieve higher accuracy rates. The literature is rich in this regard. However, Ishibuchi and Yamamoto [16] have demonstrated that using appropriate CFs for each rule yields similar results to the rule bases with tuned antecedent variables. In fact, computing CFs is much easier and less computationally expensive than tuning the antecedent parameters. Hence we use these factors. In that case, IF-THEN rules are of the following form where CF_j is a real number in the interval $[0, 1]$.

$$\text{Rule } R_j; \text{If } x_1 \text{ is } A_{j1} \text{ and} \ldots\ x_n \text{ is } A_{jn}, \quad \text{Then class is } C_j \text{ with } CF_j \quad j = 1, 2, \ldots, N \quad (21)$$

CF is equivalent to the confidence of the fuzzy rule $A_q \Rightarrow C_q$ and is defined as follows [16]:

$$\boldsymbol{c(A_q \Rightarrow C_q) = \frac{\sum_{x_p \in \text{ class } C_q} \mu_{A_q}(x_p)}{\sum_{p=1}^{m} \mu_{A_q}(x_p)}} \quad (22)$$

where $\mu_{A_q}(x_p)$ is the firing degree of the data vector x_p in the rule A_q which is calculated as the product of membership values of that data vector in its entire dimensions. A schematic view of a two dimensional two-class dataset is represented in Figure 2. In this figure each dimension is represented by three triangular linguistic membership functions. It is obvious that the number of rules increases exponentially by an increase in the number of dimensions. This is another challenge that FRBCSs face. One of the main goals of this research is to handle this problem with FRBCSs.

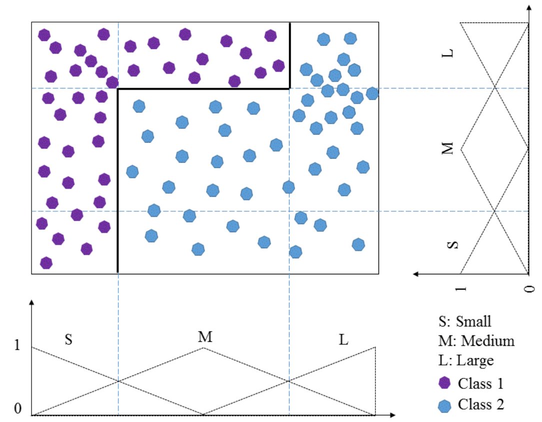

Fig. 2. A schematic view of class boundaries in an FRBCS with nine rules using three linguistic membership functions.

III. The Proposed Methodology

The proposed methodology begins with determining the near-optimal number of clusters in each class using a powerful Cluster Validiy Index (CVI) proposed by Doostparast and Zarandi [17]. This CVI is specifically designed for GT2FCM, and is formed based on the concept of α-planes in order to achieve the highest compatibility with the GT2FCM algorithm. After finding the optimal number of clusters for the data vectors in each class, the GT2FCM is applied on classes and then the cluster centers are extracted for each class. Cluster centers of each class are then used to construct the antecedent part of the rules in the rule-base. The number of rules in each class will be the same as the number of clusters within that class. This approach is advantageous over the ordinary FRBCSs identification procedures [16] since the number of rules remains reasonable, while in the current approaches the number of rules in the rule-base increases exponentially with an increase in the number of dimensions.

In our proposed approach, cluster centers are extracted and each cluster represents a rule. The cluster prototype values in each dimension are directly used in the corresponding location in the antecedent part of each rule. All the rules regarding each class have the same consequent (THEN part of the rule) as their respective class. To build the antecedent values of rules we have utilized Gaussian membership functions in such a way that the mean of the Gaussian membership function in each dimension is the cluster prototype value in that dimension. Based on our experiments, a standard deviation of 1 will result in satisfactory classification output, so we use 1 as the standard deviation of Gaussian membership functions in the antecedents of rules. For each dimension in the antecedent part of the rules, the membership function is computed using the following formulation:

$$A_{ij} = e^{-\left(\frac{x_j - m_{ij}}{\sigma_{ij}}\right)^2} \tag{23}$$

where A_{ij} is the membership function in the antecedent part of rule i for dimension j and m_{ij} (σ_{ij}) is the mean (standard deviation) value of the Gaussian membership function of rule i for dimension j. The main advantage of this approach over the currently used rule-generation mechanisms is that there is no need to generate redundant rules to cover the universe of discourse entirely. Hence, computational complexity is reduced.

After generating the rule-base for each class, the entire parts of the rule-base are aggregated in order to calculate the CF values using (22). Although the optimal number of clusters have been obtained in the initial stage, due to the randomness of the GT2FCM for finding the initial cluster prototypes we have applied a K-fold cross validation process. Also the entire process is repeated *N* times. We consider *N*=30. During each round of the cross-validation stage, a new rule-base is generated and its antecedent values, accompanied by their respective CF values are computed. Then the system is implemented on the validation dataset and validation accuracy is measured. This process is repeated until the entire folds have been tested. The proposed approach is graphically represented in Figure 3.

In order to demonstrate computational efficiency of the proposed approach, in the next section we present results of implementing the proposed approach on several datasets including 5 publicly available datasets and a real clinical dataset.

IV. Experimental Results

In order to verify the performance of the proposed framework, we have implemented this approach on several datasets with different characteristics. In the following these datasets are introduced:

1) Five publicly available datasets are used from the University of California, Irvine, machine learning repository [18]. These datasets include: Dermatology, Wine, Iris, Heart, and Wisconsin. These datasets are summarized in Table I.
2) A clinical dataset representing various static and time-series data vectors. This is a real trauma patient clinical dataset which was collected by the UCSF/San Francisco General Hospital and Trauma center and contains 1413 patients. This dataset consists of two main parts: static and hour 0 measurements, time dependent measurements on various time intervals. In this study we use the first part. In this regard the following information for each patient has been used: toxicology screening results, demographic data, co-morbidities, substance use history and time 0 physiologic and biologic (coagulation and inflammatory markers) measurements. Hour 0 measurements include basic measures such as temperature, blood pressure, heart rate, respiratory rate, some blood factors such as Factor V and VII, platelet count, and protein C level. The output of this dataset is 28 day mortality.
   In order to clean the data, we have deleted features having $\geq$35% missing values across all measurements. Patients having $\geq$40% missing values across all features were deleted. After cleaning the dataset we arrived at a set consisting of 72 features and 720 patients. In this dataset around 30% of values are still missing. Note that for the methods used in this paper we have made imputations using the SVD method. This dataset has two classes: patients who survived and patients who died after four weeks. In the cleaned data 78% of patients survived and the remaining died.

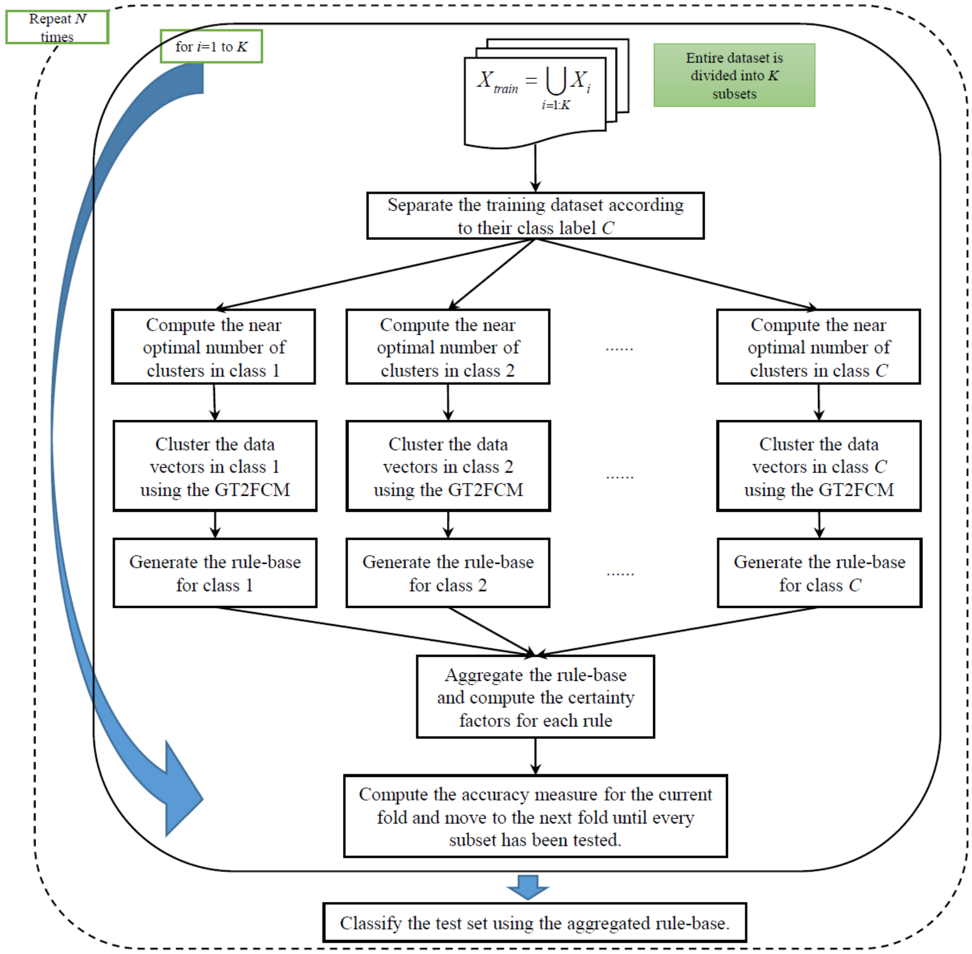

Fig. 3. The overall work flow of the proposed approach.

TABLE I. SUMMARY OF THE DATASETS USED

Dataset	#Classes	#Features	Dataset Size
Iris	3	4	150
Wine	3	13	178
Heart	2	6	270
Wisconsin	2	9	683
Dermatology	6	34	358
Clinical Data	2	72	720

To demonstrate the effectiveness of the proposed approach compared to current techniques, we have performed extensive experiments based on the following methods: IVTURS [19], FARC-HD [20], FURIA [21], and Ordinary FRBCS (OFRBCS). Note that these algorithms are based on ordinary FRBCSs and their antecedent parts have been tuned with various heuristic methods and none of them employ CF as an alternative for antecedent tuning.

We used 80% of the dataset vectors for training and 20% for testing the model. 10-fold cross validation was used in the experiments. Computational results for these experiments have been presented in Tables IV and V. For two class problems we have also made computations using Support Vector Machines (SVM), a feed-forward Artificial Neural Network (ANN) with two hidden layers each layer having 10 neurons, and Logistic Regression (LR) method. Classification results are presented in Tables II and III.

TABLE II. ACCURACY RESULTS ON THE TRAINING SETS

Dataset	The Proposed Approach	SVM	ANN	LR
Heart	**91.3**	90.85	89.9	90.80
Wisconsin	**97.3**	97	95.8	96.85
Clinical Data	**88.27**	85.3	87.2	87.3

TABLE III. ACCURACY RESULTS ON THE TEST SETS

Dataset	The Proposed Approach	SVM	ANN	LR
Heart	**85.2**	84.6	84	85.0
Wisconsin	**95.9**	94.83	93.2	94.8
Clinical Data	**86.4**	81.5	83.2	84.1

TABLE VI. NUMBER OF GENERATED RULES FOR THE CLINICAL DATASET BY THE FUZZY ALGORITHMS

Method	#Rules in the rule-base
IVTURS	19
FARC-HD	21
FURIA	21
OFRBCS	34
The Proposed Approach	10

Our proposed approach results in fewer rules in the rule-base. The final number of rules being generated by each of the five fuzzy algorithms for the clinical data is summarized in Table VI. According to the results presented in Tables II and III, the proposed approach outperforms SVM and ANN. Here accuracy is defined as the ratio of the number of correct

TABLE IV. ACCURACY OF THE RESULTS ON THE TRAINING DATASETS

Dataset	The Proposed Approach	IVTURS	FARC-HD	FURIA	OFRBCS
Iris	**98.87**	98.17	98.5	98.5	95.3
Wisconsin	97.3	98.5	98.76	**98.83**	97
Dermatology	**100**	99.86	**100**	98.88	95
Heart	91.3	93.61	**94.63**	89.72	78.2
Wine	99.4	99.30	**99.86**	99.58	97.3
Clinical Data	**88.27**	88	86.9	86	79

TABLE V. ACCURACY OF THE RESULTS ON THE TEST DATASETS

Dataset	The Proposed Approach	IVTURS	FARC-HD	FURIA	OFRBCS
Iris	**96.1**	96	94	96	93.5
Wisconsin	95.9	96.44	**96.63**	96.61	95.1
Dermatology	**95.3**	94.42	89.94	93.86	91.1
Heart	85.2	88.15	**89.44**	78.15	75.3
Wine	96.8	**97.19**	96.62	93.78	89.7
Clinical Data	**86.4**	85.1	84.8	85.2	77

results (labels) to the total number of vectors being tested. The difference between accuracy results is more obvious in the clinical dataset where the number of features is much larger than the two other datasets. Also, we can see that in most cases SVM has demonstrated slightly better classification performance than ANN.

In the other experiments which have been made between the state-of-the-art FRBCSs, the proposed approach achieves the highest accuracy rate in half of the datasets. Here accuracy rate (in percent) for each method is defined as the number of times that an underlying method has achieved the best accuracy on each of the six datasets divided by the total number of datasets (which is six here). These rates are presented in Tables IV and V and accuracy percentage rates for each method are visualized in Figure 4. For example, according to the Table IV, the proposed approach achieves the highest accuracy rate in the Iris, Dermatology, and Clinical data (shown in bold). Therefore its accuracy rate is 50% ($3 \div 6$)

In the both training and test datasets, the proposed approach has the highest accuracy rate of 50%. For the training set, the FARC-HD algorithm also yields the same rate as our method, while in the test set it shows a 33.34% accuracy rate compared to 50% of our approach. The IVTURS model does not achieve the highest rank in any of the datasets of the training group, while it has a 16.67% accuracy rate in the test datasets. A natural hypothesis before conducting the experiments was that OFRBCSs will have the lowest ranking among the participating methods. This hypothesis is proved here and OFRBCS has accuracy rate of 0 in both training and test groups for the entire datasets.

We have also provided the AUC (area under the ROC curve) value accompanied by the True Positive (TP) and True Negative (TN) rates for each method on the test clinical data. The positive class is assigned to patients who survived (TableVII). It can be observed that our proposed method achieves the highest AUC value among the methods being compared. Here are some points. We expected to obtain larger (smaller) TP (TN) values in non-fuzzy methods. SVM and ANN have a better TP value compared to our method but their TN values are the lowest among other methods. This can be due to low number of Dead patients in the test data considering large number of features. Efficiency of our method is demonstrated here where it has the highest TN value.

TABLE VII. AUC VALUES FOR THE COMPARED METHODS ON THE CLINICAL DATA

Method	AUC	True Positive Rate	True Negative Rate
IVTURS	0.79	0.81	0.79
FURIA	0.78	0.89	0.72
FARC-HD	0.73	0.80	0.71
OFRBCS	0.69	0.74	0.66
SVM	0.83	0.87	0.77
ANN	0.825	0.86	0.66
LR	0.78	0.79	0.69
The Proposed Approach	0.87	0.841	0.857

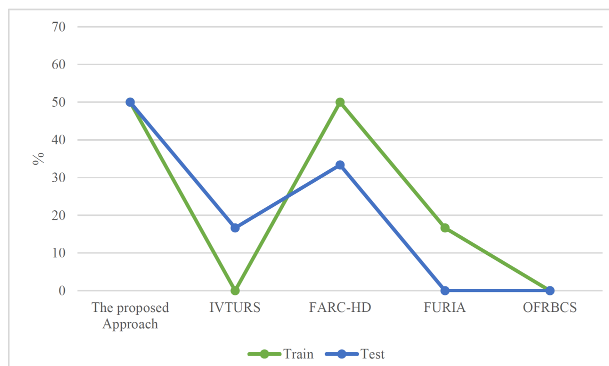

Fig. 4. Accuracy rate of each method on the clinical data.

V. CONCLUSION

This paper presents a novel direct approach in designing FRBCSs using the GT2FCM algorithm and its respective CVI using the α-plane representation method. Compared to the existing FRBCSs, which are based on complete coverage of the universe of discourse through defining linguistic variables for the entire dimensions (features), the proposed approach determines the optimum number of clusters in each class, both from a separability and compactness points of view. This reduces the total number of rules to represent the whole problem space. In each class, the number of optimal rules is equal to the number of representing rules in that class. After computing the optimal clusters in each class, antecedent variables for each dimension of the entire rules are extracted from the computed clusters. In this paper, a Gaussian membership function was considered as antecedent of the fuzzy rules since it has shown its efficacy in handling real-world classification problems compared to the traditional triangular or trapezoidal membership functions. Another major difference between our method and other available fuzzy rule-based models is its easy functionality. This is because the alternative approaches spend a huge amount of time on tuning the parameters of the antecedent part of each rule while the proposed algorithm employs a certainty factor for each rule. The robustness and computational speed of certainty factors have been verified in the literature. On the other hand, antecedent parameters are automatically tuned and there would be a need to re-tune them since the antecedent parameters are the output of an optimized clustering algorithm, themselves. The proposed algorithm has been tested on five publicly available dataset and a real clinical dataset consisting 72 clinical features and 720 samples. Numerical results demonstrate that the proposed algorithm outperforms non-fuzzy methods such as SVM, ANNs, and LR. It also shows excellent classification accuracy compared to four state-of-the-art FRBCSs with tuned antecedent parameters. Our method achieved the best accuracy on the real clinical dataset for both the training and test datasets with 88.27% and 86.4% accuracy, respectively.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by the Institute for Collaborative Biotechnologies through Coagulopathy grant W911NF-10-2-0114 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

[1] National trauma institute, Trauma statistics, http://www.nationaltraumainstitute.org, Accessed June 2015.

[2] Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS), http://www.cdc.gov/injury/wisqars, Accessed June 2015.

[3] J. A. Sanz, A. Fernandez, H. Bustince, F. Herrera, IVTURS: A Linguistic Fuzzy Rule-Based Classification System Based On a New Interval-Valued Fuzzy Reasoning Method With Tuning and Rule Selection, IEEE Trans. Fuzz. Systs, vol. 21, no. 3, pp. 399-411, 2013.

[4] Y. Y. Zhang, B. J. Daigle, L. Ferrigno, M. Cohen, L. Petzold, "Data-driven mortality prediction for trauma patients," *Proc. Annual Conf. on Neurl. Inf. Proc. Syst.*, NIPS, Montreal, Canada, 2014.

[5] E. G. Mansoori, M. J. Zolghadri, S. D. Ketabi, Protein superfamily classification using fuzzy rule-based classifier, IEEE Trans. Fuzz. Systs, vol. 8, no. 1, pp. 92-99, 2009.

[6] E. G. Mansoori, M. J. Zolghadri, S. D. Ketabi, SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzz. Systs, vol. 16, no. 4, pp. 1061-1071, 2008.

[7] R. Nowicki, Rough neuro-fuzzy structures for classification with missing data,, IEEE Trans. Syst. Man. Cybern, vol. 39, no. 6, pp. 1334-1347, 2009.

[8] C. Wei-Yuan, J. Chia-Feng, A fuzzy model with online incremental svm and margi-selective gradient descent learning for classification problems, IEEE Trans. Fuzz. Systs, vol. 22, no. 2, pp. 324-337, 2013.

[9] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Upper Saddle River, NJ: Prentice-Hall, 2001.

[10] M. H. Fazel Zarandi, A. Doostparast, I. B. Turksen, B. Rezaei, "A new indirect approach to the type-2 fuzzy systems modeling and design," Inf. Sci., vol 232, pp. 346-365, 2013.

[11] J. M. Mendel, R. I. John, "Type-2 fuzzy sets made simple," IEEE Trans. Fuzz. Syst., vol. 10, no. 2, pp. 117-127, 2002.

[12] J. M. Mendel, F. Liu, and D. Zhai, "α-plane representation for type-2 fuzzy sets: Theory and application," IEEE Trans. Fuzz. Syst., vol. 17, no. 5, pp. 1189-1207, 2009.

[13] D. Zhai and J. M. Mendel, "Computing the centroid of a general type-2 fuzzy set by means of the centroid-flow algorithm," IEEE Trans. Fuzz. Syst., vol. 19, no. 3, pp. 401-422, 2011.

[14] O. Linda, M. Manic, "General Type-2 Fuzzy C-Means Algorithm for Uncertain Fuzzy Clustering ," IEEE Trans. Fuzz. Syst., vol. 20, no. 5, pp. 883-897, 2012.

[15] H. Ishibuchi, T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," IEEE Trans. Fuzz. Syst., vol. 9, no. 4, pp. 506-515, 2001.

[16] H. Ishibuchi, T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," IEEE Trans. Fuzz. Syst., vol. 13, no. 4, pp. 428-435, 2005.

[17] A. Doostparast, M. H. Fazel Zarandi, "A New Cluster Validity Measure Based on General Type-2 Fuzzy Sets: Application in Gene Expression Data Clustering," Knowledge- based Syst., vol. 64, pp. 81-93, 2014.

[18] University of California, Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/, accessed on June 2015.

[19] J. A. Sanz, A. Fernandez, H. Bustince, F. Herrera, "IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection," IEEE Trans. Fuzz. Syst., vol. 21, no. 3, pp. 399-411, 2013.

[20] J. Alcala-Fdez, R. Alcala, and F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzz. Syst., vol. 19, no. 5, pp. 857872, 2011.

[21] J. Huhn and E. Hullermeier, FURIA: An algorithm for unordered fuzzy rule induction, Data Mining Knowl. Disc., vol. 19, no. 3, pp. 293319, 2009.