Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011

A Rough-Fuzzy Hybrid Approach on a Neuro-Fuzzy Classifier for High Dimensional Data

Chang Su Lee

***Abstract* — A new Rough-Neuro-Fuzzy (RNF) classifier is proposed in this paper for pattern classification scheme on high dimensional data as an extension of the previous work. The rough set theory is utilized to reduce the given knowledge into a compact form and to obtain a minimal set of decision rules. The proposed Rough-Neuro-Fuzzy classifier is constructed based on the structure of ANFIS (Adaptive-Network-Based Fuzzy Inference System), except its connections determined by the reduced data and the generated decision rules obtained by the rough sets-based approach. This provides the compact and minimal number of configurations for the network to adjust itself towards a faster learning. The learning scheme for the proposed approach is adopted from the one in ANFIS. The TS-type fuzzy inference model is employed to perform the decision making process. The proposed system is applied on a number of data sets for pattern classification tasks using 10-fold cross validation. The number of attributes is reduced significantly and the minimal rules are generated effectively by the rough set-based approach in the proposed system. Experimental results showed that results produced by the proposed rough-neuro-fuzzy classifier may be competitive compared to the previous work and the other existing approaches.**

I. Introduction

Hybridization approaches using the rough set theory and fuzzy systems have been a popular research interest over the past decade. Rough sets theory [1] was introduced as a powerful tool to reduce the given data into a compact form by the knowledge reduction process. It was also suggested that it is useful to generate decision rules for classification purpose. Fuzzy inference systems have been widely used in the past for better decision-making processes in a variety of practical and research areas. Approaches on classifiers such as neuro-fuzzy classifiers have been introduced to combine the adaptive mechanism from neural networks and human reasoning process from fuzzy systems towards higher accuracy for pattern classification [2], [3], [4].

However, the curse of dimensionality still remains as a big hurdle for most of the fuzzy classifiers especially when they are applied on high dimensional data. As the number of inputs increases, the number of rules in fuzzy systems increases exponentially. Also fuzzy systems do not have the adaptive mechanism to adjust its systematic parameters itself. Even the well-known neuro-fuzzy classifier, ANFIS (Adaptive-Network-Based Fuzzy Inference System) [5] still suffers from the high dimensionality on complex data.

Chang Su Lee is with the School of Computer Science, Edith Cowan University, 2 Bradford St, Mt. Lawley, Western Australia 6050, Australia (e-mail: chang-su.lee@ecu.edu.au).

The present paper proposes a new Rough-Neuro-Fuzzy (RNF) classifier for the pattern classification scheme on high dimensional data as an extension of the previous work [6], [7], and [8]. In the proposed approach, the ANFIS structure is employed as an adaptive systematic mechanism, whereas the previous work utilized the standard TS-type fuzzy inference model with least-square fit to adjust parameters. The entropy-based discretization technique is employed on the given data to generate clusters for each attribute with respect to the output information. Rough set theory is utilized to reduce the amount of features and to obtain a minimal set of decision rules on the decision table converted by clusters obtained. Then the proposed Rough-Neuro-Fuzzy system is constructed based on the structure of ANFIS, but its connections in the adaptive network is determined by the reduced data and the generated decision rules obtained by the rough sets-based approach. This provides the optimal and the minimal number of network configuration towards faster system performance. The learning scheme for the proposed approach is adopted from the one in ANFIS. The TS-type fuzzy inference model [9] is selected to perform decision-making process due to its ability to represent a general class of non-linear systems by decomposing the entire IO (input-output) domain into several partial non-linear spaces. A Multi-Input-Single-Output (MISO) TS-type fuzzy inference model is assumed for simplicity.

This paper is organized as follows: Section II provides a brief review of rough set theory and the ANFIS. Section III presents the design of the proposed Rough-Neuro-Fuzzy classifier. Experimental results are shown on a number of data sets with comparisons against other existing approaches in Section IV, and final conclusions are drawn in Section V.

II. Preliminaries

A. Entropy-based Discretization

The entropy-based discretization was introduced by Fayyad and Irani [10]. The entropy-based method uses the entropy of the given class information of candidate partitions to select optimum boundary points for discretization. The entropy of the class information is a measure of purity and it measures the amount of information that would be needed to specify to which class an instance belongs.

The goal of this method is to find the best boundary point T (selected from mid-points of the given attribute values) which produces the maximum information gain. The best mid-point T is found by examining all possible splits and then selecting the optimal one. The boundary that minimizes the entropy over all possible cases is chosen as an optimal one. Once the best T was found, given attribute values are replaced with labels corresponding to their discretized intervals.

B. Rough Set Theory

Rough set theory was developed by Pawlak [1] as a mathematical tool to deal with the classificatory analysis on imprecision, vagueness and uncertainty of the given data. The main objectives of the rough set analysis are to estimate approximations of concepts, to find the most significant attributes based on the given set of attributes, and to generate decision rules for classification purposes.

In relation to knowledge reduction, a given information system may be represented by a discernibility matrix and a discernibility function. They help to construct efficient algorithms related to a generation of minimal subsets of attributes which are sufficient to describe concepts in a given information system. The aim of the knowledge reduction is to obtain irreducible, but essential parts of the knowledge. Assume $U = \{x_1, x_2, \ldots, x_n\}$ and $Q = \{a_1, a_2, \ldots, a_m\}$. A discernibility matrix for an information system S with the set of attributes Q is a $n \times n$-matrix such that

$$m_{ij} = \{a \in Q : f(x_i, a) \neq f(x_j, a)\}. \quad (1)$$

Each entry m_{ij}, for the given i-th row and j-th column, of the discernibility matrix is a subset of attributes that discerns all objects in U.

An information system also can be represented as a decision table, $DT = < U, C \cup D >$ if the set of attributes Q can be represented by a set of condition attributes, C, and a set of decision attributes, D, in this form $Q = C \cup D$. In a decision table, every pair of condition and decision attributes determines the implication of $C \rightarrow D$ and a set of decision rules constitutes a decision algorithm. An elimination of redundant attributes is required in order to obtain a minimal set of decision rules. If $C^* = \{X_1, X_2, \ldots, X_p\}$ and $D^* = \{Y_1, Y_2, \ldots, Y_q\}$ are a C-definable set and a D-definable set of U. A set of decision rules r_{ij} for all D-definable sets Y_j is defined by (2),

$$\{Des_C(X_i) \Rightarrow Des_D(Y_j) : X_i \cap Y_j \neq \phi, \forall X_i \in C^*, \forall Y_j \in D^*\} \quad (2)$$

where $Des_C(X_i)$ and $Des_D(Y_j)$ are unique descriptions of the sets (classes) X_i and Y_j, respectively.

C. ANFIS (Adaptive-Network-Based Fuzzy Inference System)

The ANFIS [5] was proposed as a fuzzy inference system embedded in the framework of adaptive networks with a hybrid learning approach. As described in the literature [5], the adaptive network is a superset of all types of feedforward neural networks with supervised learning capability. It is a network structure consisting of nodes and directional connections through which the nodes are connected. Parts of all of nodes are adaptive, which means that their outputs depend on their parameters, and the learning rule shows how these parameters should be adjusted to minimize a selected error measure.

There are two learning schemes for adaptive networks; Off-line learning (batch learning) and On-line learning (pattern learning). In Off-line learning, the update procedure on parameters is carried out only after the whole training data set has been applied to the network, whereas parameters are updated immediately after each input-output vector has been presented in On-line learning method. Based on the adaptive network structure and the learning approaches mentioned above, there are two stages in the hybrid learning procedure for ANFIS. In the first stage which is called forward pass, the antecedent parameters are fixed and the consequent parameters are determined by the least square estimate. Once the consequent parameters are estimated, the actual ANFIS output is calculated along with the RMSE error values with respect to the target values. Then in the second stage which is backward pass, the error signals are utilized to adjust antecedent parameters using the gradient descent method while the consequent parameters are fixed.

As stated in the literature, the ANFIS can effectively predict highly non-linear models with much smaller RMSE values at the same number of iterations as compared to the conventional neural network-based models.

III. Design of a Rough-Neuro-Fuzzy (RNF) Classifier for Pattern Classification

A new Rough-Neuro-Fuzzy classifier based on the ANFIS structure is proposed here for pattern classification as an extension of the previous work [6], [7], [8] in the context of rough-fuzzy hybridization. The adaptive mechanism from the ANFIS network structure is employed in this approach, whereas the previous work utilized the conventional TS-type fuzzy inference model with least-square fit and conjugate gradient functions to adjust parameters. In this section, details of main systematic processes of the proposed RNF classifier are described in each sub-section.

A. Automatic Generation of Membership Functions

Automatic generation of membership functions is required to build the proposed hybrid TS-type fuzzy inference system as it is the design of fuzzy sets for input variables. The entropy-based discretization is selected to generate clusters for each attribute with respect to the output information, whereas the Fuzzy C-Means (FCM) clustering algorithm [11] was used in the previous work [7], [8]. The obtained cluster prototypes – the mean and the standard deviations are used to initialize antecedent parameters with a Gaussian bell-type membership function defined by (3). This design provides an optimal data distribution in terms of achieving the maximum information gain in modeling the training data with the Gaussian basis function. For consequent parameters of TS-type rules, initial values are set in a random manner in a range of [-1, 1].

$$\mu_{A_i}(x) = \frac{1}{1+\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \quad (3)$$

where a_i, b_i, c_i are the antecedent (premise) parameters which are to be adjusted in the backward pass during the hybrid learning procedure.

B. Conversion into a Decision Table

Partitioning is one of the significant factors which contribute to the accuracy of the classification results. In the proposed system, a partition method is applied by utilizing the entropy-based discretization approach.

Pattern vectors of training data set can be considered as features that compose a conditional attribute set C of an information system. The associated feature used to determine its output composes a decision attribute set D. These feature vectors and their target vectors constitute an information system as a decision table, $DT = < U, C \cup D >$.

A decision table from the training data set may be encoded using clusters obtained by the entropy-based partitioning. If partitioned regions are described as intervals of each dimension which replaced numeric attribute values of pattern vectors with labels of clusters, an original decision table may be converted into an encoded decision table as Table I.

TABLE I
ENCODED DECISION TABLE USING CLUSTERS OBTAINED

Attributes / Objects	a_1	a_2	…	a_m	DECISION
x_1	0	1	…	0	0
x_2	1	3	…	2	1
…	…	…	…	…	…
x_n	4	0	…	1	0

C. Feature Reduction and Rule Generation

The encoded decision table, however, may contain much redundant and conflicting data. One of main advantages of using the rough set methodology is that it reduces a given knowledge using the degree of dependency of attributes. This process requires finding minimal sets of condition attributes, or reducts, with respect to decision attributes in order to obtain the smallest possible number of decision rules for higher compactness. Thus the problem associated with the number of rules, which increases exponentially when more input variables are involved, can be resolved by finding a minimal set of decision rules generated using reducts. An algorithm based on the decision-relative discernibility matrix with Boolean calculation [12] is employed for the reduction of attributes. Details of the algorithm are described in [7], [8].

In addition, in order to select the *'best'* reduct amongst those candidates reducts, a fuzzy similarity measure [13] is applied. The degree of overlaps of membership functions for each feature in each reduct is measured using a fuzzy similarity measure defined in [7], [8]. This similarity is defined as a ratio of the squared sum of the degree of the intersection over the squared sum of the union between two fuzzy sets F_i, F_j. This ratio emphasizes the degree of overlaps, which is very sensitive for small changes of overlaps. As the operators of the intersection and the union calculations, the arithmetic minimum and maximum functions are used respectively. A *reduct*, which has the smallest overlap degree on average, is chosen as the final best reduct towards better accuracy in the pattern classification scheme.

Once the best reduct was obtained, a set of attributes in the reduct defines a particular relation in terms of an IO relation between selected conditional attributes and decision attributes. This IO relation generates a corresponding partition on the reduced data with respect to the decision attribute. In other words, a certain relation induced by a set of attributes generates partitioned equivalence classes on all objects in the given universe. Granules for each bio-marker in this partitioning process represent estimated range of its attribute values in terms of its unique IO mapping between each attribute and decision attribute. This remark is one of the claims to be a novel method in our proposed approach. Based on these equivalence classes, decision rules can be generated using (2). The final reduct as a minimal set of attributes will lead to the generation of minimal number of decision rules on the given training data.

D. Validation on generated Decision Rules

After the generation of a minimal set of decision rules, the validity of these rules must be ascertained to use them as TS-type fuzzy inference rules. The number of antecedent variables in the generated rules may be less than the total amount of the given input attributes. However, according to the definition of the TS-type fuzzy model [9], the TS-type fuzzy rules have a form of a combination of linear systems with all input attributes involved.

The question here is whether the generated rules provide a full IO coverage of the information inherent within the training data. The decision rules from the rough set approach were obtained by a partitioning procedure, which resulted in the corresponding equivalence classes. Since the partition algorithm divides a whole universe of the given information into disjoint equivalence classes via an indiscernibility relation, all equivalence classes are unique in terms of their IO relations. Also the partition algorithm is based on a reduct, which provides the same partition when all attributes are deployed. In other words, the minimal set of decision rules obtained offers a full IO coverage of the information and a unique set of partitions of the training data with respect to the decision attribute.

Antecedent variables in the minimal set of decision rules, however, may not show all input attributes since some of them have been eliminated in the reduction process. In order to form a complete numeric IO mapping between all IO information according to the definition of the TS-type fuzzy model, the reduced antecedent variables should be complemented in their rules. Therefore, the TS-type fuzzy rules with input variables complemented may be represented as suggested in (4),

$$y_i = y_i^{'} + y_i^{*} = c_{i0} + c_{i1}x_{k1} + \cdots + c_{il}x_{kl} + c_{i(l+1)}x_{k(l+1)} + \cdots + c_{im}x_{km} \quad (4)$$

where $y_i^{'}$ is a decision rule with reduced attributes and y_i^{*} is a complementing rule. For estimation of coefficients of the TS-type fuzzy rules, a weighted least-square algorithm can be used to minimize additional errors from complemented terms. It is crucial that this investigation should be carried out in the process of an automatic generation of fuzzy rules to provide a full IO relation of the given knowledge.

E. Construction of the proposed RNF system

Once the antecedent and consequent parameters are initialized and the minimal set of decision rules is obtained, the proposed Rough-Neuro-Fuzzy classifier can be built based on the structure of ANFIS with modified designs on the following components.

1) *Number of input attributes*: In the proposed system, the original given knowledge is effectively reduced by the rough set-based approach, which results in a set of minimal attributes. Therefore, only attributes found in the best reduct will be dominantly employed to construct antecedent parts in rules and their connections in the proposed adaptive network. A part of problems of high dimensionality can be resolved in this manner by excluding the insignificant attributes from the given data.

2) *Number of rules*: In the network structure from the original ANFIS, the rules are generated by considering all possible combinations between all MFs for all input attributes, which produces the exponential complexity in searching for firing rules – $mf\_n^{in\_n}$, where mf_n is the number of membership functions for each input and in_n is the number of inputs. However, in the proposed system rules will be generated via the rough set-based approach on the reduced data. The obtained minimal set of attributes generates a minimal set of rules with respect to the decision attributes. Thus the number of rules is significantly reduced than the one from the original ANFIS structure. This means the number of nodes in layer 2, layer 3, and layer 4 are the same with the reduced minimal number of rules, and their corresponding connections are also from the reduced rules. Thus, utilizing this approach another problem of high dimensionality can be solved towards lower computational complexity in searching process for rules.

3) *Network connections between layer 1 and layer 2*: As rules are generated via the rough set-based approach in the proposed system, the network connection between layer 1 and 2 will be determined by the decision rules generated using the reduced attributes, instead of establishing all possible cases of connections on all nodes in between layer 1 and layer 2 as designed in the original ANFIS structure. Other connections in between layer 2, 3, and 4 are determined by the reduced minimal rules which lead to the minimal configuration of network connections.

4) *Consequent parameters in nodes in layer 4*: Even though some of input attributes are eliminated in the antecedent parts in rules, all antecedent parameters are complemented back in consequent parts of rules in order to cover full IO mapping to provide a full coverage of the information inherent within the training data as described in the previous work [7], [8]. For estimation of coefficients of the TS-type fuzzy rules, a weighted least-square algorithm can be utilized to minimize additional errors from the terms complemented.

The learning procedure for the proposed system is adopted from the one in the ANFIS, which is the same hybrid learning process as described earlier. The RMSE (Root-Mean-Square-Error) measure is employed to evaluate the system performance of the proposed system on the given training set of data.

IV. EXPERIMENTAL RESULTS

This section consists of two parts: a brief description on data sets used, and experiments and results produced by the proposed Rough-Neuro-Fuzzy classifier approach along with comparisons on results with other approaches.

A. Data used Experiments

The proposed Rough-Neuro-Fuzzy classifier has been applied to a number of different data sets in order to evaluate its viability for pattern classification purposes. Most of them are obtained from the UCI machine learning repository [14]. Other high dimensional data are from some of the bio-medical research institutes. Each data set has a different number of features and samples collected from a different domain with its unique characteristics. Table II shows the number of features and sample vectors for each data set and data sets listed have a variety of range in size from a small one (Iris) up to a very high dimensional data (Leukemia). Each data set has been processed under the same conditions for removing missing data labeled '?' in the original data, carrying out the entropy-based discretization, and evaluating classification performance via 10 independent runs of 10-fold Cross Validation (CV).

TABLE II
NUMBER OF FEATURES AND SAMPLES ON DATA USED

DATA	Features	Samples
Iris	4	150
Cancer	9	699
WDBC	30	569
Sonar	60	208
Alzheimer's Disease [19]	120	83
Colon Cancer[21]	2000	62
Lymphoma [22]	4026	96
Leukemia [23]	7129	72

B. Experiments and Results

Using the proposed rough set-based approaches for feature reduction and rule generation, each data set was significantly reduced into the essential minimal size. Based on final set of reduced features, decision rules are generated employing partition algorithm as stated earlier by the rough set-based approach. The numbers of reduced features and generated rules for each data set has shown in Table III.

TABLE III
NUMBERS OF REDUCED FEATURES AND GENERATED RULES FOR DATA USED

DATA	Reduced.Features	Rules
Iris	3	23
Cancer	6	128
WDBC	17	278
Sonar	20	180
Alzheimer's Disease [19]	7	46
Colon Cancer[21]	19	120
Lymphoma [22]	36	237
Leukemia [23]	29	186

Results shown in Table III indicate that the number of attributes of the given data sets has been significantly reduced to a minimal number of features for each data set. For small data sets such as Iris, Wisconsin Breast Cancer, Wisconsin Diagnostic Breast Cancer, Sonar, and the Alzheimer's disease data, the number of features was reduced to minimal number of features. For higher dimensional data sets such as Colon cancer, Lymphoma, and Leukemia data, an additional pre-processing step was applied to reduce the complex data more efficiently before the entropy-based discretization is applied. The pre-processing consists of the followings. 1) An evaluation of the worth of each attribute by measuring the information gain with respect to the class. Attributes estimated with zero worth are eliminated. 2) Based on the estimated attributes, a search using the simple genetic algorithm proposed by Goldberg [25] is performed to select the most relevant attributes. 3) Repeat these two steps until there is no change in the final reduced set of attributes. Then apply the proposed rough set-based approach for feature reduction on top of those pre-processed data sets.

The experimental results are shown in Table IV – XI for a comparison with other existing classification approaches. The classification results shown in tables are calculated on average via 10 independent runs of 10-CV estimation. As shown in Tables, the proposed Rough-Neuro-Fuzzy classifier achieved quite compatible or competitive classification performances on the given data sets.

This contribution was achieved by the utilization of the rough set-based feature selection and rule generation, and the adaptive network structure built based on the reduced data. It can be stated that the proposed Rough-Neuro-Fuzzy classifier may be a suitable tool to perform the pattern classification on these data sets.

TABLE IV
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON IRIS

Approaches	Testing Accuracy (%)
FEBFC [15]	97.12
IRSS [16]	96.00
ARFIS [7], [8]	96.28
Proposed RNF	96.70

TABLE V
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON CANCER

Approaches	Testing Accuracy (%)
FEBFC [15]	95.14
IRSS [16]	95.89
ARFIS [7], [8]	96.63
Proposed RNF	97.20

TABLE VI
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON WDBC

Approaches	Testing Accuracy (%)
SS1 [17]	96.11
ARFIS [7], [8]	95.59
Proposed RNF	96.45

TABLE VII
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON SONAR

Approaches	Testing Accuracy (%)
FCMC [18]	86.50
ARFIS [7], [8]	80.00
Proposed RNF	85.00

TABLE VIII
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON ALZHEIMER'S

Approaches	Testing Accuracy (%)
Ray et. al [19]	89.00
Ravetti and Moscato [20]	96.00
ARFIS [7], [8]	93.68
Proposed RNF	94.60

TABLE IX
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON COLON CANCER

Approaches	Testing Accuracy (%)
Banerjee et al. [24]	90.30
ARFIS [7], [8]	87.80
Proposed RNF	89.00

TABLE X
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON LYMPHOMA

Approaches	Testing Accuracy (%)
Banerjee et al. [24]	95.80
ARFIS [7], [8]	92.80
Proposed RNF	94.69

TABLE XI
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON LEUKEMIA

Approaches	Testing Accuracy (%)
Banerjee et al. [24]	94.10
ARFIS [7], [8]	90.37
Proposed RNF	92.28

V. CONCLUSIONS

A new Rough-Neuro-Fuzzy classifier for pattern classification on high dimensional data has been proposed in this paper. It is significant that the number of attributes and the number of rules generated by the proposed approach were reduced effectively by the rough set approach. The adaptive network structure also was employed for learning systematic parameters towards better performance. The experimental results on a number of data sets indicated that the performances of the Rough-Neuro-Fuzzy classifier were found to be very competitive and satisfactory.

REFERENCES

[1] Z. Pawlak, "Rough Sets," International Journal of Computer and Information Science, vol. 11, pp. 341-356, 1982.

[2] P.K. Simpson, "Fuzzy Min-Max Neural Networks–Part 1: Classification," IEEE Transaction On Neural Networks, vol. 3, no.5, pp.776-786, Sept. 1992.

[3] G.O.A. Zapata, R.K.H. Galvao, and T. Yoneyama, "Extracting Fuzzy Control Rules from Experimental Human Operator Data," IEEE Transaction On System, Man, and Cybernetics – Part B: Cybernetics, vo. 29, no. 3, pp 25-40, Feb. 1999.

[4] S. K. Pal, S. Mitra, P. Mitra, "Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation," IEEE Transactions On Knowledge and Data Eng., vol. 15, no.1, Jan./Feb. 2003.

[5] J. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," IEEE Transactions On Systems, Man, and Cybernetics, vol. 23, no. 3, May/June 1993.

[6] C. Lee, A. Zaknich, T. Bräunl, "An Adaptive T-S type Rough-Fuzzy Inference System (ARFIS) for Pattern Classification," IEEE International Conference on North American Fuzzy Information Processing Society 2007 (NAFIPS '07), pp. 117-122, San Diego, US, June 2007.

[7] C. Lee, A. Zaknich, T. Bräunl, "A Framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)," IEEE 2008 International Conference on Fuzzy Systems (FUZZ-IEEE 2008), pp. 567-574, Hong Kong, June 2008.

[8] C. Lee, "A Framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)," Ph.D. Thesis, Department of Electrical, Electronic, and Computer Engineering, The University of Western Australia, Perth, Australia, 2008.

[9] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Transaction On Systems, Man, and Cybernetics, vol. SMC-15, pp. 116-132, Jan.-Feb. 1985

[10] U.M. Fayyad, K.B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," Proc. 13th International Joint Conference on Artificial Intelligence (IJCAI 1993), San Francisco, CA, pp. 1022-1027, 1993.

[11] James C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, pp.65-86, 1981.

[12] R. Jensen and Q. Shen, "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches," IEEE Trans On Knowledge and Data Engineering, pp. 1457-1471, vol. 16, no. 12, Dec 2004.

[13] G. Jäger and U. Benz, "Measures of Classification Accuracy Based on Fuzzy Similarity," IEEE Transactions on Geoscience and Remote sensing, vol. 38, no. 3, pp. 1462-1467, May 2000.

[14] C.L. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases," University of California, Irvine, Department of Information and Computer Science, http://www.ics.uci.edu/~mlearn /MLRepository.html, 1998.

[15] H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou, "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy," IEEE Transaction On Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 31, no. 3, pp.426-432, June 2001.

[16] A. Chatterjee and A. Rakshit, "Influential Rule Search Scheme (IRSS) – A New Fuzzy Pattern Classifier," IEEE Transaction On Knowledge and Data Engineering, vol. 16, no. 8, pp. 881-893, Aug. 2004.

[17] K. Saastamoinen, J. Ketola, "Medical Data Classification using Logical Similarity Based Measures," IEEE International Conf. on Cybernetics and Intelligent Systems, pp. 1-5, June 2006.

[18] H. Ichihashi, K. Honda, A. Notsu, T. Hattori, "FCM Classifier for High Dimensional Data," IEEE 2008 International Conference on Fuzzy Systems (FUZZ-IEEE 2008), pp. 200-206, Hong Kong, June 2008.

[19] Ray et. al, "Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins," Nature Medicine, pp. 1-4, Oct. 2007.

[20] G. Ravetti and P. Moscato, "Identification of a 5-Protein Biomarker Molecular Signature for Predicting Alzheimer's disease", PLoSONE 3(9): e3111. doi:10.1371/journal.pone. 0003111, 2008.

[21] Available from Gene expression project, Princeton University, http://microarray.princeton.edu/oncology.

[22] Available from Lymphoma/Leukemia Molecular Profiling Project, http://llmpp.nih.gov/lymphoma/data/figure1/figure1.cdt.

[23] Available from Cancer Program Data Sets, Broad Institute of MIT and Havard, http://www.genome.wi.mit.edu/MPR.

[24] M. Banerjee, S. Mitra, H. Banka, "Evolutionary Rough Feature Selection in Gene Expression Data", IEEE Transaction on Systems, Man, and Cybernetics – Part C: Applications and Reviews, vol. 37, no. 4, pp. 622-632, July 2007.

[25] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison- Wesley, 1989.