Research Article

HYEI: A New Hybrid Evolutionary Imperialist Competitive Algorithm for Fuzzy Knowledge Discovery

D. Jalal Nouri, M. Saniee Abadeh, and F. Ghareh Mohammadi

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran 14115-143, Iran

Correspondence should be addressed to M. Saniee Abadeh; saniee@modares.ac.ir

In recent years, imperialist competitive algorithm (ICA), genetic algorithm (GA), and hybrid fuzzy classification systems have been successfully and effectively employed for classification tasks of data mining. Due to overcoming the gaps related to ineffectiveness of current algorithms for analysing high-dimension independent datasets, a new hybrid approach, named HYEI, is presented to discover generic rule-based systems in this paper. This proposed approach consists of three stages and combines an evolutionary-based fuzzy system with two ICA procedures to generate high-quality fuzzy-classification rules. Initially, the best feature subset is selected by using the embedded ICA feature selection, and then these features are used to generate basic fuzzy-classification rules. Finally, all rules are optimized by using an ICA algorithm to reduce their length or to eliminate some of them. The performance of HYEI has been evaluated by using several benchmark datasets from the UCI machine learning repository. The classification accuracy attained by the proposed algorithm has the highest classification accuracy in 6 out of the 7 dataset problems and is comparative to the classification accuracy of the 5 other test problems, as compared to the best results previously published.

1. Introduction

In general, fuzzy logic is closer to human logic; thus it can deal with real-world noise and imprecise data [1, 2]. Fuzzy models have several advantages. The most important advantage is that they have flexible decision boundaries, and thus, they are characterized by their higher ability to adjust to a specific domain of application and accurately reflect its particularities. A fuzzy model can be generated by describing a fuzzy-classification rule set and then improved by decreasing the length and the number of rules. This approach is a complex task, since several issues must be resolved for the fuzzy model which are produced. First, the basic fuzzy-classification rules must be well defined. In the optimization stage, the length of rules is reduced and when length of one rule equals zero then the mentioned rule will be removed from the rule set. The resulted rule set will be more interpretable since it has fewer number of rules.

Several approaches have been suggested in the literature for the improvement of knowledge-based fuzzy models. In most of them, the model is trained using a recognized optimization technique (i.e., fuzzy rules with genetic algorithms [3]).

In this work, we propose an algorithm for generating high-quality fuzzy-classification rules which includes three stages: (i) embedded feature selection, (ii) basic rule set creation, and (iii) rule set optimization. In the first stage, based on the initial rule set it selects the best features by using the embedded feature selection; for embedding feature selection, an evolutionary algorithm and imperialist competitive algorithm for this stage are used. In the next stage, it generates fuzzy-classification rules by training data and initial rule set that are used in the previous stage. In order to rule set creation, we use an evolutionary algorithm like genetics algorithm for this stage. Finally (in the third stage), optimization for reducing the length and the number of rules and generating final rule sets is performed. We employ imperialist competitive algorithm for this stage.

HYEI is a hybrid version of Michigan and Pittsburgh approaches. Each fuzzy rule is represented by its antecedent

fuzzy sets as an integer string of fixed length. Each fuzzy rule-based classifier, which is a set of fuzzy rules, is represented as a concatenated integer string of variable length. HYEI algorithm simultaneously enhances the accuracy of rule sets and decreases their complexity. The accuracy is measured by the number of correctly classified training patterns while the complexity is measured by the number of fuzzy rules and/or the total number of antecedent conditions of the fuzzy rules. In this paper, seven benchmark classification datasets are applied to the evaluation and the reported results indicate high classification accuracy.

The rest of this paper is organized as follows. In Section 2, we describe the related work in relevant issues. We propose the HYEI in Section 3. Section 4 shows the experimental results. After that, we have discussion in Section 5. Finally, we conclude this paper in Section 6.

2. Related Works

The authors in [4] described two series of empirical tests and explored the properties of commonly used datasets. The first tested contrast of two versions of the ideal partitioning approach with a variant of the greedy multisplitting algorithm [5, 6] and a random partitioning approach. In that experiment, they used a single version of the approaches. The second series of tests examined the influence that numerical attribute handling has on decision tree learning. They incorporated the ideal and the greedy multisplitting strategies into the C4.5 (release 5) decision tree learner and 94.9% accuracy for breast W dataset, 75.0% for diabetes, 72.7% for glass, 53.7% for heart C, 75.4% for sonar, and 94.4% for wine from their algorithm was described.

The authors in [7] proposed a generic algorithm for the automatic generation of the fuzzy models. The algorithm was realized in three stages. Initially, a crisp model was produced. In the second stage, it was converted to a fuzzy one. In the third stage, all parameters entering the fuzzy model were enhanced. The proposed algorithm was novel and generic since it could integrate different techniques in each of its stages. A specific realization of this algorithm was employed, using decision trees for the creation of the crisp model, the sigmoid function, the min-max operators and the maximum defuzzifier, for the transformation of the crisp model into a fuzzy one, and four different optimization strategies, including global and local optimization algorithms and hybrid algorithms. The proposed algorithm offered several advantages and novelties: the transformation of the crisp model to the respective fuzzy one was straightforward ensuring its full automatic nature and it presented a set of parameters, expressing the importance of each fuzzy rule. 95.71% accuracy for breast W dataset, 77.34% for diabetes, 70.39% for glass, 76.27% for heart C, 75% for sonar, and 93.27% for wine from their algorithm was described.

In [8] an algorithm was presented that was used for complex algorithms like neural networks or ensembles normally resulting in very accurate models. As an alternative many researchers have tried reducing this accuracy versus comprehensibility tradeoff by converting the high accuracy, opaque model into a transparent model, a technique termed rule extraction. Successful rule extraction requires the extracted model to be not only transparent but also understandable. This was normally transformed into that the extracted model must be fairly small, thus enabling human inspection and understanding. In [8] an extension of the rule extraction algorithm G-REX was proposed and evaluated. Their algorithm succeeded in increasing clarity by shortening extracted rules.

The authors in [9] examined the interpretability-accuracy tradeoff in fuzzy rule-based classifiers using a multiobjective fuzzy genetics-based machine learning (GBML) algorithm. The proposed GBML algorithm is another hybrid version of Michigan and Pittsburgh approaches, which was implemented in the framework of evolutionary multiobjective optimization (EMO). Each fuzzy rule was represented by its antecedent fuzzy sets as an integer string of fixed length. Each fuzzy rule-based classifier, which was a set of fuzzy rules, was signified as a concatenated integer string of variable length. The suggested process simultaneously increases the accuracy of rule sets and decreases their complexity. The accuracy was measured by the number of correctly classified training patterns while the complexity was measured by the number of fuzzy rules and/or the total number of antecedent conditions of fuzzy rules. They studied the interpretability-accuracy tradeoff for training patterns through computational experiments on some benchmark datasets. A clear tradeoff construction was imagined for each dataset. They also studied the interpretability-accuracy tradeoff to test patterns. Due to the overfitting to training patterns, a clear tradeoff structure was not always obtained in computational experiments to test patterns. The proposed technique gained 97.34% accuracy for breast W dataset, 78.2% for diabetes, 66.07% for glass, 57.43% for heart C, 82.68% for sonar, and 96.96% for wine.

Computational analysis and computing could help researchers gather a group of signature genes for a certain disease [10, 11]. Since the price of microarray chips is very high and usually there is not enough tissue samples available from cancer patients, the number of records in microarray datasets is usually too high which is not suitable for most machine learning algorithms. In addition, the processing and material used for microarray analysis are typically different between manufacturers, so it is difficult to identify a unique set of genes that can form an integrated dataset [12]. Moreover, when the number of samples for each cancer type is usually balanced for computer analysis. However, the ratio of cancer patients to normal adults is mostly much smaller in real world.

Because the number of genes is always much greater than the number of tissue samples [13, 14], so one of the important challenges would be choosing a small and a discriminative subset of effective genes among tens of thousands of genes which is a very challenging task; as a result, gene selection becomes the most necessary requirement for a microarray-based cancer classification system. However, the best mixture of classification and gene selection is understood poorly because there was another methodological trouble associated with training microarray data. This was the problem of "overfitting" [15]. Briefly, overfitting means that one can

get good accuracy using a training set, when novel data is used, but a satisfactory result cannot be obtained using the trained model. This occurs often when a small number of high-dimensional samples are applied. Unluckily, there exists exactly such a problem in cancer tumor discovery using microarray datasets [16].

The authors in [17] proposed a new memetic approach which was capable of extracting interpretable and accurate fuzzy if-then rules from cancer data. The suggested algorithm is the first suggestion of memetic algorithms with multiview fitness function approach. The new presented multiview fitness function reflects two kinds of evaluating procedures. The former procedure, which is located in the key evolutionary structure of the algorithm, evaluates each single fuzzy if-then rule according to the specified rule quality (the evaluating procedure does not consider other rules). However, the latter procedure controls the quality of each fuzzy rule according to the whole fuzzy rule set performance. The proposed algorithm obtained 69.43% accuracy for 14_Tumors dataset.

Improved binary particle swarm optimization (IBPSO) was used in [18] to implement feature selection, and the K-nearest neighbor (K-NN) method serves as an evaluator of the IBPSO for gene expression data classification problems. Experimental results showed this algorithm effectively simplifies feature selection and reduces the total number of features wanted. The classification accuracy gained by the proposed algorithm was 66.56% for 14_Tumors dataset.

3. Proposed Algorithm

High-dimension datasets are the most important challenging issues for classification and feature selection algorithms in last decade. Because of that, the runtime duration is increased; it takes so much time and makes the prediction accuracy be decreased and also a powerful computer (with at most RAM and speedy cpu) is needed to implement. This section presents the HYEI algorithm, which solves the problem which we explain in detail in two subsections. Section 3.1 briefly explains fuzzy if-then rules with continuous attributes. A heuristic process is also termed to determine the consequent class and the certainty grade of each fuzzy if-then rule from training patterns. This heuristic process is an adapted version of the one which has been firstly introduced in [19]. The main process of HYEI is presented in Section 3.2.

3.1. Fuzzy Rule Base for Pattern Classification. Let us assume that our pattern classification issue is a c-class problem in the n-dimensional pattern space with indiscrete attributes. We also assume that M real vectors $X_p = (X_{p1}, X_{p2}, \ldots, X_{pn})$, $p = 1, 2, \ldots, M$, are given as training patterns. Whole values of pattern attributes are normalized in $[0, 1]$ interval. In the provided fuzzy classifier system, we use fuzzy if-then rules of the following form [20]:

Rule R_j: If x_1 is A_{j1} and x_2 is A_{j2} and ... and x_n is A_{jn},

then Class C_j with $CF = CF_j$,

$$(1)$$

where R_j stands for the label of fuzzy if-then rule, $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets in unique interval, C_j stands for consequent class, and CF_j is the grade of certainty of the fuzzy rule. In computer simulations, we use a typical set of linguistic values in Algorithm 1 as antecedent fuzzy sets.

The whole number of if-then rules, in a classification problem with n features, is 6^n. It is impossible to use all of these rules in a single fuzzy rule base when the number of attributes (i.e., n) is large (for cancer tumor diagnosis in this paper, $n = 15009$). Our fuzzy learning system searches for a few numbers of if-then rules, which are capable of identifying the cancer class with high classification accuracy. Since the result class and certainty grade of each if-then rule can be calculated according to the training samples and a simple procedure, our learning classifier system is responsible for finding the best combinations of antecedent fuzzy sets [21]. This goal maybe appeared easy in the first stage. However, solving the classification problems with high dimensions is very challenging and tough, specifically for our problem that has a huge search space with 6^n states.

In our fuzzy-classification system, the result C_j and certainty grade of each rule (CF_j) will be computed from [19].

3.2. HYEI Algorithm. In this paper, a hybrid approach based on imperialist competitive algorithm and genetic algorithm is employed to extract a rule set, named HYEI; HYEI consists of three main stages as in Figure 1: (i) embedded feature selection, (ii) basic rule set creation, and (iii) rule set optimization. An outline of the HYEI algorithm is presented in Figure 2.

3.2.1. Initialization. Suppose that $N_{\text{population}}$ is the number of fuzzy rules in the initial population. In this paper to create initial rule sets, the numbers of $N_{\text{population}}$ if-then rules will be produced randomly based on existing samples in the dataset and initialize the antecedent of rules by linguistic values in Algorithm 1. Here we set the do not care probability to zero, so each rule would have initially the maximum length. Fuzzy rules display with an array a number of attributes as in Figure 3. In this array number used for linguistic values. Number 1 means small, number 2 means medium small, number 3 means medium, number 4 means medium large, number 5 means large, and number 0 means do not care.

3.2.2. Feature Selection Condition Test. There are different approaches to decide needs to use the feature selection process. One of the approaches to use the feature selection process of the algorithm is considered a prespecified number of attributes. Another approach is the use of a training algorithm to estimate the best number of attributes. In our proposed algorithm, we have considered the first approach.

3.2.3. Embedded Feature Selection. The first problem that exists in some of datasets such as DNA microarrays is its high dimensionality; therefore, to reduce the dimensions of this problem, a preanalysis process stage for using a feature selection approach will be running. There are different approaches to feature selection process such as wrappers,

> (1) Small: $M = -4 \times A + 1$
> (2) Medium Small: $M = 4 \times A$ and $M = -4 \times A + 2$
> (3) Medium: $M = 4 \times A - 1$ and $M = -4 \times A + 3$
> (4) Medium Large: $M = 4 \times A - 2$ and $M = -4 \times A + 4$
> (5) Large: $M = 4 \times A - 3$
>
> M: Membership function; A: Attribute value

ALGORITHM 1: The antecedent fuzzy set equations used in this paper.
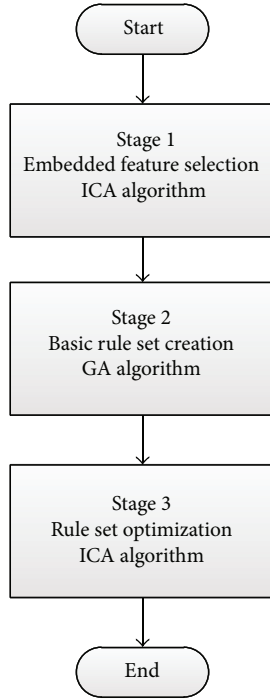

FIGURE 1: Three main stages in HYEI.

filters, and embedded algorithms. In our proposed algorithm, we have considered proposed embedded feature selection algorithms in preanalysis process stage.

In this stage, to feature selection, an array is needed that contains a set of best attribute's index for applying to all fuzzy rules in the initial population. Also an imperialist competitive algorithm is needed to find the best attribute's index. A standard imperialist competitive algorithm is employed and customized initialization, assimilation, and power function.

(1) Initialization Function. The goal of feature selection is to find an array that contains a subset of best attribute's index. This array is called "country" in ICA. In an N_{var}-dimensional feature selection problem, a country is a $1 \times N_{\text{var}}$ array. This array is defined as follows:

$$\text{Country} = [p1, p1, p1, \ldots, p_{\text{Nvar}}], \tag{2}$$

where $P_i s$ are the variables to be optimized. The variable values in the country are represented as integer numbers.
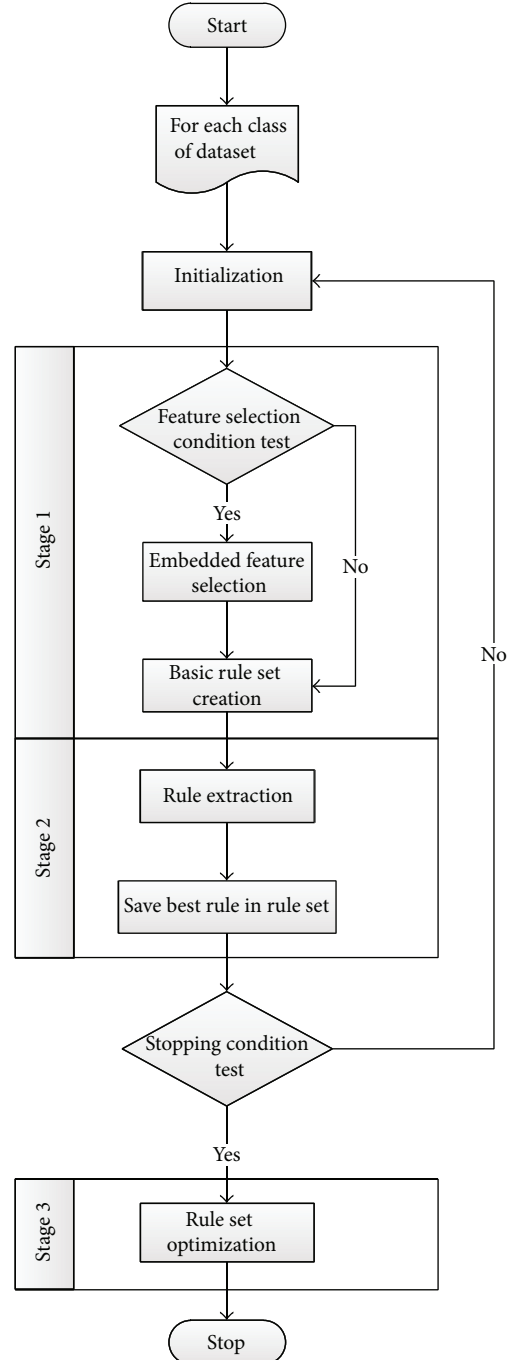

FIGURE 2: Proposed algorithm flowchart.

| 0 | 1 | 2 | 0 | 3 | 4 | ⋯ | 1 |

FIGURE 3: A fuzzy rule with an array.

| 1 | 15 | 203 | 385 | 409 | 1006 |

FIGURE 4: An array containing a subset of attributes index.

Each variable in the country can be interpreted as an index of an attribute. In our proposed algorithm, the length of each array is a prespecified number used in the second stage of HYEI algorithm.

In initialization function this array as shown in Figure 4 is produced randomly based on a nonrepetitive attribute's index. In other words, each number in each array must introduce an attribute's index and each array should not contain repetitive numbers.

(2) Assimilation Function. In assimilation function, the imperialist states try to absorb their colonies and make them a part of themselves. More precisely, the imperialist states make their colonies move toward them. In imperialist competitive algorithm this process is modelled by moving all of the colonies toward the imperialist. In our proposed algorithm, this process is modelled by substituting random attribute's index from imperialist array for colonies array. For example, in Figure 5 randomly selects the bold numbers in imperialist and replaces them in colony with assimilation function and produces changed colony.

(3) Power Function. The power of each country, the counterpart of fitness value in the GA, is inversely proportional to its cost. In power function, calculate a number for each country containing imperialists and colonies. In our proposed algorithm, this function is modelled by changing initial population that was generated in the first stage of the HYEI algorithm with select attributes that exist in country array and calculating the classification rate of new population using

$$\text{Classification Rate} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FN} + \text{FP})}, \quad (3)$$

where TP, true positives, stands for the number of cases in this training set covered by the rule that have the class predicted by the rule; FP, false positives, stands for the number of cases covered by the rule that have a class different from the class predicted by the rule; FN, false negatives, stands for the number of cases that are not covered by the rule but that have the class predicted by the rule; TN, true negatives, stands for the number of cases that are not covered by the rule and that do not have the class predicted by the rule. Figure 6 displays the approach which selects attributes in the initial population with a country array. In changed population, other attributes that do not exist in this array mean do not care.

3.2.4. Basic Rule Set Creation. When the best attribute's index was found in the previous stage, in this stage those attributes index arrays are applied to initial population, and the basic

Imperialist | 1 | **15** | 203 | **385** | 409 | 1006 |

Colony | 8 | 44 | 158 | 290 | 899 | 1250 |
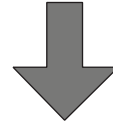
Assimilation

Changed colony | 8 | 15 | 158 | 385 | 899 | 1250 |

FIGURE 5: Assimilation function.

rule sets are generated which are shown in Figure 6. If the result of the feature selection condition test is no then embedded feature selection stage does not run them in this stage; initial population will be not changed and use it as basic rule set.

3.2.5. Rule Extraction. This stage is an evolutionary fuzzy system that learns fuzzy if-then rules in an incremental fashion, in that the evolutionary algorithm optimizes one fuzzy classifier rule at a time. The learning mechanism declines the weight of those training instances that are correctly classified by the new rule; therefore, the next rule generation cycle focuses on fuzzy rules that account for the current uncovered or misclassified instances. At each iteration the fuzzy rule that can classify the recent distribution of training samples better than other rules of the population is selected to be included in the final classification fuzzy rule base. This evolutionary fuzzy system used a basic rule set that is generated in the previous stage as initial population. In our proposed algorithm, genetic algorithm is applied for this stage.

3.2.6. Save the Best Rules in Rule Sets. During this stage, the best rules which are the results of several generations will be added to the final learned rule sets. However, this operation would be possible only if the new rule does not exist in the final learned rule set.

3.2.7. Stopping Condition Test. Since the most of the operations in this algorithm deal with random parameters, the algorithm repeats the previous stages until all samples of every class are covered by the final learned rule set.

3.2.8. Rule Set Optimization. During this stage an imperialist competitor algorithm is employed to discover a do not care mask for rules. We employ a standard imperialist competitive algorithm and customize initialization, assimilation, and power function.

(1) Initialization Function. In initialization function, arrays with a length of fuzzy rule arrays and a multiple number of rules in the rule set are created and fill it randomly with zero
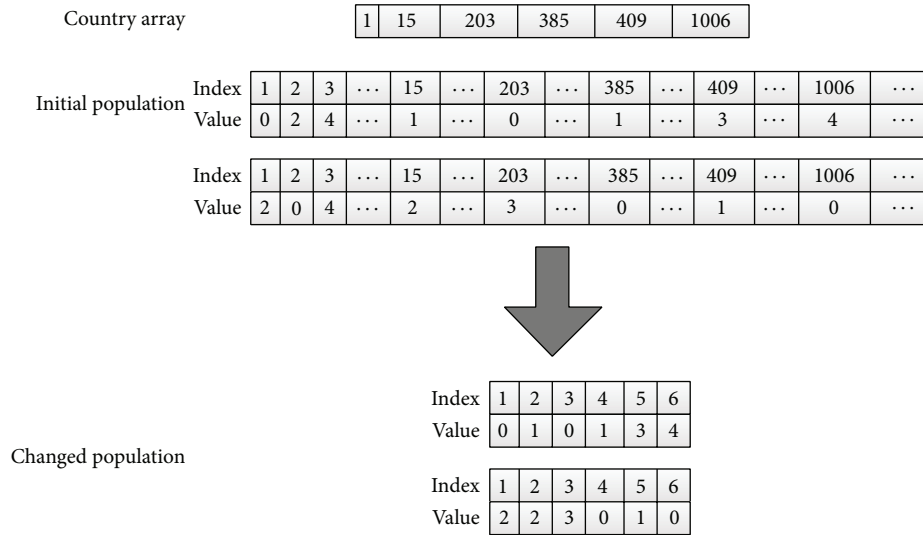
Country array

1	15	203	385	409	1006

Initial population

Index	1	2	3	⋯	15	⋯	203	⋯	385	⋯	409	⋯	1006	⋯
Value	0	2	4	⋯	1	⋯	0	⋯	1	⋯	3	⋯	4	⋯

Index	1	2	3	⋯	15	⋯	203	⋯	385	⋯	409	⋯	1006	⋯
Value	2	0	4	⋯	2	⋯	3	⋯	0	⋯	1	⋯	0	⋯

Changed population

Index	1	2	3	4	5	6
Value	0	1	0	1	3	4

Index	1	2	3	4	5	6
Value	2	2	3	0	1	0

Figure 6: Initial population is changed in power function.

0	1	0	0	1	0	0

Figure 7: A sample array in the first stage.

and one which is shown in Figure 7 in which the number 0 means old value and number 1 means do not care.

(2) Assimilation Function. In our proposed algorithm, this process is modelled by replacing random attribute's index from imperialist array to colonies array with the same embedded feature selection assimilation function.

(3) Power Function. In our proposed algorithm, this function is modelled by changing the rule set that is generated in "save best rule in rule set" with converting attribute's value that has number "1" value in country array in "rule set optimization" to apply do not care and calculate the classification rate of rule set.

4. Experimental Results

We have applied seven datasets with many numerical attributes: 14_Tumors, Wisconsin breast cancer, Cleveland heart disease, glass, wine, diabetes, and sonar to evaluate the performance of the HYEI classification system. These datasets are available from UCI. Number of classes, the number of patterns, and the number of attributes in each dataset are given in Table 1. Figure 8 shows number of patterns in each dataset. Some datasets include incomplete patterns with missing values. Those patterns are not used in our computational experiments. This is because the performance of classification algorithms usually depends on the choice of a handling algorithm of missing values.

The main parameters of the HYEI algorithm are listed in Table 2. As it shows, according to the proposed algorithm, we
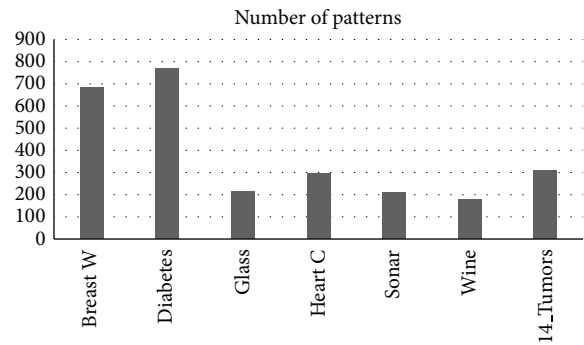

Figure 8: Number of attributes in each dataset.

Table 1: Datasets are used in experimental results.

Dataset	Number of attributes (n)	Number of classes (M)
Breast W	9	2
Diabetes	8	2
Glass	9	6
Heart C	13	5
Sonar	60	2
Wine	13	3
14_Tumors	15009	26

set the optimal value to every parameter which exists in every algorithm separately and properly.

Figure 9 shows classification rate of HYEI versus several famous classification algorithms.

We also analyze the results of proposed algorithm in Figure 9 which shows the performance of using different feature sets in the algorithm. These results show that this algorithm obtains the highest classification accuracy. As mentioned, we try to explain the better result of proposed method in detail per datasets. In case of Breast W, as shown in Figure 9 we can analyze that the proposed accuracy is the
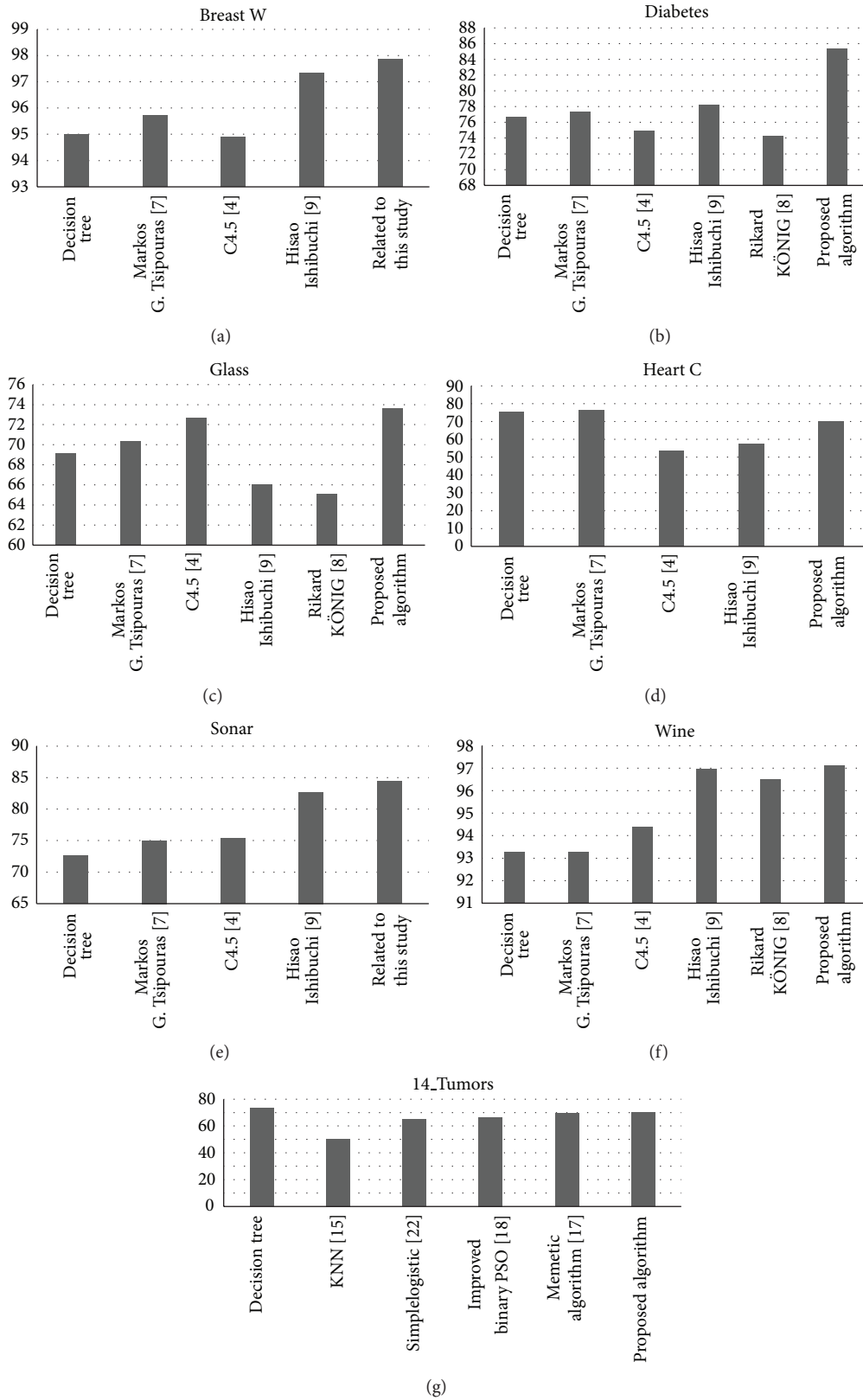
FIGURE 9: Classification rate of HYEI versus several famous classification algorithms.

Table 2: Parameter specification in computer simulations.

Stage	Parameter	Value
1 ICA	Number of initial countries	30
	Number of initial imperialists	3
	Revolution rate	0.1
2 GA	Population size	100
	Crossover probability	0.95
	Mutation probability	0.1
3 ICA	Number of initial countries	20
	Number of initial imperialists	2
	Revolution rate	0.1

highest and near 98% and also is more than the approach [9] which is the second highest accuracy. The next dataset belongs to diabetes; the figure shows that the algorithm [9] is about 78% but our algorithm's accuracy is about 85% and also in this dataset the accuracy is the highest one.

According to Figure 9, we can conclude that the proposed algorithm is done properly in most of the cases. But in case of tumor as it shows, just a little bit is less than the decision tree.

5. Discussion

The proposed algorithm has some advantages and disadvantages. The proposed algorithm is able to apply and use various datasets with alternative dimension. In other words, HYEI is able to apply to high-dimension dataset, and this is the important characteristic of HYEI. It means that HYEI works independently and does not depend on any specific dataset. However it computes properly and the results show that HYEI outperforms other relevant algorithms. On the other hand, if we focused on specific dataset we would enhance the performance of proposed algorithm. On the whole, novelty of our algorithm is to apply to every multidimension dataset and also to high dimension properly.

6. Conclusion

In this paper, an HYEI algorithm based on an imperialist competitive algorithm and genetic algorithm that can classify small datasets and large datasets like gene expression data with an accurate set of fuzzy if-then rules are presented. It begins with low quality fuzzy if-then rules and results in the high-quality rule set. This paper proposes an algorithm that is evaluated on 14_Tumors, Wisconsin breast cancer, Cleveland heart disease, glass, wine, diabetes, and sonar datasets and compared with other classification systems. Results indicate that this proposed algorithm outperforms several well-known and recent classification algorithms.
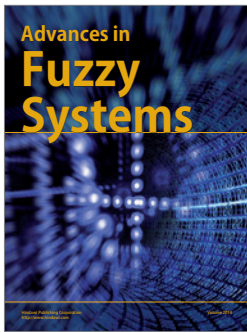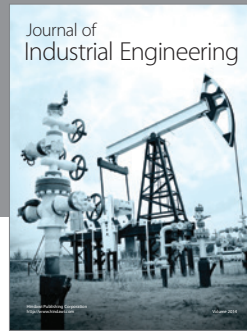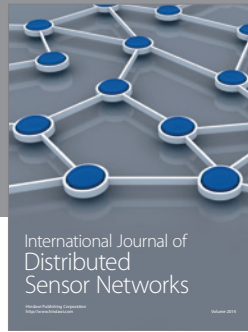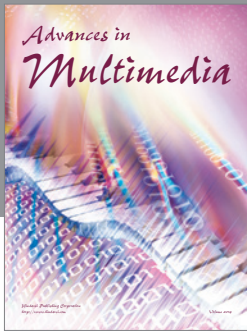
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*, John Wiley & Sons, New York, NY, USA, 1996.

[2] L.-X. Wang, *A Course in Fuzzy Systems*, Prentice-Hall Press, New York, NY, USA, 1999.

[3] W. Pedrycz and J. V. de Oliveira, "An algorithmic framework for development and optimization of fuzzy models," *Fuzzy Sets and Systems*, vol. 80, no. 1, pp. 37–55, 1996.

[4] T. Elomaa and J. Rousu, "General and efficient multisplitting of numerical attributes," *Machine Learning*, vol. 36, no. 3, pp. 201–244, 1999.

[5] J. Catlett, "On changing continuous attributes into ordered discrete attributes," in *Machine Learning-EWSL-91*, pp. 164–178, 1991.

[6] U. Fayyad and K. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," in *Proceedings of the International Joint Conference on Uncertainty in AI*, pp. 1022–1027, 1993.

[7] M. G. Tsipouras, T. P. Exarchos, and D. I. Fotiadis, "A methodology for automated fuzzy model generation," *Fuzzy Sets and Systems*, vol. 159, no. 23, pp. 3201–3220, 2008.

[8] R. Konig, U. Johansson, and L. Niklasson, "Increasing rule extraction comprehensibility," *International Journal of Information Technology and Intelligent Computing*, vol. 1, no. 2, pp. 303–314, 2006.

[9] H. Ishibuchi and Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning," *International Journal of Approximate Reasoning*, vol. 44, no. 1, pp. 4–31, 2007.

[10] Y. Wang, F. S. Makedon, J. C. Ford, and J. Pearlman, "HykGene: a hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data," *Bioinformatics*, vol. 21, no. 8, pp. 1530–1537, 2005.

[11] L. J. Buturović, "PCP: a program for supervised classification of gene expression profiles," *Bioinformatics*, vol. 22, no. 2, pp. 245–247, 2006.

[12] L. Ein-Dor, I. Kela, G. Getz, D. Givol, and E. Domany, "Outcome signature genes in breast cancer: is there a unique set?" *Bioinformatics*, vol. 21, no. 2, pp. 171–178, 2005.

[13] K. Y. Yeung, R. E. Bumgarner, and A. E. Raftery, "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2394–2402, 2005.

[14] H. Yu, J. Ni, and J. Zhao, "ACOSampling: an ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data," *Neurocomputing*, vol. 101, pp. 309–318, 2013.

[15] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.

[16] T. Prasartvit, A. Banharnsakun, B. Kaewkamnerdpong, and T. Achalakul, "Reducing bioinformatics data dimension with ABC-kNN," *Neurocomputing*, vol. 116, pp. 367–381, 2012.

[17] A. Zibakhsh and M. S. Abadeh, "Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1274–1281, 2013.

[18] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Computational Biology and Chemistry*, vol. 32, no. 1, pp. 29–37, 2008.

[19] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and Systems*, vol. 52, no. 1, pp. 21–32, 1992.

[20] D. Meng and Z. Pei, "Extracting linguistic rules from data sets using fuzzy logic and genetic algorithms," *Neurocomputing*, vol. 78, no. 1, pp. 48–54, 2012.

[21] I. Gadaras and L. Mikhailov, "An interpretable fuzzy rule-based classification methodology for medical diagnosis," *Artificial Intelligence in Medicine*, vol. 47, no. 1, pp. 25–41, 2009.