Computers in Biology and Medicine 40 (2010) 430–437

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A method of tumor classification based on wavelet packet transforms and neighborhood rough set

Shan-Wen Zhang, De-Shuang Huang *, Shu-Lin Wang

Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui 230031, China

ARTICLE INFO

Article history:
Received 21 September 2008
Accepted 15 February 2010

Keywords:
Tumor classification
Gene expression profiles (GEP)
Wavelet packet transforms (WPT)
Neighborhood rough set (NRS)

ABSTRACT

Tumor classification is an important application domain of gene expression data. Because of its characteristics of high dimensionality and small sample size (SSS), and a great number of redundant genes not related to tumor phenotypes, various feature extraction or gene selection methods have been applied to gene expression data analysis. Wavelet packet transforms (WPT) and neighborhood rough sets (NRS) are effective tools to extract and select features. In this paper, a novel approach of tumor classification is proposed based on WPT and NRS. First the classification features are extracted by WPT and the decision tables are formed, then the attributes of the decision tables are reduced by NRS. Thirdly, a feature subset with few attributes and high classification ability is obtained. The experimental results on three gene expression datasets demonstrate that the proposed method is effective and feasible.

1. Introduction

DNA microarray technology has already found many applications in the field of gene discovery, disease diagnosis, drug discovery, and toxicology research. Among them, tumor classification is an important application domain of gene expression data, which has a promising future in clinical medicine. Due to the natural characteristics of small-size-sample and high dimensionality of gene expression data, more and more new prediction, classification and clustering techniques are being used for analysis of the data [1–4].

Many clustering algorithms have been proposed for the analysis of gene expression data. Clustering algorithms attempt to partition the genes into groups exhibiting similar patterns of variation in expression level, but little guidance is available to help choose among them. The choice of suitable method for a given experimental dataset is not straightforward. Kerr et al. [5] surveyed the state of the art applications and provided a framework for the evaluation of clustering in gene expression analyses. Yang et al. [6] provided a systematic framework for assessing the results of clustering algorithms. Because of the small-size-sample problem, many clustering methods are not effective for gene expression data analysis.

Dimension reduction and feature extraction has many applications in bioinformatics and computational biology. Independent component analysis (ICA) has been developed in GEP analysis. In [7], the sequential floating forward selection (SFFS) technique is used to select the independent components of the GEP data for classification. At the same time, feature selection is a process that selects a subset of original features by reducing the number of features, removing the irrelevant, redundant, or noisy data. Feature selection has been extensively applied to the tumor classification using GEP data.

Statistical learning is useful for gene expression analysis. Some of the informative gene selection methods are based on gene ranking such as *T*-statistic rank criterion [8,9] and Fisher's discriminant criterion [10]. Jaeger [11] compared classification performance with five different test statistics: Fisher (in [12]), Golub (in [13]), Wilcoxon (in [14]), TNoM (in [15]), and *T*-test on three tumor datasets. Among these, many gene-ranking methods require the dataset following Gaussian distribution. To avoid the assumption of the normality condition, Deng et al. [16] proposed a rank sum test method for informative gene, and the corresponding experiments showed that this method outperforms the previous gene ranking methods. The general method to solve this problem is wrapper method that combines gene selection with classifier, which is a family of methods combining sequential forward search with different classifiers [17]. With regard to how to evaluate the goodness (quality) of a subset of features, the feature selection methods fall into two broad categories: the filter approach and the wrapper approach [18]. In the filter approach, a good feature set is selected as a result of pre-processing based on properties of the data itself and independent of the classification algorithm. The wrapper approach requires one predetermined mining algorithm in feature selection and uses its performance to evaluate and determine which feature is selected. It tends

* Corresponding author.
E-mail address: dshuang@iim.ac.cn (D.-S. Huang).

0010-4825/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.compbiomed.2010.02.007

to give the superior performance as it finds features better suited to the predetermined classification algorithm, but it is more computationally expensive than the filter approach [19]. For this reason the filter model is widely used in gene selection from GEP.

Wavelet transforms (WT) [20] seem more suitable for scaling biological structures than other mathematical transforms. WT provides economical and informative mathematical representations of many objects of interest. In recent years, WT has been applied to a large variety of biomedical signal analysis [21,22]. Moreover, WT is capable of providing analysis in a global fashion, which is necessary in case of GEP data analysis. Surveys of wavelet applications in GEP data are presented at [23,24]. Local wavelet power spectrum (WPS) is calculated by summing the squares of the coefficient values for each band, while global WPS [25] is the average of such local power spectra. The nature of genes in different classes is different and in varying amount. So, it may be observed to analysis the WPS or entropy that it may not be same in all classes. Based on this observation, a method to select important features relevant to each category against others can be devised. Hence, there is a possibility for the tumor class discovery and prediction by monitoring gene expression using WPS.

The multi-resolution wavelet packet transforms (WPT) can process both stationary and nonstationary data and has good multi-resolution capabilities. Because of these advantages, it has been effectively used in many bioinformatics applications such as microarray data analysis and genomic data analysis. WT and WPT have been applied to the microarray data analysis. Klevecz [25] used wavelet decomposition and denoising techniques to analyze GEP and found that the expression of most yeast genes oscillate, including both cell cycle regulated genes and ones not related to the cell cycle. Wang et al. [26] proposed a method of feature extraction and tumor classification based on WPT and SVM. The features for the tumor classification can be extracted from GEP by WPT and used as the input of SVM classifier to classify the tumor classes. The experiments proved that their method can meet real-time application requirements in clinical domain. Myasnikova et al. [27] used WT to analyze gene expression measured by using tagged antibodies in a set of embryos. They obtained a detailed gene expression map of a morphogenetic field from fragmentary data. Efron et al. [28] showed that the 'False Discovery Rate' (FDR) is a very useful approach, which might lead to other applications of WT in GEP data analysis.

Rough set (RS) theory, proposed by Pawlak [29], can be seen as a new mathematical approach for vague questions. RS have been applied mainly in mining tasks like classification, clustering and feature selection. A primary use of RS is to reduce the number of attributes in databases thereby improving the performance of applications in a number of aspects including speed, storage, and accuracy [30,31]. The microarray data often consists of small number of samples and large number of genes. The ultra high dimension of GEP makes it necessary to develop effective feature selection methods in order to reduce the computation cost and improve the classification accuracy. RS provides a feasible way to deal with redundancy. The aim of reduction in RS is to find out a minimum set of relevant attributes (features) that describe the dataset as well as all the original attributes do. Zhou et al. [32] proposed a novel feature selection method based on mutual information and RS. They selected some top-ranked features, which have higher mutual information for the target class to predict. Then RS is applied to remove the redundancy among these selected genes. Binary particle swarm optimization is first proposed for attribute reduction in RS. Banerjee et al. [33] presented an evolutionary rough feature selection algorithm, which is used for classifying GEP patterns.

Although feature reduction in classical RS is an effective reduction method, attributes must be discretized before reduction, which leads to information loss. However, neighborhood rough sets (NRS) model [34,35] omits the discretization procedure, so no information loss occurs. NRS based feature selection algorithm is able to delete most of the redundant and irrelevant features. A quick search of biological literatures shows that NRS is still seldom used in bioinformatics. Based on WPT and NRS, a tumor classification method is proposed in this paper. The main contributions of the paper are stated as follows: (1) the GEP data is decomposed by WPT and the classification features are extracted from the WPT coefficients, which are robust against the noises; (2) classification feature subsets are reduced and selected by NRS, which is able to delete most of the redundant and irrelevant features; and (3) the experiment results are analyzed in detail.

The rest of the paper is organized as follows. WPT is introduced in Section 2.1. In Sections 2.2 and 2.3, a lot of concepts of NRS and an attribute reduction algorithm are introduced. In Section 2.4, based on WPT and NRS, the tumor classification algorithm is described. In Section 2.5, three classifiers are discussed. The experimental results are presented in Section 3. Finally, concluding remarks and future works are given in Section 4.

2. Methods

2.1. Wavelet packet transforms (WPT)

WPT is an extension of the discrete WT to the full binary tree. In the discrete wavelet packet transform (DWPT), both the scaling and wavelet coefficients are subject to the high-pass and low-pass filtering when computing the next layer scaling and wavelet coefficients.

With the standard transforms, scaling coefficients identify the frequency band $[0, 1/2^{J+1}]$, with J, the coarsest layer, while wavelet coefficients at j-layer describe the frequency band $[1/2^{j+1}, 1/2^{j}]$. DWPT induces a finer partition of the frequency space. It is a collection of the functions $\{2^{-j/2}u_n(2^{-j}t-k),\ j,k\in Z, n\in Z_+\}$ generated from the following sequence of functions [36]:

$$u_{2n}(t)=\sqrt{2}\sum_{k\in z}h_k u_n(2t-k);\quad u_{2n+1}(t)=\sqrt{2}\sum_{k\in z}g_k u_n(2t-k) \tag{1}$$

where h and g are the quadrature mirror filters, $\sum_{n\in Z}h_{n-2k}h_{n-2l}=\delta_{kl}$, $\sum_{n\in Z}h_n=\sqrt{2}$, $g_k=(-1)^k h_{1-k}$, $k\in Z$, $u_0(t)$, and $u_1(t)$ are the scaling function and basic wavelet, respectively. The wavelet packet $\{2^{-j/2}u_n(2^{-j}t-k), j,k\in Z, n\in Z_+\}$ is a localized function of unit energy with scale 2^j, translation $2^j k$ and an oscillation parameter of n.

For a discrete signal, the decomposition coefficients of wavelet packets can be computed iteratively by

$$x^k_{2n,j+1}=\sum_l h_{l-2k}x^l_{n,j};\quad x^k_{2n+1,j+1}=\sum_l g_{l-2k}x^l_{n,j} \tag{2}$$

where $x^k_{n,j}$ is the decomposition coefficient sequence of the nth node at j-layer of the wavelet packet tree, generally, $j=1,2,3,4$.

The original signal can be reconstructed iteratively by

$$x^l_{n,j}=\sum_k h_{l-2k}x^k_{2n,j+1}+\sum_k g_{l-2k}x^k_{2n+1,j+1} \tag{3}$$

The wavelet packet decomposition tree is shown in Fig. 1 in the supplementary material. Each node corresponds to a frequency band. The leaf nodes of any connected sub-tree that has the same root node as the full tree form an orthonormal basis and can represent a signal of finite energy completely.

Input: $RD =< U, C, D >$ and δ

Step 1: $red = \varnothing$; //Initial red is set to empty set and is the pool to contain the informative attribute.

Step 2: For each $\forall a_i \in C - red$

Computing $SIG(a_i, red, C) = \gamma(C, red \cup a_i) - \gamma(C, red)$

// where, define $\gamma(C, \varnothing) = 0$, the dependency of subtype table

//the set C with respect to empty set is fixed to zero.

Step 3: Selecting the informative gene a_k satisfying

$$SIG(a_k, red, C) = \max_i (SIG(a_i, red, C))$$

Step 4: If $SIG(a_k, red, C) > 0$

// Add the optimal gene a_k to the reduction set red

$$red = red \cup a_k$$

go to Step2;

else

return red ;

Step 5: Algorithm end;

Output: red //reduction of condition attribute set C

Fig. 1. Forward attribute reduction based on neighborhood mode (FARNeM algorithm).

2.2. Neighborhood rough set (NRS)

In this subsection, we will introduce the basic concepts of RS theory and NRS algorithm related to our attribute reduction approach. The basic concepts of neighborhood and NRS are explained as follows [34,37–42].

Formally, the structural data for classification can be written as $RD = \langle U, A, V, f \rangle$, where U is the nonempty set of samples, $U = \{x_1, x_2, \ldots, x_n\}$, $A = \{a_1, a_2, \ldots, a_m\}$ is the nonempty set of variables (also called as features, inputs, attributes) to characterize the samples, V_a is the value domain of attribute a, and f is an information function, $f : U \times A \to V$. More specially, RD is also called a decision table if $A = C \cup D$ and $C \cap D = \emptyset$, when C is the set of condition attribute; D is the output, also called decision attribute set.

Definition 1. Given arbitrary sample $x_i \in U$ and a subset $B \subseteq C$, the neighborhood $\delta_B(x_i)$ of x_i in the subspace B is defined as

$$\delta_B(x_i) = \{x_j | x_j \in U, \Delta_B(x_i, x_j) \le \delta\} \tag{4}$$

where Δ_B is a metric function. For $\forall x_1, x_2, x_3 \in U$, it satisfies

- $\Delta_B(x_1, x_2) \ge 0$;
- $\Delta_B(x_1, x_2) = 0$, if and only $x_1 = x_2$;
- $\Delta_B(x_1, x_2) = \Delta_B(x_2, x_1)$;
- $\Delta_B(x_1, x_3) \le \Delta_B(x_1, x_2) + \Delta_B(x_2, x_3)$.

Definition 2. Given a set of samples U, R is a neighborhood relation on U, $\{\delta_B(x_i) | x_i \in U\}$ is the family of neighborhood granules. Then we call $RD = \langle U, R \rangle$ a neighborhood approximation space.

Definition 3. Given $RD = \langle U, R \rangle$, for arbitrary subset $X \subseteq U$, two subsets of objects, called lower and upper approximations of X in terms of relation R, are defined as

$$\underline{RD}(X) = \{x_i | \delta_B(x_i) \subseteq X, x_i \in U\}$$

$$\overline{RD}(X) = \{x_i | \delta_B(x_i) \cap X \ne \emptyset, x_i \in U\} \tag{5}$$

Given a neighborhood decision table $RD = \langle U, A = C \cup D, V, f \rangle$, $X_1, X_2, \ldots, X_N$ are the object subsets with decisions from 1 to N, $\delta_B(x_i)$ is the neighborhood information granules including x_i and generated by attribute subset $\forall B \subseteq C$, then the lower and upper approximations of the decision D with respect to attributes B are defined as

$$\underline{RD}(X) = \bigcup_{i=1}^{N} \underline{RD}(X_i); \quad \overline{RD}(X) = \bigcup_{i=1}^{N} \overline{RD}(X_i) \tag{6}$$

where

$$\underline{RD}(X_i) = \{x_i | \delta_B(x_i) \subseteq X, x_i \in U\}$$

$$\overline{RD}(X_i) = \{x_i | \delta_B(x_i) \cap X \ne \emptyset, x_i \in U\}$$

The boundary region of X in the approximation space is formulated as

$$BN(X) = \overline{RD}(X) - \underline{RD}(X) \tag{7}$$

The size of the boundary region reflects the degree of roughness of the set X in the approximation space. Assuming that X is the sample subset with a decision table, usually we hope that the boundary region of the decision is as little as possible for decreasing uncertainty in decision. The sizes of the boundary regions depend on X, B and U, and the threshold δ. Decision boundary is the object subset whose neighborhoods come from more than one decision class. On the other hand, the lower approximation of the decision, also called positive region of decision, denoted by $Pos_B(D)$, is the subset of objects whose neighborhoods consistently belong to one of the decision classes.

The neighborhood model divides the samples into two groups: positive region and boundary. Positive region is the sample set, which can be classified into one of the classes without uncertainty with the existing attributes, while boundary is the set of samples, which cannot be determinately classified. The samples in different feature subspaces will have different boundary regions. The size of the boundary region reflects the resolving-power of the classification problem in the corresponding subspaces. It also reflects the recognition power or characterizing power of the condition attributes. The greater the boundary region is, the weaker the characterizing power of the condition attributes will be. It can be formulated as follows.

Definition 5. Given $RD = \langle U, C \cup D \rangle$, $B \subseteq C$, the dependency degree of D to B is defined as the ratio of consistent objects:

$$\gamma_B(D) = |Pos_B(D)| / |U| \tag{8}$$

where $\gamma_B(D)$ reflects the ability of B to approximate D.

Definition 6. Given $RD = \langle U, C \cup D \rangle$, $B \subseteq C$, we define the significance of a arbitrary attribute a as

$$SIG(a, B, D) = \gamma_B(D) - \gamma_{B-a}(D) \tag{9}$$

The attribute's significance is the function of variables: a, B and D. One thing should be explained that an attribute may be of great significance in subset B_1 but of little significance in subset B_2. What is more, the attribute's significance may be different for each decision if there are multiple decision attributes in a decision table.

2.3. Attribute reduction

Although adopting attribute reduction in classic RS theory to select classification features is an effective method, its classification accuracy rate is usually not higher compared with other tumor-related feature selection and tumor classification approaches, for gene feature values must be discretized before data reduction,

which leads to information loss in classification. Therefore, the NRS model is introduced to tumor classification, which omits the discretization procedure, so no information loss occurs before attribute reduction. An algorithm of forward attribute reduction based on neighborhood model (FARNeM) [34,35] is designed, which is used as our decision table reduction approach in following experiments.

In FARNeM algorithm, $\gamma_B(D) = |Pos_B(D)|/|U|$ denotes the dependency of decision attribute D to condition attribute subset $B \subseteq C$, where $Pos_B(D)$ is the subset of tumor samples whose neighborhoods consistently belong to one of the decision classes, and $SIG(a,B,D) = \gamma_B(D) - \gamma_{B-a}(D)$ denotes the significance condition attribute a with respect to condition attribute subset $B \subseteq C$. The FARNeM algorithm is described in Fig. 1.

The FARNeM algorithm is to find the positive region samples for evaluating the significance of attributes in the decision table that is formed in step 2. The most important problem in neighborhood-based classification is to set the threshold δ, which determines the size of the neighborhood. If δ is too small, no sample will be included in the neighborhood. On the other hand, if δ is too great, the neighborhood cannot reflect the local information of the test.

2.4. Tumor classification algorithm description

In the tumor classification problem, an important goal of the analysis of GEP is to extract features and then reduce the features. Let $=\{x_{i,j}\}$ be tumor gene expression data, where $1 \le i \le M$, $1 \le j \le N$, $N \gg M$. We extract the tumor classification features by using WPT as follows:

- *Mean*:

$$M_{n,j}^k = \frac{1}{K_{n,j}^k} \sum_{k=0}^{K_{n,j}^k - 1} x_{n,j}^k;$$

- *Energy*:

$$ER_{n,j}^k = \frac{1}{K_{n,j}^k} \sum_{k=0}^{K_{n,j}^k - 1} [x_{n,j}^k]^2;$$

- *Entropy*:

$$ET_{n,j}^k = -\sum_{k=0}^{K_{n,j}^k - 1} [x_{n,j}^k]^2 \cdot \log_2 [x_{n,j}^k]^2$$

where $K_{n,j}^k$ is the length of the decomposition coefficient sequence $X_{n,j}^k$.

The mean, energy and entropy denote the classification features, respectively, which can be constructed the classification decision table, denoted $RD = \langle U, C \cup D \rangle$, where U is a tumor sample set, C is a condition attribute set formed by $M_{n,j}^k, ER_{n,j}^k, ET_{n,j}^k$, and D is a decision attribute set or tumor subtype.

Our algorithm model can be described as follows:

Step 1: Compute the J-layer full wavelet packet tree decomposition of GEP;

Step 2: Extract classification feature vectors, and form the decision table;

Step 3: Reduce the decision table by NRS;

Step 4: Evaluate the classification results using classifier.

2.5. Classifier

Three kinds of classifiers: support vector machines (SVM), K-nearest neighbor (K-NN) and neighborhood classifier (NEC) are now in widespread use. They are introduced simply as follows.

- SVM is a relatively new type of statistic learning theory. It builds up a hyper-plane as the decision surface to maximize the margin of separation between two-class samples.
- K-NN is a most common and non parametric method. To classify an unknown sample x, K-NN extracts k closest vectors from the training set using similarity measures, and makes decision for the table of the unknown sample x using the majority class label of the k nearest neighbors. Here we adopt Euclidean distance to measure the similarity of samples.
- NEC similar to K-NN, is also based on the general idea of estimating the class of unknown sample according to its neighbors, but differing from K-NN, NEC considers a kind of neighbor within a sufficiently small and near area around the sample, in other words, all training samples surrounding the test sample take part in the classification decision process.

The MATLAB toolbox implementing SVM is freely available for academic purposes. We can download SVM from http://www.isis.ecs.soton.ac.uk/resources/svminfo/. Because there are only few samples in GEP achieved in general, SVM is often used as the classifier to classify tumor samples using GEP, which have been proven to be very useful for tumor classification [43].

3. Experimental results

In this section, we will carry out a set of experiments on three public GEP datasets to show the effectiveness of our proposed method for tumor classification. Although all data samples in these three datasets have already been assigned to a training set or testing set, in order to obtain reliable experimental results, we adopt 4-fold cross-validation to classify tumor samples in all experiments.

3.1. Experiments on the SRBCT

From http://research.nhgri.nih.gov/microarray/Supplement, we can download the small round blue cell tumor (SRBCT) dataset which contain 88 samples with 2308 genes in each sample. Among them, five non tumor-related samples are removed in following experiments. To avoid over-fitting in the classifier, we design experiments on all 83 samples using 4-fold cross-validation to evaluate the classification model.

Firstly, we compute 4-layer full wavelet packet tree decomposition of GEP of 83 samples. The wavelet functions tested for this project are Haar, Debauchies 2, Debauchies 4, Debauchies 8, Biorthogonal 2.2, Biorthogonal 3.7, and Biorthogonal 6.8, respectively. The Haar wavelet is chosen because it is the most simple wavelet basis. The Debauchies (“Db”) family bases are chosen because of their properties of compact support and orthonormality. The Biorthogonal (“Bior”) wavelets are chosen for their property of exact reconstruction. In experiments, we choose different wavelet functions, and utilize wavelet packet decomposition function ‘wpdec’ and ‘wpcoef’ in MATLAB toolbox to decompose the GEP data, respectively. For k-th tumor sample, the wavelet packet decomposition coefficient of the nth node at layer j of the wavelet packet tree is denoted as $x_{n,j}^k$, $j=1,2,3,4$. For example, the gene expression profiles ‘EWS-T11’ and its wavelet coefficients at note (4.1)–(4.4) are, respectively, shown in Figs. 2 and 3 in the supplementary material.

Secondly, the feature vectors can be extracted from $\{x_{n,j}^k\}$, denoted as $M_{n,j}^k, ER_{n,j}^k, ET_{n,j}^k$, which, respectively, corresponds to the

mean, energy and entropy of the nth node at layer j of the wavelet packet tree. Before they are used, they must be normalized with mean zero and variance one. For training samples, we form four decision tables (U denotes tumor sample set and D is a decision attribute set, i.e. tumor subtype), denoted as follows:

$$RD1 = \langle U, ER^k_{n,j}, D \rangle$$

$$RD2 = \langle U, ET^k_{n,j}, D \rangle$$

$$RD3 = \langle U, (ER^k_{n,j}, ET^k_{n,j}), D \rangle$$

$$RD4 = \langle U, (M^k_{n,j}, ER^k_{n,j}, ET^k_{n,j}), D \rangle \quad (10)$$

Since 4-layer WPT decomposition graph has 30 crunodes, as shown in Fig. 1 in the supplementary material, the decision tables $RD1$, $RD2$, $RD3$ and $RD4$ have 30, 30, 60 and 90 attributes, respectively.

Thirdly, we conduct attribute reduction by using FARNeM algorithm, which requires setting the neighborhood parameter δ, $0 < \delta < 1$. For a constant δ, an attribute subset can be obtained by using FARNeM algorithm, so 100 subsets can be obtained as δ varying from 0 to 1 by step 0.01.

Fourthly, we adopt three classifiers, SVM-RBF (SVM with radial basis function $K(x,y) = \exp(-\beta\|x-y\|^2)$), K-NN and NEC, to select the optimal attribute subsets with the highest accuracy. SVM-RBF needs two parameters: α and β; K-NN classifier needs a parameter K; NEC classifier requires a parameter w varying from 0 to 0.6.

Finally, for test samples, we extract the optimal feature subset through the attribute reduction results on training set.

Having done many experiments, by 4-fold cross-validated method, we have obtained a lot of classification results. The classification results on table $RD1$ using classifiers SVM-RBF, K-NN and NEC are listed in Tables 1–3, respectively.

From Tables 1–3, it is found that SVM-RBF outperforms K-NN and NEC. So in the following experiments, we choose SVM-RBF as classifier to classify the tumor samples. The classification results on the decision tables $RD2$, $RD3$ and $RD4$ using SVM-RBF are listed in Tables 4–6, respectively.

From Tables 1–6, it is found that the experiment results in Table 5 are higher than that in other five Tables in general, and selecting 'Db8' as wavelet function is more suitable. So in the

Table 1
The classification accuracy rates (%) of decision table $RD1$ using SVM-RBF classifier on SRBCT.

Wavelet function	α	β	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	200	0.002	0.72	24	14	83.83
Db2	200	0.0003	0.75	30	12	84.67
Db4	200	0.0002	0.84	26	13	84.87
Db8	200	0.0004	0.88	26	10	85.37
Bior2.2	200	0.0003	0.74	30	12	85.25
Bior3.7	200	0.0002	0.82	28	14	84.93
Bior6.8	200	0.0004	0.90	28	13	84.85

Table 2
The classification accuracy rates (%) of decision table $RD1$ using K-NN classifier on SRBCT.

Wavelet function	K	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	3	0.78	30	13	80.36
Db2	3	0.78	30	15	82.44
Db4	3	0.78	30	15	79.89
Db8	3	0.85	30	14	82.28
Bior2.2	3	0.74	30	16	81.34
Bior3.7	3	0.80	30	14	79.86
Bior6.8	3	0.80	30	12	79.63

Table 3
The classification accuracy rates (%) of decision table $RD1$ using NEC classifier on SRBCT.

Wavelet function	w	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	0.02	0.81	30	14	79.62
Db2	0.02	0.85	30	15	79.64
Db4	0.02	0.83	30	13	81.21
Db8	0.02	0.88	30	12	80.87
Bior2.2	0.02	0.74	30	14	80.16
Bior3.7	0.02	0.80	30	14	80.43
Bior6.8	0.02	0.88	30	13	79.64

Table 4
The classification accuracy rates (%) of decision table $RD2$ using SVM-RBF classifier on SRBCT.

Wavelet function	α	β	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	200	0.002	0.71	30	12	90.23
Db2	200	0.0003	0.75	30	12	100
Db4	200	0.0002	0.80	30	13	94.58
Db8	200	0.0004	0.88	30	12	95.46
Bior2.2	200	0.0003	0.74	30	12	95.13
Bior3.7	200	0.0002	0.82	30	12	94.57
Bior6.8	200	0.0004	0.85	30	13	91.26

Table 5
The classifying accuracy rates (%) of decision table $RD3$ using SVM-RBF classifier on SRBCT.

Wavelet function	α	β	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	200	0.0002	0.71	60	12	98.42
Db2	200	0.0003	0.75	60	14	100
Db4	200	0.0002	0.80	60	13	100
Db8	200	0.0004	0.88	60	14	100
Bior2.2	200	0.0003	0.74	60	13	100
Bior3.7	200	0.0002	0.82	60	14	99.94
Bior6.8	200	0.0004	0.90	60	14	99.89

Table 6
The classification accuracy rates (%) of decision table $RD4$ using SVM-RBF classifier on SRBCT.

Wavelet function	α	β	δ	Original attribute number	Reduced attribute number	Classifying accuracy
Haar	200	0.0002	0.70	90	20	99.86
Db2	200	0.0003	0.72	90	20	100
Db4	200	0.0002	0.75	90	20	99.76
Db8	200	0.0004	0.72	90	20	99.85
Bior2.2	200	0.0003	0.75	90	22	100
Bior3.7	200	0.0002	0.75	90	22	99.57
Bior6.8	200	0.0004	0.70	90	22	99.69

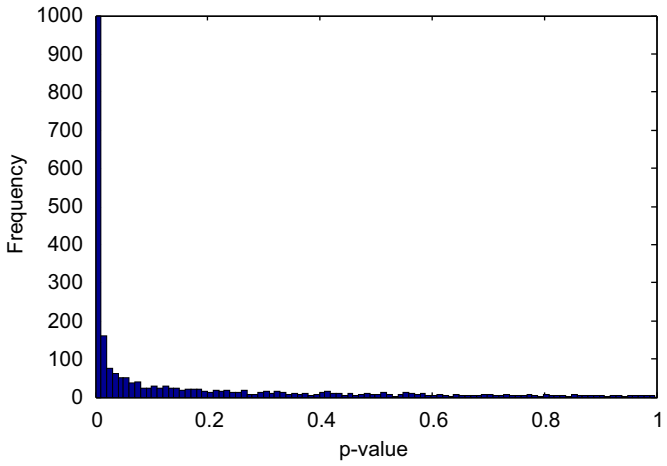

Fig. 2. The distribution of p-value for genes of SRBCT.

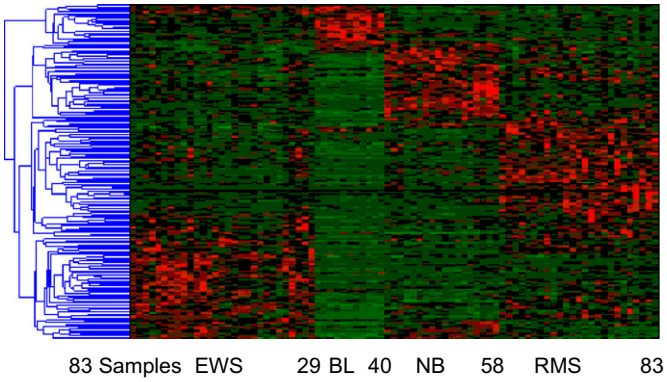

Fig. 3. Gene expression profiles of SRBCT with 200 top-ranked genes.

following experiments, we will choose *RD*3, 'Db8' and SVM-RBF to implement the tumor classification experiments.

Though the WPT is very useful for noise reduction, it is found that if we use all 2308 genes in each sample, the experiment results are not steady sometimes. In order to solve the problem, we rank the genes of the gene expression data by adopting Kruskal–Wallis rank sum test [44] to select the informative genes. The distribution of p-value for every gene of SRBCT is shown in Fig. 2.

From Fig. 2, we can see that the gene number with larger p-value is about 200, so we simply select 200 top-ranked genes as initial informative genes. The heat map of the 200 top-ranked genes is shown in Fig. 3, in which the consistent differentia is obvious among the four tumor subtypes, i.e., 29 Ewing family of tumors (EWS), 11 Burkitt lymphomas (BL), 18 Neuroblastoma (NB) and 25 Rhabdomyosarcoma (RMS). Therefore, we simply select 200 top-ranked genes used as the initial informative genes that contain complete classification information, which also indicates that the method of adopting Kruskal–Wallis rank sum test to rank gene is very effective.

Using 200 top-ranked genes of each sample, we implement repeatedly above experiments on the decision table *RD*3, where the δ-value vary from 0 to 1 with step 0.1; the classifier is SVM-RBF; wavelet function is 'Db8'; $\alpha=200$, $\beta=0.0003$, and the 4-fold cross-validated is adopted. The mean classification results versus the neighborhood δ are shown in Fig. 4. The mean classification results versus the number of feature attributes are shown in Fig. 5.

From Figs. 4 and 5, the classification results indicate that selecting 200 top-ranked genes used as informative genes by Kruskal–Wallis rank sum test is effective.

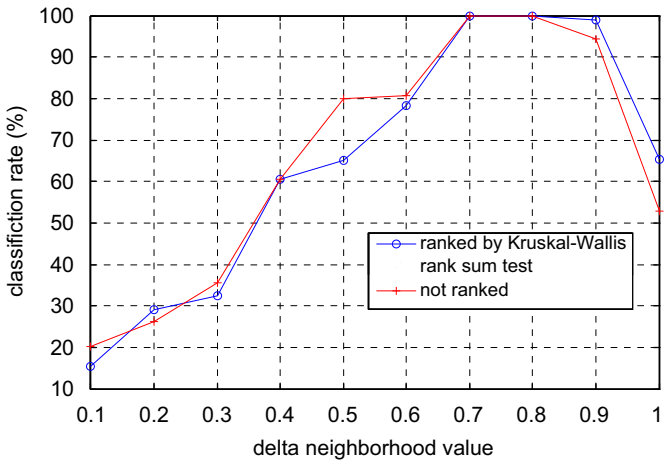

Fig. 4. The classifying accuracy versus δ neighborhood on the SRBCT dataset. The blue line is accuracy rate which ranked by Kruskal–Wallis rank sum test, the red line is accuracy rate which is not ranked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

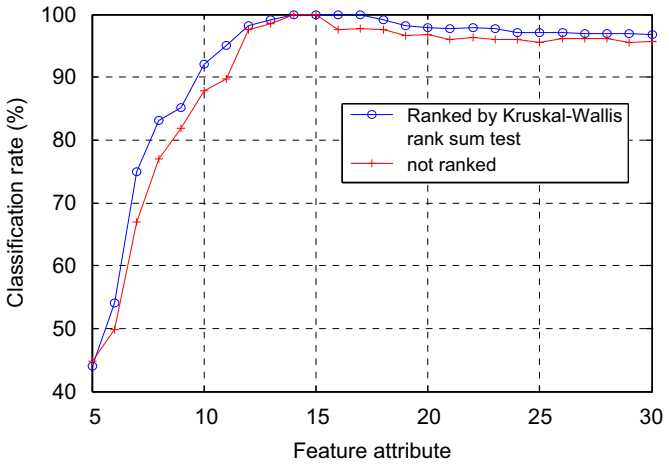

Fig. 5. The classifying accuracy versus the number of feature attribute on the SRBCT dataset. The blue line is accuracy rate which ranked by Kruskal–Wallis rank sum test, the red line is accuracy rate which is not ranked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Comparing the above experimental results, we find that the optimal scheme is to use wavelet Db8, decision table *RD*3, SVM-RBF classifier, $\alpha=200$, $\beta=0.0002$–0.0004, $\delta=0.6$–0.8, and choose top-ranked genes by Kruskal–Wallis rank sum test. We have also found that the mean feature attribute M_{n_j} is redundant.

From the above experiments, we draw some interesting conclusions. It is easy to rank the genes by Kruskal–Wallis rank sum test, as shown in Fig. 2. It takes about 2–4 s to rank all genes of SRBCT data. Ranking gene is only regarded as the first preprocessing. FARNeM is an effective attribute reduction algorithm, but when the number of attributes is large, it will cost more CPU time to reduce the attributes. The optimal classification method is described as follows: (1) rank the gene expression profiles by Kruskal–Wallis rank sum test and obtain 200 top-p-value genes; (2) compute the wavelet packet transforms and extract the feature vectors from the wavelet packet decomposition coefficients, then form the decision tables; and (3) reduce the attributes by FARNeM; finally classify the tumor samples by SVM classifier.

Table 7
Descriptions of two tumor datasets in our experiments.

Tumor dataset	Gene	Sample	Subtype 1	Subtype 2
Leukemia	7129	72	47 (ALL)	25 (AML)
Colon	2000	62	40 (Tumor)	22 (Normal)

Table 8
Comparison results of mean classification accuracy rate using SVM-RBF.

Dataset	Method				
	WPT+SVM [26] (%)	DWT+SVM [45] (%)	NRS+SVM [46] (%)	MSVM [47] (%)	WPT+NRS+SVM (Ours) (%)
SRBCT	100	98.52	100	100	100
Leukemia	94.54	93.04	93.41	93.82	95.51
Colon	90.14	89.12	90.87	90.26	90.46

3.2. Experiments on leukemia and colon data

In this subsection, we further investigate the performance of our proposed method on two well-known datasets: the leukemia dataset [13] and the colon dataset [45]. Two datasets contain only two subclasses, presented in Table 7.

After selecting the 200 top-ranked gene set by Kruskal–Wallis rank sum test, we conduct wavelet packet transforms and extract the feature vectors, then form the decision table *RD*3; reduce the attributes by the FARNeM algorithm; classify the tumor samples by SVM classifier, where wavelet function is 'Db8', parameters $\alpha=200$, $\beta=0.0003$, $\delta=0.7$. In the experiments, 4-fold cross-validation method is adopted. The mean classification rates on the leukemia data and colon data are 95.53% and 90.47%, respectively.

3.3. Performance comparison with other methods

To validate the effectiveness of the proposed method, we compare with other tumor classification methods [26,45–47]. Similar to the method in [26], Liu [45] used wavelet transform (WT) and SVM to classify tumor samples into different diagnostic classes. Wang et al. [46] proposed a *neighborhood rough set model* (*NRS*) *based gene selection for* multi-subtype tumor classification. They adopt Kruskal–Wallis rank sum test to rank all genes and then apply NRS model to gene reduction to obtain gene subsets with fewer genes and more classification ability. Lee et al. [47] proposed a multi-category SVM (MSVM), which is a recently proposed extension of the binary SVM. They applied MSVM to multi-class cancer classification problems.

We select 'Db8' as wavelet function, and also set $\alpha=200$, $\beta=0.0003$, $\delta=0.7$ and adopt SVM-RBF classifier and 4-fold cross-validation in experiments. Table 8 presents the performance comparison with other related work on the same three datasets: SRBCT, Leukemia and Colon tumor. From Table 8, we can find that our proposed method always outperforms the other methods. The comparison results show that our method is effective and feasible.

4. Conclusion

Although DNA microarray experiments provide us with huge amount of gene expression information, only a few of genes are related to tumor. It is difficult to select informative genes related to tumor from GEP because of the characteristics, such as high dimensionality, small-size-sample and noise of GEP. How to analyze and handle these data, and mine out valuable biological and medical knowledge, has become a bottleneck and hotspot in the research of post-genomic age. The ultra high dimension of GEP data makes it necessary to develop effective feature extraction and selection methods in order to reduce the computation cost and improve the classification accuracy. This paper proposed a tumor classification method based on the WPT and NRS. Using WPT, we can simply extract the feature vectors, which are robust against the noise, and reduce the attributes by NRS which speed up the tumor classification by SVM-RBF classifier. Various wavelet functions, three kind of classifiers and four decision tables were tested and their performances were measured. In most cases the project by using wavelet Db8, decision table *RD*3, SVM-RBF classifier, $\alpha=200$, $\beta=0.0003$, $\delta=0.7$ performed better than other programs in terms of classification accuracy. The experiments on the SRBCT, Leukemia and Colon show that the proposed method can achieve high classification rates. The comparison results show that our method is effective.

In experiments, there are several problems need to further research, such as how to select wavelet function, δ neighborhood and parameters of the SVM classifier, etc. In future works the performance can be enhanced by tuning the parameter values and making the classification better.

Conflict of interest statement

None conflict.

Acknowledgments

This work was supported by the Grants of the National Science Foundation of China, nos. 60705007, 60472111.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compbiomed.2010.02.007.

References

[1] M. Granzow, D. Berrar, W. Dubitzky, A. Schuste, F. Azuaje, R. Eils, Tumor classification by gene expression profiling: comparison and validation of five clustering methods, ACM SIGBIO Newsletter 21 (2001) 16–22.
[2] D. Jiang, C. Tang, A. Zhang, Cluster analysis for gene expression data: a survey, IEEE Transactions on Knowledge and Data Engineering 11 (2004) 1370–1386.
[3] T.D. Pham, C. Wells, D.I. Crane, Analysis of microarray gene expression data, Current Bioinformatics (2003) 37–53.
[4] H.A. Musa, C. Dilek, D. Omer, S.I. Mehmet, Gene expression profile classification: a review, Current Bioinformatics 1 (2006) 55–73.
[5] G. Kerr, H.J. Ruskin, M. Crane, P. Doolan, Techniques for clustering gene expression data, Computers in Biology and Medicine 38 (3) (2008) 283–293.
[6] K.Y. Yeung, D.R. Haynor, W.L. Ruzzo, Validating clustering for gene expression data, Bioinformatics 17 (4) (2001) 309–318.
[7] C.H. Zheng, D.S. Huang, L. Shang, Feature selection in independent component subspace for microarray data classification, Neurocomputing 69 (2006) 2407–2410.
[8] D. Chen, Z. Liu, X. Ma, et al., Selecting genes by test statistics, Journal of Biomedicine and Biotechnology 2 (2005) 132–138.
[9] T.S. Furey, N. Christianini, N. Duffy, et al., Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinformatics 10 (2000) 906–914.
[10] M.M. Xiong, W.J. Li, J.Y. Zhao, et al., Feature (gene) selection in gene expression-based tumor classification, Molecular Genetics and Metabolism 73 (2001) 239–247.
[11] J. Jaeger, R. Sengupta, W.L. Ruzzo, Improved gene selection for classification of microarrays, Pacific Symposium on Biocomputing 8 (2003) 53–64.
[12] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.

[13] T.R. Golub, D.K. Slonim, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[14] P.J. Park, M. Pagano, M. Bonetti, A nonparametric scoring algorithm for identifying informative genes from microarray data, PSB (2001) 52–63.
[15] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, Z. Yakhini, Tissue classification with gene expression profiles, RECOMB, 2000.
[16] D. Lin, J.W. Ma, J. Pei, Rank sum method for related gene selection and its application to tumor diagnosis, Chinese Science Bulletin 15 (2004) 1652–1657.
[17] M.M. Xiong, X.Z. Fang, J.Y. Zhao, Biomarker identification by feature wrappers, Genome Research 11 (2001) 1787–1887.
[18] R. Kohavi, G. John, Wrappers for feature subset selection, Artificial Intelligence 1–2 (1997) 273–324.
[19] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.
[20] F. Abramovich, T. Bailey, T. Sapatinas, Wavelet analysis and its statistical applications, JRSSD 48 (2000) 1–30.
[21] Aldroubi, M. Unser, Wavelets in Medicine and Biology, CRC Press, Boca Raton, 1996.
[22] P. Lio, Wavelets in bioinformatics and computational biology, state of art and perspectives, Bioinformatics 19 (2003) 2–9.
[23] T. Li, Q. Li, S.H. Zhu, M. Ogihara, A survey on wavelet applications in data mining, SIGKDD Explorations 4 (2) (2002) 49–68.
[24] T.A. Kestin, D.J. Karoly, J.I. Yano, et al., Time-frequency variability of ENSO and stochastic simulations, Journal of Climate 11 (1998) 2258–2272.
[25] R.R. Klevecz, Dynamic architecture of the yeast cell cycle uncovered by wavelet decomposition, Functional and Integrative Genomics 1 (2000) 186–192.
[26] S.L. Wang, J. Wang, H.W. Chen, et al., Feature extraction and classification of tumor based on wavelet package and support vector machines, in: 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 5007 Nanjing, PEOPLES R CHINA, vol. 4426, 2007, pp. 871–878.
[27] E. Myasnikova, A. Samsonova, K. Kozlov, et al., Registration of the expression patterns of Drosophila segmentation genes by two independent methods, Bioinformatics 17 (2001) 3–12.
[28] B. Efron, J.D. Storey, R. Tibshirani, Microarray, empirical Bayes methods, and false discovery rates, Technical Report, Department of Statistics, Stanford University, see http://www-stat.stanford.edu/~tibs/research.html, 2001.
[29] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[30] C.C. Huang, T.L. Tseng, Rough set approach to case-based reasoning application, Expert Systems with Applications 26 (2004) 369–385.
[31] N. Zhong, J.Z. Dong, S. Ohsuga, Using rough sets with heuristics for feature selection, Journal of Intelligent Information Systems 16 (2001) 199–214.
[32] W.G. Zhou, C.G. Zhou, Feature selection for microarray data analysis using mutual information and rough set theory, in: IFIP International Federation for Information Processing, vol. 204, Artificial Intelligence Applications and Innovations, 2006, pp. 492–499.
[33] M. Banerjee, S. Mitra, H. Banka, Evolutionary rough feature selection in GEP, systems, man, and cybernetics, Part C: applications and reviews, IEEE Transactions on Issue 37 (2007) 622–632.
[34] Q. Hu, D. Yu, Z. Xie, Neighborhood classifiers, Expert Systems with Applications 34 (2008) 866–876.
[35] Q. Hu, D. Yu, Z. Xie, Numerical attribute reduction based on neighborhood granulation and rough approximation, Journal of Software 3 (2008) 640–649.
[36] M.V. Wickerhauser, Adapted Wavelet Analysis—From Theory to Software, A.K. Peters, Welleslay, MA, 1994.
[37] T.Y. Lin, Neighborhood systems and relational database, in: Proceedings of 1988 ACM Sixteenth Annual Computer Science Conference, February 1988, pp. 23–25.
[38] T.Y. Lin, Neighborhood systems-application to qualitative fuzzy and rough sets, in: P.P. Wang (Ed.), Advances in Machine Intelligence and Soft-computing, Department of Electrical Engineering, Duke University Durham, North Carolina, USA, 1997, pp. 132–155.
[39] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Transactions of Knowledge and Data Engineering 16 (2004) 1457–1471.
[40] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (2003) 833–849.
[41] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Information Sciences 111 (1998) 239–259.
[42] W.Z. Wu, W.X. Zhang, Neighborhood operator systems and approximations, Information Sciences 144 (2002) 201–217.
[43] T.S. Furey, N. Cristianini, N. Duffy, et al., Support vector machines classification and validation of cancer tissue samples using microarray expression data, Bioinformatics 16 (2000) 906–914.
[44] H. William Kruskal, W. Allen Wallis, Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association 260 (1952) 583–621.
[45] Y.H. Liu, Cancer Identification based on DNA Microarray Data, vol. 4819, Springer, Berlin/Heidelberg, 2007, pp. 153–161.
[46] S. Wang, X. Li, S. Zhang, Neighborhood rough set model based gene selection for multi-subtype tumor classification, ICIC 2008, Lecture Notes in Computer Science, vol. 5226, 2008, pp. 146–158.
[47] Y.K. Lee, C.K. Lee, Classification of multiple cancer types by multicategory support vector machines using GEP, Bioinformatics 9 (2003) 1132–1139.

Shan-wen Zhang, from Xi'an City, China, Born on August 20, 1965. From September 1984 to July 1988, studied for B.Sc. degree in Mathematics at Northwest University, Xi'an, China, and obtained B.Sc. degree in Jul.1988. From July 1988 to September 2007, worked at Missile Institute at Air Force Engineering University, Xi'an, China. From September 1992 to April 1995, in pursuit for M.Sc. degree in Computer Science and Technology at Northwest Polytechnic University, Xi'an, China, and obtained M.Sc. degree in July 1995. From September 1998 to July 2001, in pursuit of Ph.D degree in Electromagnetic Field and Microwave Technology at Air Force Engineering University, Xi'an, China, and obtained Ph.D. degree in July 2001. From October 2007 on, work as Postdoctor at Intelligent Computing Lab, Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China. Research interests: Wavelet Transforms, Rough Sets, Genetic Algorithm, Pattern Recognition and Artificial Neural Networks.

De-Shuang Huang (SM'98) received the B.Sc., M.Sc. and Ph.D. degrees all in electronic engineering from Institute of Electronic Engineering, Hefei, China, National Defense University of Science and Technology, Changsha, China and Xidian University, Xian, China, in 1986, 1989 and 1993, respectively. During 1993–1997 period he was a postdoctoral student, respectively, in Beijing Institute of Technology and in National Key Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, China. In September 2000, he joined the Institute of Intelligent Machines, Chinese Academy of Sciences as the Recipient of "Hundred Talents Program of CAS". From September 2000 to March 2001, he worked as Research Associate in Hong Kong Polytechnic University. From April 2002 to June 2003, he worked as Research Fellow in City University of Hong Kong. From August to September 2003, he visited the George Washington University as Visiting Professor, Washington DC, USA. From October to December 2003, he worked as Research Fellow in Hong Kong Polytechnic University. From July to December 2004, he worked as the University Fellow in Hong Kong Baptist University. Dr. Huang is currently a senior member of the IEEE. He has published over 190 papers. Also, in 1996, he published a book entitled "Systematic Theory of Neural Networks for Pattern Recognition", which won the SecondClass Prize of the 8th Excellent High Technology Books of China and in 2001 another book entitled "Intelligent Signal Processing Technique for High Resolution Radars".

Research interests: Biology Feature Recognition, Pattern Recognition and Artificial Neural Networks.

Shu-lin Wang: Currently, he is working as Postdoctor in Intelligent Computing Lab, Heifei Institute of Intelligent Machines, Chinese Academy of Sciences, China. He obtained his Ph.D degree in the National University of Defense Technology, China, in 2007. He received his M.Sc. degree in Computer Application from the National University of Defense Technology, China, in 1997, and obtained his B.Sc. degree in Computer Application from China University of Geosciences in 1989. He also worked in Hunan University from 2000 to 2007.

Research interests: Biology Feature Recognition.