Thresholding Method for 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*—Invention of the microarray technology has facilitated considerable improvement in the survival rate of the cancer patients. Microarray gene expression data has a small sample size and a large dimension. In this paper we suggest a hybrid combination of feature selection and feature extraction methods to reduce the size of gene expression data. Thresholding method is used for feature selection and discrete wavelet transform is used for feature extraction. The classification is performed using neural network algorithms. The results of classification are compared for different values of thresholds, wavelets and classification algorithms.**

***Index Terms*—Discrete wavelet transform, Resilient back propagation algorithm, Support vector machine, Error back propagation algorithm, Conjugate gradient back propagation with Fletcher-Reeves update algorithm, Conjugate gradient back propagation with Polak-Ribiére update, Conjugate gradient back propagation with Powell-Beale restarts algorithm.**

I .INTRODUCTION

Cancer sub-classification is one of the important research areas in the biomedical field. The survival rate of the cancer patients can be enhanced by precise diagnosis of cancer sub-type using microarray technology. The microarray cancer gene expression data set contains a matrix of data, every column of which specifies one cancer sample and every row denotes the expression value of a specific gene of different samples.

Microarray cancer sub-classification is usually carried out in two steps. First the feature selection or feature extraction of microarray data is carried out. Then classification of microarray data is performed using different classifiers. Feature selection or feature extraction causes the dimension reduction of the microarray data. [1]. Feature extraction methods transform the signal into new domain using various methods like discrete cosine transform, principal component analysis [2], discrete wavelet transform (DWT) [3] etc.

The feature selection methods are basically classified as filter, wrapper, embedded, hybrid [4] and Ensemble method [5]. Feature selection methods select the most significant set of features for classification with filter methods [6], [7], [8] wrapper method [9], [10], [11] embedded method [12], [13], [14] hybrid method [15], [16], [17] and ensemble methods [18], [19], [20].

Many a time's marker genes of a particular cancer type are used for the purpose of classification. Analysis of expression values of marker genes always do not result into correct identification of cancer [21].

The different classification algorithms used are, Support Vector Machines (SVM), K-means clustering, [10], [13], Naïve Bayes [22], Error Back Propagation algorithm (EBPA) [23] etc.

Shen, Qi [24] has employed Particle Swarm Optimization method for feature selection and SVM for classification. For GDS1976 dataset with 400 genes, the classification accuracy achieved is 92.67%.

Heba [25] has applied eight different gene selection techniques like, information gain, t-statistics etc. The different classifiers used are SVM, K-mean clustering and Random forest algorithms. For GDS1975 and GDS1976, data sets, the maximum classification accuracy attained is 94.59%, 90.81% respectively, with twenty genes and Random forest algorithm.

Each method proposed [2-25] has its own merits and demerits. However, there is scope for improvement in accuracy of classification with further reduction in the size of microarray data.

It is proposed to implement the classification of glioma grade III and grade IV datasets- GDS1975 and GDS1976, downloaded from Gene Expression Omnibus Database [26], [27]. We propose to perform feature selection by using thresholding method while feature extraction by using DWT. It is proposed to use Resilient back propagation algorithm (RPROP), and conjugate gradient algorithms for the classification.

In the following, Section II describes the feature selection method, section III shows the details of the feature extraction method, section IV describes different classification algorithms and implementation is explained in section V, followed by result analysis in section VI.

II. FEATURE SELECTION

In the thersholding method, two or three values of threshold are chosen .The list of genes chosen according to range of values defining a threshold is prepared for each class. The genes those are common to all the lists are eliminated. This leaves the lists containing genes forming mutually exclusive set of genes. The latter can be considered for further processing. The process is repeated for all the threshold values. Thresholding method helps in the dimension reduction of microarray gene expression data.

III. FEATURE EXTRACTION

In the case of DWT, a time scale representation of the digital signal is computed using digital filtering techniques. The DWT is calculated by successive high pass and low pass filtering of the discrete time-domain signal. As a result of dyadic decimation of filtered data at each level, detailed and approximation coefficients are obtained.

The detailed coefficients $dj(k)$ and the approximation coefficients $sj(k)$ are given as below:

$$d_j(k) = \sum_{m=2k}^{2k+N-1} g(m-2k)\, d_{j+1}(m) \quad \text{Eq.1}$$

$$s_j(k) = \sum_{m=2k}^{2k+N-1} h(m-2k)\, s_{j+1}(m) \quad \text{Eq.2}$$

where,

$g(n)$ = impulse response of high pass filter

$h(n)$ =impulse response of low pass filter.

k = translation parameter.

j = level of decomposition.

N= no of wavelet coefficients.

Approximation coefficients constitute the low frequency part of signal and detailed coefficients constitute the high frequency part of signal. Either the approximation coefficients or detailed coefficients or both can be used for the purpose of classification [28].

.

IV. CLASSIFICATION ALGORITHMS

While dealing with the nonlinear data, artificial neural network based classifier offers significant advantage. For a particular application the multi-layer neural network can be trained using number of examples.

A) Error Back Propagation Algorithm:

One of the most commonly used classification algorithms for multilayer perceptron is EBPA, in which the individual weight change is made proportional to the slope of error curve.The slope of error curve is proportional to the learning constant, difference between input and output, and the derivative of the output of corresponding neuron. The equation for the individual weight update is given as,

$$w' = w + (c(d-o)o') \quad \text{Eq.3}$$

where,

w' = modified weight

w = weight at the previous instance

c = learning constant

d = expected output of neuron

o = actual output of neuron

o' = derivative of actual output of neuron.

For larger inputs, the actual output of neuron increases and derivative of the output drops off. Hence the weight change reduces for large value of the difference between input and output. It affects the classification accuracy. To avoid the effect of the magnitude of gradient on the weight update, RPROP algorithm can be used.

B) Resilient Back Propagation Algorithm:

RPROP algorithm considers the sign of the gradient of the error instead of the magnitude of the gradient of the error. The size of weight update is increased (decreased), if the sign of slope of the error curve in two consecutive iterations remains same (different). The weight update is retained if slope of the error curve becomes zero. [29].

C) Conjugate Gradient Back Propagation Algorithms:

EBPA performs a linear search, to arrive at the global minimum of error curve. 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 [30]. It increases speed of convergence of conjugate gradient algorithms. The new search direction is determined as

$$\text{new search direction} = (p \times \text{the previous direction}) + \text{the steepest descent direction} \quad \text{Eq.4}$$

The multiplicative factor ‘b’ is calculated in different ways for various conjugate gradient algorithms. Conjugate gradient back propagation with Fletcher-Reeves update algorithm (CGF) calculates ‘p’ as in [31], [35].

$$p = \frac{CE}{PE} \quad \text{Eq.5}$$

where,

PE = energy in the previous gradient

CE= energy in the current gradient

Conjugate gradient back propagation with Polak-Ribiére update algorithm (CGP) calculates ‘P’ as in [32], [35]

$$P = \frac{PE - CE}{PE} \quad \text{Eq. 6}$$

When the number of iterations equal to the number of network parameters, conjugate gradient algorithms converge. If the algorithms do not converge within the number of iterations equal to the number of neural network parameters, the search direction is reset. In the case of conjugate gradient back propagation with Powell-Beale restarts, the algorithm (CGB), the search direction is reset, when there is very little

orthogonality left between the current gradient and a previous gradient [33], [34], [35].

V. IMPLEMENTATION

There are 26 samples of glioma grade III and 59 samples of glioma grade IV in the proposed classification.

Many a time's number of copies of identical same gene are attached to the microarray chip at different positions. As it increases the dimension of microarray data, the expression values of such genes are replaced by a single value that is average of all values of that individual gene in corresponding sample. The process is repeated for each gene of a sample and for all samples, which reduces the size of the microarray data.

With the help thresholding method, the features are selected from the microarray gene expression data. The different values of threshold selected are- T1 (<2000), T2 (2000, 10000), T3 (10000, 100000).The mutually exclusive set of genes formed as explained in section II. After normalization of this data, the features are extracted using DWT. The different wavelets used are Sym2, Sym4, Db2, Db4, Bior1.3, Bior2.4, followed by the classification algorithms like RPROP and Conjugate gradient algorithms. The classification accuracy obtained from different wavelets and algorithms is compared. The process is repeated for all thresholds.

VI. RESULT ANALYSIS

The best of the results for GDS1975 dataset with T1 (<2000), T2 (2000, 10000), T3 (10000, 100000) of microarray gene expression data, using approximation (A) or detailed (D) wavelet coefficients as input for different neural network algorithms are as shown in Table I, Table II and Table III.

TABLE I.GDS1975 DATASET T1 (<2000)

Sr. No	Algorithm	Wavelets	Accuracy
1.	RPROP	Sym4(A)	100 %
		Sym4,Bior1.3, Bior2.4(D)	100 %
2.	CGP	Bior2.4(D)	100 %
3.	CGF	Bior1.3(D)	100 %
		----	100 %

TABLE II. GDS1975 DATASET T2 (2000, 10000)

Sr. No	Algorithm	Wavelets	Accuracy
1	RPROP	Bior1.3 (A)	100 %
		Bior2.4 (D)	100 %
		------	100 %
2.	CGP	Db2, Sym2, Bior1.3 (A)	100 %
3.	CGB	Bior1.3, Bior2.4(A)	100 %
		Sym4(D)	100 %
4.	CGF	Db2, Sym2(A)	100 %

TABLE III. GDS1975 DATASET T3 (10000, 100000)

Sr. No	Algorithm	Wavelets	Accuracy
1	RPROP	Db2(A)	96.5 %
		------	96.5 %
2.	CGP	------	96.5 %
3.	CGF	------	96.5%

The best of the results for GDS1976 dataset with T1 (<2000), T2 (2000, 10000), T3 (10000, 100000) of microarray gene expression data, using approximation (A) or detailed (D) wavelet coefficients as input for different neural network algorithms are as shown in Table IV, Table V and Table VI.

TABLE IV.GDS1976 DATASET T1 (< 2000)

Sr. No	Algorithm	Wavelets	Accuracy
1	RPROP	Sym4(D)	98.8 %
2.	CGF	Sym4(D)	98.8 %

TABLE V. GDS1976 DATASET T2 (2000, 10000)

Sr. No	Algorithm	Wavelets	Accuracy
1	RPROP	Bior1.3, Bior2.4(D)	100 %
		------	100 %
2.	CGP	Db2, Sym2, Bior1.3, Sym4, Bior2.4 (D)	100 %
3.	CGB	Db4 (A)	100 %
		Sym2,Db2 (D)	100 %
4.	CGF	Db4, Sym4, Bior1.3 (A)	100 %
		Db2, Sym2, Db4,Bior2.4(D)	100 %

TABLE VI.GDS1976 DATASET T3 (10000, 100000)

Sr. No	Algorithm	Wavelets	Accuracy
1.	RPROP	------	97.6 %

CONCLUSION

The combination of thresholding method and wavelet transform shows significant reduction in the dimensions of GDS 1975 and GDS 1976 datasets. For both the data sets, the gene expression values with threshold range between 2000 and 10000, consistently gives 100% classification accuracy. The 100% classification accuracy is achieved using approximation or detailed wavelet coefficients as opposed to classification accuracies obtained using the methods suggested by Shen, Qi [24] and Heba [25] .The combination of type of the wavelet and the neural network algorithm that gives the best results mainly depends on the nature of variations in data for the particular data set.

REFERENCES

[1] 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)

[2] 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.

[3] 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.

[4] Y. Leung and Y. Hung, "A Multiple-Filter-Multiple-Wrapper Approach to Gene Selection and Microarray Data Classifica-tion," IEEEACM Trans ComputBiolBioinforma., vol. 7, no. 1, pp. 108–117, Jan. 2010.

[5] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A Survey on Filter Techniques for Feature Selection in Gene Ex-pression Microarray Analysis," IEEEACM Trans ComputBiol Bi-oinforma., vol. 9, no. 4, pp. 1106–1119, Jul. 2012.

[6] X. Li and M. Yin, "Multiobjective Binary Biogeography Based Optimization for Feature Selection Using Gene Expression Data," IEEE Trans. NanoBioscience, vol. 12, no. 4, pp. 343–353, Dec. 2013.

[7] B. Liao, Y. Jiang, W. Liang, W. Zhu, L. Cai, and Z. Cao, "Gene Selection Using Locality Sensitive Laplacian Score," IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 11, no. 6, pp. 1146–1156, Nov. 2014.

[8] J. Ferreira and M. A. T. Figueiredo, "Efficient feature selection filters for high-dimensional data," Pattern Recognit. Lett., vol. 33, no. 13, pp. 1794–1804, Oct. 2012.

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

[10] 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.

[11] A. Sharma, S. Imoto, and S. Miyano, "A Top-r Feature Selection Algorithm for Microarray Gene Expression Data," IEEEACM Trans ComputBiolBioinforma., vol. 9, no. 3, pp. 754–764, May 2012.

[12] Y. Liang, C. Liu, X.-Z. Luan, K.-S. Leung, T.-M. Chan, Z.-B. Xu, and H. Zhang, "Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification," BMC Bioinformatics, vol. 14, no. 1, p. 198, Jun. 2013.

[13] 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.

[14] 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.

[15] P. A. Mundra and J. C. Rajapakse, "SVM-RFE with MRMR Filter for Gene Selection," IEEE Trans. NanoBioscience, vol. 9, no. 1, pp. 31–37, Mar. 2010.

[16] S. S. Shreem, S. Abdullah, M. Z. A. Nazri, and M. Alzaqebah, "Hybridizing ReliefF, mRMR filters and GA Wrapper Approaches for Gene Selection," J. Theor. Appl. Inf. Technol., vol. 46, no. 2, 2012.

[17] 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.

[18] S. Zhang, H.-S. Wong, Y. Shen, and D. Xie, "A New Unsupervised Feature Ranking Method for Gene Expression Data Based on Consensus Affinity," IEEEACM Trans ComputBiolBioinforma., vol. 9, no. 4, pp. 1257–1263, Jul. 2012.

[19] Z. Yu, H. Chen, J. You, H.-S. Wong, J. Liu, L. Li, and G. Han, "Double Selection Based Semi-Supervised Clustering Ensemble for Tumor Clustering from Gene Expression Profiles," IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 11, no. 4, pp. 727–740, Jul. 2014

[20] P. L. Tan, S. C. Tan, C. P. Lim, and S. E. Khor, A modified two stage SVM-RFE model for cancer classification using microarray data, vol. 7062 LNCS. 2011.

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

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

[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.(ebpa)

[24] Q. Shen, Z. Mei, and B.-X. Ye, "Simultaneous genes and training samples selection by modified particle swarm optimization for gene expression data classification," Comput. Biol. Med., vol. 39, no. 7, pp. 646–649, Jul. 2009.

[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] LatSoman, K. P. Insight into wavelets: from theory to practice. PHI Learning Pvt. Ltd., 2010.
[29] 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.
[30] http://www.cs.cmu.edu/~./quake-papers/painless-conjugate -gradient.pdf
[31] Fletcher, R., and C.M. Reeves, "Function minimization by conjugate gradients," Computer Journal, Vol. 7, 1964, pp. 149–154.
[32] Scales, L.E., Introduction to Non-Linear Optimization, New York, Springer-Verlag, 1985
[33] Powell, M.J.D., "Restart procedures for the conjugate gradient method," Mathematical Programming, Vol. 12, 1977, pp. 241–254
[34] Beale, E.M.L., "A derivation of conjugate gradients," in F.A. Lootsma, Ed., Numerical methods for nonlinear optimization, London: Academic Press, 1972.
[35] Matlab Neural Network Toolbox.