Neural Network Based Classification of Glioma Grade III and Grade IV.

Supriya Patil
Research Scholar,
Electronics Dept.,
Goa University
Goa, India
supriya_adi@rediff.com

G. M. Naik
Professor and HOD,
Electronics Dept.,
Goa University,
Goa, India
gmnaik@unigoa.ac.in

K. R. Pai
Professor and HOD,
ETC Dept.,
P.C. Engg College,
Goa, India
hod_etc@yahoo.co.in

R. S. Gad
Associate Professor
Electronics Dept.,
Goa University,
Goa, India
rsgad@unigoa.ac.in

***Abstract*—Microarray technology is a widely accepted for cancer subtype detection. Microarray gene expression data being a very high dimensional data, the greater challenge for microarray analysis is to identify the optimal set of genes for the purpose of classification. In this paper, we suggest a combination of ratio of mean values of a particular gene of both the classes, t-statistics and standard deviation to obtain the optimal subset of genes. The size of this data is further cut down with the help of wavelet transform. Finally, the classification is performed using neural network algorithms.**

***Index Terms*—Discrete Wavelet Transform, Principal component analysis, Resilient Back Propagation Algorithm, Support Vector Machine, Error Back Propagation Algorithm.**

I .INTRODUCTION

Various reasons like genetic factors, use of tobacco, certain types of infections lead to mutations in the genes and causes the uncontrolled growth of abnormal cells in a human body. This in turn results into benign or malignant cancer. To lessen the death rate of cancer patients it is important to recognize the cancer subtype accurately, which maximizes the efficiency of the cancer therapy. [1]. The microarray cancer gene expression data set comprises a matrix of data, in which one cancer sample is identified by a column and the expression value of a specific gene of different samples is specified by a row.

To perform the microarray based cancer sub-classification, the optimum subset of genes is derived from original high dimensional microarray data using various methods [2] like filter [3], [4], [5], wrapper [6], [7], [8] embedded [9], [10], [11] hybrid [12], [13], [14] and Ensemble methods [15], [16], [17]. Alternatively, the dimension of microarray data can be reduced by converting the microarray signal into different domain using discrete cosine transform, Principal component analysis [18], discrete wavelet transform (DWT) [19] etc.

For a particular type of cancer using the expression values of marker genes of respective cancer can be used for the classification. It always does not lead to precise recognition of cancer subtype [20].

The subset of genes obtained by the above mentioned methods can be used for cancer classification. The different classification algorithms used are, K-means clustering, support vector machines (SVM) [21], [22], error back propagation algorithm (EBPA) [23], Naïve Bayes [24] etc.

Heba [25] has utilized 8 gene selection techniques like, t-statistics, information gain, etc. The classifiers used are Random forest algorithm, SVM and K-mean clustering. For GDS1975 data sets, the maximum classification accuracy attained is 94.59%.

Every method suggested [3-25] has its advantages and disadvantages. However, there is possibility for enhancing accuracy of cancer classification using microarray analysis.

In the proposed work , we have introduced a system for the classification of glioma grade III and grade IV datasets-GDS1975, downloaded from Gene Expression Omnibus Database [26], [27]. We propose to obtain the subset of informative genes using ratio of average values of both classes, t-statistics and standard deviation. The dimensions of microarray data are further reduced by converting signal into frequency domain with the help of discrete wavelet transform (DWT). It is proposed to perform the classification using Resilient back propagation algorithm (RPROP) and conjugate gradient algorithms.

In the following, Section II describes the method for extracting the required gene subset, section III shows the details of the DWT, section IV describes various classification algorithms and implementation is described in section V, followed by result analysis in section VI.

II. FEATURE SELECTION

Number of methods can be used to select the subset of genes for classification. In one of the method, the ratio of average intensity for a particular gene of one class with that of the other class can be used to select the genes. For the genes up regulated by some value the ratio will be above one and for genes down regulated by same value as that of up regulated genes the ratio will be between 0 and 1. Application of log transform (base 2) to the ratio, yields the ratio of same magnitude but opposite sign. The genes whose magnitude of logarithmic transformed ratio more than one can be considered for classification.

In the second method, according to t-statistics the t value can be calculated for all genes by using the formula given below [28]

$$t = \frac{avg1 - avg2}{\sqrt{\frac{var1}{n} + \frac{var2}{l}}} \quad \text{Eq. 1}$$

where,
for a particular gene,
$avg1$= average intensity for class 1
$avg2$ = average intensity for class 2
$var1$= variance of class 1
$var2$= variance of class 2
n = number of samples in class 1
l = number of samples in class 2.

For a given gene, higher the difference between average values and smaller the variance of both classes, larger will be the t value. Following this, the ambiguity in the classification is resolved by calculating probability that the gene is not affected by cancer using statistical method [28]. The negative of logarithmic transform (base 10) of probability (p) can be calculated as given by the formula,

$$q = -\log(p) \quad \text{Eq.2}$$

The genes having the q value above some particular value can be considered for further processing. Also the subset of genes can be refined by considering genes with less standard deviation for both the classes.

III. DISCRETE WAVELET TRANSFORM

In the case of signal analysis DWT plays, an important role. It offers number of advantages like multi resolution analysis, sub band coding etc. The DWT decomposes the signal into low and high frequency coefficients by successively passing it through high pass and low pass filter. The low frequency (approximation) coefficients $cAi(p)$ and the high frequency (detailed) coefficients c$Di(p)$ at each level are given as below:

$$cD_i(p) = \sum_{m=2k}^{2k+N-1} D(m-2p)\, cD_{i+1}(m) \quad \text{Eq.3}$$

$$cA_i(p) = \sum_{m=2k}^{2k+N-1} A(m-2p)\, cA_{i+1}(m) \quad \text{Eq.4}$$

where,
$D(n)$ = Impulse response of high pass filter
$A(n)$ = Impulse response of low pass filter.
p = translation parameter.
i = level of decomposition
N= number of wavelet coefficients.

The approximation coefficients or detailed coefficients can be used for the purpose of classification [29].

.

IV. CLASSIFICATION ALGORITHMS

Artificial neural network has an ability to handle highly nonlinear and chaotic data. It can be efficiently used for classification of such data.

A) Error Back Propagation Algorithm (EBPA):

For multilayer neural network, one of the most commonly used classification algorithms is EBPA. In the case of EBPA, the individual weight change is made proportional to the gradient of error curve. The equation for the individual weight update for EBPA is given as,

$$w(t+1) = w(t) + (c(dk - ok)\, ok') \quad \text{Eq.5}$$

Where,
$w(t+1)$ = new weight
$w(t)$ = weight at the previous instance
c = learning constant
dk = expected output of neuron
ok = actual output of neuron
ok'= derivative of actual output of neuron.

For large values of inputs, the output of neuron increases and derivative of the output decreases. Hence the weight change decreases for a large difference between the input and the output. It disturbs the accuracy of classification. RPROP algorithm efficiently deals with this problem.

B) Resilient Back Propagation Algorithm:

In the weight update process, RPROP algorithm considers the sign of the derivative of error. The amount of weight update is raised (lowered), if the sign of the derivative of the error in two consecutive epochs remains same (dissimilar). The weight update is kept same, in case the derivative of the error diminishes to zero. [30].

C) Conjugate Gradient Back Propagation Algorithms:

In order to reach the global minimum of error curve, EBPA implements a linear search. The next search direction is orthogonal to the former search direction. In the case of Conjugate gradient algorithms the new search direction is A-orthogonal to previous search direction [31]. It results in faster speed of conjugate gradient algorithms. The next search direction is calculated as

$$\text{next search direction} = (r \times \text{the former direction}) + \text{the direction of steepest descent} \quad \text{Eq.6}$$

Conjugate gradient back propagation with Polak-Ribiére update algorithm calculates ‘r’ as [32], [33].

$$r = \frac{\text{PGE} - \text{CGE}}{\text{PGE}} \quad \text{Eq. 7}$$

where,
PGE = former gradient energy
CGE= current gradient energy

Conjugate gradient back propagation with Fletcher-Reeves update algorithm calculates ‘r’ as [33], [34].

Eq.8

$$r = \frac{CGE}{PGE}$$

When the number of epoch becomes the same as the number of neural network parameters, conjugate gradient algorithms completes the classification. The search direction is reset, in case, the algorithms fails to converge with above mentioned condition. The conjugate gradient back propagation with Powell-Beale restarts algorithm resets the search direction, when there is hardly any orthogonality left between the consecutive gradients [33], [35],[36].

V. IMPLEMENTATION

In the proposed classification, there are 26 samples of glioma grade III and 59 samples of glioma grade IV.

The ratio of mean intensity values of grade III with respect to grade IV is calculated. The logarithm (base2) of the ratio is calculated. Then t value is calculated for each gene using t statistics formula as explained in section II. Then using statistical technique, the probability that the gene is not affected is calculated. For every gene the q value is calculated as explained in section II. The genes having log (base2) transformed ratio of average of both classes more than one, q value more than one and standard deviation less than 500 for both the classes are considered for further processing. The data is normalized and wavelet transform is applied to the data using wavelet transforms Bior2.4, Bior1.3, Sym4, Sym2, Db4, and Db2. The wavelet transformed data is used for classification. Finally the classification is implemented using classification algorithms like RPROP and Conjugate gradient algorithms. The performance of different wavelets and neural network algorithms is compared.

VI. RESULT ANALYSIS

The results for GDS1975 microarray gene expression dataset, without using DWT, using approximation (Approx.) and detailed (Det.) wavelet coefficients at level 1 as input for different neural network algorithms are as shown in Table I.

TABLE I.GDS1975 DATASET

Sr. No	Algorithm	With/without wavelet	Accuracy
1.	Resilient Back Propagation Algorithm	Without wavelet	98.8
		Bior2.4, Bior1.3(Approx.)	100
		Bior1.3(Det.)	100
2.	Conjugate gradient back propagation with Polak-Ribiére updates	Without wavelet	98.8
		Db4, Sym4(Approx.)	98.8
		Db2, Sym2(Det.)	96.5
3.	Conjugate gradient back propagation with Powell-Beale restarts	Without wavelet	98.8
		Db4, Sym4(Approx.)	97.6
		Db2, Sym4(Det.)	96.5
4.	Conjugate gradient back propagation with Fletcher-Reeves updates	Without wavelet	98.8
		Bior2.4, Db4(Approx.)	98.8
		Db2, Sym2(Det.)	96.5

CONCLUSION

The combination of ratio of means, t-statistics and standard deviation along with wavelet transform shows considerable decrease in the size of GDS 1975 dataset. RPROP algorithm with Bior1.3 and Bior2.4 wavelets gives 100% classification accuracy. The 100% classification accuracy is achieved using 25 genes as opposed to classification accuracies obtained using the methods suggested by Heba [25] .The blend of type of the wavelet and the neural network algorithm that gives the 100% classification accuracy largely depends on the nature of the data for the specific data set.

REFERENCES

[1] Zarita zainuddin,ong Pauline,"Improved Wavelet Neural Network for Early Diagonosis of Cancer Patients Using Microarray Gene Expression Data".Proceedings of International joint Conference on Neural Networks,Atlanta,Georgia ,USA,June 14-19,2009

[2] Golub, Todd R., Donna K. Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P. Mesirov, Hilary Coller et al. "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring."science 286, no. 5439 (1999): 531-537.

[3] L. Lan and S. Vucetic, "Improving accuracy of microarray classification by a simple multi-task feature selection filter," Int. J. Data Min. Bioinforma., vol. 5, no. 2, pp. 189-208, Jan. 2011.

[4] Y. Zheng and C. K. Kwoh, "A Feature Subset Selection Method Based On High-Dimensional Mutual Information," Entropy, vol. 13, no. 4, pp. 860–901, Apr. 2011.

[5] R. Xu, S. Damelin, B. Nadler, and D. C. Wunsch II, "Clustering of high-dimensional gene expression data with feature filtering methods and diffusion maps," Artif. Intell. Med., vol. 48, no. 2–3, pp. 91–98, Feb. 2010.

[6] Y.Ai-Jun and S. Xin-Yuan, "Bayesian variable selection for disease classification using gene expression data," Bioinformatics, vol. 26, no. 2, pp. 215–222, Jan. 2010.

[7] M. Perez, D. M. Rubin, T. Marwala, L. E. Scott, and W. Stevens, "A Population-Based Incremental Learning approach to microarray gene expression feature selection," in 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI), 2010, pp. 10–14.

[8] L. Yu, Y. Han, and M. E. Berens, "Stable Gene Selection from Microarray Data via Sample Weighting," IEEEACM Trans Comput Biol Bioinforma., vol. 9, no. 1, pp. 262–272, Jan. 2012.

[9] S. Maldonado, R. Weber, and J. Basak, "Simultaneous feature selection and classification using kernel-penalized support vector machines," Inf. Sci., vol. 181, no. 1, pp. 115–128, Jan. 2011.

[10] H. Pang, S. L. George, K. Hui, and T. Tong, "Gene Selection Using Iterative Feature Elimination Random Forests for Survival Outcomes," IEEEACM Trans Comput Biol Bioinforma., vol. 9, no. 5, pp. 1422–1431, Sep. 2012.

[11] A. Anaissi, P. J. Kennedy, M. Goyal, and D. R. Catchpoole, "A balanced iterative random forest for gene selection from microarray data," BMC Bioinformatics, vol. 14, no. 1, p. 261, Aug. 2013.

[12] Q. Hu, W. Pan, S. An, P. Ma, and J. Wei, "An efficient gene selection technique for cancer recognition based on

neighborhood mutual information," Int. J. Mach. Learn. Cybern., vol. 1, no. 1–4, pp. 63–74, Dec. 2010.

[13] P. Saengsiri, S. N. Wichian, P. Meesad, and U. Herwig, "Comparison of hybrid feature selection models on gene expression data," in Knowledge Engineering, 2010 8th International Conference on ICT and, 2010, pp. 13–18.

[14] Y. B. Kim and J. Gao, "Unsupervised Gene Selection For High Dimensional Data," in Sixth IEEE Symposium on BioInformatics and BioEngineering, 2006. BIBE 2006, 2006, pp. 227–234.

[15] M. A. Gaafar, N. A. Yousri, and M. A. Ismail, "A novel ensemble selection method for cancer diagnosis using microarray datasets," in IEEE 12th International Conference on BioInformatics and BioEngineering, BIBE 2012, 2012, pp. 368–373.

[16] S. Ghorai, A. Mukherjee, S. Sengupta, and P. K. Dutta, "Cancer Classification from Gene Expression Data by NPPC Ensemble," IEEEACM Trans Comput Biol Bioinforma., vol. 8, no. 3, pp. 659– 671, May 2011.

[17] T. Abeel, T. Helleputte, Y. V. de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," Bioinformatics, vol. 26, no. 3, pp. 392–398, Feb. 2010.

[18] Mahapatra, Rajat, BanshidharMajhi, and Minakhi Rout. "Development and performance evaluation of improved classifiers of microarray data." In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on, pp. 519-523. IEEE, 2012.

[19] Li, Shutao, Chen Liao, and James T. Kwok. "Wavelet-based feature extraction for microarray data classification." In Neural Networks, 2006. IJCNN'06. International Joint Conference on, pp. 5028-5033. IEEE, 2006.

[20] www.cancer.gov/about-cancer/diagnosis-staging/ diagnosis/ tumor-markers-fact-sheet

[21] Q. Liu, Z. Zhao, Y. Li, X. Yu, and Y. Wang, "A Novel Method of Feature Selection based on SVM," J. Comput., vol. 8, no. 8, Aug. 2013.

[22] X. Cai, F. Nie, H. Huang, and C. Ding, "Multi-Class L2, 1-Norm Support Vector Machine," in 2011 IEEE 11th International Conference on Data Mining (ICDM), 2011, pp. 91–100.

[23] S.W. Chang, S. Abdul-Kareem, A. F. Merican, and R. B. Zain, "Oral cancer prognosis based on clinicopathologic and ge-nomic markers using a hybrid of feature selection and machine learning methods," BMC Bioinformatics, vol. 14, no. 1, p. 170, May 2013

[24] X. Wang and O. Gotoh, "A Robust Gene Selection Method for Microarray-based Cancer Classification," Cancer Inform., vol. 9, pp. 15-30, Feb. 2010.

[25] Abusamra, Heba. "A comparative study of feature selection and classification methods for gene expression data of glioma." Procedia Computer Science 23 (2013): 5-14.

[26] http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS1976

[27] http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS1975

[28] www.doc.ic.ac.uk/~yg/course/bioinformatics/5_MicroArray-Ttest.ppt

[29] LatSoman, K. P. Insight into wavelets: from theory to practice. PHI Learning Pvt. Ltd., 2010.

[30] Riedmiller, Martin, and Heinrich Braun. "A direct adaptive method for faster backpropagation learning: The RPROP algorithm." Neural Networks, 1993, IEEE International Conference on. IEEE, 1993.

[31] http://www.cs.cmu.edu/~./quake-papers/painless- conjugate - gradient.pdf

[32] Scales, L.E., Introduction to Non-Linear Optimization, New York, Springer-Verlag, 1985

[33] Matlab Neural Network Toolbox.

[34] Fletcher, R., and C.M. Reeves, "Function minimization by conjugate gradients," Computer Journal, Vol. 7, 1964, pp. 149–154.

[35] Powell, M.J.D., "Restart procedures for the conjugate gradient method," Mathematical Programming, Vol. 12, 1977, pp. 241–254

[36] Beale, E.M.L., "A derivation of conjugate gradients," in F.A. Lootsma, Ed., Numerical methods for nonlinear optimization, London: Academic Press, 1972