A Genetic Fuzzy Approach for Rule Extraction for Rule-Based Classification with Application to Medical Diagnosis

Rahil Hosseini[1,2], Tim Ellis[1], Mahdi Mazinani[1,2], Salah Qanadli[3], Jamshid Dehmeshki[1]

[1]Quantitative Imaging lab, (Research & Development), Faculty of Computing, Information Systems and Mathematics,
Kingston University, London, UK,
[2]Azad University in Oxford, Oxford, UK
[3]Quantitative Imaging lab, (Clinical Development), at the Department of Radiology and Cardiology and the Intensive Care Unit, Lausanne Hospital, Switzerland
(rahil.hosseini@kingston.ac.uk, j.dehmeshki@kingston.ac.uk)

Abstract— Rule extraction is an important task in knowledge discovery from imperfect training dataset in uncertain environments such as medical diagnosis. In a medical classification system for diagnosis, we cope with expensive or lack of expert knowledge in the design of the classifier. This paper presents an evolutionary fuzzy approach for tackling the problem of uncertainty in the process of rule extraction for classification. A type-2 fuzzy logic (T2FL) in combination with genetic algorithm is proposed for modeling uncertainty a long with rule extraction process. The approach has been applied on the Wisconsin breast cancer diagnostic dataset (WBCD) with an average classification accuracy of 96% which is competitive with the best results to date. Furthermore, the proposed T2FLS maintains the tradeoffs between interpretability and accuracy by producing the most comprehensive rule set with only one linguistic term per rule in compare to other methods proposed for the Wisconsin breast cancer diagnosis.

Keywords—Rule extraction, type-2 fuzzy logic system, Genetic algorithm, Classification, Modeling Uncertainty

1 Introduction

Learning rules in an uncertain environment is a major task for developing a fuzzy rule-based system. Fuzzy logic systems (FLSs) as an expert rule-based system has been concentrated for managing uncertainties associated with linguistic expert knowledge. However, designing a FLS would be challenging when dealing with uncertain environment with imperfect and lack of expert knowledge. The idea of uncertain rule-based fuzzy logic systems were introduced by Mendel in [1]. This system takes advantage of type-2 fuzzy sets for tackling uncertainty issues such as imprecision in the input data, noisy measurements, inter- and intra- uncertainties, and word perception and non-stationary features [1]. A type-2 fuzzy logic system (T2FLS) utilizes a three-dimensional type-2 fuzzy membership function (T2MF) and the footprint of uncertainty (FOU) for managing the problem of uncertainty. The rule set contains a collection of linguistic rules (extracted from experts or defined by a learning algorithm) whilst the database includes linguistic term sets in the rules and their associated MFs. The accuracy of a fuzzy rule-based system is affected by rule sets as well as membership functions.

There are three common types of rules in a fuzzy rule-based system, as follows:

For a given $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$ in an n-dimensional pattern space, the ith fuzzy rule ($\mathrm{R_i}$) type can be:

Ishibuchi rule type with a class label ($\mathrm{C_i}$) in the consequent part [3]:

$$Rule\ R_i{:}\ If\ x_1\ is\ A_{i1}\ and\ \ldots and\ x_n\ is\ A_{in}\ then\ Class\ C_i$$

where $\mathrm{A_{i1}}, \ldots, \mathrm{A_{in}}$ are antecedent fuzzy sets associated to linguistic terms. There is also another extension of this rule type which assigns each rule a weight ($\mathrm{W_i}$) or a soundness degree [2] which shows the degree of certainty of the rule:

$$Rule\ R_i\ :\ If\ x_1\ is\ A_{i1}\ and\ \ldots and\ x_n\ is\ A_{in}\ then\ \ Class\ Ci\ with\ W_i$$

The Ishibuchi rule type is more applicable for pattern classification problems and has higher interpretability than the other types of rules such as Mamdani and TKS[2] and is the focus of this paper. We have extended the Ishibuchi classification rule type for rule-based pattern classification. The proposed rule structure in this paper for uncertain rule-based pattern classification, for a given pattern $\mathrm{X} = (\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n})$ is as follow:

$$Rule\ R_i{:}\ If\ x_1\ is\ \tilde{A}_{i1}\ and\ \ldots and\ x_n\ is\ \tilde{A}_{in}\ then\ Class\ C_i\ with\ GC_i$$

where $\widetilde{\mathrm{A}}_{\mathrm{i1}}, \ldots, \widetilde{\mathrm{A}}_{\mathrm{in}}$ are type-2 fuzzy sets, i=1..M is the number of rules, C_i is the class label and the GC_i is the grade of certainty. This rule type has the capacity to capture the uncertainties both in the training dataset and the mathematical models applied for measuring the features (different methods with different error rate may lead to different results) in the antecedent part of the rule. The T2FLS classifier makes the deterministic decision, in terms of the class label, with the degree of certainty of the rules. The degree of certainty is estimated based on the imprecision in the training dataset.

Our proposed approach for extracting linguistic rules from training data extends the method proposed in [4] for handling uncertainty in rule-based pattern classification problems. The proposed learning rule algorithm is applicable for multi-dimensional pattern classification problems with high degree of uncertainty in numerical measurements and linguistic terms, such as medical diagnosis applications.

The rest of the paper is organized as follow: a brief overview of the theory and the concepts of type-2 fuzzy set theory are described in Section 2. Section 3 describes proposed approach for rule extraction for fuzzy rule-based classifier followed by a genetic algorithm for uncertain linguistic rules selection. The result of the proposed uncertain rule-based pattern classifier applied to the Wisconsin breast cancer diagnostic dataset

(WBCD) is demonstrated in Section 4 and the paper is concluded in Section 5.

2 An Overview of the Theory and Concepts of the Type-2 Fuzzy Sets

In this section, a brief overview of a T2FS, according to [1], is demonstrated.

2.1 Type-2 fuzzy set theory

The membership function of a type-1 fuzzy set is a crisp number, whereas in a type-2 fuzzy set (T2FS), the MF is a subset of a fuzzy set and is itself fuzzy. The membership function of a type-2 fuzzy set (see Figure 1), is defined by the function , is called the secondary membership function of a type-2 fuzzy set, where [1]. Fuzzy set can be defined as [1]:

The domain of a secondary membership function is called the primary membership function of x and in (1) is a primary membership function at , (see Figure 1). For simplicity can be written as [1].

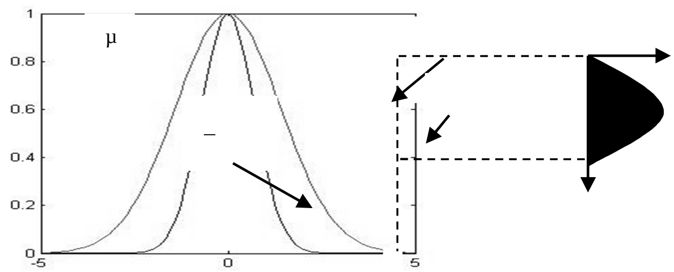

Fig. 1. Primary Membership Function of Type-2 Fuzzy Set at

Uncertainty in the primary membership function of a type-2 fuzzy set includes a bounded and blurred region called the Footprint of Uncertainty (FOU) [2]. It can be written as the union of all primary memberships:

An upper and a lower membership function bound the FOU of a type-2 fuzzy set (see Figure 1) and are defined as follows, A type-2 fuzzy membership function can be any type such as Gaussian, trapezoidal, triangular or interval. The name that we choose to describe the type-2 membership function is related to the name of its secondary membership function. When the secondary membership functions are interval sets, we have an interval type-2 membership function (IT2MF) defined as:

$$\mu_{\tilde{A}}(x, u) = 1 \qquad \forall\, u \in J_x \subseteq [0,1] \tag{5}$$

Interval type-2 fuzzy sets are the practical type-2 fuzzy sets and are frequently used because of lower computational complexity and greater simplicity. Using Zadeh's extension principle [5], union, intersection and negation of type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ can be defined [6].

2.2 Type-2 Fuzzy Logic System

The general architecture of a T2FLS [2] is demonstrated in Figure 2. The important components in the architecture of the T2FLS are: fuzzifier, inference engine, rules, and output producer, which includes a type reducer and a defuzzifier. The fuzzifier converts crisp inputs into a type-2 fuzzy set, mapping a crisp point $X_i = (x_{i1}, x_{i2,}, \dots, x_{im})$ into a type-2 fuzzy set $\tilde{F}$ in the universe of discourse X. In a type-2 FLS, the inference engine combines rules and gives a mapping from the input type-2 fuzzy sets to the output type-2 fuzzy sets. Multiple antecedents are connected by the t-norm and the membership of input and output sets are combined using the sup-star composition. Multiple rules are combined using the t-conorm operation [6].

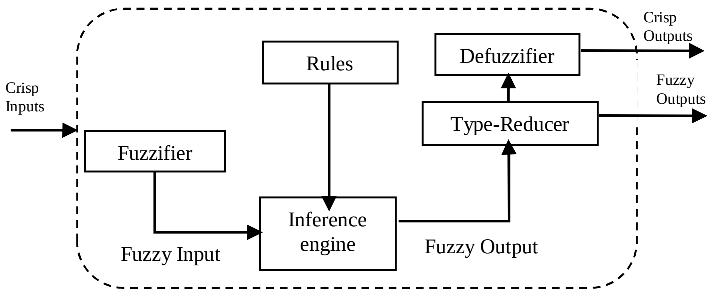

Fig. 2. General Architecture of Type-2 Fuzzy Logic System

The output producer consists of two blocks: type reduction and defuzzification. The type reducer maps a type-2 fuzzy set (T2FS) to a type-1 fuzzy set (T1FS) and the defuzzifier defines the crisp output of the T2FLS. The centroid type reducer [2] is more commonly applied and is the method used in this paper.

3 Rule Extraction for Type-2 Fuzzy Pattern Classification System

In this section, our genetic algorithm for rules extraction for a type-2 fuzzy pattern classification system is demonstrated. The approach extracts systems rules from an uncertain, noisy and imperfect training dataset. Our proposed approach for learning uncertain linguistic rules from training data extends the method proposed in [3] and improves it for uncertain rule-based pattern classification systems.
Interval type-2 Gaussian membership function is automatically generated using the proposed approach in [8]. In this method the FOU of an interval type-2 fuzzy set is considered to have a Gaussian distribution. The lower and upper bound parameters of an interval type-2 Gaussian membership function with mean m and standard deviation s are considered as follows:

$$\overline{m} = m + k_m s\,,\ \underline{m} = m - k_m s,\ \ k_m \in [0, 1] \tag{6}$$

$$\overline{s} = s \times k_v,\ \ \underline{s} = s\,/\,k_v, \qquad k_v \in [0.3, 1] \tag{7}$$

where the parameters k_m and k_v are considered as the parameters for tuning the FOU, $\overline{m}$, and $\underline{m}$ are the lower and upper bound of the mean and $\overline{s}$ and $\underline{s}$ are the lower and upper bound of the standard deviation of the interval type-2 lower and upper bound membership functions, respectively [8].

Then the procedure explained in the rest of this section is applied to generate linguistic classification rules for our uncertain classification system. The uncertainty in the rules is managed through type-2 fuzzy sets in the antecedent part. The structure of a rule in a type-2 fuzzy rule-based pattern classification for a given pattern $X_p = (x_{p1}, \ldots, x_{pn})$ is given as

$$Rule\ R_i{:}\ If\ x_{p1}\ is\ \tilde{A}_{i1}\ and\ \ldots and\ x_{pn}\ is\ \tilde{A}_{in}\ then\ Class\ C_i\ with\ GC_i$$

where $\tilde{A}_{i1}, \ldots, \tilde{A}_{in}$ are interval type-2 fuzzy sets, i=1..M is number of rules. The steps of the proposed algorithm for selecting rules are explained as follows:

Step1: The grade of compatibility of pattern X_p to the jth linguistic classification rule R_j is calculated as [1]

$$\tilde{A}_j(X_p) = \tilde{A}_{j1}(x_{p1}) \cap \tilde{A}_{j_2}(x_{p2}) \cap \ldots \cap \tilde{A}_{j_n}(x_{pn}) \tag{8}$$

where $\cap$ is the type-2 fuzzy intersection which can be implemented by a meet operation [1]. The total grade of compatibility ($\tilde{\beta}_{\text{Class h}}$) of the given patterns in Class h to the jth classification rule R_j is an interval type-2 fuzzy set $\tilde{\beta}_{\text{Class h}} = \left[\underline{\beta}_{\text{Class h}} \overline{\beta}_{\text{Class h}}\right]$ which is calculated as

$$\tilde{\beta}_{\text{Class h}} = \sum_{\forall X_p \in \text{Class h}} \mu_j(X_p) \tag{9}$$

where μ_j is the firing strength of the jth classification rule as

$$\mu_j(X_p) = \mu_{\tilde{A}_{j_1}}(x_{p1}) \cdot \mu_{\tilde{A}_{j_2}}(x_{p2}) \cdot \ \ldots \cdot \mu_{\tilde{A}_{j_n}}(x_{pn}) \tag{10}$$

Step2: The consequent C_j of the rule R_j is the class with maximum grade of compatibility (Class $\tilde{h}$) [1].

If $\tilde{\beta}_{\mathrm{Class\ h}}$ is not unique, it means two or more classes have the same $\tilde{\beta}_{\mathrm{Class\ h}}$, then it is assigned to $\emptyset$, which means empty class. A rule with the empty class in the consequent part is called a dummy rule [4].

Step3: The grade of certainty $\mathrm{GC_j}$ of a dummy rule is assigned to 0. The grade of certainty of a non-dummy rule is defined as

$$\mathrm{GC_j} = \frac{\beta_{\mathrm{Class\ h}} - \acute{\beta}}{\sum_{\mathrm{k=1}}^{\mathrm{c}} \beta_{\mathrm{Class\ k}}} \tag{11}$$

$$\acute{\beta} = \left(\sum_{\mathrm{h \neq k}}^{\mathrm{c}} \beta_{\mathrm{Class\ k}} \right) \Big/ (c-1) \tag{12}$$

Fuzzy reasoning is applied for classifying new patterns. Consider we are given a subset S of a set of generated rules, a new pattern $\mathrm{x_p}$ is classified by the Class h in rule $\mathrm{R_j} \in \mathrm{S}$, $(\mathrm{j} = 1..\mathrm{M})$ which has the maximum α_{j}:

$$\alpha_{\mathrm{j}}(\mathrm{x_p}) = \max\{\, \mu_{\mathrm{j}}(\mathrm{X_p})\, .\, \mathrm{GC_j}\} \tag{13}$$

The centroid type reduction following defuzzification in [8] of an interval type-2 fuzzy set $\tilde{\alpha}$ is used to calculate α_{j}. Therefore, the class label (h) is defined by the rule with maximum $\alpha_{\mathrm{j}}(\mathrm{x_p})$ as [4]

$$\alpha_{\mathrm{Class\ h}} = \max\{(\alpha_{\mathrm{j}}(\mathrm{x_p}) \big| \ \mathrm{Class\ h} = \mathrm{C_j} \text{ and } R_j \in S\} \tag{14}$$

Furthermore, if two or more classes take the maximum $\alpha_{\mathrm{Class\ h}}$, then we assign -1 to $\alpha_{\mathrm{Class\ h}}$ which means the given pattern $\mathrm{x_p}$ cannot be classified by rule set S and is an unclassified pattern [4]. The value of the $\alpha_{\mathrm{Class\ h}}$ can be considered as the confidence measure of assigning pattern $\mathrm{x_p}$ to class h. The interval of $[\underline{\alpha}_{\mathrm{j}} \quad \overline{\alpha}_{\mathrm{j}}]$ of the rule with maximum $\alpha_{\mathrm{j}}(\mathrm{x_p})$ defines the lower and upper bounds of the confidence measure of the classifier decision. The next section describes the GA approach for rule extraction and tuning for our presented uncertain rule-based pattern classification based on imprecise and noisy training dataset. The GA linguistic rule selection approach presented by Ishibuchi et al. [4] has been applied for rule selection. The general steps of their GA rule selection method are described in this section.

Initialization - the structure of the chromosome is similar to the method proposed in [4]. In this structure, each rule is treated as an individual and a population consists of a fixed number of classification rules (N_{pop}) [4]. The solution is a set of rules ($R_1 R_2\ \dots\ R_n$) which has the maximum fitness value. The antecedent part of rules in the initial population is randomly selected from various type-2 fuzzy partitions in the fuzzy pattern space (Section 3.1). The consequent and grade of certainty of these rules are defined through the rule generation method (Section 4.1).

Rule elimination - in this step, dummy rules are first extracted from the list of rules. Then the rules which do not classify any patterns (non-active rules) are removed from the list.

Fitness evaluation - A fitness value is assigned to each linguistic classification rule R_j in the current generation as

$$\text{Fitness}(R_j) = \left(W_{NCC} * N_{CC}(R_j) - W_{NMC} * N_{MC}(R_j)\right) \quad (15)$$

where N_{CC} is the number of patterns correctly classified by classification rule R_j, N_{MC} is the number of misclassified patterns, and W_{NCC} and W_{NMC} are the classified and non-classified weights.

Rule Selection - In this step, the $N_{pop/2}$ of the rules with the higher selection probability (P) are selected. The probability of rule j in the gth generation is

$$P(R_j, g) = \frac{f(R_j) - f_{min}(S)}{\sum_{R_i \in S}\{ f(R_i) - f_{min}(S)\}} \quad (16)$$

$$f_{min}(S) = \min\{f(R_i) \mid R_i \in S\} \quad (17)$$

where $f(R_j)$ is the fitness of rule j, and $f_{min}(S)$ is the minimum fitness in the current generation (g).

Crossover and Mutation - In this step, for each pair of selected parents in the previous step, the crossover is applied. In crossover, a position bit is randomly selected. Mutation then randomly chooses a bit and changes it to a new value. The mutation and crossover operation apply only on the antecedent part of a rule. Then the consequent and the grade of certainty of rules must be defined according to the rule generation method in Section 3.2.

Rule Replacement - the generated rules in the previous step are replaced with some of the rules in the current population. The replacement strategy is based on the fitness value of the rules which means new rules are replaced by rules with lower fitness values.

Termination - the GA algorithm stopping criteria is either a desirable fitness value or the total number of the generations.

4 Experimental Results and Performance Evaluation

The Wisconsin breast cancer dataset (WBCD) has been frequently used for the evaluation of classifiers. The dataset was computed from fine needle aspiration (FNA) of a breast mass through image processing and was collected at the University of Wisconsin [19]. The samples contain visual assessment of the nuclear features of fine needle aspirates (FNAs) taken from patients breasts by Dr. Wolberg and can be obtained from UCI (University of California at Irvine) machine learning depository [19]. This dataset contains some noise and residual variations in its 699 samples. The 16 examples with missing features were removed as was done in previous studies [20]-[27] and [29]-[31]. There are nine integer features. Each feature has a value between 1 and 10, value 1 corresponding to a normal state and 10 to a most abnormal state. The two outputs are benign and malignant samples, are distributed 444 benign (65%) and 239 malignant (35%).These features describe characteristics as follows

1. Clump Thickness (CT)
2. Uniformity of Cell Size (UC)

3. Uniformity of Cell Shape (UCS)
4. Marginal Adhesion (MA)
5. Single Epithelial Cell Size (SEC)
6. Bare Nuclei (BN)
7. Bland Chromatin (BC)
8. Normal Nucleoli (NN)
9. Mitoses (M)

For the reason of the evaluation of the generated classifier using an uncertain linguistic rule learning approach, our proposed method was applied to the WBCD dataset.

4.1 Estimating the parameters of the FOU

In order to generate the interval type-2 fuzzy membership functions for the breast cancer features, first the FOU parameters need to be estimated. These parameters are defined according to the model proposed in Section 3. We applied the genetic algorithm for all possible sets of k_m and k_v. The average fitness of the 10-fold cross-validation result was calculated for a fixed number of generations (60). The average $[k_m, k_v]$ of the top 10 fitness's was used to estimate the FOU parameters, Table1 summarizes the results obtained for IT2FLS with three linguistic terms.

Table 1. Estimation of the FOU Parameters

No	k_m	k_v	10- fold CV	75/25	50/50
1	1	0.9	96.97	98.24(69)*	97.07(52)
2	0.8	1	96.88	98.82(15)	97.07(86)
3	0.4	0.9	96.78	97.65(64)	96.48(51)
4	0.4	0.6	96.76	96.08(27)	97.65(56)
5	0.7	0.5	96.74	96.47(64)	96.19(49)
6	0.1	1	96.68	96.47(63)	97.36(68)
7	0.7	0.8	96.66	98.82(39)	98.24(93)
8	0.2	0.7	96.65	97.65(74)	96.19(48)
9	0.9	1	96.63	98.24(80)	98.53(35)
10	0.6	0.8	96.63	97.06(83)	96.48(50)
		average=	96.74	97.55	97.12

*The numbers in parenthesis includes the number of runs

4.2 Learning the system rules

The proposed learning- rule approach was applied to the training dataset in order to extract uncertain linguistic rules. The GA was applied for exhaustive search in various rule sets. In our GA uncertain rule learning approach, we specified the parameters as follow; the crossover rate was 0.25 and the mutation rate was 0.11 and $\mathrm{W_{NCC}}$ and $\mathrm{W_{NMC}}$ weights were defined 1, and 5, respectively as were suggested in [3] after applying several examination for various parameters. The GA was converged to the best fitness after a maximum of 100 generations. The learning rule approach generates the best set of rules according to the fitness function as well as the IT2FMs parameters. For the selected $[k_m, k_v]$ parameter the genetic algorithm for learning rules were applied. The 10-fold cross-validation method was applied to evaluate the performance of the uncertain rule-based pattern classification system. In order to have a robust and unbiased view of the performance of the classification system, the 10-fold cross validation was repeated 100 times and the average accuracy of 100 runs is considered as the average accuracy of the classification system. The proposed genetic type-2 pattern classification system was applied to the Wisconsin breast cancer diagnosis dataset. IT2FLS with two linguistic terms sets “Low” and “High” (IT2FLS) has been developed. Also, we considered “none” for the time that the feature is not involve in the rule.

The generated rules for the IT2FLS are shown in Table2. The genetic learning rule approach has been applied to the breast cancer dataset converged after 100 generations as shown in Figure 3. The $[k_m, k_v]$ parameters were selected as [0.8, 0.6] by the genetic algorithm. Three rules with only one variable (type-2 fuzzy set term set) are extracted by the algorithm, between them two rules are generated for classifying malignant samples as follows:

If (UCS is Low) then object is Malignant

Also there are two rules for classifying benign objects as follows:

If (BN is High) then object is Benign

If (UC is High) then object is Benign

This set of rule has the high interpretability and provides a classifier with the best accuracy (96.38%). The average accuracy of this system calculated from an average of 100 runs of 10 fold cross-validation on WBCD dataset which is 95.96% with a 95% confidence interval [95.87 96.05]. In this 100 runs, the average of all of the 10 fold cross-validation runs are more than 95%. The generated IT2 Gaussian membership functions of IT2FLS for nine features of the breast cancer dataset are shown in Figure 4.

Table 2. Extracted Rules for Type-2 Fuzzy Classifier for WBCD dataset

No	Antecedent	Class	Rule accuracy (%)	NCP	NMP
R1	[0 0 1 0 0 0 0 0 0]	0	99.97	430	13
R2	[0 0 0 0 0 3 0 0 0]	1	99.78	157	7
R3	[0 3 0 0 0 0 0 0 0]	1	99.95	69	7
Sum =				656	27

*0, 1, and 2 are equivalent to none, Low, and High linguistic terms

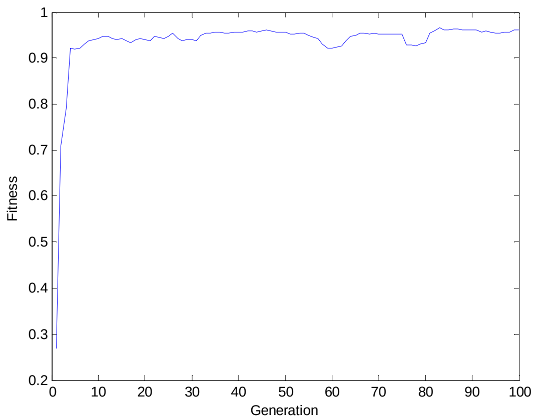

Fig. 3. Genetic learning of the uncertain rules for IT2FLS

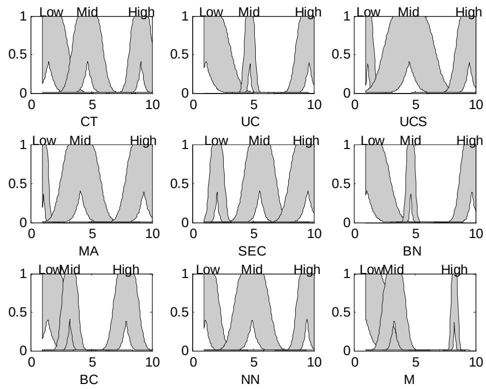

Fig. 4. Generated IT2 Gaussian membership functions of IT2FLS for nine features of the breast cancer dataset

4.3 Performance Comparison with Other fuzzy based rule extraction methods

Several methods have been applied for the Wisconsin breast cancer diagnostic problem. Here we present an overview of recent methods and compare the accuracy of their results. The breast cancer diagnosis dataset is one of the popular datasets for evaluation of the performance of classification systems. Various techniques have been reported in the literature to classify breast cancer dataset such as naive Bayesian, neural network, support vector machine, fuzzy approaches and a combination of learning and data mining approaches with these methods, such as neuro fuzzy, fuzzy decision tree and genetic fuzzy approaches. A review of neural network classification methods can be found in [20]-[21]. A survey of improved naive Bayes methods for classification is presented in [22]. The rest

of this section presents a review of various rule extraction methods in the rule-based classifiers applied on the WBCD and their performance comparison.
A rule extraction method from a pruned neural network is presented by Setiono for breast cancer diagnosis in [28]. In this method network pruning is used to remove the redundant connections and decrease the complexity of the network [28]. Their rules achieved more than 95% accuracy on training and testing dataset [28]. In another work [29], the samples with missing values were removed and feature extraction was applied as a pre-processing method on the dataset to improve the accuracy of the classifier. Taha and Ghosh presented an approach for extracting rules from a trained neural network [30]. In their method the final decision is made with a confidence measure. They introduced three methods for extracting the rules: first, a binarized Input-Output Rule Extraction (BIO-RE) which extracts binary rules from a trained neural network from binary inputs [30]. Second, a Partial Rule Extraction (Partial-RE) searches for the incoming link that activates a hidden or output node and calculates a measure of belief. This method has a lower computational cost and a lower number of premises per rule [30]. The third method is a Full-RE which generates intermediate rules by considering the effect of each of the inputs on the consequent [30] using linear programming and an input discretization method [30]. All of the rule extraction methods based on the trained neural network need a default rule to classify new samples and their accuracy drops significantly without the default rule [30]. Recently, support vector machines (SVM) have been applied for the classification of the WBCD. Mehmet presented a SVM technique combined with a feature selection method [31]. This method provides a maximum 98.51% ROC accuracy [31]. An overview of the SVM rule extraction methods are provided in [32]. Martensa et al. added comprehensibility to SVM by extracting symbolic rules from the trained model in order to makes it appropriate for medical diagnosis. This method provides an accuracy of 96.6% [32].

Table 3. Comparison of the Best Accuracy Results of Rule Extraction Approaches for WBCD classification; the accuracy of the proposed method is comparable to other methods in terms of the tradeoffs between accuracy and interpretability

Method	**Accuracy (%)**	**Rules**
Neuro Rules [29]	98.24	5
SVM Rules [31]	96.6	5
T1 Fuzzy Rules [23]	97.8	5
T2 Fuzzy Rules [this work]	96.38	3

As shown in Table 3, the accuracy of the proposed method is comparable to the best reported accuracy of the counterpart methods although the main goal of the proposed approach is to maintain the tradeoffs between accuracy and interpretability. The main advantages of our type-2 fuzzy genetic approach for learning rules in comparison to the other approaches are:

1. It models uncertainties of the noisy and imperfect training dataset in the antecedent part of the rules
2. It is applicable for classification problem in vague environment such as medical imaging
3. There is no need for expert knowledge for initialization of the genetic algorithm and the fuzzy system parameters
4. The IT2FLS rule set has higher interpretability than the rule set of the other method with only one linguistic term per rule
5. The reported accuracy result is more robust as it is based on the average of 100 runs of 10-fold cross-validation on the dataset
6. It is more reliable; results in a smaller confidence interval.
7. It defines the degree of certainty of the rules as well as the confidence interval, which leads to more reliability in the system's decisions
8. It converges to a maximum accuracy very rapidly

5 CONCLUSION

This paper has proposed a genetic fuzzy approach for rule extraction for classification from uncertain and imprecise datasets such as medical diagnostic dataset. The main advantages of the proposed rule extraction method in this paper is higher interpretability of the rule set than the other methods, robust and reliable accuracy result using 100 runs of 10-fold cross-validation and learning the degree of certainty of the rules.

To evaluate our approach we applied it to the Wisconsin breast cancer diagnostic dataset. The average accuracy of our classifier, using three rules with only one variable per rule, after applying 10-fold cross-validation is 95.96% with a 95% confidence interval [95.87 96.05], which is comparable with the average result of the previous methods. In addition, our approach maintains the best trade-off between accuracy and interpretability. The proposed method has the capability to manage more uncertainties in the rules and is applicable for pattern classification problems which exhibit an expensive or a lack of expert knowledge with only an imprecise and noisy training dataset.

Our future work is to apply our approach to other pattern classification problems in medical image analysis systems and improve the overall system accuracy.

REFERENCES

[1] J. M. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions,* Prentice-Hall, 2001.

[2] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms: with applications to image processing and pattern recognition,* Advances in Fuzzy Systems–Applications and Theory, World Scientist, vol. 10, 1996.

[3] H. Ishibuchi, T. Nakashima and T, Murata, "A fuzzy classifier system that generates linguistic rules for pattern classification problems", *In Fuzzy Logic, Neural Networks, and Evolutionary Computation Lecture Notes in Computer Science*, vol. 1152, pp. 35-54, 1996.

[4] H. Ishibuchi, Y. Nojima, “Pattern Classification with Linguistic Rules”, *In Fuzzy Sets and Their Extension: Representation, Aggregations and Models, Studies in Fuzziness and Soft Computing*, vol. 220, pp. 377-395, Springer, 2008.

[5] L. A. Zadeh, The Concept of A Linguistic Variable and Its Application to Approximate Reasoning – 1, *Information Sciences J.* , vol. 8, pp. 199- 249, 1975.

[6] N. N. Karnik, J. M. Mendel, Operations on type-2 fuzzy sets, *J. of Fuzzy Sets and Systems*, vol. 122, pp.327–348, 2001.

[7] N. N. Karnik, J. M. Mendel. and Qilian Liang, Type-2 Fuzzy Logic Systems, *IEEE trans. on fuzzy systems*, vol. 7, no. 6, pp. 643–658, 1999.

[8] R. Hosseini, J. Dehmeshki, S. Barman, M. Mazinani, S. Qanadli, "A Genetic Type-2 Fuzzy Logic System for Pattern Recognition in Computer Aided Detection Systems," In proc. IEEE International Conference on Fuzzy Systems (FUZZ), no. 5584773, pp.1–7, Barcelona , July 2010.

[9] D. Hidalgo, O. Castillo, and P. Melin, “Interval type-2 fuzzy inference systems as integration methods in modular neural networks for multimodal biometry and its optimization with genetic algorithms”, *Int. Journal of Biometrics*, vol. 1, no. 1, pp. 114 – 128, 2008.

[10] H. B. Mitchell, “Pattern recognition using type-II fuzzy sets”, *Information Sciences*, vol. 170, pp. 409–418, 2005.

[11] J. R. Castro, O. Castillo, P. Melin, L. G. Martínez, S. Escobar, and I. Camacho, “Building Fuzzy Inference Systems with the Interval Type-2 Fuzzy Logic Toolbox,*” in Analysis and Design of Intelligent Systems using Soft Computing Techniques*, Springer Berlin, 2007, vol. 41, pp. 53–62.

[12] D. Wua, and W. Wan Tan, Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers, *Engineering Applications of Artificial Intelligence J.*, vol. 9, pp. 829–841, Dec. 2006.

[13] A. I. Martinez, O. Castillo, and M. Garcia, “Comparative Study of Type-1 and Type-2 Fuzzy Systems Optimized by Hierarchical Genetic Algorithms,” *in Soft Computing for Hybrid Intelligent Systems*, Springer Berlin, 2008, vol. 154, pp. 53–70.

[14] N. R. Cazarez-Castro a, L. T. Aguilar, O. Castillo, Fuzzy logic control with genetic membership function parameters optimization for the output regulation of a servomechanism with nonlinear backlash, *Expert Systems with Applications J.*, vol. 37, pp. 4368–4378, 2010.

[15] J. Zenga, L. Xieb, and Zhi-Qiang Liuc, Type-2 fuzzy Gaussian mixture models, *Pattern Recognition J.*, vol. 41, no. 12, pp. 3636–3643, 2008.

[16] R. R. Picard, and R. D. Cook, Cross-Validation of Regression Models, *J. of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, 1984.

[17] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, *Evolutionary Intelligence*, pp. 27–46, Jan. 2008.

[18] J. Beutel, M. Sonka, H. L. Kundel, J. M. Fitzpatrick, and R. L. V. Metter, *Handbook of Medical Imaging: Medical image processing and analysis*, SPIE, 2000.

[19] C.L. Blake and C.J. Merz., UCI repository of machine learning databases, http://archive.ics.uci.edu/ml/, University of California, Irvine, Dept. of Information and Computer Sciences, 1998.

[20] H. A. Abbass, An evolutionary artificial neural networks approach for breast cancer diagnosis, *Artificial Intelligence in Medicine*, vol. 25, Issue 3, pp. 265, July 2002.

[21] D.B. Fogel, E.C. Wasson, and E.M. Boughton, Evolving neural networks for detecting breast cancer, *Cancer letters*, vol. 96, Issue 1, pp. 49–53, 1995.

[22] L. Jiang, D. Wang, Z. Cai1 and X. Yan, “Survey of Improving Naive Bayes for Classification”, In *Advanced Data Mining and Applications*, Springer Berlin -Heidelberg, vol. 4632, pp. 134-145, 2007.

[23] C. A. Pena-Reyes, M. Sipper, A fuzzy-genetic approach to breast cancer Diagnosis, Carlos Andres Pen a-Reyes, Moshe Sipper, *Artificial Intelligence in Medicine*, vol. 17, pp. 131–155, 1999.

[24] X. Chang, and J. H. Lilly, Evolutionary Design of a Fuzzy Classifier From Data, *IEEE Trans. on Sys., Man, and Cybernetics—Part B: Cybernetics*, vol. 34, no. 4, pp. 1894-1906, August 2004.

[25] H. Ishibuchi and T. Yamamoto, "Effects of three-objective genetic rule selection on the generalization ability of fuzzy rule-based systems," in Proc. of Int. Conf. Evolutionary Multi-Criterion Optimization, Portugal, vol. 2632, pp. 608–622, Apr 2003.

[26] J. S. Wang, and C. S. G. Lee, Self-adaptive neuro-fuzzy inference systems for classification applications, *IEEE Transactions on Fuzzy Systems*, vol. 10, Issue 6, pp. 790 - 802 , 2002.

[27] J. Abonyi, J. A. Roubos, and F. Szeifert, Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *International journal of approximate reasoning*, vol. 32, Issue 1, pp. 1-21, Jan 2003.

[28] R. Setiono, Extracting rules from pruned neural networks for breast cancer diagnosis, *Artificial intelligence in medicine J.*, vol. 8, Issue 1, pp. 37-51, Feb. 1996.

[29] R. Setiono, Generating concise and accurate classification rules for breast cancer diagnosis, *Artificial intelligence in medicine*, vol. 18, Issue 3, pp. 205-219, 2000.

[30] I. Taha, and J. Ghosh, Characterization of the Wisconsin Breast cancer Database Using a Hybrid Symbolic-Connectionist System, Tech. Rep. UT-CVIS-TR-97-007, Center for Vision and Image Sciences, University of Texas, Austin, 1997.

[31] M. F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, *Expert Systems with Applications*, vol. 36, Issue 2, part 2, pp. 3240–3247, March 2009.