

Research Article

Sparse Representation for Classification of Tumors Using Gene Expression Data

Xi Yi Hang¹ and Fang-Xiang Wu^{2,3}

¹Department of Electrical and Computer Engineering, California State University, Northridge, CA 91330, USA

²Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK, Canada S7N 5A9

³Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK, Canada S7N 5A9

Correspondence should be addressed to Fang-Xiang Wu, faw341@mail.usask.ca

Received 24 November 2008; Accepted 12 January 2009

Recommended by Dechang Chen

Personalized drug design requires the classification of cancer patients as accurate as possible. With advances in genome sequencing and microarray technology, a large amount of gene expression data has been and will continuously be produced from various cancerous patients. Such cancer-alerted gene expression data allows us to classify tumors at the genomewide level. However, cancer-alerted gene expression datasets typically have much more number of genes (features) than that of samples (patients), which imposes a challenge for classification of tumors. In this paper, a new method is proposed for cancer diagnosis using gene expression data by casting the classification problem as finding sparse representations of test samples with respect to training samples. The sparse representation is computed by the l_1 -regularized least square method. To investigate its performance, the proposed method is applied to six tumor gene expression datasets and compared with various support vector machine (SVM) methods. The experimental results have shown that the performance of the proposed method is comparable with or better than those of SVMs. In addition, the proposed method is more efficient than SVMs as it has no need of model selection.

Copyright © 2009 X. Hang and F.-X. Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The treatment of cancer greatly depends on the accurate classification of tumors. In spite of its effectiveness in classifying tumors by microscopic tissue examination, traditional histopathological approach fails to classify many cancer cases. The number of unclassified cancer cases can reach up to 40 000 per year just in the United States [1]. DNA microarray technology, on the other hand, has the potential to provide a more accurate and objective cancer diagnosis due to its high throughput capability of measuring expression levels of tens of thousands genes simultaneously. Since Golub et al. [2] successfully classified between acute myeloid leukemia (AML) and acute lymphocytic leukemia (ALL), many other types of cancer have been classified using gene expression data including breast cancer [3], lymphoma [4], lung cancer [5], bladder cancer [6], colon cancer [7], ovarian cancer [8], prostate cancer [9], melanoma [10], and brain tumors [11].

The successful application of microarray technology in cancer diagnosis greatly depends on the careful design of two important components of a gene data classification system: gene selection and sample classification, shown in Figure 1. Gene selection mainly serves two purposes: (i) to reduce dramatically the number of genes used in classification to manage the “curse of dimensionality” and (ii) selected genes might be biologically relevant, allowing further biological exploration which may lead to better understanding of underlying molecular mechanism associated with tumorigenesis and progression. Gene selection can be made by test statistics [12]. An excellent review on gene selection methods can be found in [13].

The second component, sample classification, is a challenging issue for a problem with a small number of learning samples and yet a large number of features (genes). The number of samples available for analysis ranges from tens to hundreds. Many established methods have been proposed to address the challenge. According to Lee et al. [14], they can

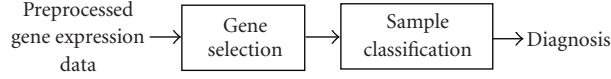


FIGURE 1: The pipeline of cancer diagnosis using gene expression data.

be classified into four categories: (i) classical methods such as Fisher's linear discriminant analysis, logistic regression, K-nearest neighbor, and generalized partial least square, (ii) classification trees and aggregation methods such as CART, random forest, bagging, and boosting, (iii) machine learning methods such as neural network and support vector machines (SVMs), and (iv) generalized methods such as flexible discriminant analysis, mixture discriminant analysis, and shrunken centroid method.

In this paper, we propose a novel approach for classification, called sparse representation, inspired by the recent progress in l_1 -norm minimization-based methods such as basis pursuit denoising [15], compressive sensing for sparse signal reconstruction [16–18], and Lasso algorithm for feature selection [19]. Ideally, a testing sample can be represented just in terms of the training samples of the same category. Hence, when the testing sample is expressed as linear combination of all the training samples, the coefficient vector is sparse, that is, the vector has relatively few nonzero coefficients. Testing samples of same category will have similar sparse representation, while different categories will result in different sparse representations. In order to recover the sparse coefficient vector, l_1 -regularized least square [20] is used.

Unlike general supervised learning methods, where a training procedure is used to create a classification model for testing, the sparse representation approach does not contain separate training and testing stages. Instead, classification is achieved directly out of the testing sample's sparse representation in terms of training samples. Another unique feature of the new method is no model selection needed. It is well known that the performance of a classifier, such as SVM, relies upon careful choice of the model parameters via model selection procedure.

2. Materials and Methods

2.1. Sparse Representation. Consider a training dataset $\{(\mathbf{x}_i, l_i); i = 1, \dots, n\}$, $\mathbf{x}_i \in R^d$, $l_i \in \{1, 2, \dots, N\}$, where \mathbf{x}_i represents the i th sample, a d -dimensional column vector containing gene expression values with d as the number of genes, and l_i is the label of the i th sample with N as the number of categories. For a testing sample $\mathbf{y} \in R^d$, the problem of sparse representation is to find a column vector $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$ such that

$$\mathbf{y} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_n \mathbf{x}_n, \quad (1)$$

and $\|\mathbf{c}\|_0$ is minimized, where $\|\mathbf{c}\|_0$ is l_0 -norm, and it is equivalent to the number of nonzero components in the vector \mathbf{c} .

Defining a matrix by putting \mathbf{x}_i as the i th column $A = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, the problem of sparse representation can be converted into

$$\mathbf{c} = \min_{\mathbf{c}' \in R^n} \|\mathbf{c}'\|_0 \quad \text{subject to } \mathbf{y} = A\mathbf{c}. \quad (2)$$

Finding the solution to sparse representation problem is NP-hard due to its nature of combinational optimization. Approximation solution can be obtained by replacing the l_0 -norm in (2) by the l_p -norm

$$\mathbf{c} = \min_{\mathbf{c}' \in R^n} \|\mathbf{c}'\|_p \quad \text{subject to } \mathbf{y} = A\mathbf{c}, \quad (3)$$

where the l_p -norm of a vector \mathbf{v} defined as $\|\mathbf{v}\|_p = (\sum_i |v_i|^p)^{1/p}$. A generalized version of (3), which allows for certain degree of noise, is to find a vector \mathbf{c} such that the following objective function is minimized:

$$J(\mathbf{c}, \lambda) = \min_{\mathbf{c}} \{\|\mathbf{A}\mathbf{c} - \mathbf{y}\|_2 + \lambda \|\mathbf{c}\|_p\}, \quad (4)$$

where the positive parameter λ is a scalar regularization that balances the tradeoff between reconstruction error and sparsity.

Since l_1 -norm minimization can efficiently recover sparse signal [20] and are robust against outliers, this study takes $p = 1$ in (4). Therefore, the problem is reduced to solve (3) an l_1 -regularized least square problem:

$$J(\mathbf{c}, \lambda) = \min_{\mathbf{c}} \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_2 + \lambda \|\mathbf{c}\|_1. \quad (5)$$

A truncated Newton interior-point method (TNIPM) proposed in [20] can be used to solve the above optimization problem in (5). For the convergence of the algorithm, the regularization parameter must satisfy the following condition:

$$\lambda \leq \|2A^T \mathbf{y}\|_\infty. \quad (6)$$

Please refer to [20] for more information about l_1 -regularized least square and the specialized interior-point method.

Another approach to determine the sparse solution to (2) is to use the framework of compressive sensing, which requires the system to be underdetermined. Including the construction errors \mathbf{e} in (1) yields

$$\mathbf{y} = A\mathbf{c} + \mathbf{e}. \quad (7)$$

In compressive sensing approach, we need to rewrite (7) as

$$\mathbf{y} = \begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{e} \end{bmatrix} = B\mathbf{d}, \quad (8)$$

Input: $\{(\mathbf{x}_i, y_i); i = 1, \dots, n\}$, and \mathbf{y}

1. Normalize \mathbf{x}_i , $i = 1, 2, \dots, n$, and \mathbf{y}
2. Create matrix \mathbf{A}
3. Solve the optimization problem defined in (5)
4. Compute $g_k(\mathbf{y})$, $k = 1, 2, \dots, N$

Output: $\arg \min_k g_k(\mathbf{y})$

ALGORITHM 1: Classification by sparse representation.

where $\mathbf{B} = [\mathbf{A} \ \mathbf{I}] \in R^{d \times (n+d)}$ and $\mathbf{d} = [\mathbf{c}^T \ \mathbf{e}^T]^T \in R^{n+d}$. With these notations, the sparse representation can be obtained by the following constrained l_1 -norm minimization problem:

$$\min_{\mathbf{d}} \|\mathbf{d}\|_1 \quad \text{subject to } \mathbf{B}\mathbf{d} = \mathbf{y}. \quad (9)$$

The above linear programming problem can be solved by a specialized interior-point method called l_1 -magic [21]. The approach in (9) is used in [22] for face recognition by sparse representation.

Both approaches do generate nearly the same classification performance in our experiments. Our approach, based on l_1 -regularized least square, however, is much faster. First, the optimization problem scale in our approach is much smaller. For example, when the training dataset contains 300 samples and the gene number is 10 000, the matrix in our approach is $\mathbf{A} \in R^{10000 \times 300}$ while $\mathbf{B} \in R^{10000 \times 10300}$. Secondly, TNIPM is $O(n^{1.2})$ while l_1 -magic is $O(n^{1.3})$ [20]. In addition, it is noticed that basis pursuit, compressive sensing, and Lasso algorithm can also be converted into l_1 -regularized least square problems [20].

Let $\hat{\mathbf{c}}$ denote the sparse representation obtained by l_1 -regularized least square. Ideally, the nonzero entries in $\hat{\mathbf{c}}$ are associated with the columns in \mathbf{A} corresponding to those training samples of the same category as the testing sample \mathbf{y} . However, noises may cause the nonzero entries to be linked with multiple categories [22]. Simple heuristics, such as assigning \mathbf{y} to the category with the largest entry in $\hat{\mathbf{c}}$, are not dependable. Instead, we define N discriminate functions

$$g_k(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}\hat{\mathbf{c}}_k\|_2, \quad k = 1, 2, \dots, N, \quad (10)$$

where $\hat{\mathbf{c}}_k$ is obtained by keeping only those entries in $\hat{\mathbf{c}}$ associated with category k and assigning zeros to other entries. Thus g_k represents the approximation error when \mathbf{y} is assigned to category k , and we can assign \mathbf{y} to the category with the smallest approximation error. The classification algorithm is summarized (see Algorithm 1).

2.2. Numerical Experiments. Numerical experiments are designed to quantitatively verify the performance of sparse representation method for cancer classification using gene expression data. The performance metric used in this study is accurate, obtained by stratified 10-fold cross-validation. We compare our approach with a few variants of multi-category SVMs. SVMs, as state-of-the-art machine learning algorithms, have been successfully applied in gene profile classification [23, 24]. The comprehensive study in [25]

also shows that SVMs outperform K-nearest neighbors and neural network in gene expression cancer diagnosis.

All experiments are done on a PC with duo Intel 2.33 G CPU and 4 G memory under Windows XP (SP2). MATLAB R14 is used to implement sparse representation method. The optimization is done by l_1 -ls MATLAB package, which is available online (http://www.stanford.edu/~boyd/l1_ls/). The results of SVMs are obtained by gene expression model selector (GEMS), a software with graphic user interface for classification of gene expression data, which is freely available at <http://www.gems-system.org/> and used in [25] for the comprehensive study of the performance of multiple classifiers on gene expression cancer diagnosis. Besides standard binary SVM, GEMS has implemented the following multiclass SVMs: one-versus-rest (OVR) [26], one-versus-one (OVO) [26], directed acyclic graph (DAG) [27], all-at-once method by Weston and Watkins (WW) [28], and all-at-once method by Crammer and Singer (CS) [29], which are used in comparison with sparse representation approach. Polynomial and RBF kernels are used for SVMs.

For fair comparison, the partition file of cross-validation generated by GEMS is used in sparse representation approach. As for model selection, 9-fold cross validation is used for SVMs.

The comparison is done with and without gene selection. Two popular gene selection methods are used in this study: Kruskal-Wallis nonparametric one-way ANOVA (KW) [30] and the ratio of between-groups to within-groups sum of squares (BW) [31].

2.3. Datasets. In the experiment, we use six datasets, which are among 11 datasets used in the comprehensive study [25]. For easy comparison, we adopt the name used in [25]. The information about the six datasets is summarized below.

- (i) 9_Tumors [32]: the dataset comes from a study of 9 human tumor types: NSCLC, colon, breast, ovary, leukemia, renal, melanoma, prostate, and CNS. There are 60 samples, each of which contains 5726 genes.
- (ii) 11_Tumors [23]: the dataset includes 174 samples of gene expression data of 11 various human tumor types: ovary, bladder/ureter, breast, colorectal, gastro-esophagus, kidney, liver, prostate, pancreas, adeno lung, and squamous lung. The number of genes is 12 533.
- (iii) 14_Tumors [24]: the dataset contains 308 samples of 14 various human tumor types including leukemia, prostate, lung, colorectal, lymphoma, bladder, melanoma, uterus, breast, renal, pancreas, ovary, mesothelioma, and CNS, and 12 normal tissues including breast, prostate, lung, colon, germinal center, bladder, uterus, peripheral blood, kidney, pancreas, ovary, and brain. Each sample has 15 009 genes.
- (iv) Brain_Tumor1 [11]: the dataset comes from a study of 5 human brain tumor types: medulloblastoma, malignant glioma, AT/RT, normal cerebellum, and PNET, including 90 samples. Each sample has 5920 genes.

TABLE 1: Results without gene selection.

Methods	Prostate_Tumor	9_Tumors	11_Tumors	14_Tumors	Brain_Tumor1	Brain_Tumor2	
SVM	OVR	93.27%	67.06%	94.99%	75.29%	90%	75.5%
	OVO	93.27%	54.63%	90.22%	46.39%	90%	73.83%
	DAG	93.27%	54.63%	90.22%	45.10%	90%	73.83%
	WW	93.27%	68.17%	94.31%	65.84%	90%	77.17%
	CS	93.27%	68.17%	94.31%	75.38%	90%	75.5%
SR	92.27%	68.79%	95.02%	74.04%	90%	80.83%	

TABLE 2: Results with gene selection.

Method	Prostate_Tumor	9_Tumors	11_Tumors	14_Tumors	Brain_Tumor1	Brain_Tumor2
SVM	94.36%	72.89%	96.66%	75.38%	90%	82.83%
	OVR	CS	OVR	CS	WW	OVR
	BW50	BW3000	KW1000	NG*	NG*	KW500
SR	94.18%	72.40%	96.10%	76.69%	90%	80.83%
	BW800	BW3000	KW2000	BW5000	NG*	NG*

* NG: no gene selection.

- (v) Brain_Tumor2 [33]: there are 4 types of malignant glioma in this dataset: classic glioblastomas, classic anaplastic oligodendrogliomas, nonclassic glioblastomas, and nonclassic anaplastic oligodendrogliomas. The dataset has 50 samples, and the number of genes is 10 367.
- (vi) Prostate_Tumor [9]: the binary dataset contains gene expression data of prostate tumor and normal tissues. There are 10 509 genes in each sample and 102 samples.

According to [25], 9_Tumors, 14_tumors, and Brain_Tumor2 are the most difficult datasets which make all the classifiers, including SVMs, generate low classification performance.

All the gene expression data are normalized by being rescaled between 0 and 1. It is also for the purpose of speeding up the training of SVMs.

3. Results and Discussion

Table 1 shows the classification results of the experiment without gene selection for both sparse representation (SR) and SVMs. The results of SVMs are slightly differently from [25]. A possible explanation is that the distribution file of cross validation is different in our study from [25]. From Table 1, the proposed SR approach performs better than all SVM variants on 9_Tumors, 11_Tumors, and Brain_Tumor2, and most SVM variants on 14_Tumors, while the SR approach performs comparably with SVM variants on Prostate_Tumor and Brain_Tumor1. In addition, similar to SVMs, the SR approach also finds it difficult to classify three multicategory datasets: 9_Tumors, 14_Tumors, and Brain_Tumor2. However, the SR approach performs better than all SVM variants on these datasets except CS and OVR on 14_Tumors. The difficulty may mainly be caused by the small number of total samples and even the smaller number of samples for each category. For example, the 9_Tumors

dataset only has 60 samples, and category 7 (prostate tumor) just has two samples.

Table 2 shows the results of sparse representation when KW and BW methods are used for gene selection, along with the best results achieved by SVMs with the corresponding gene selection methods. From Table 2, the performance of the proposed SR is comparable with the best SVM variant on all six datasets. In addition, since gene selection generate limited improvement for both methods, sparse representation approach, similar to SVMs, seems less sensitive to curse of dimensionality than non-SVM methods such as neural network and k-nearest neighbors.

It is worth mentioning that the results of SVMs for both with and without gene selection are obtained by careful model selection using 9-fold cross validation. Sparse representation approach, on the other hand, has no need of adjusting model parameters for different datasets.

As for the computing efficiency, sparse representation approach is very fast when sample number is less than 100. For example, without gene selection, it needs less than 10 seconds for Brain_Tumor2 dataset, which has only 50 samples. The efficiency, however, is dramatically reduced for relatively large sample cases. The dataset 14_Tumors, which has 308 samples, needs more than 3000 seconds! The main reason lies in the fact that the current implementation needs solving one optimization problem defined in (5) for classification of each testing sample. As a result, the number of optimization problems to be solved equals to the number of samples in the dataset. When compared with SVMs, however, the proposed SR is still faster, at least, than GEMS implementations when model selection is counted for SVMs.

4. Conclusion

In this paper, we have described a new approach for cancer diagnosis using gene expression data. The new method

expresses each testing sample as a linear combination of all the training samples. The coefficient vector is obtained by l_1 -regularized least square. Classification is achieved by defining discriminating functions from the coefficient vector for each category. Since l_1 -norm minimization leads to sparse solution, we call the new approach sparse representation.

Numerical experiments show that sparse representation approach can match the best performance achieved by SVMs. Furthermore, the new approach has no need of model selection. One direction of our future work is to investigate how to classify multiple testing samples by solving only one optimization problem to improve the efficiency.

Acknowledgments

The second author would like to thank Natural Science and Engineering Research Council of Canada (NSERC) for supporting this research. Both authors thank the editor and reviewers for their kind comments and suggestions.

References

- [1] R. Rifkin, S. Mukherjee, P. Tamayo, et al., "An analytical method for multiclass molecular cancer classification," *SIAM Review*, vol. 45, no. 4, pp. 706–723, 2003.
- [2] 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.
- [3] L. J. van't Veer, H. Dai, M. J. van de Vijver, et al., "Expression profiling predicts outcome in breast cancer," *Breast Cancer Research*, vol. 5, no. 1, pp. 57–58, 2003.
- [4] M. A. Shipp, K. N. Ross, P. Tamayo, et al., "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," *Nature Medicine*, vol. 8, no. 1, pp. 68–74, 2002.
- [5] A. Bhattacharjee, W. G. Richards, J. Staunton, et al., "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 24, pp. 13790–13795, 2001.
- [6] L. Dyrskjot, T. Thykjaer, M. Kruhoffer, et al., "Identifying distinct classes of bladder carcinoma using microarrays," *Nature Genetics*, vol. 33, no. 1, pp. 90–96, 2002.
- [7] F. Bertucci, S. Salas, S. Eysteries, et al., "Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters," *Oncogene*, vol. 23, no. 7, pp. 1377–1391, 2004.
- [8] G. P. Sawiris, C. A. Sherman-Baust, K. G. Becker, C. Cheadle, D. Teichberg, and P. J. Morin, "Development of a highly specialized cDNA array for the study and diagnosis of epithelial ovarian cancer," *Cancer Research*, vol. 62, no. 10, pp. 2923–2928, 2002.
- [9] D. Singh, P. G. Febbo, K. Ross, et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.
- [10] N. H. Segal, P. Pavlidis, W. S. Noble, et al., "Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling," *Journal of Clinical Oncology*, vol. 21, no. 9, pp. 1775–1781, 2003.
- [11] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, et al., "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, no. 6870, pp. 436–442, 2002.
- [12] D. Chen, Z. Liu, X. Ma, and D. Hua, "Selecting genes by test statistics," *Journal of Biomedicine and Biotechnology*, vol. 2005, no. 2, pp. 132–138, 2005.
- [13] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [14] J. W. Lee, J. B. Lee, M. Park, and S. H. Song, "An extensive comparison of recent classification tools applied to microarray data," *Computational Statistics & Data Analysis*, vol. 48, no. 4, pp. 869–885, 2005.
- [15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [16] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [17] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [18] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [19] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [20] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1 -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [21] E. Candès and J. Romberg, " l_1 —magic: a collection of MATLAB routines for solving the convex optimization programs central to compressive sampling," 2006, <http://www.acm.caltech.edu/l1magic>.
- [22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [23] A. I. Su, J. B. Welsh, L. M. Sapinoso, et al., "Molecular classification of human carcinomas by use of gene expression signatures," *Cancer Research*, vol. 61, no. 20, pp. 7388–7393, 2001.
- [24] S. Ramaswamy, P. Tamayo, R. Rifkin, et al., "Multiclass cancer diagnosis using tumor gene expression signatures," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 26, pp. 15149–15154, 2001.
- [25] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.
- [26] U. Kressel, "Pairwise classification and support vector machines," in *Advances in Kernel Methods: Support Vector Learning*, chapter 15, pp. 255–268, MIT Press, Cambridge, Mass, USA, 1999.
- [27] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multiclass classification," in *Advances in Neural Information Processing Systems 12*, pp. 547–553, MIT Press, Cambridge, Mass, USA, 2000.
- [28] J. Weston and C. Watkins, "Support vector machines for multi-class pattern recognition," in *Proceedings of the 7th European Symposium on Artificial Neural Networks (ESANN '99)*, pp. 219–224, Bruges, Belgium, April 1999.

- [29] K. Crammer and Y. Singer, "On the learnability and design of output codes for multiclass problems," in *Proceedings of the 13th Annual Conference on Computational Learning Theory (COLT '00)*, pp. 35–46, Stanford University, Palo Alto, Calif, USA, June–July 2000.
- [30] J. D. Gibbons, *Nonparametric Statistical Inference*, CRC Press, Boca Raton, Fla, USA, 4th edition, 2003.
- [31] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 77–87, 2002.
- [32] J. E. Staunton, D. K. Slonim, H. A. Collier, et al., "Chemosensitivity prediction by transcriptional profiling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 19, pp. 10787–10792, 2001.
- [33] C. L. Nutt, D. R. Mani, R. A. Betensky, et al., "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification," *Cancer Research*, vol. 63, no. 7, pp. 1602–1607, 2003.