Research Article

A Novel Weighted Support Vector Machine Based on Particle Swarm Optimization for Gene Selection and Tumor Classification

Mohammad Javad Abdi,[1] Seyed Mohammad Hosseini,[1] and Mansoor Rezghi[2]

[1] *Department of Computer Sciences, Faculty of Mathematical Sciences, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran*
[2] *Department of Applied Mathematics, Sahand University of Technology, Tabriz, Iran*

Correspondence should be addressed to Mohammad Javad Abdi, mj.abdi@modares.ac.ir

We develop a detection model based on support vector machines (SVMs) and particle swarm optimization (PSO) for gene selection and tumor classification problems. The proposed model consists of two stages: first, the well-known minimum redundancy-maximum relevance (mRMR) method is applied to preselect genes that have the highest relevance with the target class and are maximally dissimilar to each other. Then, PSO is proposed to form a novel weighted SVM (WSVM) to classify samples. In this WSVM, PSO not only discards redundant genes, but also especially takes into account the degree of importance of each gene and assigns diverse weights to the different genes. We also use PSO to find appropriate kernel parameters since the choice of gene weights influences the optimal kernel parameters and vice versa. Experimental results show that the proposed mRMR-PSO-WSVM model achieves highest classification accuracy on two popular leukemia and colon gene expression datasets obtained from DNA microarrays. Therefore, we can conclude that our proposed method is very promising compared to the previously reported results.

1. Introduction

Microarray technology is a tool for analyzing gene expressions consisting of a small membrane containing samples of many genes arranged in a regular pattern. Microarrays may be used to assay gene expression within a single sample or to compare gene expression in two different cell types or tissue samples, such as in healthy and cancerous tissue. The use of this technology is increased in recent years to identify genes involved in the development of diseases. Various clustering, classification, and prediction techniques have been utilized to analyze, classify, and understand the gene expression data such as Fisher discriminant analysis [1], artificial neural networks [2], and support vector machines (SVM) [3]. Briefly, SVM is a supervised learning algorithm based on statistical learning theory introduced by Vapnik [4]. It has great performance since it can handle a nonlinear classification efficiently by mapping samples from low dimensional input space into high dimensional feature space with a nonlinear kernel function. It is useful in handling classification tasks for high-dimensional and sparse microarray data and has been recommended as an effective approach to treat this specific data structure [5–8]. Due to its many attractive characters, it has been also widely used in various fields such as image recognition, text classification, speaker identification, and medical diagnosis, bioinformatics. Therefore, our study intends to investigate the application of SVM in tumor classification problem and suggests an effective model to minimize its error rate.

It is well known that SVM assumed that all the available genes of certain gene expression data have equal weights in classification process. However, for a real tumor classification problem each gene may possess different relevance to the classification results. Thus, the genes with more relevance are more important than those with less relevance. Usually, there are two approaches to tackle this issue. One strategy is gene selection aiming at determination of a subset of genes which is most discriminative and informative for classification. The other is gene weighting which seeks to estimate the relative importance of each gene and assign it a corresponding weight [9–11]. Gene selection has attracted increasing interests in bioinformatics in recent years because

its results can effectively help cancer diagnosis and clinical treatment. In this case, many outstanding methods based on particle swarm optimization (PSO) have been developed. PSO is a new evolutionary computation technique proposed by Kennedy and Eberhart [12] which was motivated by simulations of bird flocking or fish schooling. Shen et al. [8] introduced a combination of PSO and support vector machines (SVMs) for gene selection and tumor classification problem. In their work, the modified discrete PSO was applied to select genes and SVM to diagnose colon tumor. They also proposed a combination of PSO and tabu search (TS) approaches for gene selection problem [13]. The combination of TS as a local improvement procedure and PSO enabled their algorithm to overleap local optima and showed satisfactory performance. In 2008, Chuang et al. [14] suggested an improved binary PSO. The main contribution of their work was resetting all the global best particle positions after no change in three consecutive iterations. Li et al. [15] introduced a novel hybrid of PSO and genetic algorithms (GA) for the same purpose, overcoming the local optimum problem.

On the other hand, instead of making a binary decision on a genes' relevance, gene weighting utilizes a continuous value and hence has a finer granularity in determining the relevance. The strategy proposed in this work is a combination of gene selection and gene weighting. The proposed method consists of two stages. First, we apply minimum redundancy-maximum relevance (mRMR) method, proposed by Hanchuan et al. [16], to preselect genes having the highest relevance with the target class and being maximally dissimilar to each other. Then, PSO is employed to form a novel weighted SVM (WSVM) to classify samples. In this WSVM, PSO not only discards redundant genes (gene selection), but also especially takes into account the degree of importance of each gene and assigns diverse weights to the different genes (gene weighting). To construct an accurate SVM, we also use PSO to find appropriate kernel parameters, since the choice of gene weights influences the optimal kernel parameters and vice versa. Experimental results show that our proposed method (called mRMR-PSO-WSVM) achieves higher classification rate than previously reported results.

The rest of this paper is organized as follows. The following section provides a brief description of the well-known mRMR filter method, SVM classifier, weighted SVM and PSO besides the proposed method, respectively. Experimental results and conclusions are demonstrated in Sections 3 and 4, respectively.

2. Method

2.1. Minimum Redundancy-Maximum Relevance (mRMR).
In this work a well-designed filter method, mRMR, is employed to enhance the gene selection in achieving both high accuracy and fast speed. In high-dimensional microarray data, due to the existence of a set of several thousands of genes, it is hard and even infeasible for SVM to be trained accurately. Alternative methods should be effectively applied to tackle this problem. Therefore, first of all, mRMR is applied to filter noisy and redundant genes. More

specifically, mRMR method [16] is a criterion for first-order incremental gene selection, which is warmly being studied by a great number of researchers. In mRMR, genes which have both minimum redundancy for input genes and maximum relevancy for disease classes should be selected. Thus this method is based on two important metrics. One is mutual information between disease classes and each gene, which is used to measure the relevancy, and the other is mutual information between every two genes, which is employed to compute the redundancy. Let S denote the subset of selected genes, and Ω is the set of all available genes; the minimum redundancy can be computed by

$$\min_{S \subset \Omega} \frac{1}{|S|^2} \sum_{i,j \in S} I\left(g_i, g_j\right), \qquad (1)$$

where $I(g_i, g_j)$ is the mutual information between ith and jth genes which measures the mutual dependence of these two variables. Formally, the mutual information of two discrete random variables g_i and g_j can be defined as

$$I\left(g_i, g_j\right) = \sum_{m \in g_i} \sum_{n \in g_j} p(m, n) \log \frac{p(m, n)}{p(m)p(n)}, \qquad (2)$$

where $p(m, n)$ is the joint probability distribution function of g_i and g_j, and $p(m)$ and $p(n)$ are the marginal probability distribution functions of g_i and g_j, respectively [17]. In (4), $|S|$ is the number of genes of S. In contrast, mutual information $I(T, g_j)$ is usually employed to calculate discrimination ability from gene g_i to class $T = \{t_1, t_2\}$, where t_1 and t_2 denote the healthy and tumor classes. Therefore, the maximum relevancy can be calculated by

$$\max_{S \subset \Omega} \frac{1}{|S|} \sum_{i \in S} I(T, g_i). \qquad (3)$$

Combined (5) with (6), mRMR feature selection criterion can be obtained as below in difference form:

$$\max_{S \subset \Omega} \left\{ \sum_{i \in S} I(T, g_i) - \left[\frac{1}{|S|} \sum_{i,j \in S} I\left(g_i, g_j\right) \right] \right\}. \qquad (4)$$

2.2. Support Vector Machines (SVM).
SVM classifier is briefly described as follows [18, 19]. Assume $\{x_i, y_i\}_{i=1}^{N}$ is a training dataset, where x is the input sample, and $y \in \{+1, -1\}$ is the label of classes. The SVM aim is to determine a hyper plane that optimally separates two classes using training dataset. This hyper plane is defined as $w \cdot x + b = 0$, where x is a point lying on the hyper plane, w determines the orientation of the hyper plane, and b is the bias of the distance of hyper plane from the origin. To find the optimum hyper plane, $\|w\|^2$ must be minimized under the constraint $y_i(w \cdot x_i + b) \geq 1$, $i = 1, 2, \ldots, n$. Therefore, it is required to solve the optimization problem given by

$$\min \quad \frac{1}{2} \|w\|^2$$
$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \ldots, n. \qquad (5)$$

Now, the positive slack variables ξ_i are introduced to substitute in the optimization problem and allow the method to extend for a nonlinear decision surface. The new optimization problem is given as

$$\min_{w, \xi} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \ i = 1, 2, \ldots, n, \tag{6}$$

where C is a penalty parameter which manages the tradeoff between margin maximization and error minimization. Thus, the classification decision function becomes

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} L_i y_i K\left(x_i, x_j\right) + b\right), \tag{7}$$

where L_i are Lagrange multipliers, and $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ is a kernel function which can map the data into a higher dimensional space through some nonlinear mapping function $\varphi(x)$ for a nonlinear decision system. In present work, we use radial basis function (RBF) kernel function. Consider two samples $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$ and $x_j = [x_{j1}, x_{j2}, \ldots, x_{jd}]^T$. The RBF kernel is calculated using $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, where $\gamma > 0$ is the width of Gaussian.

2.3. Weighted Support Vector Machines (WSVM).

Traditional SVMs assume that each gene of a sample contributes equally to the tumor classification results. However, this is not desirable since the quality of genes has a significant impact on the performance of a learning algorithm, and the quality of different genes is not the same. In this work, we propose a novel WSVM based on PSO. Section 2.5 describes this process in more details give the training set $\{x_i, y_i\}_{i=1}^{N}$ and the weighted vector $\alpha \in R^d$ which fulfills $\sum_{i=1}^{d} \alpha_i = 1$ for $\alpha_i \geq 0$. With respect to (5), this optimization problem can be written as follows:

$$\min \quad \frac{1}{2} \|w\|^2$$

$$\text{s.t.} \quad y_i(w \cdot \text{diag}(\alpha) \cdot x_i + b) \geq 1, \quad i = 1, 2, \ldots, n, \tag{8}$$

where $\text{diag}(\alpha) = \begin{bmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \alpha_d \end{bmatrix}$.

Substituting (8) into (6) yields the following new optimization problem

$$\min_{w, \xi} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad y_i(w \cdot \text{diag}(\alpha) \cdot x_i + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \ i = 1, 2, \ldots, n \tag{9}$$

$$\sum_{i=1}^{d} \alpha_i = 1, \ \alpha_i \geq 0.$$

1	2	3	4	$\cdots$	$n+2$
C	γ	a_1	a_2	$\cdots$	a_n

Figure 1: Solution representation.

Finally, the classification decision function becomes

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} L_i y_i K'\left(x_i, x_j\right) + b\right), \tag{10}$$

where $K'(x_i, x_j) = \exp(-\gamma \sqrt{\sum_{k=1}^{d} a_k (x_{ik} - x_{jk})^2})$ is the weighted RBF kernel.

2.4. Particle Swarm Optimization (PSO).

PSO, proposed by Kennedy and Eberhart [12], is inspired by social behavior among individuals like the birds flocking or the fish grouping. PSO consists of a swarm of particles that search for the best position according to its best solution. During each iteration, every particle moves in the direction of its best personal and global position. The moving process of a particle is described as

$$v_{id}^{t+1} = w * v_{id}^{t} + C_1 \cdot \text{rand} \cdot \left(p_{id}^{k} - x_{id}^{t}\right)$$

$$+ C_2 \cdot \text{rand} \cdot \left(p_{g\text{best}}^{t} - x_{id}^{t}\right), \tag{11}$$

$$X_{id}^{t+1} = \alpha V_{id}^{t} + x_{id}^{t},$$

where t denotes the tth iteration; C_1 and C_2 are learning factors; rand is positive random number between 0 and 1 under normal distribution. α is the constraint factor which can control the velocity weight. w denotes the inertial weight coefficient; x_{id} denotes the velocity of a particle i; v_{id} denotes the velocity of a particle i; p_{id} is the personal best position of particle i; $p_{g\text{best}}$ denotes the best one of all personal best positions of all particles within the swarm [19, 20].

2.5. Proposed Method.

In this section, we introduce the proposed mRMR-PSO-WSVM method. The aim of this system is to optimize the SVM classifier accuracy by automatically (1) preselecting the number of genes using mRMR method, (2) estimating the best gene weights and optimal values for C and γ by PSO. First, the original microarray dataset is preprocessed by the mRMR filter. Each gene is evaluated and sorted according to mentioned mRMR criterions in Section 3, and the first fifty top-ranked genes are selected to form a new subset. In fact, mRMR is applied to filter out many unimportant genes and reduces the computational load for SVM classifier. Then, a PSO-based approach is developed for determination of kernel parameters and genes weight. Gene weighting is introduced to approximate the optimal degree of influence of individual gene using the training set. Without gene weighting, two decision variables C and γ are required to be optimized. If n genes are required to decide for gene weighting, then $n + 2$ decision variables must be adopted (see Figure 1). The value of n variables ranges between 0 and 1, where sum of them is equal to 1.

The range of parameter C is between 0.01 and 5,000, while the range of γ is between 0.0001 and 32. Figure 2 illustrates the solution representation. We used this representation for particles and allowed PSO to find right value for each variable.

We also define a threshold function $U_\delta(\cdot)$ to avoid using noisy genes with lower predictive power and to put more importance on the genes with higher discriminative power. In fact, $U_\delta(\cdot)$ works as gene selector which omits the redundant genes in the final step again. The domain of this function is the set of gene weights and the range is a revised weight for each gene

$$U_\delta(a_i) = \begin{cases} 0 & \text{if } a_i \le \delta, \\ a_i & \text{if } a_i > \delta, \end{cases} \tag{12}$$

where $0 \le \delta \le 1$ and a_i is the degree of importance of ith gene. Finally, the weighted vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is determined by normal form as

$$\alpha_k = \frac{U_\delta(a_k)}{\sum_{i=1}^{d} U_\delta(a_i)}, \quad k = 1, 2, \ldots, d. \tag{13}$$

Therefore, as mentioned in Figure 2 the training process can be represented as follows

(1) Use the mRMR method to preselect fifty top-ranked genes. These selected genes then utilized in next stages where the PSO was employed to obtain optimal gene weights and kernel parameters.

(2) Involve the cross-validation method to separate dataset into training and testing set.

(3) Then, for each training set set up parameters of PSO. Generate randomly all particles' positions and velocity and set up the learning parameters, the inertia weight and the maximum number of iterations.

(4) Train WSVM classifier according to particles values.

(5) Calculate the corresponding fitness function formulated by (*classified/total*) (*total* denotes the number of training samples, and *classified* denotes the number of correct classified samples) for each particle.

(6) Update the velocity and position of each particle using (11).

(7) If the specified number of generations is not yet satisfied, produce a new population of particles and return to step (4).

(8) Select the gene weights and kernel parameters values from the best global position $p_{g\text{best}}$ and discard redundant genes with threshold function $U_\delta(\cdot)$.

(9) Train WSVM classifier with obtained parameters.

(10) Classify patients with the optimal model.

3. Experimental Results

The proposed mRMR-PSO-WSVM was implemented using the MATLAB software package version 7.2. We compared our

TABLE 1: Detailed information of gene expression datasets.

Dataset name		Number of		
		Samples	Categories	Genes
Leukemia	Acute myeloid leukemia	25	2	7129
	Acute lymphoblastic leukemia	47		
Colon	Cancerous colon tissues	40	2	2000
	Normal colon tissues	22		

TABLE 2: PSO parameters.

Parameters	Values
Swarm size	50
The inertia weight	0.9
Acceleration constants C_1 and C_2	2
Maximum number of iterations	70

suggested method with SVM, mRMR-SVM, mRMR-PSO-SVM classifiers to consider the effect of each component on classification results. We also extend our experiments by employing the classifiers that have been suggested before by Shen et al, [8] and Abdi and Giveki [18] which were denoted by PSO-SVM[1] and PSO-SVM[2] in Table 3, respectively. The discrete PSO was applied to select genes in PSO-SVM[1]. Each particle was encoded to a string of binary bits associated with the number of genes, which is made up of an SVM classifier with all its features. A bit "0" in a particle represented the uselessness of corresponding gene. Also, in PSO-SVM[2] Abdi and Giveki utilized PSO to determine SVM kernel parameters based on the fact that kernel parameters setting in training procedure significantly influence the classification accuracy [18].

The classifiers are evaluated on two popular public datasets: leukemia [21] and colon [22] datasets both of which consist of a matrix of gene expression vectors obtained from DNA microarrays for a number of patients. The first set was obtained from cancer patients with two different types of leukemia, acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The complete dataset contains 25 AML and 47 ALL samples. The second set was obtained from cancerous and normal colon tissues. Among them, 40 samples are from tumors, and 22 samples are from healthy parts of the colons of the same patients [23]. The detailed information of them is collected in Table 1.

To calculate the accuracy of classifiers, the leave-one-out cross-validation (LOOCV) was involved using a single observation from the original sample as the testing data, and the remaining observations as the training data. This was repeated such that each observation in the sample was used once as the testing data. Moreover, in order to make experiments more realistic, we conducted each experiment 10 times on each dataset, and the average of classification accuracies of ten independent runs besides the average of number of selected genes as considered to evaluate the performance of classifiers. The related parameters of PSO algorithm applied in the experiments are also shown in Table 2.
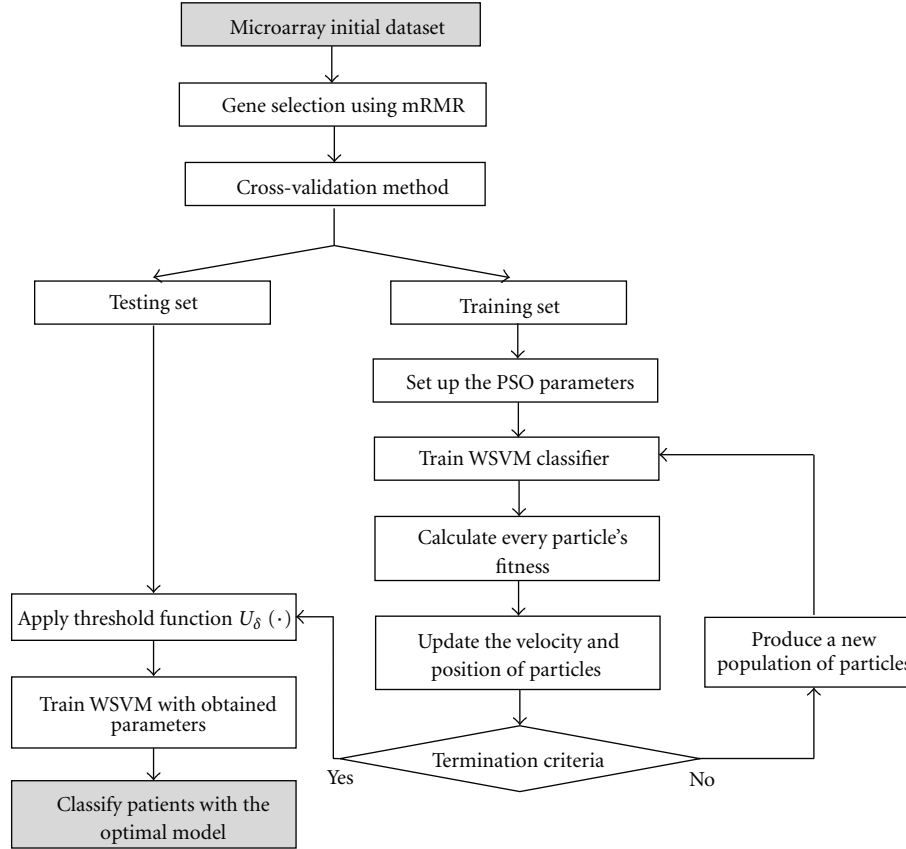
FIGURE 2: The process of classification by mRMR-PSO-WSVM.

TABLE 3: The values of the statistical parameters of the classifiers.

Methods/datasets	Leukemia		Colon	
	Acc (%)	Selected genes	Acc (%)	Selected genes
SVM	90.28	7129	83.87	2000
mRMR-SVM	97.22	50	83.87	50
PSO-SVM[1]	94.44	22.5	85.48	20.1
PSO-SVM[2]	93.06	7129	87.01	2000
mRMR-PSO-SVM	100	17.7	90.32	10.3
mRMR-PSO-WSVM	*100*	*3.8*	*93.55*	*6.2*

In addition, we filtered out all those genes having the PSO weight equal to or less than a quality threshold δ in the proposed method. To find the best value for this, we started from 0.2 and kept increasing this threshold value by 0.1 and saved the classification results. We found that for leukemia and colon datasets $0.3 \leq \delta \leq 0.5$ is always the best choice. Table 3 shows the classification accuracy of classifiers. As it can be observed, the classification accuracy of SVM on two datasets is not very interesting. Furthermore, the accuracy when the mRMR filter is employed generally outperforms the accuracy without gene selection. This implies that gene selection is able to improve the classification accuracy and mRMR is an effective tool to omit the redundant and noisy genes. In addition, the accuracy of PSO-SVM[1] shows that the selection of genes that are really indicative for tumor classification is a key step in developing a successful gene expression-based data and PSO is a promising tool for handling this. Also, the result of PSO-SVM[2] emphasizes on the fact that kernel parameters setting significantly influences the classification accuracy of SVM. Classification accuracy of the mRMR-PSO-SVM explains well the benefits of both gene selection and kernel parameters determination using PSO. In final, the proposed mRMR-PSO-WSVM achieves the highest classification accuracy together with lowest average of selected genes on test sets. This confirms that the suggested PSO-based gene weighting achieves better performance compared to binary PSO. Also, the average of selected genes shows that using the threshold function $U_\delta(\cdot)$ is very effective to reduce the number of selected genes.

TABLE 4: Classification accuracy of our method with other methods from literature (under 10-fold cross validation).

(Authors, year)	Method	Leukemia		Colon	
		Acc (%)	S. G.	Acc (%)	S. G.
(Ruiz et al., 2006) [24]	NB-FCBF	95.9	48.5	77.6	14.6
(Shen et al., 2007) [8]	PSOSVM	N. C.	N. C.	91.67	4.00
(Li et al., 2008) [15]	Single PSO	94.6	22.3	87.1	19.8
(Li et al., 2008) [15]	Single GA	94.6	23.1	87.1	17.5
(Li et al., 2008) [15]	Hybrid PSO/GA	97.2	18.7	91.90	18.00
(Shen et al., 2008) [17]	HPSOTS	98.61	7.00	93.32	8.00
(Abdi et al., 2012) [13]	*mRMR-PSO-WSVM*	*98.74* [18]	*4.1*	*93.55*	*6.8*

*S. G. and N. C. denote selected genes and not considered, respectively.

TABLE 5: Classification accuracy of our method with other methods from literature (under LOOCV).

(Authors, year)	Method	Leukemia		Colon	
		Acc (%)	S. G.	Acc (%)	S. G.
(Mohamad et al., 2007) [25]	IG + NewGASVM	94.71	20.00	N. C.	N. C.
(El Akadi et al., 2011) [17]	mRMR-GA	100	15.00	85.48	15.00
(Abdi et al., 2012) [18]	*mRMR-PSO-WSVM*	*100*	*3.8*	*93.55*	*6.2*

*S. G. and N. C. denote selected genes and not considered, respectively.

Tables 4 and 5 present the results of previously suggested methods besides the proposed mRMR-PSO-WSVM classifier. In order to make a more reliable comparison we try to carry out experiments with two cross-validation methods since some previously reported results were obtained under 10-fold cross validation and the other under LOOCV. Tables 4 and 5 show the results under 10-fold and LOO, respectively. We can see that the proposed classifier can obtain far better classification accuracy than previously suggested methods under both the cross-validation methods. Therefore, we can conclude that our method obtains promising results for gene selection and tumor classification problems.

4. Conclusion and Future Researches

This work presented a PSO-based approach to construct an accurate SVM in classification problems dealing with high-deminsional datasets especially gene expressions. This novel approach was a two-stage method in which, first of all, the mRMR filter technique was applied to preselect an effective genesubset from the candidate set. Then it formed a novel SVM in which PSO not only discarded redundant genes, but also especially took into account the degree of importance of each gene and assigned diverse weights to the different genes. It also used PSO to find appropriate kernel parameters since the choice of gene weights influences the optimal kernel parameters and vice versa. The experiments conducted using two different datasets for cancer classification show that the proposed mRMR-PSO-SVM outperforms the previously reported results. Experimental results obtained from UCI datasets or other public datasets and real-world problems can be tested in the future to verify and extend this approach.

References

[1] D. Hwang, W. A. Schmitt, G. Stephanopoulos, and G. Stephanopoulos, "Determination of minimum sample size and discriminatory expression patterns in microarray data," *Bioinformatics*, vol. 18, no. 9, pp. 1184–1193, 2002.

[2] J. Khan, J. S. Wei, M. Ringnér et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.

[3] M. P. S. Brown, W. N. Grundy, D. Lin et al., "Knowledge-based analysis of microarray gene expression data by using support vector machines," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 1, pp. 262–267, 2000.

[4] V. Vapnik, *The Nature of Statistical Learning Theory*, New York, NY, USA, 1995.

[5] E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider, "Comparison of support vector machine and artificial neural network systems for drug/nondrug classification," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1882–1889, 2003.

[6] C. Z. Cai, W. L. Wang, L. Z. Sun, and Y. Z. Chen, "Protein function classification via support vector machine approach," *Mathematical Biosciences*, vol. 185, no. 2, pp. 111–122, 2003.

[7] H. X. Liu, R. S. Zhang, F. Luan et al., "Diagnosing breast cancer based on support vector machines," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 3, pp. 900–907, 2003.

[8] Q. Shen, W.-M. Shi, W. Kong, and B.-X. Ye, "A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification," *Talanta*, vol. 71, no. 4, pp. 1679–1683, 2007.

[9] B. Jin and Y. Q. Zhang, "Support vector machines with evolutionary feature weights optimization for biomedical data classification," in *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS '05)*, pp. 177–180, June 2005.

[10] H. J. Xing, M. H. Ha, D. Z. Tian, and B. G. Hu, "A novel support vector machine with its features weighted by mutual information," in *Proceedings of IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (IJCNN '08)*, pp. 315–320, June 2008.

[11] T. Wang, "Improving SVM classification by feature weight learning," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA '10)*, pp. 518–521, May 2010.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, vol. 1944, pp. 1942–1948, December 1995.

[13] Q. Shen, W.-M. Shi, and W. Kong, "Hybrid particle swarm optimization and tabu search approach for selecting genes for tumor classification using gene expression data," *Computational Biology and Chemistry*, vol. 32, no. 1, pp. 53–60, 2008.

[14] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Computational Biology and Chemistry*, vol. 32, no. 1, pp. 29–38, 2008.

[15] S. Li, X. Wu, and M. Tan, "Gene selection using hybrid particle swarm optimization and genetic algorithm," *Soft Computing*, vol. 12, no. 11, pp. 1039–1048, 2008.

[16] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[17] A. El Akadi, A. Amine, A. El Ouardighi, and D. Aboutajdine, "A two-stage gene selection scheme utilizing MRMR filter and GA wrapper," *Knowledge and Information Systems*, vol. 26, no. 3, pp. 487–500, 2011.

[18] M. J. Abdi and D. Giveki, "Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules," *Engineering Applications of Artificial Intelligence*. http://www.sciencedirect.com/science/article/pii/S0952197612000218.

[19] J. Wei, Z. Jian-Qi, and Z. Xiang, "Face recognition method based on support vector machine and particle swarm optimization," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4390–4393, 2011.

[20] M. J. Abdi and H. Salimi, "Farsi handwriting recognition with mixture of RBF experts based on particle swarm optimization," *International Journal of Information Science and Computer Mathematics*, vol. 2, no. 2, pp. 129–136, 2010.

[21] T. R. Golub, D. K. Slonim, P. Tamayo et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.

[22] U. Alon, N. Barka, D. A. Notterman et al., "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 12, pp. 6745–6750, 1999.

[23] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[24] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Incremental wrapper-based gene selection from microarray data for cancer classification," *Pattern Recognition*, vol. 39, no. 12, pp. 2383–2392, 2006.

[25] M. S. Mohamad, S. Omatu, S. Deris, and S. Z. M. Hashim, "A model for gene selection and classification of gene expression data," *Artificial Life and Robotics*, vol. 11, no. 2, pp. 219–222, 2007.