Simultaneous Optimization of Gene Selection and Tumor Classification Using Intelligent Genetic Algorithm and Support Vector Machine

Hui-Ling Huang[1] **Shinn-Ying Ho**[2]
[1]*Department of Information Management, Jin-Wen Institute of Technology, Taipei, Taiwan*
[2] *Department of Biological Science and Technology, and Institute of Bioinformatics, National Chiao Tung University, Hsinchu, Taiwan*
Email : hlhuang@jwit.edu.tw, syho@mail.nctu.edu.tw

ABSTRACT: Microarray gene expression profiling technology is one of the most important research topics in clinical diagnosis of disease. Given thousands of genes, only a small number of them show strong correlation with a certain phenotype. To identify such an optimal subset from thousands of genes is intractable, which plays a crucial role when classify multiple-class genes express models from tumor samples. This paper proposes an efficient classifier design method to simultaneously select the most relevant genes using an intelligent genetic algorithm (IGA) and design an accurate classifier using Support Vector Machine (SVM). IGA with an intelligent crossover operation based on orthogonal experimental design can efficiently solve large-scale parameter optimization problems. Therefore, the parameters of SVM as well as the binary parameters for gene selection are all encoded in a chromosome to achieve simultaneous optimization of gene selection and the associated SVM for accurate tumor classification. The effectiveness of the proposed method IGA/SVM is evaluated using four benchmark datasets. It is shown by computer simulation that IGA/SVM performs better than the existing method in terms of classification accuracy.

1 INTRODUCION

Microarray gene expression profiling technology is one of the most important research topics in clinical diagnosis of disease. The practical applications of microarray gene expression profiles include management of cancer and infectious diseases [1]. The normal cells can evolve into malignant cancer cells through a series of mutation in genes that control the cell cycle [2]. However, given thousands of genes, only a small number of them show strong correlation with a certain phenotype [3]. To identify such an optimal subset from thousands of genes is intractable, which plays a crucial role when classify multiple-class genes express models from tumor samples. How to design an accurate tumor classifier with a smallest subset of genes from microarray gene expression data is investigated in this paper.

Generally, feature selection methods can be categorized into two classes: filter and wrapper. For the filter approach, Golub et al. [4] and Furey et al. [5] employed an individual gene ranking score and a weighting factor to perform gene selection prior to classification. Liu et al. [6] proposed a feature selection method which combines top-ranked, test-statistic, and principle component analysis (PCA) in conjunction with ensemble neural networks to improve classification. Zhou and Mao [7] suggested a filter-like evaluation criterion, called LS Bound measure, which provides gene subsets leading to more accurate classification. The wrapper approach involves the computation overhead of evaluating candidate gene subsets by executing a selected classification algorithm on the dataset represented using each gene subset under consideration. Li *et al.* [8] proposed a hybrid method of the genetic algorithm (GA) based gene selection and k-nearest neighbor classifier (GA/KNN) to assess the importance of genes for classification.

Support Vector Machine (SVM) [9], a supervised machine learning technique, is one of the methods successfully applied to cancer diagnosis problems in the previous studies [5], [7], [10]-[14]. To build an efficient and effective model for classification, it is indicated that SVM performs better than some existing classification algorithms [7]. Statnikov *et al.* [15] investigated classification algorithms which can handle multiple classes and a large number of variables, and compared multi-category SVM to Neural Networks and K-Nearest Neighbor classifier. The results indicate that the multi-category SVM is the most effective classifier for tumor classification

Most of the above-mentioned methods except GA/KNN [8] comprise two separated stages: gene selection and classifier design. To advance the classification performance, it is better to take the two stages into account simultaneously [16], [17]. Liu *et al.* [16] proposed a PCA-based two-class classifier. Ooi and Tan [17] proposed a GA/MLHD (maximal likelihood)-based methodology for multi-class prediction using gene expression data.

This paper proposes an efficient classifier design method to simultaneously select the most relevant genes and the parameters of SVM using an intelligent genetic algorithm (IGA) for designing an accurate tumor classifier. IGA is a specific version of intelligent evolutionary algorithms [18] which can efficiently solve large-scale parameter optimization problems. Some of the IEA-based classifier design methods can be referred to [19]-[21]. Therefore, the parameters of SVM as well as the binary parameters for feature selection are all encoded in a chromosome to achieve simultaneous optimization of feature selection and SVM model. IGA with an intelligent

crossover operation is efficient in solving the resultant optimization problem with a large number of parameters.

The effectiveness of the proposed method IGA/SVM is evaluated using four benchmark datasets: 9_Tumors, 14_Tumors, Brain_Tumors1, and Brain_Tumors2 [ ]. It is shown by computer simulation that IGA/SVM performs better than the existing method [ ] in terms of classification accuracy with the same number of selected features.

2 Support Vector Machine

Support Vector Machine (SVM) is a very popular method to deal with classification, prediction, and regression problems. Various SVMs introduced by Vapnik and other co-workers [9], [22] are powerful classifiers. For the binary SVM, the training data consist of n pairs (x_1, y_1), (x_2, y_2),..., (x_n, y_n), with $x_i \in \Re^m$ and $y_i \in \{-1,1\}$, $i = 1, 2, \ldots, n$. The standard SVM formulation [23] is as follows:

$$\min_{w,b,\xi} \ \tfrac{1}{2} w^T w + C\sum_{i=1}^{n} \xi_i \quad \text{subject to} \qquad (1)$$

$$y_i(w^T \phi(x_i)+b) \geq 1-\xi_i,\ \xi_i \geq 0,\ i = 1,\ldots, n,$$

where $w \in \Re^m$ is a vector of weights of training instances; b is a constant; C is a real-valued tradeoff (cost) parameter; ξ_i is a penalty parameter; and ϕ is to map x_i into a higher dimensional space. The SVM of (1) is called a linear kernel SVM when $\phi(x_i) = x_i$. The SVM finds a linear separating hyperplane with the maximal margin in the higher dimensional space. $C > 0$ is the penalty parameter of error term. The SVM of (1) is called a nonlinear SVM when ϕ maps x_i into a higher dimensional space.

For the nonlinear SVM, the number of variables w can be vary large or even infinite, so it is very difficult to solve using (1). The general method is to use the following dual formulation:

$$\min_{\alpha} \ \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \quad \text{subject to} \qquad (2)$$

$$y^T\alpha=0,\ 0 \leq \alpha_i \leq C,\ i = 1,\ldots, n,$$

where e is the vector of all ones, $C>0$ is the supper bound, Q is an $n \times n$ positive semidefinite matrix, $Q_{ij} \equiv y_i\, y_j\, K(x_i\, x_j)$, and $K(x_i\ x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel function. Some commonly-used kernel functions are: $e^{-\gamma\|x_i - x_j\|^2}$ (Radial basis function), $(x_i^T x_j/\gamma+\delta)^d$ (Polynomial), and $\tanh(\gamma x_i^T x_j+\delta)$ (sigmoid), where γ, d, and δ are kernel parameters. The number of variables in (2) is the size n of the training dataset which is smaller than the dimensionality of $\phi(x)$. Given w and b, one can classify an instance x using the decision function is $\operatorname{sgn}(\sum_{i=1}^{n} y_i\alpha_i K(x_i, x) + b)$.

Chang and Lin [24] develop a software tool LIBSVM (Library for Support Vector Machine) for support vector classification, regression and distribution estimation. LIBSVM uses the "one-against-one" approach [25] for multiclass classification. In the one-against-one approach, $k(k\text{-}1)/2$ classifiers are established where k is the number of classes. The classifiers between each pair of k classes are optimized using the following dual formulation:

$$\min_{i,j} \ \frac{1}{2}(w^{i,j})^T w^{i,j} + C\sum_t \xi_t^{ij} \quad \text{subject to} \qquad (3)$$

$$(w^{ij})^T \phi(t) + b^{ij} \geq 1 - \xi_t^{ij}, \quad \text{if } y_t = i$$

$$(w^{ij})^T \phi(t) + b^{ij} \leq -1 + \xi_t^{ij}, \quad \text{if } y_t = j$$

$$\xi_t^{ij} \geq 0.$$

After solving the optimization problem using (3), $k(k\text{-}1)/2$ decision functions can be obtained. To predict a class label of a given instance x, the prediction for each of the $k(k\text{-}1)/2$ classifiers is calculated using a voting strategy [8]. If there is a class, say j, that receives the largest number of votes, the instance x is assigned to class j, where a tie is broken randomly. One advantage of using this method is that each classifier is easy to train since only the binary SVM is needed. Another approach to multiclass classification is called "one-against-all". In this approach, k models of SVM are established. For each class j, the SVM is trained using all the instances in the class j as positives and the rest of instances as negatives. Previous research has shown that one-against-one outperforms one-against-all for multiclass classification [28].

3 The proposed Method

The proposed hybrid method IGA/SVM simultaneously selects the most effective genes and control parameters of SVM using IGA, and designs an accurate SVM classifier for tumor classification.

3.1 Gene Expression Data

Consider the training data $\{(x_i,y_i) \mid x_i \in X, y_i \in Y, i=1,\ldots n\}$ where X is a vector space of dimension m and Y is a finite set of tumor class labels. The gene expression data be formulated as the following matrix G:

$$G = \begin{array}{c} \\ y_1 \\ y_2 \\ \mathrm{M} \\ y_n \end{array} \begin{array}{c} \begin{array}{cccccc} gene_1 & gene_2 & \wedge & gene_m \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix} \end{array} \qquad (6)$$

where x_{ij} is the measurement of expression level of the jth gene for the ith pattern and y_i is the tumor class label, $j=1,\ldots, m$, $i=1,\ldots n$. Microarray datasets are characterized by a small number (n) of patterns and a large number (m) of genes for each pattern. Of the thousands of genes, only a small number of them show strong correlation with a certain phenotype.

The problem of feature selection can be formally addressed as follows. From the given m features, select $l \ll m$ feature that give the smallest expected generalization error of tumor classification.

3.2 The used SVMs model

The formulation in Section 2 can take nonlinearly separable cases into account by letting C be finite values. SVM has shown good performance in data classification that depends on tuning of several parameters. The

parameters affect the generalization ability. The basic approach to SVM classification may be extended to allow for nonlinear decision surfaces. For this, the input data are mapped into a high dimensional space through a nonlinear mapping function which has effect of spreading the distribution of the data points in a way that facilitates the fitting of a linear hyperplane. The classification decision function is as follows:

$$\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i y_i K(x_i,x)+b) \tag{4}$$

where α_i, $i = 1,\ldots,n$, are Lagrange multipliers. The magnitude of α_i is determined by the parameter C and lies on a scale of 0-C [28]. The kernel used must meet Mercer's condition [23] and RBF kernels will be adopted as a default throughout the paper. In the RBF kernels function,

$$k(x_i\text{-}x)=e^{-\gamma\|x_i-x\|^2} \tag{5}$$

where γ is the parameter controlling the width of the Gaussian kernel. It has been proved [29] that one can classify any consistent training set with zero errors by using a sufficiently large value of γ. Moreover, we apply nonlinear kernel to solve classification problems, so we need to select the cost parameter C and kernel parameter γ.

Parameters of the classification algorithm [15] were chosen by nested cross validation procedures to optimize performance. Statnikov et al. construct SVM with RBF kernel using the following ranges for optimization of SVM parameters: cost C ={0.0001, 0.01, 1, 100} and values of γ ={0.0001, 0.001, 1}. In this paper we extended ranges for optimization of SVM parameters: cost C ={0.0001×2^d, 0.001×2^d, 0.01×2^d, 0.1×2^d, 1×2^d, 10×2^d, 100×2^d} and values of γ ={0.0001×2^d, 0.001×2^d, 1×2^d}, d=1, 2, 3. In order to select the optimal values of parameters, they are encoded into chromosomes of IGA. We can represent this search space by a positive integer encoding where the integer value indicates the selection of the parameters at the corresponding sequence position.

3.3 Chromosome representation and fitness function

Let S be the set of parameters $\{g_1,\ldots, g_l, c, r\}$. The parameter $g_i\in[1, m]$ is the index of the selected gene, which are used for feature selection. The parameters $c\in$ {1, 2,... , 21} and $r \in$ {1, 2,..., 9}, which are used select an effective SVM model. All the parameters are encoded into a chromosome using integer values, as shown in Fig. 1.

Fitness value guides IGA to choose offspring for the next generation from the current parents. If S represents the set of parameters to be evolved by IGA, then the fitness function $F(S)$ is to maximize classification accuracy.

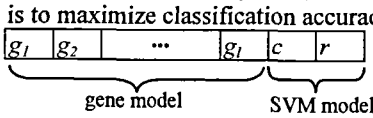

Fig. 1. Chromosome representation.

3.4 The proposed method IGA/SVM

The main power of IGA arises from intelligent crossover based on orthogonal experimental design (OED) [30]. In the crossover operation of the conventional GA, two parents generate and test two chromosomes using a random combination of their chromosomes, and the best two chromosomes among the four chromosomes are selected as the children using an elitist strategy. In IGA, the generate-and-test search for children using a random combination of GA is replaced with a systematic reasoning search method using an intelligent combination. The intelligent crossover with OED can economically estimate the contribution of individual genes to a fitness function and consequently pick up the better one of two parents to form chromosomes of children.

It is well realized that the parameters g representation to the same column in the matrix G which are feature subset. A given feature subset as an input and return the estimated generalization performance of the learning machine as an evaluation of feature subset. It needs to be repeated for each feature subset taken into consideration. We should leave out the parameters with identical genes between two chromosomes before we assign the factor in OED. The repair procedure can avoid producing the unreasonable solutions, as shown in Fig. 2.

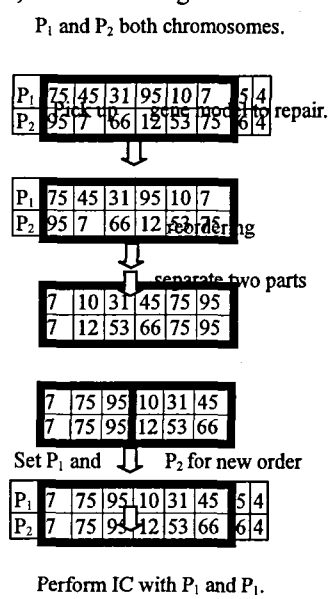

Fig. 2. The repair procedure.

Once the repair procedure is done, the intelligent crossover (IC) can be easily implemented by referring the reference [18]. The used IGA with the proposed chromosome encoding and SVM classifier is described as follows:

Step 1: Initialization: Feasible chromosomes of IGA are randomly generated where each gene g_i is unique in a chromosome.

Step 2: Evaluation: According to the selected subset of genes, establish the SVM classifier using training data: n pairs (x_1, y_1), (x_2, y_2),..., (x_n, y_n), where $x_i\in\Re^l$, $y_i\in\{1, \ldots, k\}$, $i = 1, 2, \ldots, n$, and k is the number of tumor classes. Decode the parameters and set the SVM model. Using the n pairs vectors and SVM model calculate the fitness value. Generate an additional individual I_{best} which is the same as the best individual in the population.

Step 3: Selection: Use the simple ranking selection that replaces the worst $P_s \cdot N_{pop}$ individuals with the best

$P_s \cdot N_{pop}$ individuals to form a new population, where P_s is a selection probability.

Step 4: Crossover: Randomly select $P_c \cdot N_{pop}$ individuals I_i (i=1,2, ..., $P_c \cdot N_{pop}$) to perform IC, where P_c is a crossover probability, N_{pop} is population size. Repeat the following steps for i=1,2, ..., $P_c \cdot N_{pop}$:

4a) Use I_{bsest} and I_i as parents, which will produce two children I_{C1} and I_{C2}.

4b) Repair: Reorder the two I_{bsest} and I_i individuals. Compare I_{bsest} with I_i, there are two parts to be obtained. One part is same genes B_s, and another part is different genes B_d.

4c) B_d is assigned to the factors in ODE. Two children I_{C1} and I_{C2} are produced after to perform IC.

4d) Replace I_{bsest} and I_i with the best and the second best individuals, respectively, among I_{bsest}, I_i, I_{C1} and I_{C2} according to fitness performance.

Step 5: Mutation: Randomly change $P_m \cdot N_{pop}$ genes in each chromosome, which constraint is same as initialization. P_m is mutation probability.

Step 6: Termination test: If a prespecified condition is met, stop the algorithm. Otherwise, go to Step 2.

4. Experimental Results

4.1. Data set

In this section, IGA/SVM is evaluated by four datasets: 9_Tumors, 14_Tumors, Brain_Tumors1, and Brain_Tumors2. The datasets are described in Table 1. The expression genes were excluded from the analysis to reduce the amount of noise in the datasets [32][33]. The four multicategory datasets are available for download from [34]. The four datasets had 4-26 distinct diagnostic categories, 50-308 patients, and 5920-15009 genes.

We relied instead on standard normalization and data preparatory steps performed by authors of the primary dataset studies. Moreover, we performed a simple rescaling of gene expression values to be between –1 to 1 for speeding up SVM training.

The summary of application in [15], the four datasets in Table 1 are most challenging. We experimented on these challenging datasets [15] to shown that our approach is practicable and efficient.

Table 1. The four tumor-gene express datasets are evaluated in this paper.

dataset name	distinct diagnostic categories	genes	patients
9_Tumors	9	5726	60
14_Tumors	26	15009	308
Brain_Tumors1	5	5920	90
Brain_Tumors2	4	10367	50

4.2 Experiments

The parameters of IGA are as follows: population size N_{pop} = 50, crossover rate P_c = 0.5, selection rate P_s = 0.2, and mutation rate P_m = 0.1 The stopping condition is 50 generations of IGA. For IGA/SVM, the classification accuracy for each dataset is calculated from results of 10-fold cross-validation. In the experiments, the number l of selected features is set to 10, 20, ..., 100. Comparison between two methods in terms of accuracy is shown in Tables 2-5. However, IGA/SVM outperforms MC-SVM.

Table 2. Results with 9_Tumor

Dataset: 9_Tumor		
Number of Selected gene(l)	Accuracy	
	MC-SVM	IGA/SVM
10	68.71%	90%
20	74.44%	93.33%
30	68.98%	88.33%
40	77.09%	96.67%
50	75.66%	95%
60	75.66%	95%
70	76.55%	93.33%
80	77.66%	95%
90	78.27%	96.67%
100	77.48%	96.67%
wins	0	10

Table 3. Results with 14_Tumor

Dataset: 14_Tumor		
Number of Selected gene(l)	Accuracy	
	MC-SVM	IGA/SVM
10	39.11%	57.46%
20	49.72%	66.83%
30	52.60%	70.41%
40	56.28%	72.03%
50	57.88%	74.63%
60	57.70%	77.88%
70	57.65%	80.16%
80	60.09%	80%
90	60.06%	80.81%
100	61.84%	82.76%
wins	0	10

Table 4. Results with Brain_Tumors1

Dataset: Brain_Tumors1		
Number of Selected gene(l)	Accuracy	
	MC-SVM	IGA/SVM
10	68.71%	100%
20	74.44%	96.67%
30	68.98%	98.89%
40	77.09%	98.89%
50	75.66%	100%
60	75.66%	98.89%
70	76.55%	98.89%
80	77.66%	98.89%
90	78.27%	98.89%
100	77.48%	98.89%
wins	0	10

Table 5. Results with Brain_Tumors2

Dataset: Brain_Tumors1		
Number of Selected gene(l)	Accuracy	
	MC-SVM	IGA/SVM
10	39.11%	100%
20	49.72%	100%
30	52.60%	100%
40	56.28%	100%
50	57.88%	100%
60	57.70%	100%
70	57.65%	100%
80	60.09%	100%
90	60.06%	100%
100	61.84%	100%
wins	0	10

The Statnikov *et al.* (2004) have mentioned varieties of experiments. The best accurate results for the four datasets: 9_Tumors, 14_Tumors, Brain_Tumors1, and Brain_Tumors2 are 74.86%, 76.60%, 92.67%, and 85.67%, respectively. The best results of IGA/SVM in Tables 2~5 is more accurate than the result of MC-SVM in [15]. The comparison is shown in Table 6.

Table 6. Comparison of the two approaches using the best results from Tables 2-5.

dataset name	MC-SVM	IGA/SVM
9_Tumors	74.86%	96.67%
14_Tumors	76.60%	82.76%
Brain_Tumor1	92.67%	100%
Brain_Tumor2	85.67%	100%
wins	0	4

REFERENCES

[1] P. Fortina, et al.. Molecular diagnostics: hurdles for clinical implementation. *Trends Mol Med.* Jun; 8(6):264-6, 2002.

[2] A. Ben-dor, et al.. Tissue classification with gene expression profiles. *In Proceeding of the Fourth International Conference on Computational Molecular Biology (RECOMB2000)*, ACM Press, New York, 2000.

[3] Chris H.Q. Ding. Unsupervised feature selection via two-way ordering in gene expression analysis. *Bioinformatics*, vol. 19 no. 10, p.p. 1259-1266, 2003.

[4] T. R. Golub, D. K. Slonim, et al.. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286, 531-537, 1999.

[5] T. Furey, N. Cristianini, et. al.. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, Vol. 16, no. 10, pp. 906-914, 2000.

[6] Bing Liu, Qinghua Cui, Tianzi Jiang and Songde Ma. A combinational feature selection and ensemble neural network method for classification of gene expression data. *BMC Bioinformatics* , 5:136, 2004.

[7] Xin Zhou and K. Z. Mao. LS Bound based gene selection for DNA microarray data. *Bioinformatics*, Vol. 21, no. 8, pp. 1559-1564, 2005.

[8] L. Li, C. R., T. A. Darden and L. G. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, 17, p.p. 1131-1142, 2001.

[9] Vapnik, V. N. Statistical Learning Theory. *New York*, Wiley, 1998.

[10] S. Mukherjee, P. Tamayo, et. al.. Support vector machine classification of microarray data. *Technical Report AI Memo 1677*, MIT, 1999.

[11] Yoonkyung Lee and Cheol-Koo Lee. Classification of multiple cancer types by multicategory support vector machines using gene express data. *Bioinformatics*, Vol. 19, no. 9, pp. 1132-1139, 2003.

[12] Nathalie Pochet, Frank De Smet, Johan A. K. Suykens, and Bart L. R. De Moor. Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics*, Vol. 20, no.17, pp. 3185-3195, 2004.

[13] Tao Li, Chengliang Zhang, and Mitsunori Ogihara. A Comparative Study of feature Selectin and Multiclass Classification Method for Tissue Classification Based on Gene Expression. *Bioinformatics*, 2004.

[14] D. Komura, H. Nakamura, et. al.. Multidimesional support vector machines for visualization of gene expression data. *Bioinformatics*, Vol. 21, no.4, pp. 439-444, 2005.

[15] Statnikov, Alexander et al.. A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis. *Bioinformatics Advance* Access published September 16, 2004.

[16] Juan Liu and Hitoshi Iba. Selecting Informative genes with parallel genetic algorithms in tissue classification. *Genome Informatics*, 12: 14-23, 2001.

[17] C. H. Ooi anf Patrick Tan. Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics*, Vol. 19, no.1, pp. 37-44, 2003.

[18] Ho, S.-Y., Shu, L.-S., Chen H.-M.. Intelligent genetic algorithm with a new intelligent crossover using orthogonal arrays. *In Proc. Genetic and Evolutionary Computation Conference*, pp. 289-296, 1999.

[19] S.-Y. Ho, H.-M. Chen, S.-J. Ho, and T.-K. Chen. Design of Accurate Classifiers with a Compact Fuzzy-Rule Base Using an Evolutionary Scatter Partition of Feature Space. *IEEE Trans. Systems, Man, and Cybernetics—Part B*, vol. 34, no. 2, pp. 1031-1044, April. 2004.

[20] S.-Y. Ho, Y.-C. Chen. An efficient evolutionary algorithm for accurate polygonal approximation. *Pattern Recognition*, 34 (12):2305-2317, 2001.

[21] S.-Y. Ho, C.-C. Liu, S. Liu. Design of an optimal nearest neighbor classifier using an intelligent genetic algorithm. *Pattern Recognition Letter*, 23:1495-1503, 2002.

[22] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimum margin classifiers. In the Fifth Annual Workshop on Computational Learning Theory,

Pittsburgh. ACM. pp. 144-152, 1992.
[23] C. Cortes and V. Vapnik. Support-vector network. *Machine Learning*, 20:273-297, 1995.
[24] C. Chang and C. Lin. A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[25] S. Knerr, L. Personnaz, and G. Dreyfus. Single layer learning revisited: a stepwise procedure for building and training a neural network. *Neurocomputing: Algorithms*, Architectures and Applications. J. Fogelman (Ed.), Springer-Verlag, 1990.
[26] J. Friedman. Another approach to polychotomous classification. *Technical report*, Department of Statistics, Stanford University, 1996.
[27] J. C. Palatt, N.Cristianini, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. *Advances in Neural Information Processing Systems* 12:547-533. MIT Press, 2000.
[28] A. I. Belousov, S. A. Verzakov, and J. Von Frese. A flexible classification approach with optimal generalization performance: Support vector machines. *Chemometrices and Intelligent Laboratory Systems*, 64,15-25, 2002.
[29] C. Burges. A tutorial on support vector machines for pattern recognition. D*ata Mining Knowledge Discovery*, 2 (2) (1998) 121-167.
[30] S.H. Park. Robust Design and Analysis for Quality Engineering. *Chapman and Hall*, 1996.
[31] J. Lu, et al.. Classical statistical approaches to molecular classification of cancer from gene expression profiling. *In Methods of Microarray Data Analysis* : Papers from CAMDA'00, eds. S.M. Lin, K. F. Johnson, pp. 97-107, Kluwer Academic Publishers, Boston/Dordrecht/London, 2002.
[32] Wouters L, et al.. Graphical exploration of gene expression data: a comparative study of three multivariate mwthods. *Biometrics*, Dec; 59(4):1131-9, 2003.
[33] Statnikov A, Aliferis C, Tsamardinos I, Hardin D, and Levy S. http://www.gems-system.org, 2004.