

On the Classification of Cancer Cell Gene via Expressive Value Distance (EVD) Algorithm and Its Comparison to the Optimally Trained ANN Method

Tong Zhang¹

Faculty of Science and Technology, University of Macau
MSAR, PROC
ma96530@umac.mo

Chi-Hsu Wang²

Department of Electrical Engineering, National Chiaotung
University
Hsinchu, Taiwan
cwang@cn.nctu.edu.tw

Sik Chung Tam^{2*}

Faculty of Science and Technology, University of Macau
MSAR, PROC
fststct@umac.mo

^{2*} Corresponding author

C.L. Philip Chen³

Faculty of Science and Technology, University of Macau
MSAR, PROC
philipchen@umac.mo

Abstract—In recent years, cancer can be detected and recognized by analyzing the sample's expression profile. The cancer gene expression data are high dimensional, high variable dependent, and very noisy. The dimension reduction method is often used for processing the high dimensional data. In this study, a new statistical dimension reduction method called Expressive Value Distance (EVD) is developed and proposed for the practical high-dimensional gene expression cancer data. The feature genes data extracted by EVD are arranged for training the optimally trained Artificial Neural Network (ANN). The trained ANN is then used to classify whether the unseen gene data is cancer or not. In comparison of ANN classification with and without EVD, it is found that both of the ANN can classify the cancer data in good accuracy. With the EVD method, the great amount of data (2000 genes) can be effectively reduced to 16 genes. Therefore, EVD is an effective dimension reduction method. Even the EVD method is not used, the optimally trained ANN is also an advanced method for classifying the high dimensional and complicated cancer data. Briefly, it proves that optimally trained ANN is a very robust classification technique.

Keywords—Gene expression profile; Dimension reduction; Artificial neural network; Expressive value distance; Classification of cancer;

I. INTRODUCTION

Classification has become one of the most promising learning techniques. It is crucial in many applications such as cancer classification [1]. Such classifications have an immense impact on cancer prognosis and therapy. In contrast with clustering, classification techniques start with a collection of labeled (pre-classified) expression patterns. The purpose is to train a classification model that would be able to classify a new expression pattern. In recent years, the development of microarray technology has allowed the simultaneous monitoring of thousands of genes [2, 3], which have motivated

the progress in cancer classification by using gene expression [4, 5, 6].

Known as a parallel distributed processing network, artificial neural network is a robust tool recently used as classification. It has been applied to a very broad range of tasks in many different disciplines. It consists of interconnected processing elements called nodes or neurons that work together to produce an output function. The output of a neural network relies on the cooperation of the individual neurons within the network.

It becomes a reality to apply neural network with bioinformatics technology to cancer diagnosis for the development of DNA microarray technology. It has attracted tremendous interest in both the scientific community and in industry. Several studies have recently reported on the application of microarray gene expression analysis for the classification of cancer [7]. Most of these applications assumed a predefined network architecture (including connectivity and node transfer functions) and used a training algorithm to learn weights.

In this study, 62 colon cells with 2000 genes expression data are analyzed (22 normal and 40 cancer cells). On the basis of the feature of gene expression data (high dimensional, high variable dependent, and very noisy), feature extraction by dimension reduction is necessary before the classification from the original data set [8, 9]. To classify the high dimensional gene expression cancer data, the dimension reduction methods were often used, such as principal component analysis (PCA), generalized topographic mapping, projection pursuit and so on [10]. After the dimension is reduced, the feature genes can be obtained and analyzed more easily. A new dimension reduction method (EVD) based on the concept of probability distribution and hierarchical clustering method is proposed for cancer data. After the feature gene data is obtained, the feature genes data will be arranged as the input for training the optimal trained Artificial Neural Network (ANN) [11]. The trained ANN is

used to classify whether the unseen gene data is cancer or not. The benchmark classification tests have been performed to show that the EVD can yield a slightly better result than that of the optimally trained ANN method. For computational complexity comparison, it is obvious that the optimally trained ANN has less computational complexity than that of EVD, and yet provides a very robust classification technique with good accuracy.

II. THE ALGORITHMS

2.1 Classification method

The structure of two layers ANN is shown in Fig. I, which will be used as a cancer gene classifier in this paper.

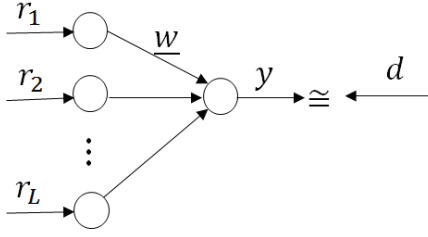


Fig. I. Two-layers ANN

Where

$$\underline{r} = [r_1, r_2, \dots, r_L]^T \in \mathbb{R}^L \quad (2.1)$$

is the input vector of the network

$$\underline{w} = [w_1, w_2, \dots, w_L]^T \in \mathbb{R}^L \quad (2.2)$$

is the weighting vector $y \in \mathbb{R}$ is the network output, and “ T ” is the transpose operator of matrices.

Hence, according to the structure of the two layers ANN, we obtain the following formula

$$y = \sum_{l=1}^L r_l w_l = \underline{r}^T \underline{w} \quad (2.3)$$

Given a training dataset with P samples

$$\{(\underline{r}_1, d_1), (\underline{r}_2, d_2), \dots, (\underline{r}_P, d_P)\},$$

We define the sum of square error which is a function of the weight vector $\underline{w}$ by

$$\begin{aligned} J(\underline{w}) &= \frac{1}{2P} (\underline{y} - \underline{d})^T (\underline{y} - \underline{d}) \\ &= \frac{1}{2P} (\underline{R}^T \underline{w} - \underline{d})^T (\underline{R}^T \underline{w} - \underline{d}) \end{aligned} \quad (2.4)$$

Where

$$\underline{R} = [\underline{r}_1, \underline{r}_2, \dots, \underline{r}_P] \in \mathbb{R}^{L \times P} \quad (2.5)$$

$$\underline{y} = [y_1, y_2, \dots, y_P] \in \mathbb{R}^P \quad (2.6)$$

$$\underline{d} = [d_1, d_2, \dots, d_P] \in \mathbb{R}^P \quad (2.7)$$

Indeed to train the ANN is to minimize the error function $J(\underline{w})$ in equation (2.4). The classical approach to find the minimum weight vector is using steepest descent method. Precisely, we will construct an iteration sequence which is expected to converge to the minimum weight vector as below:

$$\underline{w}(t+1) = \underline{w}(t) - \beta(t) \frac{\partial J(\underline{w})}{\partial \underline{w}} \bigg|_{\underline{w}(t)} \quad (2.8)$$

Where t denotes the t th iteration and the number $\beta(t)$ is the learning rate. By chain rule, the gradient $\frac{\partial J(\underline{w})}{\partial \underline{w}}$ is equal to $\frac{1}{P} \underline{R}(\underline{R}^T \underline{w} - \underline{d})$. Hence, we have

$$\underline{w}(t+1) = \underline{w}(t) - \frac{\beta(t)}{P} \underline{R}(\underline{R}^T \underline{w}(t) - \underline{d}) \quad (2.9)$$

It was shown that the relatively large or small learning rates may affect the learning progress and even may lead to failure of the process [12]. So, to avoid these situations, dynamic optimal learning rate training for ANN is employed to train the two layers ANN in this paper. It is not only guaranteed that the maximum error reduction can be achieved by choosing the optimal learning rate for the next training iteration, but also that the initial choice of the weights become not essential. According to Theorem 1 in [11], the optimal learning rate $\beta_{opt}(t)$ can be found from the minimum of a quadratic polynomial $A\beta^2 + B\beta$ where

$$\begin{aligned} A &= \frac{1}{2P^3} \underline{R}^T \underline{R} \underline{e}_t \underline{e}_t^T \underline{R}^T \underline{R} \\ B &= \frac{1}{P^2} \underline{R}^T \underline{e}_t \underline{e}_t^T \underline{R} \\ \underline{e}_t &= \underline{R}^T \underline{w}(t) - \underline{d} \end{aligned} \quad (2.10)$$

The dynamic optimal learning rate training is summarized as follow:

Algorithm 1: Classification by using ANN with Dynamic Optimal Learning Rates

1) Training the classifier

step1. Given P training data and obtain $\underline{R}$ and $\underline{d}$ as above.

step2. Set the iteration counter $t = 0$ and the initial weight to be $\underline{w}(0)$. Find the optimal learning rate $\beta_{opt}(0)$. Start the steepest descend to the error function $J(\underline{w})$.

step3. Find if the error function $J(\underline{w})$ is smaller than a given tolerant ε ? If yes, go to Step 7.

step4. Update the weighting matrix to yield $\underline{w}(t+1)$ according to (2.9)

step5. Find the optimal learning rate $\beta_{opt}(t+1)$ for the next iteration.

step6. Set $t = t + 1$. Go back to Step 3.

step7. Store the current $\underline{w}(t)$ to $\underline{w}_{opt}$ and end.

2) Testing the classifier:

step8. Given another set of testing data and form $\underline{R}_{test}$ and $\underline{d}_{test}$ as if in step 1 of the training process.

step9. Obtain the trained network output vector $\underline{y}_{test}$ by using the formula $\underline{y}_{test} = \underline{R}_{test}^T \underline{w}_{opt}$.

step10. Convert the output $\underline{y}_{test}$ into the classification label (1/-1).

step11. Compare $\underline{y}_{test}$ and $\underline{d}_{test}$ and have the diagnostic accuracy.

step12. End.

2.2 Dimension reduction method

Gene expression profile of a DNA is very complicated and it can be represented by a high dimensional vector. For example, in this paper, the dimension of such vector is up to 2000. However, most the genes within DNAs are not relevant

for cancer classification and they should be eliminated [13, 14]. Besides, the rest of the genes can be grouped into clusters. The genes within a cluster actually are highly correlated. Therefore, one of the genes in a cluster can be chosen as its representative. Based on this theory, we are able to reduce the dimension of the gene expression profile vectors. In our study, we have actually implemented such dimension reduction by vertical process and horizontal process. The vertical process is to eliminate those irrelevant genes by using EVD method, while the horizontal process is to classify the resulted genes from vertical process into clusters by the hierarchical clustering method and to choose one representative gene from each cluster [15].

A. Vertical process

First of all, we assume that the condition distribution of a gene expression given that it is from a normal DNA (or a cancer DNA) follows normal distribution. Therefore, the conditional probability density function of a gene expression is

$$f(x|normal) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2}$$

and

$$f(x|cancer) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2}$$

Note that practically, the mean and the variance can be replaced by its sample mean and sample variance respectively in order to obtain the above density functions.

Definition 1: The expressive value distance between $f(x|normal)$ and $f(x|cancer)$ of a gene expression is defined to be **EVD**:

$$EVD = \int_{\min(x)-\delta}^{\max(x)+\delta} |f(x|normal) - f(x|cancer)| dx \quad (2.11)$$

Where δ is called the generalization error of the distribution area and

$$\delta = \frac{\min(x) - \min(x)}{5} \quad (2.12)$$

Algorithm 2: Vertical process of EVD

- step1.** Calculate the EVD of all the gene expressions by using equation (2.11).
- step2.** Set the EVD threshold α .
- step3.** Select the i^{th} gene if its EVD is bigger than α to form a new feature gene set and its size is denoted by L_1 .
- step4.** End.

B. Horizontal process

In order to have further reduction of the numbers of genes, we are going to group the genes with similar gene expression. Assume a gene expression profile x_k of the resulted dataset of vertical process. Therefore,

$$x_k = (x_{k1}, x_{k2}, \dots, x_{kL_1}) \in \mathbb{R}^{L_1}$$

Where $k = 1, 2, \dots, n$ and n is the number of samples.

The correlation between i^{th} genes and j^{th} gene is defined as below

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{[\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2]^{\frac{1}{2}}} \quad (2.13)$$

Where $\bar{x}_i$ and $\bar{x}_j$ are the sample means of the gene expression of i^{th} and j^{th} genes respectively.

Then, we get the relation matrix of the feature genes as below (2.14):

$$RM = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1L_1} \\ r_{21} & r_{22} & \dots & r_{2L_1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{L_11} & r_{L_12} & \dots & r_{L_1L_1} \end{bmatrix} \quad (2.14)$$

The average related value represent as (2.15)

$$\overline{RM}(C_1, C_2) = \frac{1}{n_{C_1} \times n_{C_2}} \sum_{x_k \in C_1} \sum_{x_m \in C_2} r_{km} \quad (2.15)$$

Where C_1 and C_2 stand for two classes of genes and n_{C_1} and n_{C_2} are the sizes of C_1 and C_2 respectively.

Algorithm 3: Horizontal process of EVD

- step1.** Fix a integer number L_2
- step2.** Clustering the output gene set of vertical process by classical hierarchical clustering method (bottom-up), refer to [15], with distance function $\overline{RM}(C_1, C_2)$ and $K = L_2$.
- step3.** Choose one gene from every sample of the final L_2 clusters to form a reduce list of genes for ANN classification.
- step4.** End.

III. EXAMPLE

In this paper, the expression of colon dataset of 62 cells were used, 22 normal cells and 40 cancer cells (Fig. II and Table I). The dataset was free for download in the website [16].

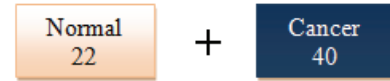


Fig. II. Colon cancer dataset

TABLE I

COLON CANCER DATASET

Name Sample		1_Has.3004	...	2000_Hsa.9683
S1	N	12.77219	...	5.888287
S2	N	10.90289	...	4.252098
S3	N	11.83636	...	4.942221
⋮	⋮	⋮	⋮	⋮
S61	C	13.15361	...	5.007476
S62	C	12.60609	...	4.540089

Then 6 of 22 normal cells and 16 of 40 cancers randomly was selected as the training set. The remains were arranged into the testing set (see Fig. III).

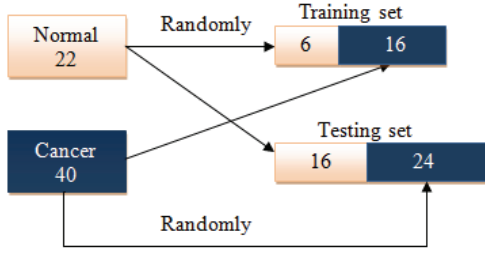


Fig. III. Colon cancer dataset

For comparison purpose, we did two kinds of methods experiments with ANN for classification in this study. The first method, the ANN was treated as the classifier directly. The second method, apply the dimension reduction method EVD before classification by ANN. (See Fig. IV).

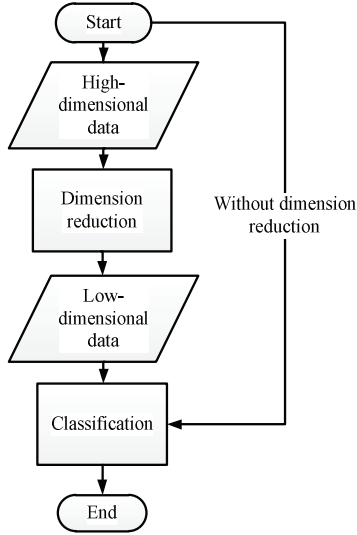


Fig. IV. Flow chart of two groups of classification

By using algorithm 1, the variables were listed as below:

R : refers to the expression dataset matrix of colon cancer, treated as the input matrix. And $L = 2000, P = 22$.

$\underline{d}$: refers to the desired output vector. $\underline{d} = [d_1, d_2, \dots, d_P]$ $d_1 \sim d_6$ equal to 1, $d_7 \sim d_{22}$ equal to -1 . 1 stand for normal, and -1 stands for cancer.

R_{test} : refers to the testing data. And $L = 2000, P_{test} = 40$.

$\underline{d}_{test}$: refers to desired output of the testing input R_{test} . $\underline{d}_{test} = [d'_1, d'_2, \dots, d'_{40}]$, $d'_1 \sim d'_{16}$ equal to 1, $d'_{17} \sim d'_{40}$ equal to -1 .

The tolerant $\epsilon = 0.008$, i.e. when $J(\underline{w}) \leq 0.008$, $\underline{d}$ and $\underline{y}$ are closed enough. Then, the current $\underline{w}(t)$ was stored to $\underline{w}_{opt}$.

Thus, we test the classifier by comparing $\underline{y}_{test}$ and $\underline{d}_{test}$ we have the diagnostic accuracy of classification using ANN listed in Table IV.

By using dimension reduction method mentioned as EVD, we get the EVD bar chart (see Fig. V) of the initial genes.

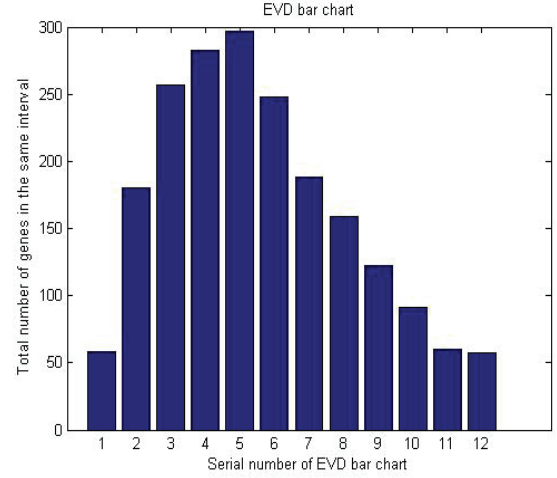


Fig. V. Statistical graph of EVD

Considering the statistical result of the EVD, in Table II, the threshold of the genes was set as $\alpha = 0.8$. Thus, L_1 ($L_1 = 330$) genes ($EVD > 0.8$) were selected and form the feature dataset in the vertical process.

TABLE II

THE STATISTICAL DATA OF EVD

EVD	0~0.1	0.1~0.2	0.2~0.3	0.3~0.4
Number of genes	58	180	257	283
EVD	0.4~0.5	0.5~0.6	0.6~0.7	0.7~0.8
Number of genes	297	248	188	159
EVD	0.8~0.9	0.9~1.0	1.0~1.1	1.1~1.5
Number of genes	122	91	60	57

In horizontal process, dimensions of the feature genes from the vertical process were reduced by using algorithm 3. Here, L_2 was set to be 16. That is to say, 16 clusters were clustered in the final. Then, choose the gene, whose EVD is the largest in each cluster.

Since the indexes of the genes have been stored in every step, the sequence number of the genes was laid out in Table III, i.e. these 16 genes form the feature genes dataset.

Then we apply algorithm 1 by using ANN as a cancer gene classifier. Here, $\underline{d}'$ and $\underline{d}'_{test}$ are same with $\underline{d}$ and $\underline{d}_{test}$ in the first experiment. However, L has been reduced to L_2 applying the EVD algorithm. Then, the variables were listed as below:

R' : refers to the expression dataset matrix of colon cancer, treated as the input matrix. And $L_2 = 16, P' = 22$.

$\underline{d}'$: refers to the desired output vector. $\underline{d}' = \underline{d}$.

R'_{test} : refers to the testing data. And $L_2 = 16, P'_{test} = 40$.

$\underline{d}'_{test}$: refers to desired output of the testing input R'_{test} , $\underline{d}'_{test} = \underline{d}_{test}$.

Thus, the trained network output vector $\underline{y}'_{test}$ can be obtained by using the formula $\underline{y}'_{test} = R'^T \underline{w}'_{opt}$. By comparing $\underline{y}'_{test}$ and $\underline{d}'_{test}$ we have the diagnostic accuracy of classification using ANN with EVD. (Refer to Table IV).

TABLE III

SEQUENCE NUMBER OF THE FEATURE GENES IN THE INITIAL DATASET

sequence number	1680	177	1559	260
name	Hsa.1380	Hsa.8831	Hsa.1731	i
sequence number	686	1792	1655	1895
name	Hsa.1920	Hsa.1057	Hsa.30310	Hsa.8192
sequence number	1025	715	673	1636
name	Hsa.3215	Hsa.44686	Hsa.10908	Hsa.345
sequence number	765	1505	1504	1732
name	Hsa.692	Hsa.862	Hsa.44244	Hsa.888

TABLE IV

THE ACCURACY FOR CLASSIFICATION OF THE TWO METHODS

Method Test times	Accuracy of classification use ANN	Accuracy of classification using ANN with EVD
1	0.9000	0.9500
2	0.9000	0.9750
3	0.9500	0.9000
4	0.8500	0.8750
5	0.9500	0.9500
6	0.9000	0.9000
7	0.9500	0.9500
8	0.8750	0.8750
9	0.8500	0.9000
10	0.9000	0.9500
11	0.9500	0.9000
12	0.8750	0.8750
13	0.9000	0.9250
14	0.9500	1.0000
15	0.9000	0.9500
16	0.8500	0.9000
17	0.9250	0.8500
18	0.9000	0.9500
19	0.9250	0.9250
20	0.8750	0.8750
Average accuracy	0.9038	0.91875

IV. EXPERIMENTAL RESULTS

The program in this study was written in the Matlab® code and operated in the laptop PC with Windows® platform. The accuracy of classification is described as Table IV.

We compute 20 times by using optimal trained Artificial Neural Network (ANN) and ANN method with the Expressive Value Distance (EVD) algorithm separately. Then we get the results of the accuracy respectively (see Table IV). It suggests that both of the two kinds of methods could be used for colon cancer classification efficiently. By the comparison of these two ways of training methods, we found that the average accuracy is slightly better by using EVD. So, it has been shown that EVD is an effective method in the area of dimension reduction. For computational complexity comparison, it is obvious that the optimally trained ANN has less computational complexity, and yet provides a very robust classification technique with good accuracy.

V. CONCLUSION

In this study, the new statistical dimension reduction algorithm EVD is developed successfully. It is applied to the practical colon cancer dataset. After using the EVD method, the dimension of the input data to ANN was largely reduced from 2,000 to 16. The reduction of the number of inputs does not affect the subsequent optimally trained ANN approach because both of the ANNs with or without the EVD method can provide the very accurate classification results. It is also represented that the important feature of the data can be extracted by the EVD method, although the optimally trained ANN approach can provide less computational complexity.

ACKNOWLEDGMENT

This research is supported by the research grant of the National Grand Fundamental Research 973 Program of China under Grant 2011CB302801.

REFERENCES

- [1] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, Chih-Jen Lin, "LIBLINEAR: A Library for Large Linear Classification," Journal of machine learning, pp. 1871-1874, Sept. 2008.
- [2] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," Nature, vol. 403, pp. 503-511, 2000.
- [3] X. Ma, R. Salunga, J. T. Tuggle, J. Gaudet, E. Enright, P. McQuary, T. Payette, M. Pistone, K. Stecker, B. M. Zhang, et al., "Gene expression profiles of human breast cancer progression," in Proc. Natl. Acad. Sci. USA, vol. 100, pp. 5974-5979, 2003.
- [4] J. M. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," Nature Medicine, vol. 7, pp. 673-679, 2001.
- [5] A. B. Olshen, and A. N. Jain, "Deriving quantitative conclusions from microarray expression data," Bioinformatics, vol. 18, pp. 961-970, 2002.

- [6] R. Tibshirani, T. Hastie, B. Narashiman, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," in *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 6567-6572, 2002.
- [7] Sridhar Ramaswamy, Pablo Tamay, Ryan Rifkin, Sayan Mukherjee, Chen-Hsiang Yeang, Michael Angelo, et al., Multiclass cancer diagnosis using tumor gene expression signatures, the national academy of sciences of the United States of America, 2001. 98(26): pp. 15149-15154.
- [8] Kyeong Eun Lee, Naijun Sha, Edward R. Dougherty, Marina Vannucci and Bani K. Mallick, "Gene selection: a Bayesian variable selection approach," *BIOINFORMATICS*, 2003. 19(1): pp. 90-97.
- [9] Deutsch, J.M., "Evolutionary algorithms for finding optimal gene sets in microarray prediction," *BIOINFORMATICS*, 2003. 19(1): pp. 45-52.
- [10] M. C. Mozer, M. I. Jordan, and T. Petsche, eds., *A principled alternative to the self-organizing map*, in *Advances in Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA, 1997.
- [11] Chi-Hsu Wang, Han-Leih Liu, and Chin-Teng Lin, "Dynamic Optimal Learning Rates of a Certain Class of Fuzzy Neural Networks and its Applications with Genetic Algorithm," *IEEE transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, VOL. 31, NO. 3, JUNE 2001.
- [12] Xiao-Hu Yu, Guo-An Chen, and Shi-Xin Cheng, "Dynamic learning rate optimization of the back propagation algorithm," *IEEE Trans. Neural Networks*, vol. 6, pp. 669-677, May 1995.
- [13] U. Alon, N. Barkai, D. A. Nottierman, K. Gish, S. Ybarra, D. Mack, AND A. J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, national academy of sciences of the United States of America, 1999, 96(12): pp. 6745-6750.
- [14] Zhang H, Yu C-Y, Singer B, Xiong M., Recursive partitions for tumor classification with gene expression microarray data, the national academy of sciences of the United States of America, 2001. 98(12): pp. 6730-6735.
- [15] Jain, A. and Dubes, R., *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [16] Kaufman, L. and Rousseeuw, P., *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley and Sons, New York, NY. 1990.
- [17] (2011) The Princeton University website. [Online]. Available: <http://genomics-pubs.princeton.edu/oncology/>
- [18] P. Berkhin, *A Survey of Clustering Data Mining Techniques, Grouping Multidimensional Data*, Springer, 2006, 25-71.
- [19] R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Class predication by nearest shrunken centroids with applications to DNA microarrays, *Statistical Science*, 2003, vol. 18, pp. 104-117.
- [20] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science*, 1995, vol. 270, pp. 467-470.
- [21] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, et al., Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science*, 1999. 286(5439): pp. 531-537.
- [22] M. P. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Jr. Ares, D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines, the national academy of sciences of the United States of America, 2000, vol. 97, pp. 262-267.
- [23] Lillian Y. Chang, Nancy S. Pollard, Tom M. Mitchell, and Eric P. Xing, "Feature selection for grasp recognition from optical markers," in *Proc. IEEE International conference on Intelligent Robots Systems*, VOLS 1-9, 2007: pp. 2950-2956.