

A Classification of Microarray Gene Expression Data Using Hybrid Soft Computing Approach

Vaishali P Khobragade¹, Dr.A.Vinayababu²

¹ Jyothishmathi Institute of Technology & Science, CSE Department, JNTU Hyderabad
Karimnagar, AndhraPradesh 505001, India

² Principal JNTUH College of Engineering, JNT University
Hyderabad, AndhraPradesh 500085, India

Abstract

In this paper, an efficient technique is proposed for the precise classification of microarray genes from the microarray gene expression dataset. The proposed classification technique performs the classification process with the aid of three phases namely, dimensionality reduction, feature selection and gene classification. Initially, the proposed technique reduces the dimensionality by utilizing Genetic Algorithm (GA). The main objective of dimensionality reduction is to select the optimal number of genes from the microarray gene expression dataset. Next, in the feature selection process the features are extracted from the column gene values. Here, probability of GA-indexed gene and new statistical features are selected for each column gene values and these selected features are given to the Feed Forward Back propagation Neural Network (FFBNN). The FFBNN network is trained using the selected features and then this well trained FFBNN network performance is tested with the column gene values. The FFBNN network classifies the microarray gene values into their corresponding cancer class types. The performance of the classification technique is evaluated by the performance measures such as accuracy, specificity and sensitivity. **Keywords:** *Micro array gene expression data, Classification, Feed Forward Back Propagation Neural Network (FFBNN), Statistical measures.*

1. Introduction

Deoxyribonucleic acid (DNA) microarray technology provides tools for monitoring the expression levels of huge number of different genes simultaneously [9]. It is possible for the biologists to concurrently evaluate the expressions of thousands of genes in a single experiment by the aid of microarray technologies [5] [15] [19]. This technology provides a unique tool, which has been currently used for medical diagnosis and gene analysis, especially to inspect how a cell's gene expression pattern changes in different conditions. A microarray method also plays an important role in personalized medicine because it can be used to identify the individual's unique genetic vulnerability to treat the diseases [1]. High-throughput microarray technology presents a robust tool in biomedical research. Particularly, DNA microarray profiling technology is very useful in disease diagnosis and forecast, as well as in subtype detection [6] [16]. A standard microarray experiment dataset contains expression levels of a large number of genes in a number of experimental samples or conditions [10]. The expression data is represented in a matrix form, where the rows are denoting genes and the columns are denoting samples. This matrix is termed as gene expression matrix [11]. Gene expression data is often used in disease analysis, especially for cancer diagnosis [8]. Gene expression data from DNA microarrays are described by several variables (genes) with only a small number of observations (experiments) [7][17]. Prediction, classification, and clustering methods are employed

for analysis and understanding of the data [2]. One salient application of gene expression microarray data is the classification of biological samples or prophecy of clinical and other outcomes [3]. Microarray technology is used to categorize the tissue samples by using their gene expression profiles as one of the several types (or subtypes) of cancer. The gene expression profiles measured by microarray technology have given an exact, consistent and objective cancer classification than the standard histopathological tests. The DNA microarray data for cancer classification contains huge number of genes (dimensions) than the number of samples or feature vectors [4] [18]. The gene expression variation of different tumor types is measured by using the genome-wide expression data obtained from the cancer tissues, which further provides hints for cancer classification of individual samples. Extracting biological insights from the original amount of data on gene expression patterns is the main challenging tasks in microarray studies [12]. An efficient model is necessary for predicting the class membership of data, creating an exact label on training data, and predicting the label for any anonymous data correctly in order to achieve a high classification accuracy [8]. Classification analysis of microarray gene expression data has been performed extensively to find out the biological features and to differentiate intimately related cell types that usually appear in the diagnosis of cancer [13]. Some of the classification techniques for gene expression data analysis are classification decision tree, k-nearest neighbor classifier (KNN), support vector machine (SVM), neural network etc. Generally, the techniques used for the classification of microarray gene expression data are divided into two classes: one is based on clustering and the other is based on machine-learning approach [14]. Plenty of researches have been performed for the successful microarray gene cancer classification. A few recent works available in the literature are reviewed in the following section.

2. Related work

Ahmad M. Sarhan [20] has proposed a robust system for the identification of stomach cancer using Artificial Neural Network (ANN) and Discrete Cosine Transform (DCT). In this proposed system, the DCT has been used to extract the classification features from the stomach microarrays. Subsequently, the features extracted from the DCT coefficients have been applied to an ANN for the classification in order to find whether the microarray contains tumor or non-tumor. Here, the microarray images have been taken from the database called Stanford Medical Database (SMD), which is one of the famous microarray databases. Simulation results have proved that the proposed system has achieved a very high success rate. Bharathi *et al.* [21] have attempted to identify the minimum set of genes that can provide a precise classification of cancer from microarray data via supervised machine learning algorithms. The importance of determining the smallest gene subset has three advantages: i) The computational burden and noise occurred from irrelevant genes have been minimized considerably. ii) The gene expression tests have been simplified to comprise only a very less number of genes instead of thousands of genes, so the expense for cancer testing has been reduced greatly. iii) Further analysis has been made into the possible biological relationship between this small number of genes and cancer development and treatment. Their simple yet very efficient method contains two steps. In the first step, a 2 way Analysis of Variance (ANOVA) ranking scheme has been employed to select some significant genes. In the second step, a good classifier such as Support Vector Machines has been utilized to analyze the classification potency of all simple combinations of those significant genes. Their approach has achieved a very high precision with only two genes. Bo Li *et al.* [22] have discussed that the gene expression data gathered from DNA microarray are characterized by a huge number of variables (genes), but with only a small number of observations i.e., experiments. They have proposed a manifold learning technique to map the gene expression data to a low dimensional space, and then the basic structure of the features have been analyzed in order to categorize the microarray data more precisely. Their proposed algorithm has projected the gene expression data into a subspace with high intra-

class compactness and inter-class separability. Xiaosheng Wang *et al.* [23] have analyzed the properties of one feature selection approach that was proposed in their previous work, which was the simplification of the feature selection technique based on the depended amount of attribute in rough sets. The feature selection technique has been compared with the conventional methods regarding the depended degree, chi-square, information gain, Relief-F and symmetric uncertainty, and its properties have been evaluated through a series of classification experiments. Mallika *et al.* [24] have presented a technique for enhancing the classification performance for cancer classification with a small number of microarray gene expression data. In this proposed approach, individual gene ranking and gene subset ranking has been performed. Here, the same classifier has been employed for both selection and classification purpose. Chanda Ray [25] has proposed an algorithm to inspect the DNA microarray gene expression patterns robustly for large amount of DNA microarray data. Graphical representation has been presented for the experimental results of DNA microarray gene pattern analysis for enhanced visibility and understanding. An eight-directional chain code sequence has been employed to represent the shape of each graph related to a DNA microarray gene expression pattern. Seeja *et al.* [26] DNA microarrays allow the biologist to evaluate the performance of thousands of genes concurrently on a small chip. These microarrays produce large number of data and new techniques are required to analyze them. In this paper, a new classification technique based on support vector machine is proposed. The proposed technique is employed to categorize the gene expression data recorded on DNA microarrays. The proposed technique is tested by using benchmark datasets and it is found that the proposed technique is faster than neural network and the classification performance is also high when compared to neural network. Most of the existing research works have achieved the cancer classification process by extracting the features using threshold and statistical features based methods. In such threshold and statistical features based methods, the features are selected based on the defined threshold value, but some genes are not suitable for this threshold value. Also, the method has more chance for missing the significant features from the classification process. If

the number of genes is decreased, then it provides only a low accuracy result. So, it is a major drawback of this method. Hence, all these drawbacks have degraded the performance of the microarray gene expression cancer classification. To shun this drawback, we propose an efficient approach to classify the microarray cancer genes more accurately. This classification technique executes the classification process by reducing the dimension of the dataset and selecting the significant features. The FFBNN network is utilized for microarray gene classification. The outline of the paper is as follows: The proposed gene classification process is briefly explained in section 3. In section 3.1 and 3.2, the dimensionality reduction process and the feature selection process are described respectively. The microarray gene classification process is performed in section 3.3. The experimental result and conclusion of this paper are given in section 4 and 5 respectively.

3. Proposed Micro Array Gene Cancer Classification

Here, we propose an efficient technique to classify the microarray gene cancer data. The proposed classification technique is comprised of three phases, namely, (i) dimensionality reduction, (ii) feature selection and (iii) cancer gene classification. The pictorial representation of the proposed technique is shown in figure 1.

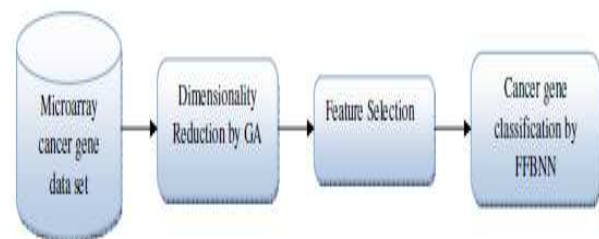


Fig 1: Proposed Microarray Gene Classification Technique Structure

In figure 1, the dimensionality reduction process reduces the dimension of microarray cancer gene dataset by utilizing Genetic Algorithm (GA) and the dimensionality reduced data's are given to the feature selection phase. In the feature selection phase, the features are selected from the dimensionality

reduced microarray cancer gene dataset. After that, the selected features are given to the feed forward back propagation neural network (FFBNN) to perform the gene cancer classification process. The proposed cancer gene classification process is explained in the following subsections. Let, M_{ij} ; $1 \leq i \leq S$, $1 \leq j \leq G$ be the microarray cancer gene data, where, S represents the number of samples and G represents the number of genes.

3.1 Dimensionality Reduction using GA

Initially, the dimensionality reduction process is performed on the microarray cancer gene dataset for reducing the complexity in the gene classification. Because the dataset size is high dimensional, which increases the processing time and also it will not produce accurate result for the classification process. The high dimensional M_{ij} dataset is converted into low dimensional dataset by selecting the optimal number of genes. To accomplish the optimal gene selection process, we are using Genetic algorithm. The procedures that are involved in GA operation are discussed below.

Initial population:

The chromosome's gene value is randomly generated between the intervals $[1, r_G]$ in the dataset M_{ij} , and N_p is the population size.

Fitness Function:

The fitness function is done to select the best chromosomes among the generated chromosomes. The fitness function calculation plays a major role in GA. The following fitness formula is calculated for each chromosome to select the best chromosomes for feature selection process. The fitness function formula is given below

$$f1 = \frac{S^{(C_m^{(s)})} * S^{(l_1)} * S^{(l_2)} * \dots * S^{(l_N)}}{E^{(C_m)} * T^{(C_m)} * t_1} \quad (2)$$

The standard deviation is calculated of the N_p chromosomes.

Rank Selection:

The best $N_p/2$ chromosomes having minimum fitness values are selected. The selected chromosomes are used for performing the genetic operations such as crossover and mutation.

Termination:

The process is continued until it reaches the utmost number of iterations I . Once it reaches I , the $N_p/2$ chromosomes containing minimum fitness value are selected. Then, the feature selection process is performed by using one of the $N_p/2$ best chromosomes.

3.2 Feature Selection

The best chromosomes that are selected from the GA process are given to the feature selection phase. The features like Standard Deviation, Probability of GA-indexed gene and new statistical features are selected from dimensionality reduced data set.

Probability of GA-indexed gene

Probability of GA-indexed gene feature is computed by utilizing each column gene values in the reduced set. Each gene value in the column is compared with other gene values in the same column. In this comparison, each gene value computes the number of gene values which are greater than the current gene values and this number of greater value count is the Probability of GA-indexed gene feature of each gene value. The Probability of GA-indexed gene feature is computed for all the columns.

3.2.2 New Statistical Features

Mean (A), Median (D), Population and Sample standard deviations (PS &SS) features are extracted from the gene values in the reduced data set. The features computation initially performs distinction process between the genes in column. The new statistical feature extraction process is different from

the normal statistical feature extraction process. Here, the new statistical feature extraction process is performed after computing the distinction process between genes in the column and then, we select the genes that are highly deviated between the class types. The selection of highly deviated genes for the new statistical feature extraction process gives high classification result.

3.3 Classification using Feed Forward Neural Network

The classification process is performed by utilizing the SD, Probability of GA-indexed gene and new statistical features from the previous phase. To perform the classification process, here we utilize Feed Forward Back Propagation Neural Network (FFBNN). Each column in P_{uv} contains six features and these features are given to the FFBNN to perform the training process. Network training and testing process is performed by using these extracted features. In the training phase, the extracted features (feature selection described in the section 3.2) are given to FFBNN network. The FFBNN network is well trained by these extracted features. The network is created with six input unit, H_d hidden units and one output unit. The basic structure of FFBNN network is shown in fig.2.

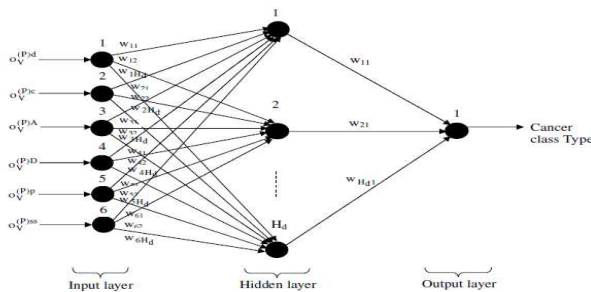


Fig 2: Proposed Gene Cancer Classification FFBNN Structure

The FFBNN network is well trained by the extracted features and this network classifies the particular column gene values to which the class types it belong. Testing process is performed for the column gene values in the dimensionality reduced dataset P_{uv} . Column values are selected from P_{uv} and the feature selection process is performed (described in section 3.2). The extracted Standard

Deviation, Probability of GA-indexed gene and new statistical features are given to the FFBNN network. The well trained FFBNN network classifies the columns gene values into any of the cancer class type by using the extracted features. In this manner, the microarray gene expression dataset gene values are classified.

4. Results and Discussion

The proposed classification technique is implemented in MATLAB platform (version 7.8) and it is evaluated using the microarray gene expression dataset. The dataset contains 675 genes and 156 samples. The high dimensional dataset is subjected to dimensionality reduction using a Genetic Algorithm (GA) and so the dataset dimension is reduced to 10 genes, and 156 samples. Among these 156 samples, 1 to 139 samples are belongs to the AD class type and 140 to 156 samples are belongs to the NL class type. The dimensionality reduced dataset with 10 genes are used in the feature selection process. In the feature selection phase standard deviation, Probability of GA-indexed gene and new statistical features are extracted. These extracted features are then given to the FFBNN training and testing process. The training dataset contains 156 training samples and testing dataset contains 156 testing samples i.e., initially 155 samples are utilized for training and the 156th sample is used for testing, subsequently, 1 to 154 and 156th samples are exploited for training process and 155th sample is used for testing process. This procedure is repeated until all the 156 samples are involved in both the training and testing process. The performance of proposed method is analyzed by using the statistical measures. The statistical measures [27] are used to measure the classification performance. The performance analyses have shown that the proposed method has been successfully classified the genes in their specified gene data types. The performance of proposed gene classification method is analyzed with GA parameters by tuning these GA parameter values. The GA parameters crossover rate and mutation rate are involved in the tuning process. The population size is sets 10 and the chromosome length to be changed 10, 20 and 30. Moreover the existing SVM result values are obtained by changing the gene values 10, 20 and 30.

Table 1 tabulates the TP, TN, FP and FN values from the GA parameter tuning process

Population size	Crossover Rate	Mutation Rate	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	ACC
10	0.2	0.1	136	11	6	3	91.03
	0.3	0.1	133	17	0	6	85.26
	0.4	0.1	138	17	0	1	88.46
	0.5	0.1	135	11	6	4	90.38
	0.2	0.2	134	13	4	5	88.46
	0.3	0.2	137	14	3	2	89.74
	0.4	0.2	137	17	0	2	87.82
	0.5	0.2	135	17	0	4	86.54
	0.2	0.3	137	14	3	2	89.74
	0.3	0.3	127	12	5	12	84.62
	0.4	0.3	134	9	8	5	91.03
	0.5	0.3	134	16	1	5	86.54
	0.2	0.4	127	9	8	12	86.54
	0.3	0.4	124	10	7	15	83.97
	0.4	0.4	131	14	3	8	85.90
	0.5	0.4	133	14	3	6	87.18
	0.2	0.1	134	14	3	5	87.82
	0.3	0.1	136	9	8	3	92.31
	0.4	0.1	134	12	5	5	89.10
	0.5	0.1	137	8	9	2	93.59
	0.2	0.2	136	11	6	3	91.03
	0.3	0.2	131	8	9	8	89.74
	0.4	0.2	121	8	9	18	83.33

20	0.5	0.2	135	9	8	4	91.67
	0.2	0.3	136	15	2	3	88.46
	0.3	0.3	134	13	4	5	88.46
	0.4	0.3	136	9	8	3	92.31
	0.5	0.3	132	10	7	7	89.10
	0.2	0.4	137	9	8	2	92.95
	0.3	0.4	137	11	6	2	91.67
	0.4	0.4	125	12	5	14	83.33
	0.5	0.4	138	6	11	1	95.51
30	0.2	0.1	134	9	8	5	91.03
	0.3	0.1	131	9	8	8	89.10
	0.4	0.1	134	9	8	5	91.03
	0.5	0.1	134	15	2	5	87.18
	0.2	0.2	131	9	8	8	89.10
	0.3	0.2	134	12	5	5	89.10
	0.4	0.2	134	10	7	5	90.38
	0.5	0.2	136	10	7	3	91.67
	0.2	0.3	132	9	8	7	89.74
	0.3	0.3	136	14	3	3	89.10
	0.4	0.3	137	9	8	2	92.95
	0.5	0.3	138	7	10	1	94.87
	0.2	0.4	131	4	13	8	92.31
	0.3	0.4	135	13	4	4	89.10
	0.4	0.4	122	12	5	17	81.41
	0.2	0.1	124	8	9	15	85.26

Table 1: GA parameters tuning results of TP, FP, TN and FN values

True positive (TP): AD class type correctly identified as AD. False positive (FP): NL class type incorrectly identified as AD class type. True negative (TN): NL class type correctly identified as NL. False negative (FN): AD incorrectly identified as NL class type. Based on the accuracy values in Table 1, the best and worst case TP, TN, FP and FN values are identified. The best and worst case values in table 1 are represented as bold black, italic formats. These best and worst cases statistical measures values are tabulated in the following Table 2

Statistical Measures	Number of GA indexed genes	Best Case	Worst Case
Sensitivity (%)	10	96.40	89.21
	20	99.28	89.93
	30	99.28	87.77
FPR (%)	10	52.94	58.82
	20	35.29	70.59
	30	41.18	70.59
Accuracy (%)	10	91.03	83.97
	20	95.51	83.33
	30	94.87	81.41
Specificity (%)	10	47.06	41.18
	20	64.71	29.41
	30	58.82	29.41
PPV (%)	10	93.71	92.54
	20	95.83	91.24
	30	95.17	91.04
NPV (%)	10	61.54	31.82
	20	91.67	26.32
	30	90.91	22.73

FDR (%)	10	6.29	7.46
	20	4.17	8.76
	30	4.83	8.96
MCC (%)	10	49.00	27.20
	20	74.83	18.43
	30	70.72	15.38

Table 2: Performance of proposed FFBN best, worst cases

In analyzing the statistical performance, the proposed method in the best case has given 93.8 % of mean accuracy, and in the worst case it has given 82.9 % of mean accuracy as result. When compared to the existing classifier, the proposed method has provided high accuracy and when compared to the old statistical features, the proposed method performance is slightly varied. The statistical measures of proposed method in the best case are compared with existing classifier. The accuracy of proposed and existing method comparison results are shown in the Figure 3.

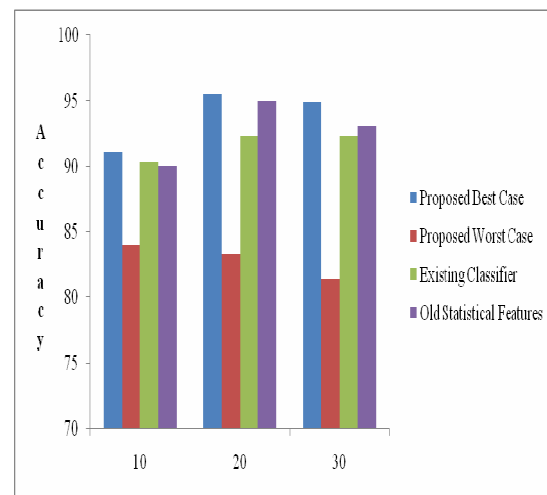


Fig 3: Comparison of proposed method with existing and old statistical features classifier

The comparison graph shows that the existing method has low classification performance than our proposed method in all 10, 20 and 30 GA indexed gene values.

5. Conclusion

In this paper, an efficient classification technique was introduced to classify the microarray genes into their specified cancer class type. The performance of the proposed classification technique was analyzed by performing statistical measures in terms of true positive and true negative values and compared with the existing classifier and old statistical features. For the performance analysis process the GA parameters were tuned. The parameter tuning process best and worst case results were analyzed to acquire the better result. The comparative results have shown that the proposed classification technique has performed better in terms of accuracy than the existing classifier and old statistical features. Hence, from the experimental results and analysis, it was clearly understood that the performance of the proposed classification technique was superior.

References

- [1] Mark A. Iwen, Willis Lang and Jignesh M. Patel, "Scalable Rule-Based Gene Expression Data Classification", In Proceedings of the IEEE 24th International Conference on Data Engineering, 2008
- [2] Nguyen and Rocke, Classification of Acute Leukemia based on DNA Micro array Gene Expressions using Partial Least Squares, Kluwer Academic, Dordrecht, 2001
- [3] Jian J. Dai, Linh Lieu, and David Rocke, "Dimension Reduction for Classification with Gene Expression Micro array Data", Statistical Applications in Genetics and Molecular Biology: Vol. 5, No. 1, 2006
- [4] Alok Sharma and Kuldip K. Paliwal, "Cancer classification by gradient LDA technique using micro array gene expression data", Data & Knowledge Engineering, Vol. 66, pp. 338-347, 2008
- [5] Yuh-Jye Lee and Chia-Huang Chao, "A Data Mining Application to Leukemia Micro array Gene Expression Data Analysis", International Conference on Informatics, Cybernetics and Systems (ICICS), Kaohsiung, Taiwan, 2003
- [6] Donald Geman, Bahman Afsari, Aik Choon Tan and Daniel Q. Naiman, "Micro array Classification from Several Two-Gene Expression Comparisons", In Proceedings of ICMLA, pp.583-585, 2008
- [7] Chun-Hou Zheng, Bo Li, Lei Zhang and Hong-Qiang Wang, "Locally Linear Discriminant Embedding for Tumor Classification", In Proceedings of ICIC, pp.1093-1100, 2008
- [8] Cheng-San Yang, Li-Yeh Chuang, Chao-Hsuan Ke and Cheng-Hong Yang, "A hybrid Feature Selection Method for Micro array Classification", International Journal of Computer Science, Vol. 35, No. 3, 2008
- [9] James J. Chen and Chun-Houh Chen, "Micro array Gene Expression", Encyclopedia of Biopharmaceutical Statistics, 2nd Edition, Marcel Dekker, Inc., pp. 599-613, 2003
- [10] Gad Getz, Erel Levine and Eytan Domany, "Coupled two-way clustering analysis of gene micro array data", In Proceedings of the National Academy of Sciences, Vol. 97, No. 22, pp. 12079-12084, 2000
- [11] Andrej Kastrin, "Item Response Theory Modeling for Micro array Gene Expression Data", Advances in methodology and statistics, Vol. 6, No. 1, pp. 51-67, 2009
- [12] Srinivas Mukkamala, Qingzhong Liu, Rajeev Veeraghattam and Andrew H. Sung, "Computational Intelligent Techniques for Tumor Classification (Using Micro array Gene Expression Data)", In Proceedings of World Congress on Lateral Computing, 2005
- [13] Fei Pana, Baoying Wanga, Xin Hub and William Perrizoa, "Comprehensive vertical sample-based KNN/LSVM classification for gene expression analysis", Journal of Biomedical Informatics, Vol. 37, pp. 240-248, 2004
- [14] Peng Qiu, Z. Jane Wang and Ray Liu, "Ensemble Dependence Model-based Cancer Classification using Gene Micro array Data", IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS), 2005
- [15] Seeja and Shweta, "Microarray Data Classification Using Support Vector Machine", International Journal of Biometrics and Bioinformatics (IJBB), Vol. 5, No. 1, pp. 10-15, 2011
- [16] Manoj Kumar, "The Widely Used Diagnostics "DNA Microarray"-A Review", American Journal of Infectious Diseases, Vol. 5, No. 3, pp. 207-218, 2009
- [17] Danh V. Nguyen, David M. Rocke, "Tumor Classification by Partial Least Squares Using Micro array Gene Expression Data", Bioinformatics, Vol. 18, No. 1, pp. 39-50, 2002

[18] Pengyi Yang and Zili Zhang, "An Embedded Two-Layer Feature Selection Approach for Microarray Data Analysis", IEEE Intelligent Informatics Bulletin, Vol.10, No.1, pp. 24-32, 2009

[19] Yee Hwa Yang and Natalie P. Thorne, "Normalization for Two-color cDNA Microarray Data", Science and Statistics: A Festschrift for Terry Speed, Vol. 40, pp. 403-418, 2003

[20] Ahmad M. Sarhan, "Cancer Classification Based on Micro array Gene Expression Data Using DCT and ANN", Journal of Theoretical and Applied Information Technology, Vol. 6, No. 2, pp. 208-216, 2009

[21] Bharathi and Natarajan, "Cancer Classification of Bioinformatics data using ANOVA", International Journal of Computer Theory and Engineering, Vol. 2, No. 3, pp. 369-373, June 2010

[22] Bo Li, Chun-Hou Zheng, De-Shuang Huang, Lei Zhang and Kyungsook Han, "'Gene expression data classification using locally linear discriminant embedding", Computers in Biology and Medicine, Vol. 40, pp. 802–810, 2010

[23] Xiaosheng Wang and Osamu Gotoh, "A Robust Gene Selection Method for Micro array-based Cancer Classification", Journal of Cancer Informatics, Vol. 9, pp. 15-30, 2010

[24] Mallika and Saravanan, "An SVM based Classification Method for Cancer Data using Minimum Micro array Gene Expressions", World Academy of Science, Engineering and Technology, Vol. 62, No. 99, pp. 543-547, 2010

[25] Chhanda Ray, "Cancer Identification and Gene Classification using DNA Micro array Gene Expression Patterns", International Journal of Computer Science Issues, Vol. 8, Issue 2, pp. 155-160, March 2011

[26] Seeja and Shweta, "Microarray Data Classification Using Support Vector Machine", International Journal of Biometrics and Bioinformatics (IJBB), Vol. 5, No. 1, pp. 10-15, 2011

[27]
http://en.wikipedia.org/wiki/Sensitivity_and_specificity