Neuro-Fuzzy Ensemble Approach for Microarray Cancer Gene Expression Data Analysis

Zhenyu Wang, *Student Member, IEEE,* Vasile Palade, and Yong Xu

Abstract— **A Neuro-Fuzzy Ensemble model (NFE) is proposed in this paper for analysing the gene expression data from microarray experiments. The proposed approach was tested on three benchmark cancer gene expression data sets. Experimental results show that our NFE model can be used as an efficient computational tool for microarray data analysis. In addition, compared to some current most widely used approaches, Neuro-Fuzzy(NF)-based models not only supply good classification results, but their behavior can also be explained and interpreted in human understandable terms, which provides the researchers with a better understanding of the data.**

I. INTRODUCTION

Many methods have been designed and used so far for discovering the mystery behind the DNA data. One of the most revolutionary techniques is the high-density DNA microarray chips, or commonly called DNA chips, or gene chips. This technique can measure the activities of thousands of genes simultaneously under different experimental environments and conditions. It allows us to have a "global" view of the cell [1].

The gene expression profiles from particular microarray experiments have been recently used for cancer classification [2] [3] [4]. This approach promises to give a better therapeutic measurement to cancer patients by diagnosing cancer types with improved accuracy [4]. However, the amount of data produced by this new technology is usually too large to be manually analysable. Hence, the need to automatically analyse the microarray data offers an opportunity for Machine Learning (ML) methods to have a significant impact on cancer research.

Unsupervised methods, such as Clustering [5], and Self-Organizing Maps (SOMs) [6] were initially used to analyse the relationships among different genes. Recently, supervised methods, such as Support Vector Machines (SVMs) [7], Multi-Layer Perceptrons (MLP or NNs) [8] [9], K Nearest Neighbor (KNN) method [10] [11], and Decision Trees (DTs) [9] have been successfully applied to classify different tissues. But most of the current methods in microarray analysis can not completely bring out the hidden information in the data. Meanwhile, they are generally lacking robustness with respect to noisy and missing data.

Different from black-box methods, fuzzy rule-based models can not only provide good classification results, but also easily

Z. Wang and V. Palade are with the Computing Laboratory, Oxford University, Oxford, OX1 3QD, UK (e-mail: zhen.yu.wang@comlab.ox.ac.uk; vasile.palade@comlab.ox.ac.uk).

Y. Xu is with the Centre of Excellence for Research in Computational Intelligence and Applications, University of Birmingham, Birmingham, B15 2TT, UK (e-mail:y.xu@cs.bham.ac.uk).

be explained and interpreted by human understandable fuzzy rules. This provides the researchers an insight into the models. Meanwhile, fuzzy systems adapt numerical data (input/output pairs) into human linguistic terms, which offer very good capabilities to deal with noisy and missing data.

However, how to define the rules and membership functions requires a lot of prior knowledge. This can be usually obtained from human expert, especially in the case of large amount of gene expression data, which is not an easy task. Hybrid Neuro-Fuzzy (NF) models, which combine the learning ability of neural systems and the interpreting ability of fuzzy systems, can automatically generate and adjust the membership functions and linguistic rules directly from data.

Unfortunately, NF methods have suffered some well-known limitations in dealing with high dimensional data. Although some fuzzy-rule-based applications for microarray analysis have already been presented [12] [13], all these reported systems are small models and only perform well on simple data sets. Because large rule-based models imply huge computational cost, they sometimes are unacceptable in practice. In order to improve the inherent weakness of individual NF models, a Neuro-Fuzzy Ensemble (NFE) model is developed in this paper. The proposed NFE model is tested on three benchmark microarray cancer data sets, including leukemia cancer data set, colon cancer data set, and lymphoma cancer data set.

Compared to some other benchmark problems in machine learning, microarray data sets may be problematic. The number of features (genes), usually in the range of 2,000-30,000, is much larger than the number of samples (usually in the range of 40-200). But not all of these genes are needed for classification. Most genes do not influence the performance of the classification task. Taking such genes into account during classification increases the dimension of the classification problem, poses computational difficulties and introduces unnecessary noise in the process. A major goal for diagnostic research is to develop diagnostic procedures based on inexpensive microarrays that have enough probes to detect certain diseases. This requires the selection of some genes which are highly related to the particular classification problem, i.e., the informative genes. This process is called gene selection, or feature selection in machine learning in general. A most widely used gene selection method, Information Gain (IG), is adopted for the purpose of this work.

Because a small number of available data usually can not sufficiently represent the whole search space, the traditional training and testing strategies do not perform well. As suggested in previous work [8] [14], we will be using Leave One

Out Cross Validation strategy (LOOCV) to evaluate models, as described in Section IV-C.

The rest of this paper is organized as follows: how to classify different types of cancer by using microarray technology is described in Sections II and III; why single NF approaches do not perform well for this particular application is explained in Section IV-A; a NFE model is introduced in Section IV-B; experimental results and analytical work are presented in Section V; some conclusions are drawn in Section VI.

II. MICROARRAY TECHNOLOGY

Applying microarray experiments to study biological problems was firstly introduced by a research team at Stanford University in 1995. It is currently one of the best tools for biomedical research. Microarray techniques allow simultaneous measuring of the expression of thousands of genes under different experimental environments and conditions. It enables us to analyse the gene information very rapidly by managing them at one time.

In the experiments, different DNA samples are fixed to a glass microscope slide, each at a predefined position in the array, known as "gene chip". mRNAs isolated from different tissue samples, or under different conditions, are labeled with two different fluorochromes (generally the green Cy3 and the red Cy5), then hybridized with the arrayed DNA probes on the slide. Using a fluorescent microscope and image analysis, the gene expression data (denoted as G) are measured by computing the log ratio between the two intensities of each dye.

$$G = log_2 \frac{Int(Cy5)}{Int(Cy3)} \qquad (1)$$

where $Int(Cy5)$ is the intensity of red color, and $Int(Cy3)$ is the intensity of green color.

Generally, microarray experiments can be divided into two types. One focuses on time series data which contains the gene expression data of various genes during the time span of an experiment. Another type of microarray experiment consists of gene expression data of various genes taken from different tissue samples or under different experimental conditions. Different conditions can be used to answer such questions as which genes are changed under certain conditions. Meanwhile, different tissues under the same experiment conditions are helpful in the classification of different types of tissues. The data from a series of n such experiments can be represented as a $m \times n$ gene expression matrix, see Table I. Each row represents a sample that consists of m genes from one experiment. Each sample belongs to a certain class (normal or tumor). In each data set, the researchers repeated the same experiment on n different volunteers, each line in this data set representing a volunteer.

From Table I, we can see that classifying microarray gene expression data can be looked as a high-dimensional-low-sample problem. Common approaches are to select a subset of the most useful features, then classify the different samples as cancer or non-cancer, according to the selected features, by

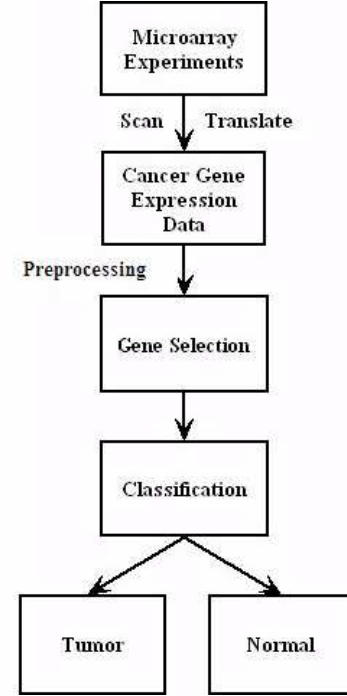|	Gene 1	Gene 2	...	Gene m-1	Gene m	Class
Sample 1	165.1	276.4	...	636.6	784.9	1
Sample 2	653.6	1735.1	...	524.1	104.5	-1
...	...	...	...	...	...	...
Sample n-1	675.0	45.1	...	841.9	782.8	-1
Sample n	78.2	893.8	...	467.9	330.1	1

Fig. 1. Cancer Classification System.

using certain classifiers. This can be summarized as follows. Let the given gene expression data be denoted as

$$\tilde{D} = \{(g_1, t_1,), ..., (g_n, t_n,)\}, \qquad (2)$$

where an input vector $g_i = (g_1, ..., g_m)$ denotes a gene expression pattern, m is the number of genes in this pattern, t_i represents which class the pattern belongs to (see Section III), and n denotes the number of patterns in the data set. Now, let us choose $\tilde{m}$ genes out of m according to certain algorithms. Then, select $\tilde{n}$ patterns with $\tilde{m}$ genes to train the classifier, and leave $n - \tilde{n}$ patterns (with $\tilde{m}$ genes) out to test the performance of trained model. Figure 1 gives a typical cancer classification system.

III. GENE SELECTION

Recent researches have shown that a small number of genes is sufficient for accurate diagnosis of most cancers, even though the number of genes vary greatly between different diseases [15]. Indeed, a large set of gene expression features will not only significantly bring higher computational cost and slow down the learning process, but also decrease the

classification accuracy due to the phenomenon known as curse of dimensionality, in which the risk of over-fitting increases as the number of selected genes grows [15]. More importantly, by using a small subset of genes, we can not only get a better diagnostic accuracy, but also get an opportunity to further analyse the nature of the disease and the genetic mechanisms responsible for it. Therefore, the microarray cancer classification problem can be classified as a combinational optimization problem with two main objectives: minimizing the number of selected genes and maximizing the classification accuracy.

Filter method and wrapper method are two basic feature selection approaches in machine learning. Filtering procedures are widely used in the area of microarray analysis. A common way is to rank all features according to the values of a univariate scoring metric, and the top ranked features are selected for classification. The wrapper approach, which is popular in many machine learning applications, is not largely used in DNA microarray tasks. Therefore, in our work, we will adopt the Information Gain, the most widely used filtering gene selection method.

A. Information Gain

Given entropy E as a measure of impurity in a set of training samples, it is possible to define a measure of the effectiveness of a feature/gene in classifying the training data. This measure is simply the expected reduction in entropy caused by partitioning the data according to this feature, so-called Information Gain [16]. Assuming a given set of microarray gene expression data M, the information gain of a gene i is defined as:

$$IG(M, i) = E(M) - \sum_{v \in V(i)} \frac{M_v}{M} E(M_v),$$ (3)

where $V(i)$ is the set of all possible values of feature i, M_v is the subset of M for which feature I has the value v, $E(M)$ is the entropy of the entire set, and $E(M_v)$ is the entropy of the subset M_v. The entropy function E is defined as:

$$E = \sum_{j=1}^{c} -\frac{|C_j|}{|\sum C|} log_2 \frac{|C_j|}{|\sum C|}$$ (4)

where $|C_j|$ is the number of samples in class C_j. The entropy is supposed to give the information required in bits, and is traditionally used to deal with boolean valued features (hot/cold, true/false, etc.). Fortunately, this method can be easily extended to handle the data with continuous valued features.

IV. NEURO-FUZZY ENSEMBLE

Here we construct a Neuro-Fuzzy Ensemble model by combining several individual NF models to learn the same data using different subsets of genes.

TABLE II
THE RELATIONSHIP AMONG THE NUMBER OF INPUT FEATURES, THE NUMBER OF FUZZY RULES, AND THE NUMBER OF PARAMETERS NEEDED TO BE UPDATED IN EACH EPOCH. NoG DENOTES THE NUMBER OF SELECTED GENES.

NoG	Number of Rules	Number of Parameters
2	9	45
3	27	108
4	81	360
5	243	774
6	729	2484
7	2187	7008
...	...	...

A. Individual Neuro-Fuzzy Models

We use Adaptive-Network-based Fuzzy Inference System (ANFIS) to build the individual classifiers in the ensemble. ANFIS is a Sugeno-like fuzzy system in a five-layered network structure. Back-propagation strategy is used to train the membership functions, while the least mean squares algorithm (LSE) determines the coefficients of the linear combinations in the consequent part of the rules. It has been proven that this hybrid algorithm is much more efficient than a standard gradient method in training the ANFIS [17].

The same as other fuzzy-rule-based systems, ANFIS can not easily handle high dimensional problems, as this leads to a large number of input nodes, rules, and, hence, consequent parameters. In an ANFIS model, if the number of inputs is N, the number of membership functions for each input is K, then the number of the fuzzy rules R is:

$$R = K^N$$ (5)

The number of the adaptive parameters P is then:

$$P = K^{N+1} + K^N \times N$$ (6)

where K^{N+1} refers to the non-linear adaptive parameters, and $K^N \times N$ to the linear adaptive parameters. For example, when the number of membership functions for each input is 3, the relationship between the inputs and rules is shown in Table II. From Eqs. 5, 6 and Table II, we can see that the computational cost increases very quickly while the number of inputs grows. We simulated the models on an IBM R51 laptop (CPU: PIV-1.5G, Memory: 1G). The computer is out of memory when the number of inputs is larger than 6.

But, the collection of well-distributed, sufficient, and accurately measured input data is the basic requirement to obtain an accurate model [18]. When data sets require a relatively large number of genes to represent their properties, we need to design some strategies to enable the model to accept more inputs with less computational cost. Recommended approaches include: evaluate and select rules, delete antecedents, delete fuzzy sets, etc.

B. Ensemble Learning

Besides building individual ANFIS models as described above, we designed a Neuro-Fuzzy Ensemble (NFE), which consists of several different single ANFIS models, to deal with this problem. Each individual model learns different subsets
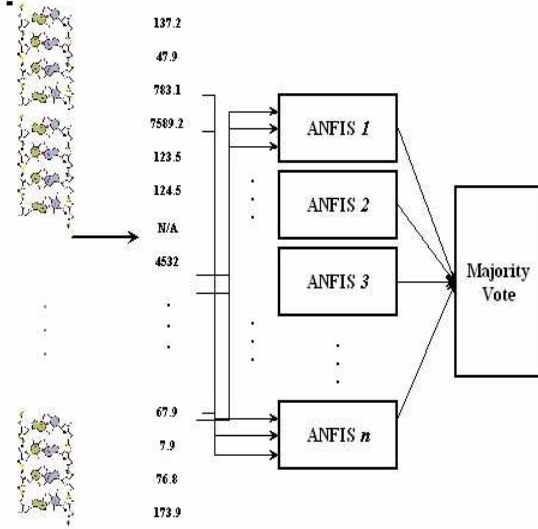
Fig. 2. The main structure of the NFE: n individual ANFIS classifiers in the ensemble, each having R inputs, so that the whole model can study $R*n$ genes. The output of the ensemble is taken by simple majority voting (MV).

TABLE III

COMPARISON OF COMPUTATIONAL COST BETWEEN INDIVIDUAL NF AND NFE MODELS. WE COMPARE THE NUMBER OF RULES AND PARAMETERS OF INDIVIDUAL NF AND NFE MODELS. EACH INPUT OF THE NF MODELS HAS 3 MEMBERSHIP FUNCTIONS. IN THIS COMPARISON, THE NFE CONTAINS TWO INDIVIDUAL NF MODELS. NoG DENOTES THE NUMBER OF SELECTED GENES, NoR DENOTES THE NUMBER OF RULES, NoP DENOTES THE NUMBER OF PARAMETERS NEEDED TO BE UPDATED IN EACH EPOCH.

	NF		NFE	
NoG	NoR	NoP	NoR	NoP
2	9	45	N/A	N/A
3	27	108	N/A	N/A
4	81	360	N/A	N/A
6	729	2484	54	216
8	6.6×10^4	7.9×10^5	162	720
12	1.7×10^7	2.7×10^8	243	1080
16	4.3×10^9	8.6×10^{10}	324	1440
20	1.1×10^{12}	2.6×10^{13}	405	1800
...	...	...	...	...

of genes, so that the overal model can work with a relatively large number of genes. Meanwhile, extra good performance can be obtained by the nature of the ensemble learning itself. We set the maximum number of inputs for an individual NF network in the NFE to be 4, due to the considerations on computational costs implied, as discussed above. Each input is randomly selected from the top $4 \times n$ ranked genes, where n is the number of singe NF models in the ensemble. The output strategy of our NFE model is Majority Voting, where the output of the majority of individual NF networks will be considered to be the output of the ensemble. The main structure of our NFE is shown in Figure 2.

The advantages of our NFE model can be summarized as follows:

- It allows the model to learn more features when the optimal subset of genes is relatively large.
- Normally, several classifiers imply higher computational costs than building a single classifier only. However, our NFE model seems to require much less computational cost than building a single large NF model, when the necessary number of inputs is relatively large. A comparison of computational costs between individual NF models and the NFE is shown in Table III.
- NFE can significantly improve the generalization ability (classification performance) compared to single NF models, and they can also help address three classic machine learning problems: lack of data, local optima, and representational limitations. Lack of data is one of the major problems in microarray analysis.
- NFE can relieve the trial-and-error process by tuning architectures.

Though, we have to point out that the ensemble structure increases the complexity of the system, which means the model will loose the transparency of the decision-making process [19].

C. Training and Testing Strategy

In order to use as many samples as possible, we adopt leave-one-out cross validation (LOOCV or jackknife strategy) to train and evaluate our models. We divide all samples, at random, into K distinct subsets, where K equals to the number of samples. We then train the model using $K - 1$ subsets (samples in our case), and test the training performance on the Kth sample. The LOOCV accuracy is obtained by:

$$LOOCV\,accuracy = \frac{Acs}{K} \qquad (7)$$

where Acs is the number of correctly classified samples in K experiments. LOOCV accuracy is strongly suggested to be used as an evaluation of microarray data classification by many other researchers. In order to compare with their work, this strategy is also adopted in our study.

V. EXPERIMENTAL RESULTS

In this study, the proposed models are tested on three data sets: leukemia cancer data set, colon cancer data set, and lymphoma cancer data set. We adopt three important criteria for empirical evaluation of the performance of our models:

- Number of selected genes,
- Predictive accuracy on selected genes,
- Extracted knowledge from the trained models.

Before the experiments, we linearly scale all data in the [0, 1] interval. If y is a gene expression value of a gene g, the scaled value will be

$$g(a') = \frac{y - min(g)}{max(g) - min(g)} \qquad (8)$$

where $min(g)$ and $max(g)$ is the minimum and maximum values of gene expressions in the database.

A. Microarray Cancer Data Sets

1) Colon Cancer Data Set: The data set we used here was firstly reported in [20]. The "cancer" biopsies were collected from tumors, and the "normal" biopsies were collected from healthy parts of the colons of the same patients [20]. This data set contains 62 samples. There are 40 tumor samples, and 22 normal samples. From about 6000 genes represented in each sample in the original data set, only 2000 genes were selected. The data is available at http://sdmc.i2r.a-star.edu.sg/rp/ColonTumor/ColonTumor.html.

2) Leukemia Cancer Data Set: The data set we used here was reported in [3]. The gene expression measurements were taken from 63 bone marrow samples and 9 peripheral blood samples [3]. This data set contains 72 samples. All samples can be divided into two subtypes: 25 samples of acute myeloid leukemia (AML) and 47 samples of acute lymphoblastic leukemia (ALL). The expression levels of 7129 genes were reported. The data is available at http://sdmc.i2r.a-star.edu.sg/rp/Leukemia/ALLAML.html.

3) Lymphoma Cancer Data Set: The data set we used here was reported in [21]. This data set contains 47 samples. B cell diffuse large cell lymphoma (B-DLCL) data set includes two subtypes: germinal center B cell-like DLCL and active B cell-like DLCL. The expression levels of 4026 genes were reported. 24 samples are germinal center B-like DLCL and 23 samples are active B cell-like DLCL. The data is available at http://www.genome.wi.mit.edu/MPR.

B. Experiment Setup

Each variable is described by three membership functions for both NF and NFE models, and the initial shape of the membership function is a bell-shaped function, see Figure 3 (Top). There are 5 individual NF networks in our NFE model, and each NF model has 4 inputs. The output of the ensemble is obtained by using Majority Voting (MV). The top 20 ranked genes obtained by the IG method were selected for classification when using the NFE model. The top 4 ranked genes are selected for classification when using the single NF models. The top 20 Leukemia genes with the highest scores were listed in Table IV.

C. Results and Comparison

The performance of our NFE model was compared with that of single NF models and some other reported approaches, by using the same training and testing strategies, see Table V. Our models obtained better results on Colon cancer data set, and similar results on Leukemia and Lymphoma data set, but both NF and NFE models use less number of genes comparing with other approaches. The performance of the NFE model is much better than that of single NF models on the three cancer data sets, see Table V.

Furthermore, different from black-box approaches, NF-based models can extract some useful knowledge from data, for example, adjusted membership function (Figure 3), and trained fuzzy rules (Table VI). All this knowledge can be presented in human understandable form. This seems very

TABLE IV

TOP 20 RANKED LEUKEMIA GENES SELECTED BY IG2

Rank	ID	Gene name	Description
1	4050	X03934	GB DEF = T-cell antigen receptor gene T3-delta
2	6510	U23852	GB DEF = T-lymphocyte specific protein tyrosine kinase p56lck (lck) abberant mRNA
3	4342	X59871	TCF7 Transcription factor 7 (T-cell specific)
4	4055	X04145	CD3G CD3G antigen, gamma polypeptide (TiT3 complex)
5	5542	M37271	T-CELL ANTIGEN CD7 PRECURSOR
6	5543	M37271	T-CELL ANTIGEN CD7 PRECURSOR
7	5466	X58072	GATA3 GATA-binding protein 3
8	6606	X00437	TCRB T-cell receptor, beta cluster
9	1694	M12886	TCRB T-cell receptor, beta cluster
10	6696	X76223	GB DEF = MAL gene exon 4
11	1893	M28826	CD1B CD1b antigen (thymocyte antigen)
12	2833	U16954	(AF1q) mRNA
13	4357	X60992	T-CELL DIFFERENTIATION ANTIGEN CD6 PRECURSOR
14	4847	X95735	Zyxin
15	1106	J04132	CD3Z CD3Z antigen, zeta polypeptide (TiT3 complex)
16	3332	U50743	Na,K-ATPase gamma subunit mRNA
17	6236	U83239	CC chemokine STCP-1 mRNA
18	4484	X69398	CD47 CD47 antigen (Rh-related antigen, integrin-associated signal transducer)
19	4291	X56468	14-3-3 PROTEIN TAU
20	2454	S65738	Actin depolymerizing factor [human, fetal brain, mRNA, 1452 nt]

TABLE V

COMPARISON OF THE CLASSIFICATION PERFORMANCE OF DIFFERENT CLASSIFIERS ON LEUKEMIA CANCER DATA SET, COLON CANCER DATA SET, AND LYMPHOMA CANCER DATA SET. GSMETHOD DENOTES THE GENE SELECTION METHOD.

	GSMethod	NoSG	Colon	Leukemia	Lymphomia
Single NF	IG	4	93.55	87.5	87.23
NFE	IG	20	100	95.85	95.65
SVM [7]	IG	50	90.30	94.10	N/A
SVM [9]	SNR	50	65.0	59.0	76.0
	LLE	50	85.0	95.0	91.0
KNN	EA	50	75.81	72.64 [11]	74.47
C4.5	ReliefF	4-60	85.48 [16]	81.94 [16]	82.98

attractive for researchers in the area, as they can better understand the data or explain how the results are obtained. Meanwhile, NF-based models can also easily incorporate prior knowledge, which helps obtaining more refined models and shortening the training process.

VI. CONCLUSION AND FUTURE WORK

The classification performance obtained by our NFE model is very competitive. But, there are still many issues that need to be considered in future research. After combination, the NFE model becomes more complex than a single NF model and therefore more difficult to analyse. While a single NF model can be easily explained and interpreted by users, an ensemble of several NF models would be more difficult to understand. The balance between the classification accuracy and interpretability should be further explored. The performance of the NFE can be further enhanced by using other ensemble training
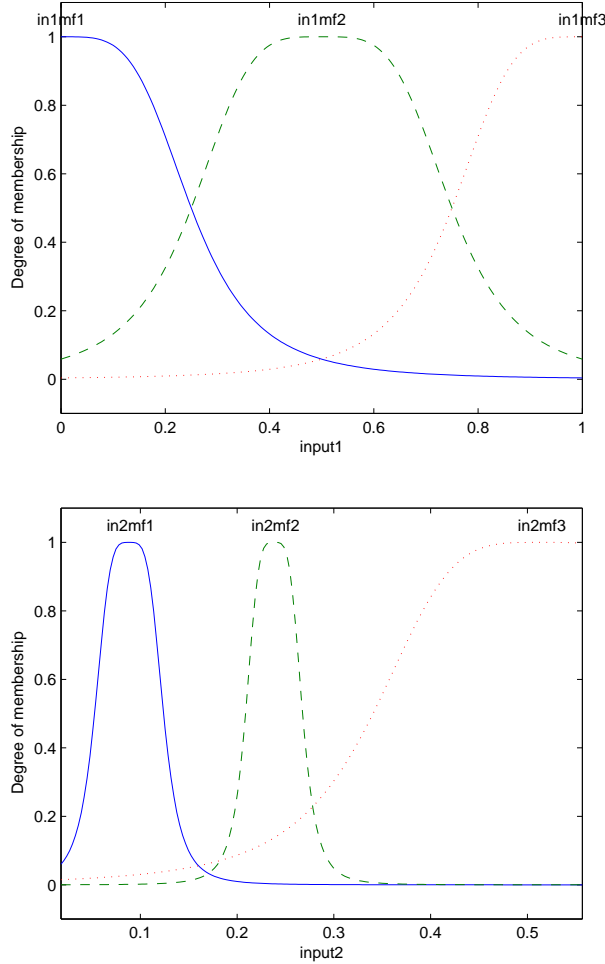
Fig. 3. Initial membership functions (Top) and Adjusted membership functions (Bottom) of the NFE model on Colon Cancer Data Set.

	Descriptions of Rules
1	If (M63391 is small) and (R87126 is small) then (output is Cancer)
2	If (M63391 is small) and (R87126 is medium) then (output is Cancer)
3	If (M63391 is small) and (R87126 is large) then (output is Normal)
7	If (M63391 is large) and (R87126 is small) then (output is Cancer)
9	If (M63391 is large) and (R87126 is large) then (output is Normal)

techniques, i.e., bagging and boosting. While LOOCV is not the only appropriate strategy, bootstrap resampling technique can also be considered for such kind of problems [19]. In addition, NF-based methods offer good potential to deal with highly noisy/missing data, which can also be considered as an important future research direction. Our experimental results show that the NFE models can be very effective tools for microarray gene expression data classification problems.

REFERENCES

[1] P.-S. Gregory and T. Pablo, "Microarray data mining: facing the challenges," *SIGKDD Explorations*, vol. 5, no. 2, pp. 1–5, 2003.

[2] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Natnl. Acad. Sci. USA.*, vol. 96, no. 12, pp. 6745–6750, 1999.

[3] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531–537, 2002.

[4] D. K. Slonim, P. Tamayo, J. P. Mesirov, T. R. Golub, and E. S. Lander, "Class prediction and discovery using gene expression data," in *RECOMB*, 2000, pp. 263–272.

[5] R. Baumgartner, C. Windischberger, and E. Moser, "Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis." *Magn Reson Imaging*, vol. 16, no. 2, pp. 115–125, 1998.

[6] T. Kohonen, Ed., *Self-organizing maps*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1997.

[7] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, pp. 906–914, 2000.

[8] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, pp. 673–679, 2001.

[9] C. Shi and L. Chen, "Feature dimension reduction for microarray data analysis using locally linear embedding," in *APBC*, 2005, pp. 211–217.

[10] L. Li, C. R. Weinberg, T. A. Darden, and L. G. Pedersen, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the ga/knn method," *Bioinformatics*, vol. 17, pp. 1131–1142, 2001.

[11] T. Jirapech-Umpai and S. Aitken, "Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes," *Bioinformatics*, vol. 6, pp. 168–174, 2005.

[12] X.-R. Jiang and L. Gruenwald, "Microarray gene expression data association rules mining based on bsc-tree and fis-tree," *Data Knowl. Eng.*, vol. 53, no. 1, pp. 3–29, 2005.

[13] H. Ressom, R. Reynolds, and R. S. Varghese, "Increasing the efficiency of fuzzy logic-based gene expression data analysis," *Physiol. Genomics*, vol. 13, pp. 107–117, 2003.

[14] C. H. Q. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data." in *CSB*, 2003, pp. 523–529.

[15] M. Xiong, W. Li, J. Zhao, L. Jin, and E. Boerwinkle, "Feature (gene) selection in gene expression-based tumor classification," *Molecular Genetics and Metabolism*, pp. 239–247, 2001.

[16] L. Yu and H. Liu, "Redundancy based feature selection for microarray data," Department of Computer Science and Engineering Arizona State University," Technical Report, 2004.

[17] D. Nauck, "Neuro-fuzzy systems: Review and prospects," in *Proc. Fifth European Congress on Intelligent Techniques and Soft Computing (EUFIT'97)*, vol. 8, no. 11, 1997, pp. 1044–1053.

[18] S. Y. Belal, A. Taktak, A. Nevill, S. Spencer, D. Roden, and S. Bevan, "Automatic detection of distorted plethysmogram pulses in neonates and paediatric patients using an adaptive-network-based fuzzy inference system," *Artificial Intelligence in Medicine*, no. 24, pp. 149–165, 2002.

[19] B. Gabrys, "Combining neuro-fuzzy classifiers for improved generalisation and reliability," in *Proceedings of the Int. Joint Conference on Neural Networks (IJCNN'2002)*, Honolulu, USA, 2002, pp. 2410–2415.

[20] S. Cho and H. Won, "Machine learning in dna microarray analysis for cancer classification," in *APBC*, vol. 34, 2003, pp. 189–198.

[21] I. S. Lossos, A. A. Alizadeh, M. B. Eisen, W. C. Chan, P. O. Brown, D. Bostein, L. M. Staudt, , and R. Levy, "Ongoing immunoglobulin somatic mutation in germinal center b cell-like but not in activated b cell-like diffuse large cell lymphomas," in *Proc. of the Natl. Acad. of Sci. USA*, vol. 97, 2000, pp. 10 209–10 213.