

Microarray Gene Expression for Cancer Classification using Fast Extreme Learning Machine with ANP

****C. Chandrasekar, *P.S. Meena,**

**Assistant Professor, Computer Applications, Sree Narayana Guru College, Coimbatore – 641 105.

*Research Scholar , Sree Narayana Guru College , Coimbatore – 641 105.

Abstract—Cancer is a class of diseases in which a set of cells display uncontrolled growth, invasion that interrupts upon and demolishes nearby tissues, and occasionally metastasis, or spreading to other locations in the body via lymph or blood. Cancer has become one of the dangerous diseases in the present scenario. DNA microarrays turn out to be an effective tool utilized in molecular biology and cancer diagnosis. Microarrays can be utilized to determine the comparative amount of particular mRNAs in two or more tissue samples for thousands of genes concurrently. As the supremacy of this technique has been identified, various open queries arise about suitable examination of microarray data. The multicategory cancer classification is playing a vital role in the field of medical sciences. As the numbers of cancer victims are increasing steadily, the necessity of the cancer classification techniques has become indispensable. For the above impenetrability and to obtain better consequences of the system with accuracy a new learning algorithm called Extreme Learning Machine (ELM) is used. ELM overcomes difficulties such as local minima, inappropriate learning rate and over fitting usually occurred by iterative learning techniques and performs the training rapidly. In this approach, the performance of the ELM is improved through the use of Analytic Network Process (ANP) and Levenberg Marquardt training algorithm. This approach also utilizes the error free ANOVA techniques in the preprocessing stage. This paper represents that ANOVA technique can be utilized to normalize microarray data and afford determination of alterations in gene expression that are corrected for potential perplexing effects. The proposed technique is evaluated with the help of Lymphoma data set. The experimental result represents that proposed technique results in better classification accuracies with lesser training time and implementation complexity compared to conventional techniques.

Keyword--- ELM, ANOVA, Cancer Classification and Gene Expression, Fast ELM, ANP.

I. INTRODUCTION

CANCER is a group of diseases distinguished by unregulated division and spread of cells [1]. The cancerous cells may happen in liquids, as in leukemia. Most, on the other hand, happen in solid tumors that initially emerge in different tissues in different parts of the body. By their original places they are classified into different types of cancer like lung, colon, breast, or prostate cancer. Localized tumors can be detached by surgery or irradiation with high survival rates. As cancer

progresses, conversely, it metastasizes invading the surrounding tissues, entering the blood stream, spreading and establishing colonies in isolated parts of the body. Only one-third of patients with metastasized cancer stay alive more than five years. Invasive expansions spreading crab-like from a tumor in the breast were illustrated by Hippocrates.

DNA microarray is a set of microscopic DNA spots connected to a solid surface. DNA microarrays are used to compute the expression levels of vast numbers of genes concurrently or to genotype multiple regions of a genome [22].

High-density DNA microarray [17] gathers the behaviors of various genes concurrently and the gene expression [18] profiles have been utilized for the cancer classification in recent times. This new technique assures to afford superior therapeutic capacity to cancer persons by means of diagnosing cancer kinds with enhanced accuracy. SVM, FNN [9, 16], etc., are the various classification techniques effectively utilized to the cancer diagnosis difficulty. On the other hand, its optimal extension to more than two classes was not evident that may enforce restrictions in its application to multiple tumor kinds. Then the Multicategory SVM that is a recently suggested extension of the binary SVM, and pertain it to multiclass cancer diagnosis problems [11][12]. The advance level of cancer prediction is ELM [10].

A blend of Integer Coded Genetic Algorithm (ICGA) and Particle Swarm Optimization (PSO), combined with the neural network based Extreme Learning Machine (ELM) is engaged for gene choosing and cancer classification. ICGA is utilized with PSO-ELM to choose an optimal set of genes that are afterwards utilized to construct a classifier to enlarge an algorithm (ICGA_PSO_ELM) that can deal with the sparse data and sample imbalance. An investigation into the functions of the selected genes, using a systems biology approach, revealed that many of the identified genes are involved in cell signaling and proliferation. An analysis of these gene sets shows a larger representation of genes that encode secreted proteins than found in randomly selected gene sets. Secreted proteins constitute a major means by which cells interact with their surroundings. Increasing biological proof has recognized the tumor microenvironment as a serious issue that identifies tumor survival and growth. Therefore, the genes detected by this examination that encode secreted proteins may afford significant insights to the nature of the critical biological characterizes in the microenvironment of every tumor kind that permits these cells to thrive and proliferate.

A chief shortcoming of all the past techniques is that they cannot offer a posterior probability of classification of the tissue to different classes. Such classification is predominantly

very vital in the context of asymmetric misclassification costs where the misclassification cost connected with some classes (false negative for cancer) may be considerably higher than that of others (false positive for cancer).

In recent days, Extreme Learning Machine (ELM) is utilized for predicting cancer cells in living organism by the method of ANOVA (Analysis Of Variance). In order to speed up the training process, Levenberg Marquardt training is used. Moreover, this approach also uses the Analytic Network Process (ANP).

This technique solves issues such as local minima, improper learning rate and over fitting usually occurs in iterative learning techniques and completes the training very fast. However, the usage of ELM will take more time when large data is used for classification. This is overcome by using the proposed ELM technique called Fast ELM. This proposed technique has the capability to perform the classification in lesser less than the conventional techniques.

II. RELATED WORK

Sridhar Ramaswamy *et al.*, [15] explains about multiclass cancer diagnosis with the help of tumor gene expression [8] signatures, that intentionally tells about, the complex grouping of clinical and histopathological data for optimal healing of patients with cancer based on establishing accurate diagnoses, it appears to be hard because of the atypical clinical presentation or histopathology. To conclude whether the diagnosis of multiple common adult malignancies could be attained entirely by molecular classification, for instance this paper uses 218 tumor samples, spanning 14 general tumor kinds, and 90 normal tissue models to oligonucleotide microarray gene expression examination. In recent times, [7] DNA microarray-based tumor gene expression profiles have been utilized for cancer diagnosis.

Runxuan Zhang *en al.*, in [6] proposed a fast and efficient classification technique called ELM algorithm. In ELM one may randomly desire and repair all the hidden node parameters and then systematically make a decision on the output weights. Examinations have indicated [2] that ELM has good simplification concert and can be executed effortlessly. Several nonlinear activation functions can be utilized in ELM, such as sigmoid, sine, hard limit [5], radial basis functions [3] [4], and complex activation functions [21].

Lipo wang *et al.*, [19] presents the accurate cancer classification with the help of expression of very few genes, the author targets at determining the smallest set of genes that can guarantee highly accurate classification of cancers from microarray data with the help of supervised machine learning techniques. The importance of determining the minimum gene subsets is in three phases as below:

- It significantly decreases the computational load and noise occurring from unrelated genes. In the illustrations examined in this paper, determining the minimum gene subsets still permits for extraction of simple diagnostic rules that directs to accurate diagnosis without the requirement for any classifiers.

- It makes things easier for gene expression examinations to contain only a very small number of genes slightly than thousands of genes that can reduce the cost for cancer testing appreciably.
- It terms for additional examinations into the probable biological relationship among these few numbers of genes and cancer expansion and treatment.

Ahmad M. Sarhan, [20] suggests the cancer classification based on microarray gene expression data using DCT and ANN. The author generally discusses about a stomach cancer identification system according to the Artificial Neural Network (ANN), and the Discrete Cosine Transform (DCT). The presented technique gathers classification characteristics from stomach microarrays with the help of DCT. The features obtained from the DCT coefficients are then used in ANN for classification. The microarray images utilized in this paper were gathered from the Stanford Medical Database (SMD).

III. METHODOLOGY

Microarray analysis is not straightforward because of the large number of genes, which are investigated simultaneously. By incorporating several factors of interest (for instance time and different treatments) in the experimental design, the interpretation of the data becomes even more difficult. The influence of the factors of interest should be separated from each other to draw sensible conclusions from the data analysis. To address these problems, a new methodology called ELM is proposed. Unlike traditional implementations and learning theory, from function approximation point of view, ELM theory shows that the hidden node parameters can be completely independent from the training data. ANOVA ranking technique used for selecting the most appropriate gene. In this approach, ANP technique is integrated with the ELM for better performance.

Extreme Learning Machine (ELM)

Extreme learning machine (ELM) [14] meant for Single Hidden Layer Feed-forward Neural Networks [13] (SLFNs) will randomly selected the input weights and analytically determines the output weights of SLFNs. This algorithm tends to afford the best generalization performance at extremely fast learning speed.

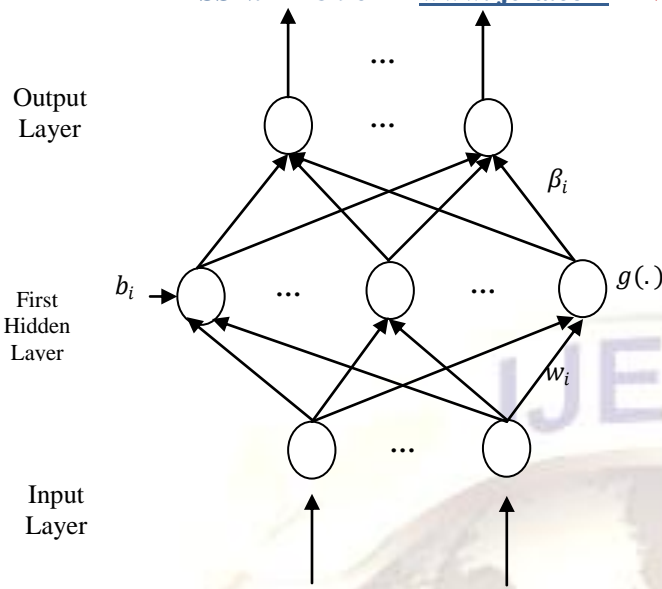


Figure 1: Structure of ELM network

The structure of ELM network is shown in figure 1. ELM contains an input layer, hidden layer and an output layer.

ELM has several interesting and significant features different from traditional popular learning algorithms for feed forward neural networks. These include the following:

The learning speed of ELM is extremely fast. The learning step of ELM can be completed in seconds or less than seconds for many applications. In the past, it seems that there exists a virtual speed barrier which most (if not all) classic learning algorithms cannot break through and it is not unusual to take very long time to train a feed-forward network using classic learning algorithms even for simple applications.

The ELM has better generalization performance than the gradient-based learning, such as, back propagation in most cases. The traditional classic gradient-based learning algorithms and some other learning algorithms may face several issues like local minima, improper learning rate and over fitting, etc. For avoiding these issues, some methods such as weight decay and early stopping methods may need to be used often in these classical learning algorithms.

The ELM likely to reach the solutions straightforward without such trivial issues. The ELM learning algorithm looks very simpler than most learning algorithms for feed-forward neural networks. Different from the traditional classic gradient-based learning algorithms which only work for differentiable activation functions, as easily observed the ELM learning algorithm could be used to train SLFNs with many non-differentiable activation functions.

Extreme Learning Machine Training Algorithm

If there are N samples (x_i, t_i) , where $x_i = [x_{i1}, x_{i2} \dots x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, then the standard SLFN with N hidden neurons and activation function $g(x)$ is defined as:

$$\sum_{i=1}^N \beta_i g(w_i \cdot x_j + b_i) = 0_j, j = 1, \dots, m, N.,$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is nothing but the weight vector that connects the i th hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector that connects the i th neuron and the output neurons, and b_i is the threshold of the i th hidden neuron. The “.” in $w_i \cdot x_j$ means the inner product of w_i and x_j . The SLFN aims to minimize the difference between o_j and t_j . This can be expressed mathematically as:

$$\sum_{i=1}^N \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, m, N.,$$

or, more in a matrix format as $H\beta = T$, where

$$H(a_1, \dots, a_N, b_1, \dots, b_N, x_1, \dots, x_N) = \begin{bmatrix} g(w_1, x_1 + b_1) & \dots & g(w_N, x_N + b_N) \\ \vdots & \ddots & \vdots \\ g(w_1, x_1 + b_1) & \dots & g(w_N, x_N + b_N) \end{bmatrix}_{N \times N}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{N \times m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

The matrix H is the hidden layer output matrix of the neural network. If the number of neurons in the hidden layer is equal to the number of samples, then H is square and invertible. Otherwise, the system of equations needs to be solved by numerical methods, concretely by solving

$$\min_{\beta} \|H\beta - T\|$$

The result that minimizes the norm of this least squares equation is

$$\hat{\beta} = H^+T$$

where H^+ is called Moore-Penrose generalized inverse. The most important properties of this solution are:

- Minimum training error.
- Smallest norm of weights and best generalization performance.
- The minimum norm least-square solution of $H\beta = T$ is unique, and is

$$\hat{\beta} = H^+T$$

Give a training set $N = \{(x_i, t_i) | x_i \in R^n, t_i \in R^m, 1 = 1 \dots \dots N\}$ activation function $g(x)$ and hidden neuron \tilde{N} , do the following

- Assigning random value to the input weight w_i and the bias $b_i, i = 1, \dots, \tilde{N}$
- Find the hidden layer output matrix H.

The ELM algorithm works as follows

After the learning process is completed by providing several conditions, the proposed technique can be able to detect the cancer occurrence in the microarray gene.

A. Levenberg Marquardt Extreme Learning Machine with ANP Technique

Initially, the input weights and hidden biases are created by with the help of ANP technique. Next, the equivalent output

weights are analytically determined with the help of ELM algorithm only in first step and randomly produce the output hidden biases. Then, the parameters (all weights and biases) are restructured with the help of LM algorithm.

The process for the Fast Extreme Learning Machine with ANP is described below:

Provided a training set $\mathfrak{X} = \{x_i, t_i\} | x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, N\}$ activation functions $f_1(x)$ & $f_2(x)$, and hidden nodes number \tilde{N} & K of hidden first and second layer.

Step 1: Randomly choose the starting values of input weight vectors w_1 and bias vector b_1 with the help of ANP technique and bias vector b_2 without using the AHP technique.

Step 2: Determine the hidden first layer output matrix a_1 . With the help of ELM algorithm, determine the output weight

$$w_2 = a_1^{-1} \cdot t \quad 5.15$$

Step 3: Determine the hidden second layer output matrix a_2 , errors

$$e_1 = t - a_2 \quad 5.16$$

and determine the sum of squared errors over all input.

Step 4: Determine the Jacobian matrix. Calculate the sensitivities with the recurrence relations.

$$S_q^m = f^m(n_q^m)(w^{m+1})^T \cdot S_q^{m+1} \quad 5.17$$

after initializing with the following equation

$$S_q^M = -f^m(n_q^m) \quad 5.18$$

Augment the individual matrices into the Marquardt sensitivities using the following equation

$$S^m = [S_1^m, S_2^m, \dots, S_Q^m] \quad 5.19$$

Determine the elements of the Jacobian matrix with the equations

$$[J]_{h,l} = S_{i,h}^m \times S_{j,k}^{m-1} \quad 5.20$$

And

$$[J]_{h,l} = S_{i,h}^m \quad 5.21$$

Step 5: Solve equation the below equation to determine Δw_k and update weight vectors w_1, w_2 and bias vectors b_1, b_2 .

$$\Delta w_k = [J^T(w_k) \cdot J(w_k) + \mu \cdot I]^T \cdot J^T(w_k) \cdot e(w_k) \quad 5.22$$

Step 6: Recalculate the sum of squared errors with the help of $w_k + \Delta w_k$. If this new sum of squared is lesser than that evaluated in step3, then multiply μ by μ_{dec} , let $w_{k+1} = w_k + \Delta w_k$ and process from step4. If the sum of squared is not decreased, then multiply μ by μ_{inc} and process from step5.

For the training process, this approach uses Levenberg-Marquardt (LM) Algorithm which speeds up the training process.

Levenberg-Marquardt (LM) Algorithm

Step 1: Initialize the weights and parameter μ ($\mu=.01$ is appropriate).

Step 2: Calculate the Sum of the Squared Errors over all inputs $F(w)$.

Step 3: Solve step (2) to get the increment of weights Δw .

Step 4: Recompute the Sum of Squared Errors $F(w)$.

Using $w + \Delta w$ as the trial w and judge

IF trial $F(w) < F(w)$ in step 2 THEN

$$w + \Delta w$$

$$\mu = \mu \cdot \beta (\beta = .1)$$

Go back to step 2

ELSE

$$\mu = \mu / \beta$$

go back to step 4

END IF

Thus, by using the ANP technique, the Extreme Learning Machine is modified as Fast Extreme Learning Machine which has the advantage of training the classifier in very less time.

ANOVA Model

Analysis of Variance (ANOVA) can be used in microarray data analysis to investigate the significance of the effects from factors which could possibly influence the gene expression. The ANOVA fixed effects model in which three of the possible factors of interest are incorporated is given by expression below. In this model the measured gene expression (X_{ijk}) is assumed to be the result of the added effects of the factors Time (T),

Treatment (S) and Gene (G) over timepoint i, treatment j, gene k and replicate r:

$$X_{ijk} = \mu + T_i + S_j + G_k + TS_{ij} + TG_{ik} + GS_{jk} + TSG_{ijk} + \varepsilon_{ijk} \quad 5.23$$

The effect of interactions between factors is also incorporated in the three factor ANOVA model shown here (TS, TG, GS and three way interaction TSG). The effects are added to a general mean expression value which is indicated with μ . Finally the remaining variation is captured in the error term. In the normal application of ANOVA, the sum of squares and mean squares are calculated for each factor and interaction. With an F-test the significance of the effect of each factor is then calculated. The main effects are vectors with a length equal to the number of levels in each factor. The interaction matrices consist of the combined effect of two factors, when corrected for the general effect of these factors. For example, the interaction matrix of the factors Gene and Treatment shows the effect of a gene and treatment after the general treatment effect and general gene effect have been corrected for. Thus, the interaction effect could be interpreted as the response of a gene to the treatments additional to general gene and treatment effects. This is of course interesting to the biologists who are looking for genes which respond to the treatments incorporated in the experiment.

First, the main effect is calculated for each of the factors of interest. The three factors used in the ANOVA model are time effect, treatment effect and the effect of each individual gene. The main effect is then calculated for each factor by subtracting the general overall mean from the mean per Time point, the mean per Gene, and the mean per Treatment

$$main\ effect = \bar{y}_r - \bar{y} \dots$$

where y is gene expression, and r is either $i \dots, j \dots$ or $\dots k \dots$, depending on the effect which is calculated. The next step is the calculation of the interaction matrices. The interaction matrices of the interaction between Gene and Treatment (GS),

Time and Treatment (TS) and Time and Gene (TG) are calculated with:

$$GS = \bar{y}_{jk.} - \bar{y}_{j..} - \bar{y}_{.k.} - \bar{y}_{...}$$

$$TS = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...}$$

$$TG = \bar{y}_{i.k.} - \bar{y}_{i..} - \bar{y}_{.k.} - \bar{y}_{...}$$

The three factor interaction (GTS) of Time and Gene can be calculated as follows:

$$GTS = \bar{y}_{ijk.} - \bar{y}_{jk.} - \bar{y}_{ij.} - \bar{y}_{i.k.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{.k.} - \bar{y}_{...}$$

The three factor interaction is in fact a data cube.

IV. EXPERIMENTAL RESULTS

This chapter experiments the proposed methodology using the Lymphoma data set. The Lymphoma data set is a data set about the three most prevalent adult lymphoid malignancies. It contains 62 samples consisting of 4,026 genes spanning three classes, which include 42 Diffuse Large B-Cell Lymphoma (DLBCL) samples, nine Follicular Lymphoma (FL) samples, and 11 B-CELL CHRONIC LYMPHOCYTIC LEUKEMIA (B-CELL) SAMPLES.

The data set can be found at <http://genome-www.stanford.edu/lymphoma/>. The 62 samples are randomly split into 50 training samples and 12 testing samples at each trial and the average performance has been obtained over 100 trials for both ELM and SVM-OVO.

The BSS/WSS method is also used for gene selection. Ten different numbers of genes, from 10 to 100 in intervals of 10, are selected and used in the simulation of both the ELM and SVM-OVO algorithms which uses ANOVA test. The average testing accuracies over 100 trials are shown in Table 5. From the table it can be clearly observed that the accuracy for the usage of ELM is better when compared to the conventional methods (Eg. SVM).

Table 1 shows the comparison of the testing accuracy of the classification techniques such as SVM, ELM and LMELM with ANP.

Table 1: Testing Accuracy (%) for the ELM and SVM Algorithms on the Lymphoma Data Set

#Gene combination	SVM	ELM	LMELM with ANP
	Accuracy	Accuracy	Accuracy
10-20	97.63	98.56	99.45
21-30	97.45	98.64	99.34
31-40	96.12	98.14	100
41-50	95.21	98.26	99.68
51-60	98.32	97.65	99.42
61-70	97.45	98.12	98.65

71-80	96.41	97.68	99.32
81-90	96.65	98.60	99.45
91-100	96.44	97.44	99.67

Figure 2 shows the graph for accuracy comparison between the SVM, ELM and the proposed LMELM with AHP technique. From the graph, it is clear that the proposed method shows better accuracy for all the gene combinations. The overall accuracy is higher for proposed LMELM with AHP when compared to the SVM and ELM techniques.

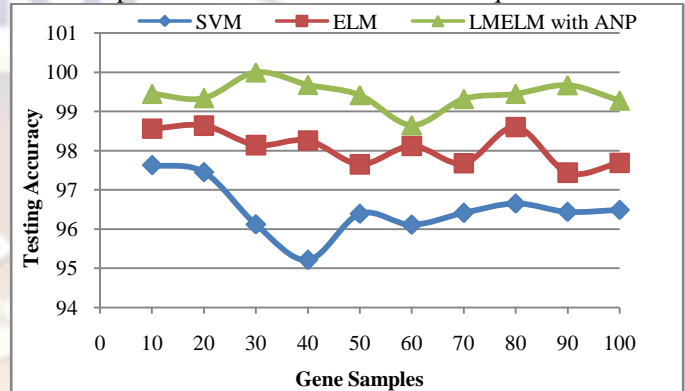


Figure 2: Accuracy Comparison

From the table it can be clearly observed that the training time for the usage of ELM is lesser when compared to the training time taken by SVM. As the table indicates, the training time taken by SVM and ELM approaches for the gene combination of 10-20 is 390.13 and 165.21 seconds respectively, where as the proposed method yields the training time of 101 seconds which is very much lesser than SVM and ELM. For the gene combination of 21-30, the proposed ELM takes 111 seconds for training, whereas, SVM and ELM takes higher training time.

The total training time taken by the classification approaches like SVM, ELM and the proposed LMELM with AHP is used in this approach. From the table it can be clearly observed that the training time for the usage of proposed LMELM with AHP is lesser when compared to the training time taken by SVM and ELM with more compact network.

Table 2: Training Time(s) and Averaged Number of Hidden Nodes for the ELM and SVM Algorithms for 100 Splits of Training and Test Set on the Lymphoma Data Set

#Gene combination	Training Time (secs)		
	SVM	ELM	LMELM with AHP
10-20	390.13	165.21	101
21-30	420.22	185.39	111
31-40	460.89	215.23	123
41-50	494.36	232.18	105
51-60	507.55	270.45	108
61-70	522.63	315.23	123
71-80	525.27	341.25	96
81-90	545.01	379.25	98
91-100	553.65	421.45	109

For the selected data set, the proposed ELM approach with ANOVA technique takes much less total training time than the SVM algorithm with ANOVA. As it is mentioned before, the SVM algorithm has to build $c(c-1)/2$ binary classifiers to distinguish between every two class combinations. For the Lymphoma data set, with the number of categories classified decreases, the difference between ELM and SVM is also decreased.

It can be seen that the number of hidden nodes for ELM is always smaller than the number of support vectors for SVM, indicating a more compact network realized by ELM.


V. CONCLUSION

Multicategory cancer diagnosis problem based on microarray data has become an active area of research. In this paper, ELM algorithm is used for classification in which the ELM is trained using Levenberg Marquardt algorithm for training. Moreover, this approach uses Analytic Network Process (ANP) for choosing the weights. ANOVA statistical ranking approach is used in this approach. Its performance has been compared with the SVM and ELM algorithms. SVM for multicategory classifications is done by modifying the binary classification method of SVM to a one-versus-all or one-versus-one comparison basis. This inevitably involves more classifiers, greater system complexities and computational burden, and a longer training time. ELM with ANOVA can perform the multicategory classification directly, without any modification. In order to improve the performance of the ELM, ANP approach is used for selecting the weights. The performance of the proposed LMELM with ANP algorithm achieves higher classification accuracy than the SVM

algorithm for the ANOVA test with less training time and a smaller network structure.

REFERENCES

- [1] Kinzler, Kenneth W.; Vogelstein, Bert, "Introduction", the genetic basis of human cancer (2nd, illustrated, revised ed.). New York: McGraw-Hill, Medical Pub. Division, 2002.
- [2] G.-B. Huang and C.-K. Siew, "Extreme Learning Machine: RBF Network Case," Proc. Eighth Int'l Conf. Control, Automation, Robotics, and Vision (ICARCV '04), Dec. 2004
- [3] D. Serre, Matrices: Theory and Applications. Springer-Verlag, 2002.
- [4] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes," IEEE Trans. Neural Networks, vol. 17, no. 4, pp. 879-892, 2006.
- [5] S. Dudoit, J. Fridlyand, and T.P. Speed, "Comparison of Discrimination Methods for Classification of Tumors Using Gene Expression Data," J. Am. Statistical Assoc., vol. 97, no. 457, pp. 77-87, 2002.
- [6] Runxuan Zhang, Guang-Bin Huang, Narasimhan Sundararajan, and P. Saratchandran, " Multicategory Classification Using an Extreme Learning Machine for Microarray Gene Expression Cancer Diagnosis," vol 4, no 3, July-september 2007.
- [7] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray," Science, vol. 270, pp. 467-470, 1995.
- [8] S. Dudoit, J. Fridlyand, and T.P. Speed, "Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data," J. Am. Statistical Assoc., vol. 97, pp. 77-87, 2002.
- [9] R. Linder, D. Dew, H. Sudhoff, D. Theegarten, K. Remberger, S.J. Poppl, and M. Wagner, "The 'Subsequent Artificial NeuralNetwork' (SANN) Approach Might Bring More Classificatory Power to ANN-Based DNA Microarray Analyses," Bioinformatics, vol. 20, no. 18, pp. 3544-3552, 2004
- [10] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks," Proc. Int'l Joint Conf. Neural Networks (IJCNN '04), July 2004.
- [11] G.-B. Huang and C.-K. Siew, "Extreme Learning Machine: RBF Network Case," Proc. Eighth Int'l Conf. Control, Automation, Robotics, and Vision (ICARCV '04), Dec. 2004.
- [12] G.-B. Huang and C.-K. Siew, "Extreme Learning Machine with Randomly Assigned RBF Kernels," Int'l J. Information Technology, vol. 11, no. 1, 2005.
- [13] G.-B. Huang, Q.-Y. Zhu, K.Z. Mao, C.-K. Siew, P. Saratchandran, and N. Sundararajan, "Can Threshold Networks Be Trained Directly?" IEEE Trans. Circuits and Systems II, vol. 53, no. 3, pp. 187-191, 2006.

- 
- The logo of the International Journal of Engineering Research and Applications (IJERA) is a large, semi-transparent watermark in the background. It features a globe with a laurel wreath encircling it. The letters 'IJERA' are prominently displayed in a stylized font across the globe.
- [14] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully Complex Extreme Learning Machine," *Neurocomputing*, vol. 68, pp. 306-314, 2005.
- [15] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub, "Multiclass Cancer Diagnosis Using Tumor Gene Expression Signatures," *Proc. Nat'l Academy Sciences, USA*, vol. 98, no. 26, pp. 15149-15154, 2002.
- [16] R. Linder, D. Dew, H. Sudhoff, D. Theegarten, K. Remberger, S.J. Poppl, and M. Wagner, "The 'Subsequent Artificial Neural Network' (SANN) Approach Might Bring More Classificatory Power to ANN-Based DNA Microarray Analyses," *Bioinformatics*, vol. 20, no. 18, pp. 3544-3552, 2004.
- [17] O. Troyanskaya et al., "Missing Value Estimation Methods for DNA Microarrays," *Bioinformatics*, vol. 17, pp. 520-525, 2001.
- [18] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J.A. Olson Jr., J.R. Marks, and J.R. Nevins, "Predicting the Clinical Status of Human Breast Cancer by Using Gene Expression Profiles," *Proc. Nat'l Academy of Sciences USA*, vol. 98, pp. 11 462-11 467, 2001.
- [19] Lipo Wan, Feng Chu and Wei Xie, "Accurate Cancer Classification Using Expressions of Very Few Genes", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 1, Pp. 40-53, 2007.
- [20] Ahmad M. Sarhan, "Cancer Classification based on Microarray Gene Expression Data using DCT and ANN", *Journal of Theoretical and Applied Information Technology*, 2009.
- [21] G. W. Wayt. The unseen genome: beyond DNA. *Scientific American*, 289(6): 106-114 (2003).
- [22] M. Ringner, C. Peterson, and J. Khan, "Analyzing Array Data Using Supervised Methods," *Pharmacogenomics*, vol. 3, no. 3, pp. 403-415, 2002.