Biomedical Signal Processing and Control 46 (2018) 67–75

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Stacked Autoencoder for classification of glioma grade III and grade IV

Supriya Patil[a,b,*], Gourish Naik[b], Radhakrishna Pai[a], Rajendra Gad[b]

[a] Padre Conceicao College of Engineering, Goa, India
[b] Electronics Dept., Goa University, Goa, India

ARTICLE INFO

Article history:
Received 4 July 2017
Received in revised form 12 October 2017
Accepted 10 July 2018

Keywords:
Resilient Back Propagation algorithm
Conjugate Gradient Back Propagation with Fletcher–Reeves Update algorithm
Conjugate Gradient Back Propagation with Polak–Ribire Update
Conjugate Gradient Back Propagation with Powell–Beale Restarts algorithm
Levenberg Marquardt algorithm
Stacked Autoencoder algorithm

ABSTRACT

Invention of the microarray technology has rendered it possible to inspect the whole genome at once in cancer classification. However, in order to curtail the computational complexity and augment the accuracy of cancer classification, it is essential to sift the vast microarray data for the informative genes. In this paper, Thresholding and Ratio methods are presented, individually as well as conjointly (hybrid method) to choose optimal gene subset from the microarray data. Moreover, Discrete Wavelet Transform (DWT) is deployed to pare the size of microarray data still further. The classification is accomplished by using various neural network algorithms and Stacked Autoencoder algorithm. The results of classification are compared for number of thresholds, ratios, wavelets and classification algorithms. It is observed that the Stacked Autoencoder network trained by Back Propagation algorithm delivers the best results in terms of classification accuracy and number of genes.

1. Introduction

Cancer develops when certain genes (cancer causing or cancer suppressing) in the human body start mutating. These gene mutations may be hereditary or acquired. As a result of the gene mutations, cells multiply rapidly and cause cancer. Knowing the subtype of a cancer plays a vital role in determining the prognosis and planning the treatment. However, conventional techniques of cancer diagnosis depend largely on the experience and skill of the physician. Microarray technology atomizes the process of cancer identification and assists in accurate cancer diagnosis. There is enormous number of genes in the microarray data. Besides, the intensity values of these genes if given directly as input to the classifier, will increase the computations and hinder the speed of cancer identification. In other words, to enhance the speed of accurate cancer classification, it is essential that the dimensions of microarray data be diminished to the maximum extent. This is achieved by filtering the informative genes from microarray data or converting the microarray data in different domain or using fusion of these methods [1]. To choose the required genes filter [2–4], wrapper [5–7], ensemble [8–10], hybrid [11–13] and embedded methods [14–16] are used. Wrapper, ensemble, hybrid and embedded methods are difficult to understand and are computationally more expensive while the filter methods are advantageous in terms of easy implementation, speed and computational complexity [17].

Zhang et al. [18] implemented distance based feature selection using Bhattacharya distance along with Support Vector Machine (SVM) classifier for colon and leukemia datasets. The average misclassification rate obtained is more than 3% with optimum number of genes. Tarek et al. [19] compared performance of Backward Elimination Hilberts Schmidt Independence Criterion, Extreme Value Distribution and Singular Value Deposition Entropy based gene selection methods. The classification is performed using five different ensemble classification algorithms for leukemia, colon and breast cancer datasets. The average classification accuracy achieved is less than 100% for larger number of genes. However, on account of using ensemble approach for classification, the method appears to be more complex and computationally expensive. Li et al. [20] compressed the size of microarray datasets using DWT. With approximation as well as detailed coefficients and SVM algorithm, classification accuracy of 100% for 100 genes and 93.5% for 250 genes is obtained for leukemia and colon datasets, respectively. Abusamra et al. [21] implemented eight different gene selection methods like Information Gain, Twoing Rule, Sum Minority, Gini Index, Sum of Variances, One Dimensional SVM and t-statistics. The classification algorithms employed are SVM, K-mean clustering and Random Forest. For GDS1975 and GDS1976 datasets, the maximum classification accuracy attained is 94.59%, 90.81%, respectively, with

* Corresponding author at: Electronics Dept., Goa University, Goa, India.
E-mail address: supriya@pccegoa.org (S. Patil).

https://doi.org/10.1016/j.bspc.2018.07.002
1746-8094/

twenty genes and Random Forest algorithm. Shen et al. [22] compared the performance of Naive SVM on GDS1976 dataset, 400 top ranked genes selected by t-statistics along with SVM classifier and simultaneous gene and sample selection by Modified Particle Swarm Optimization method. The maximum classification accuracy of 92.67% is achieved for GDS1976 dataset with 41 genes. Many a times the cancer marker genes are used for screening of the cancer but there are evidences of failure of this method [23].

Researchers implemented plenteous ways for diminishing the size of the microarray data, however, there are many open prospects for further improvement in terms of achieving 100% classification accuracy for less number of genes [17]. Therefore we are motivated to design a method to obtain 100% classification accuracy with optimum number of genes. We propose to perform the selection of features using Thresholding method and Ratio method, individually as well as conjointly (hybrid method), while the transformation of the signal is accomplished using DWT. The proposed classification algorithms are Resilient Back Propagation algorithm (RPROP), Conjugate Gradient algorithms, Levenberg Marquardt algorithm (LM) and Stacked Autoencoder algorithm (SAEN). The Thresholding method and the Stacked Autoencoder provides major contribution towards achieving 100% classification accuracy for optimum number of genes. The Thresholding method is computationally less expensive and has an ability to diminish the gene subset size by eliminating the genes which have a very large variation in their intensity value that may not be useful for classification. Based on the threshold ranges selected it is possible to obtain more than one gene subset to confirm the result of classification. In absence of the Thresholding method in the proposed work, 100% classification accuracy may be obtained due to some noisy genes (cross hybridized) with less maximum to minimum gene intensity ratio. This result will miss lead the classifier. Further, as compared to the other neural network algorithms such as Error Back Propagation (EBP), RPROP, Conjugate Gradient, LM etc., SAEN algorithm tends to outperform as a consequence of pre-training of the Sparse Autoencoder stages and the Softmax layer. The sparsity constraint of Sparse Autoencoder during its unsupervised training leads to significant reduction in the reconstruction error of the signal which in turn increases the classification accuracy [24,25].

2. Materials and methods

The details of the databases, system flow chart, different feature selection methods, feature extraction method and various classification algorithms are explained in the following subsections.

2.1. Database

In the proposed work, classification is implemented for glioma grade III/grade IV datasets- GDS1975, GDS1976, GDS1815 and GDS1816 from Gene Expression Omnibus Database [26–28]. Originally, Phillip and Freije obtained the datasets using Affymetrix Human Genome U133 Array. GDS1975, GDS1976 datasets contain 26 samples of glioma grade III and 59 samples of glioma grade IV while GDS1815, GDS1816 contain 24 samples of glioma grade III and 76 samples of glioma grade IV. Every dataset contains 22,283 features. From every dataset 70% samples are used for training and 15% are used for testing and validation each. Before using the datasets the duplicate features are eliminated by averaging and the data is normalized to have zero mean and unit standard deviation.

The proposed work is implemented using MATLAB R2017a software.

2.2. System flow chart

Fig. 1 shows the system flow chart for individual microarray data set.

Finally, to derive the identical gene subset across glioma datasets, considering optimal gene subsets, the genes common to the glioma datasets are mined. The classification is implemented using these genes.

2.3. Gene subset selection

For glioma grade III and grade IV classification, the various gene selection methods to obtain the required gene subset from the glioma datasets, Thresholding method, Ratio method and combination of two methods are explained in following subsections.

2.3.1. Thresholding method

Thresholding method is a simple method to select the consistently varying gene subset for cancer classification. It helps to narrow down the search space for selection of informative genes by discarding the genes with inconsistent intensity variation within the selected threshold range across the cancer samples. To filter the informative subset of genes for cancer classification, various values of threshold range are selected based on gene intensity variation in a specific data set. The microarray gene expression datasets for GDS1975, GDS1976, GDS1815 and GDS1816 are generated from 16 bit microarray image. As per the general observation of gene intensity values of glioma datasets, gene intensity values predominantly lie in 0 to 2000 range. If threshold ranges are considered as per the standard 1–2–5 sequence (500–1000, 1000–2000, 2000–5000, 5000–10,000, etc.), most of the threshold ranges either include no genes or very less number of genes. Therefore, the threshold ranges THD1 (500, 2000), THD2 (2000, 10,000) and THD3 (10,000, 100,000) are selected as a result of combination of two consecutive threshold ranges till 10,000 and thereafter from 10,000–100,000. The gene intensity values below 500 are not considered as it mostly includes noise. For each dataset, the genes with intensity value within these threshold ranges for both the classes of glioma are extracted. It gives three subsets of genes for every dataset that are used for classification of glioma grade III and grade IV.

2.3.2. Ratio method

Ratio method limits the search space for the selection of informative genes by eliminating the genes with inconsistent intensity variation across the cancer samples. In Ratio method, for every gene in both cancer classes, the ratio of maximum to minimum intensity is calculated. For GDS1975, GDS1976, GDS1815 and GDS1816 datasets various ratios considered are ratio <= 4, ratio <= 3.5, ratio <= 3, ratio <= 2.5. For each dataset, the genes having intensity value within the chosen ratio for both the classes of glioma are extracted and used for classification.

2.3.3. Fusion of Thresholding method and Ratio method

The fusion of Threshold and Ratio method helps us to obtain the subset of genes with less ratio within the specific range of threshold. For GDS1975, GDS1976, GDS1815 and GDS1816 datasets the threshold range and ratio that gives 100% classification accuracy almost by every algorithm are selected. For each of the dataset, the subset of genes that is common to the chosen thresholding range and the ratio are mined from both the classes of glioma and considered for classification. To get the optimal subset of genes, resultant genes are filtered depending on the difference in the average intensity of the genes for both the classes of glioma.

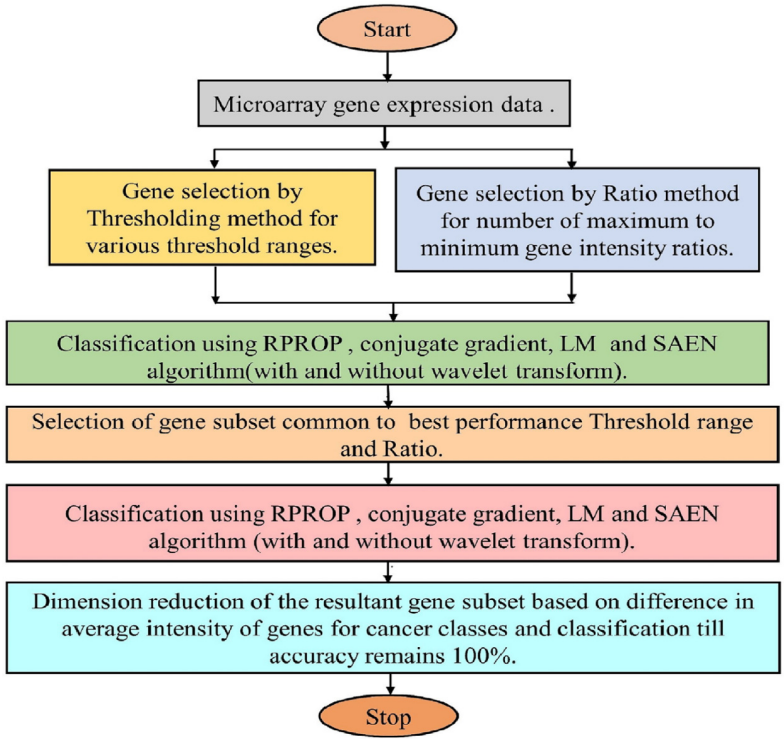


Fig. 1. System flow chart.

2.4. Feature extraction

The feature extraction is implemented using the DWT due to its ability to provide multiresolution analysis, localized time and frequency information while dealing with stationary and non-stationary signals. Due to higher energy compaction of DWT, less number of wavelet coefficients are required for classification. Moreover, availability of the several of types of wavelets like Haar, Daubechies, Coiflet, Symlet, Bio-orthogonal, Meyer, Mexican Hat etc. offer more flexibility for the selection of the wavelet basis function and ease of comparison. Hence, it can be safely concluded that, DWT performs best among all other feature extraction methods. In the case of DWT, a time scale representation of the digital signal is computed using digital filtering techniques. The DWT of a signal 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 approximation coefficients $q_j(k)$ and the detailed coefficients $p_j(k)$ are given as below:

$$q_j(k) = \sum_{m-2k}^{2k+N-1} r(m-2k)q_{j+1}(m), \tag{1}$$

$$p_j(k) = \sum_{m-2k}^{2k+N-1} s(m-2k)p_{j+1}(m), \tag{2}$$

where $s(n)$ = impulse response of high pass filter; $r(n)$ = impulse response of low pass filter; k = parameter related to time shift; j = decomposition level; N = number of wavelet coefficients.

Approximation coefficients constitute the low frequency part of the signal and the detailed coefficients constitute the high frequency part of the signal. Either the approximation coefficients or the detailed coefficients or both can be used for the purpose of classification [20,29]. For proposed work, the approximation coefficients obtained by using Db2, Db4, Sym2, Sym4, Bior1.3 and Bior2.4 wavelets are used for classification.

2.5. Classification algorithms

Artificial neural network processes information in parallel and it learns by examples. This distinctive characteristic makes the neural network appealing for solving problems with non-linear relationship between the input and the output. For a specific application the multi-layer neural network can be trained using number of examples. In the present work the classification is implemented using RPROP, Conjugate Gradient, LM and SAEN algorithm as explained in the following subsections.

2.5.1. Error Back Propagation algorithm

One of the most commonly used classification algorithm for multilayer perceptron is EBP algorithm [30]. The flow chart for EBP algorithm is as shown in Fig. 2.

The EBP algorithm makes the individual weight change 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 of the output layer neuron is given as,

$$w'_{kj} = w_{kj} + \eta(e_k - y_k)y'_k x_j, \tag{3}$$

where $w_{kj'}$ = modified weight that connects output of jth neuron in the hidden layer to kth neuron in output layer; w_{kj} = weight that connects output of jth neuron in hidden layer to kth neuron in the output layer; η = learning constant; e_k = expected output of the neuron; y_k = actual output of the neuron; y'_k = differentiation of actual output of the neuron; x_j = input to the neuron.

For larger inputs, as the actual output of neuron increases, derivative of the output drops off. As a consequence of reduction in weight change, the classification accuracy gets affected with

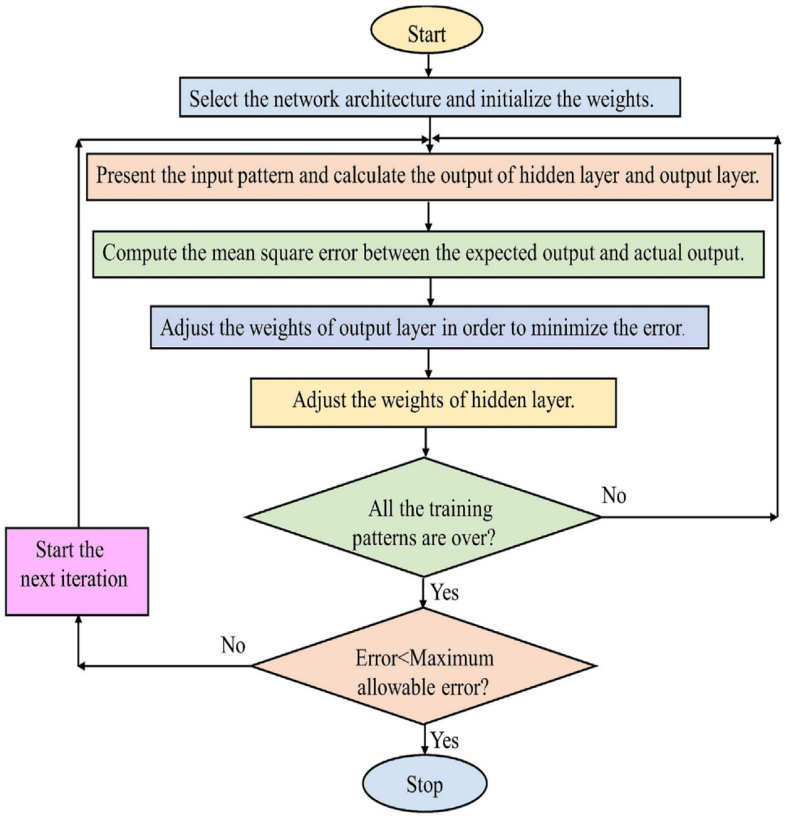


Fig. 2. Flow chart for Error Back Propagation algorithm.

increased difference between input and output. However, to avoid the effect of the magnitude of gradient on the weight update, RPROP algorithm can be used [31,32].

2.5.2. Resilient Back Propagation algorithm

For weight update, EBP algorithm considers the magnitude of the gradient of the error while, RPROP algorithm considers the sign of the gradient of the error. The size of weight update is increased, if the sign of slope of the error curve in two consecutive iterations remains same and vice-versa. Further, the weight update is retained if the slope of the error curve becomes zero [31,32].

2.5.3. Conjugate Gradient Back Propagation algorithms

EBP algorithm 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 the previous search direction [33]. It increases the speed of convergence of Conjugate Gradient algorithms. The new search direction is determined as,

$$a = (yc) + d, \tag{4}$$

where a = new search direction; y = multiplicative factor; c = previous search direction; d = the direction of steepest descent.

The multiplicative factor 'y' is calculated in different ways for various Conjugate Gradient algorithms. Conjugate Gradient Back Propagation with Fletcher–Reeves Update algorithm (CGFR) calculates 'y' as given by the equation below [32,34],

$$y = EC/EP, \tag{5}$$

where EC = energy in the current gradient; EP = energy in the previous gradient.

Conjugate Gradient Back Propagation with Polak–Ribire Update algorithm (CGPR) calculates 'y' as given by the equation below [32,35]

$$y = (EP - EC)/EP. \tag{6}$$

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 (CGPB) algorithm, the search direction is reset, when there is very little orthogonality left between the current gradient and a previous gradient [32,36,37].

2.5.4. Levenberg Marquardt algorithm

LM algorithm is one of the most competent algorithms for training multilayer perceptron algorithm. It performs like an EBP algorithm near the areas of complex part of the error curves while, near the areas of quadratic part of the error curve it performs like Newton's algorithm. The weight update rule for LM algorithm is given as [32,38]

$$w'_{kj} = w_{kj} - (J_k^T J_k + \mu I) J^T e_k, \tag{7}$$

where J_k = Jacobian matrix; u = combination coefficient; I = identity matrix; e_k = error vector.

2.5.5. Stacked Autoencoder Marquardt algorithm

An Autoencoder type of neural network makes use of an unsupervised Back Propagation training algorithm. It consists of an encoder and a decoder. An encoder converts the input into a hidden representation in order to extract the features from the input data. The decoder converts the hidden representation back to the

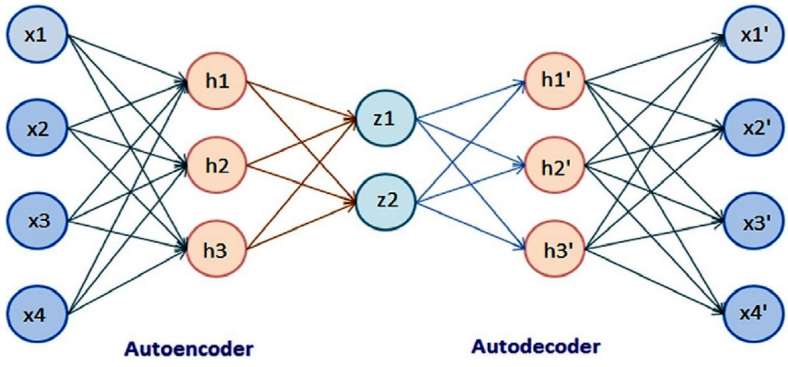


Fig. 3. Autoencoder.

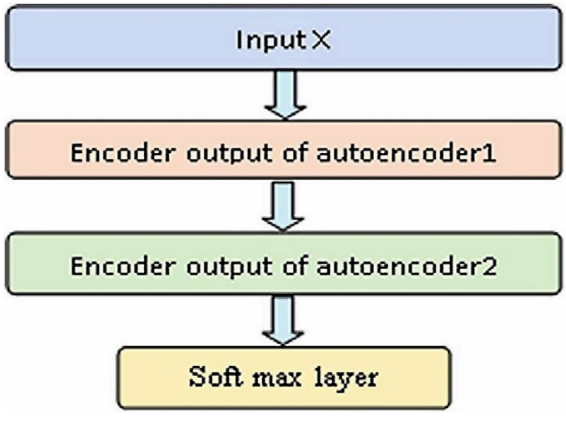


Fig. 4. Stacked Autoencoder.

input. To get the best possible representation of the input, the error between the original input and reconstructed input is used to update the weights. An Autoencoder is as shown in Fig. 3.

The Stacked Autoencoder network [39] consists of one or more Autoencoders and Softmax layer. The Stacked Autoencoder network is as shown in Fig. 4

The steps in training the Stacked Autoencoder network are as given below:

- Train the first Autoencoder to minimize the error between the original input and reconstructed input.
- Considering the output of hidden layer of first Autoencoder as input to second Autoencoder, train the second Autoencoder.
- Repeat the procedure for subsequent Autoencoders.
- Consider output of hidden layer of last Autoencoder as input to Softmax layer and train the Softmax layer using the supervised Back Propagation learning algorithm with the labeled data.
- Train the entire network using supervised Back Propagation algorithm to fine tune the weights and bias.

3. Results

For GDS1975, GDS1976, GDS1815, GDS1816 datasets, Figs. 5 and 6 present the results of Thresholding method and Ratio method, respectively.

Application of threshold THD1, results into 574, 330, 493 and 521 number of genes for GDS1975, GDS1976, GDS1815 and GDS1816 datasets, respectively. The number of genes chosen are 1301, 1413, 147 and 118 for GDS1975, GDS1976, GDS1815 and GDS1816 datasets, respectively using THD2. Threshold THD3 implementation results into 274, 185, 32 and 41 genes for GDS1975, GDS1976, GDS1815 and GDS1816 datasets, respectively. Db2 wavelet appears to be most suitable for the analysis of above mentioned datasets.

The accuracy obtained by the fusion of Thresholding and Ratio method along with u1–u2 (difference in mean intensity values of genes from both the classes of glioma) using common genes across GDS1975, GDS1976, GDS1815 and GDS1816 datasets is as shown in Fig. 7.

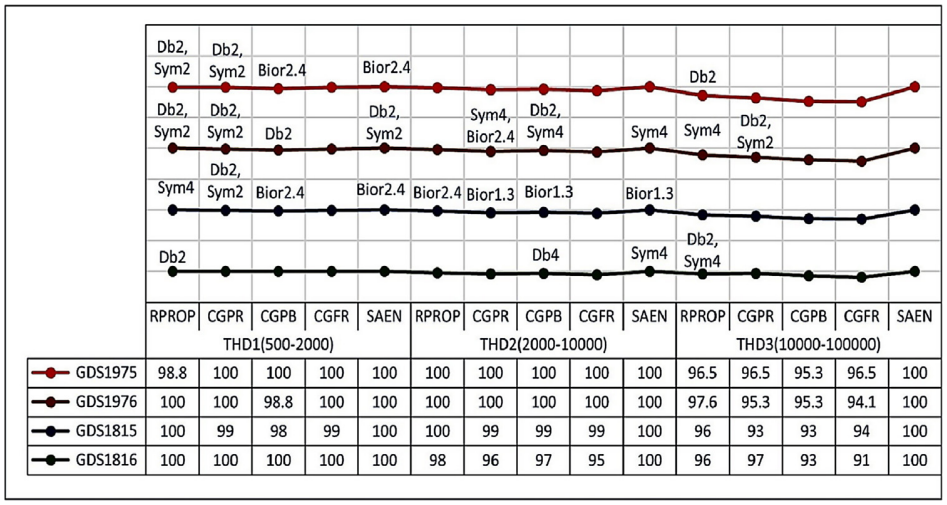
4. Discussion

Cancer classification reported in the literature, vary widely in respect of microarray datasets as well as methods employed to measure parameters defining and evaluating types of cancers. In this paper accuracy of classification and optimum number of genes obtained are compared with the results of the authors [21,22], who employed some of the same dataset as ours.

Figs. 8 and 9 demonstrates the variations in 30 average gene intensity values of malignant, benign and glioma grade III, grade IV samples, respectively.

As shown in Figs. 8 and 9, the cancerous and non-cancerous samples of brain tumor have gene intensity values wide apart from each other, making it easy to differentiate between them. However, with increase in the level of malignancy, genes become less differentially expressed, making the classification an uphill battle. Grade III and grade IV glioma brain tumors are the subtypes of the brain tumor at higher malignancy level. Consequently, it is a real challenge to singularize them.

In the proposed work, the threshold range THD2 (2000, 10,000) gives 100% classification accuracy with/without wavelet transform for GDS1975, GDS1976 and GDS1815 dataset using RPROP, Conjugate Gradient and Stacked Autoencoder algorithm while, THD1 (500, 2000) performs better for GDS1816 dataset. To effectively classify the samples using neural network, it is necessary to have smaller number of genes with considerable difference between them for both classes or else, more number of the genes with lesser difference in the intensity values. Threshold range THD1 (500, 2000) partially satisfies this criterion and at times gives 100% classification accuracy with wavelet transform for GDS1975, GDS1976 and GDS1815 datasets. The combination of the type of wavelet and neural network algorithm that gives the best result depends on the nature of variation of classification data and network parameters. Threshold range THD3 (10,000, 100,000) contains less number of genes with higher values of intensity. Since intensity values are very large, a small change often remains unnoticed especially if number of genes is lesser. However, SAEN algorithm, having the advantage of pre training and fine tuning of its stages gives 100% classification accuracy without wavelet transform for THD3 (10,000, 100,000) of all above mentioned datasets. Thresholding method is advan-



	RPROP	CGPR	CGPB	CGFR	SAEN	RPROP	CGPR	CGPB	CGFR	SAEN	RPROP	CGPR	CGPB	CGFR	SAEN
	THD1(500-2000)					THD2(2000-10000)					THD3(10000-100000)				
GDS1975	98.8	100	100	100	100	100	100	100	100	100	96.5	96.5	95.3	96.5	100
GDS1976	100	100	98.8	100	100	100	100	100	100	100	97.6	95.3	95.3	94.1	100
GDS1815	100	99	98	99	100	100	99	99	99	100	96	93	93	94	100
GDS1816	100	100	100	100	100	98	96	97	95	100	96	97	93	91	100

Fig. 5. Comparison of accuracy obtained for GDS1975, GDS1976, GDS1815, GDS1816 datasets using Thresholding method.

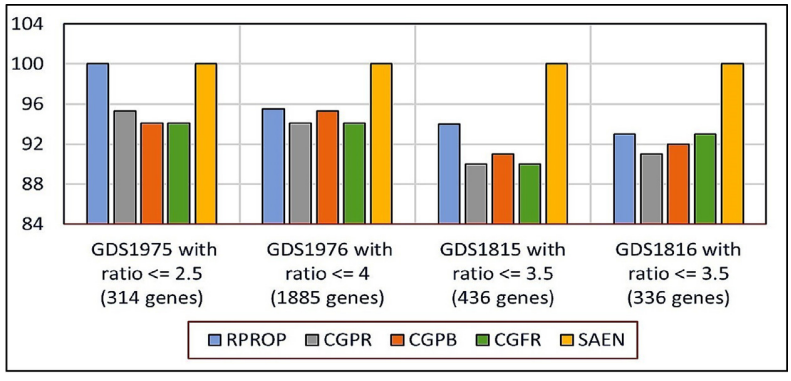


Fig. 6. Best of the results of the Ratio method for GDS1975, GDS1976, GDS1815 and GDS1816 datasets.

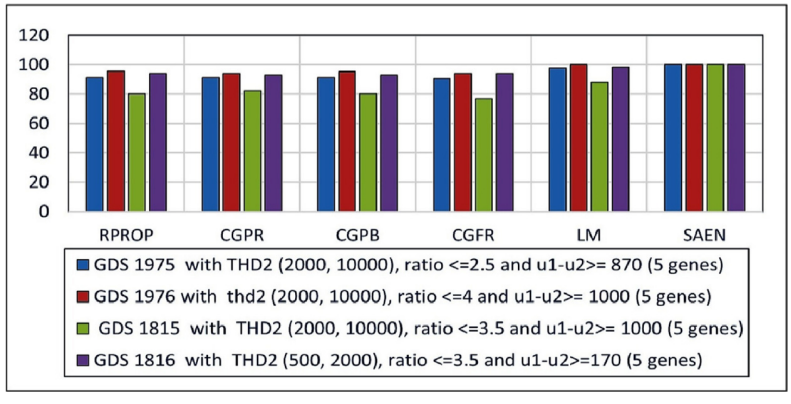


Fig. 7. Comparison of the accuracy achieved by fusion of Thresholding method and Ratio method for GDS1975, GDS1976, GDS1815 and GDS1816 datasets.

tageous in terms of providing the alternative subset of genes to confirm the result of classification. Genes with very large value of the intensity ratio exhibits inconsistent variation of intensity across the samples of that particular class. Ratio method eliminates such unreliable genes from the whole gene set. The hybrid of Thresholding method and Ratio method gives a subset of genes that are common and small in number in comparison with both Thresholding and Ratio method, carried out individually.

Further filtering of genes on the basis of difference in the average gene intensity with or without wavelet transform leads to 100% classification accuracy with less number of genes. While dealing with large size and number of samples, the increased memory requirement of LM algorithm, makes the implementation inefficient. For smaller network LM algorithm performs better than that of Conjugate Gradient algorithms. Stacked Autoencoder network trained with Back Propagation algorithm gives the best result as compare to Conjugate Gradient algorithms and LM algorithm owing to the pre training of Autoencoder stages, fine tuning of Softmax layer and finally fine tuning of entire Stacked Autoencoder.

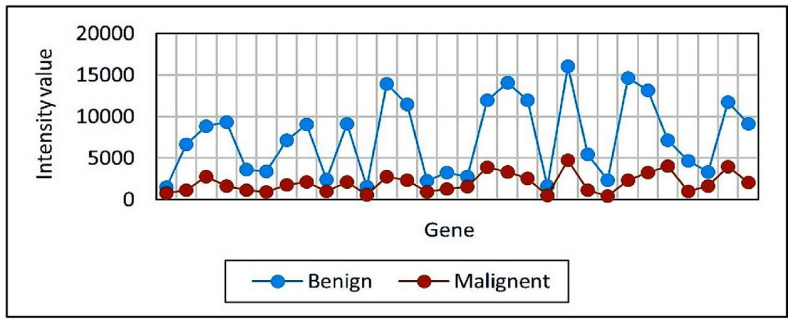


Fig. 8. Variation in 30 average gene intensity values of malignant and benign samples.

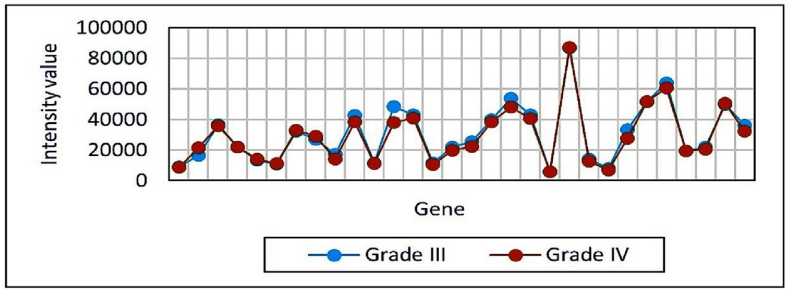


Fig. 9. Variation in 30 average gene intensity values of glioma grade III and grade IV samples.

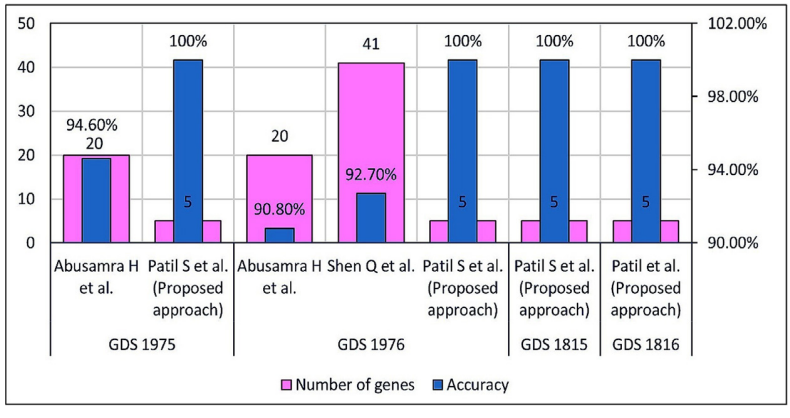


Fig. 10. Comparative results of proposed method with the existing methods.

The comparative analysis of four datasets in proposed study utilizes commonly transcribed genes, such as AKT3, MORF4L2, ANKRD17, SRP14 and ZNF550. AKT3 gene coding for serine/threonine protein kinase is involved in cell proliferation, differentiation and apoptosis. Further, AKT3 gene expression gets down regulated from grade III to grade IV [40]. It may be noted that, MORF4L2 is a vital component of NuA4 HAT and has significant role in transcriptional activation of several genes including oncogenes and proto-oncogenes [41]. An alteration in the gene expression of ANKRD17 observed from glioma grade III to grade IV may be attributed to G1/S transition [42]. SRP14 along with SRP9 and Alu RNA constitute elongation arrest domain signal recognition particle and plays a crucial role in targeting secretary protein to endoplasmic reticulum. Down regulation of SRP14 would alter signal recognition particle mediated vernacular protein transport system leading to cancer progression [43]. The uniportKB database has reviewed and annotated ZNF550 to be involved in transcriptional regulation. An Alteration in ZNF550 expression may lead to remodeling in expression pattern of cancer related genes promoting oncogenesis. The common transcriptions among four datasets and related functions of these genes leads to direct or indirect correlation of mutations in the above genes with the development of glioma grade III and IV. Hence it is opined that proposed computational approach facilitates the easy selection and precise classification of commonly transcribed genes among different transcription datasets with highest accuracy with moderate time.

The glioma grade III and grade IV classification using the fusion of Thresholding and Ratio method along with SAEN algorithm is advantageous in number of ways. It uses simple and computationally less expensive feature selection method. A classification accuracy of 100% is obtained using all the samples in the dataset as opposed to the method suggested by Shen et al. [22]. Thresholding method results into an alternative subset of genes to confirm the result of classification. It can be clearly seen that the optimal subset of genes obtained by proposed method is much smaller as compared to Abusamra et al. [21] and Shen et al. [22] for GDS1975 and GDS1976 datasets. Therefore, the proposed work provides a good trade-off between the number genes selected for classification

Table 1
Comparison of computational time of proposed method and existing methods.

Sr. No.	Method	Computational time in seconds
1	Abusamra et al.	3
2	Shen et al.	258
3	Patil et al. (proposed approach)	17

and the computational time for achieving maximum classification accuracy with respect to existing methods. However, the proposed method has certain disadvantages. The gene selection method does not consider interdependency between the genes. Also it offers moderate speed but considering the state of art it appears to be insignificant.

Fig. 10 illustrates the comparison of the accuracy and the number of genes obtained by proposed method with the existing methods.

Table 1. demonstrates the computational time of the proposed method with the existing method using Intel(R) Core(TM) i3 CPU M380 @2.53 GHz and MATLAB R2017a.

5. Conclusion

To select the informative subset of genes, Thresholding method considers the genes within the certain range of threshold values, while Ratio method considers entire gene set. From the results, it can be concluded that Thresholding method appears to have an edge, in the sense, it provides an alternative subset of genes for obtaining 100% classification accuracy. However, fusion of the two methods along with difference between mean intensity values of the both the classes of glioma gives optimum subset of required genes with less ratio within the certain threshold range. Alternatively, ratio can be chosen first and thresholding can be applied later, yielding the same subset of genes. The Stacked Autoencoder network along with the combination of Threshold and Ratio method outperforms the methods suggested by Abusamra et al. [21] and Shen et al. [22] giving 100% classification accuracy using only five common genes for GDS1975, GDS1976, GDS1815 and GDS1816 datasets.

Acknowledgements

We would like to thank Dr. Sanjeev C. Ghadi, Professor, Department of Biotechnology, Goa University, for his timely guidance in correlating the proposed study with the Genomics.

References

[1] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537, http://dx.doi.org/10.1126/science.286.5439.531.
[2] A.J. Ferreira, M.A. Figueiredo, Efficient feature selection filters for high-dimensional data, Pattern Recognit. Lett. 33 (2012) 1794–1804, http://dx.doi.org/10.1016/j.patrec.2012.05.019.
[3] Y. Zheng, C.K. Kwoh, A feature subset selection method based on high-dimensional mutual information, Entropy 13 (2011) 860–901, http://dx.doi.org/10.3390/e13040860.
[4] D. Mishra, B. Sahu, Feature selection for cancer classification: a signal-to-noise ratio approach, Int. J. Sci. Eng. Res. 2 (2011) 1–7.
[5] Q. Liu, Z. Zhao, Y.-x. Li, X. Yu, Y. Wang, A novel method of feature selection based on SVM, JCP 8 (2013) 2144–2149, http://dx.doi.org/10.4304/jcp.8.8.2144-2149.
[6] A. Sharma, S. Imoto, S. Miyano, A top-r feature selection algorithm for microarray gene expression data, IEEE/ACM Trans. Comput. Biol. Bioinf. 9 (2012) 754–764, http://dx.doi.org/10.1109/TCBB.2011.151.
[7] D.M. Witten, R. Tibshirani, A framework for feature selection in clustering, J. Am. Stat. Assoc. 105 (2010) 713–726, http://dx.doi.org/10.1016/j.ins.2010.08.047.
[8] S. Maldonado, R. Weber, J. Basak, Simultaneous feature selection and classification using kernel-penalized support vector machines, Inf. Sci. 181 (2011) 115–128, http://dx.doi.org/10.1016/j.ins.2010.08.047.
[9] M. Hindawi, K. Benabdeslem, Local-to-global semi-supervised feature selection, ACM International Conference on Information and Knowledge Management (CIKM 2013) (2013) 2159–2168, http://dx.doi.org/10.1145/2505515.2505542.
[10] S. Xiang, F. Nie, G. Meng, C. Pan, C. Zhang, Discriminative least squares regression for multiclass classification and feature selection, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1738–1754, http://dx.doi.org/10.1109/TNNLS.2012.2212721.
[11] P. Shi, S. Ray, Q.E.A. Zhu, Top scoring pairs for feature selection in machine learning and applications to cancer outcome prediction, BMC Bioinf. 12 (2011) 375, http://dx.doi.org/10.1186/1471-2105-12-375.
[12] S.-W. Chang, S. Abdul-Kareem, A.F.E.A. Merican, Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods, BMC Bioinf. 14 (2013) 170, http://dx.doi.org/10.1186/1471-2105-14-170.
[13] Y.B. Kim, J. Gao, Unsupervised gene selection for high dimensional data, Sixth IEEE Symposium on BioInformatics and BioEngineering, BIBE 2006 (2006) 227–234, http://dx.doi.org/10.1109/BIBE.2006.253339.
[14] M.A. Gaafar, N.A. Yousri, M.A. Ismail, A novel ensemble selection method for cancer diagnosis using microarray datasets, in: 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), IEEE, 2012, pp. 368–373, http://dx.doi.org/10.1109/BIBE.2012.6399652.
[15] H. Liu, L. Liu, H. Zhang, Ensemble gene selection by grouping for microarray data classification, J. Biomed. Inf. 43 (2010) 81–87, http://dx.doi.org/10.1016/j.jbi.2009.08.010.
[16] T. Abeel, T. Helleputte, Y.D. Van de Peer, et al., Robust biomarker identification for cancer diagnosis with ensemble feature selection methods, Bioinformatics 26 (2009) 392–398, http://dx.doi.org/10.1093/bioinformatics/btp630.
[17] J.C. Ang, A.H. Mirzal, et al., Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection, IEEE/ACM Trans. Comput. Biol. Bioinf. 13 (2016) 971–989, http://dx.doi.org/10.1109/TCBB.2015.2478454.
[18] J.W.W. Zhang, X. Lu, Feature selection for cancer classification using microarray gene expression data, Biostat. Biometrics Open Access J. 1 (2017) 555557.
[19] M.S.S. Tarek, R. Elwahab, Gene expression based cancer classification, Egypt. Inf. J. 18 (2017) 151–159.
[20] S. Li, C. Liao, J.T. Kwok, Wavelet-based feature extraction for microarray data classification, International Joint Conference on Neural Networks, IJCNN'06 (2006) 5028–5033, http://dx.doi.org/10.1109/IJCNN.2006.247208.
[21] H. Abusamra, A comparative study of feature selection and classification methods for gene expression data of glioma, Proc. Comput. Sci. 23 (2013) 5–14, http://dx.doi.org/10.1016/j.procs.2013.10.003.
[22] Q. Shen, Z. Mei, B.-X. Ye, Simultaneous genes and training samples selection by modified particle swarm optimization for gene expression data classification, Comput. Biol. Med. 39 (2009) 646–649, http://dx.doi.org/10.1016/j.compbiomed.2009.04.008.
[23] Tumor markers, 2011. URL https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-markers-fact-sheet. (Accessed 20 December 2011).
[24] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[25] Deep learning. Sparse autoencoders, 2014. URL http://www.ericlwilkinson.com/blog/2014/11/19/deep-learning-sparse-autoencoders. (Accessed 1 October 2017).
[26] W.A. Freije, F.E. Castro-Vargas, Z.H. Fang, et al., [Dataset] gene expression profiling of gliomas strongly predicts survival, Cancer Res. 64 (2004) 6503–6510, http://dx.doi.org/10.1158/0008-5472.CAN-04-0452.
[27] H.S. Phillips, S. Kharbanda, R.F. Chen, et al., [Dataset] molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis, Cancer Cell 9 (2006) 157–173, http://dx.doi.org/10.1016/j.ccr.2006.02.019.
[28] B.M. Costa, J.S. Smith, Y.C. Chen, et al., Reversing hoxa9 oncogene activation by pi3k inhibition: epigenetic mechanism and prognostic significance in human glioblastoma, Cancer Res. 70 (2010) 453–462, http://dx.doi.org/10.1158/0008-5472.
[29] K. Soman, Insight into Wavelets: From Theory to Practice, PHI Learning Pvt. Ltd., 2010.
[30] J.M. Zurada, Introduction to Artificial Neural Systems, vol. 8, West St. Paul, 1992.
[31] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, IEEE International Conference on Neural Networks (1993) 586–591, http://dx.doi.org/10.1109/ICNN.1993.298623.
[32] Matlab neural network toolbox, 2016.
[33] J.R. Shewchuk, An introduction to the conjugate gradient method without the agonizing pain, 2009, URL https://www.cs.cmu.edu/quake-papers/painless-conjugate-gradient.pdf. (Accessed 3 February 2014).
[34] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154, http://dx.doi.org/10.1093/comjnl/7.2.149.
[35] L. Scales, Introduction to Non-Linear Optimization, Springer-Verlag New York, Inc., 1985.
[36] M.J.D. Powell, Restart procedures for the conjugate gradient method, Math. Program. 12 (1977) 241–254, http://dx.doi.org/10.1007/BF01593790.
[37] E. Beale, A derivation of conjugate gradients, Numer. Methods Nonlinear Optim. (1972) 39–43.
[38] B.M. Wilamowski, H. Yu, Neural network learning without backpropagation, IEEE Trans. Neural Netw. 21 (2010) 1793–1803, http://dx.doi.org/10.1109/TNN.2010.2073482.

[39] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, Neurocomputing 187 (2016) 27–48, http://dx.doi.org/10.1016/j.neucom.2015.09.116.
[40] H. Mure, K. Matsuzaki, K.T. Kitazato, Y. Mizobuchi, K. Kuwayama, T. Kageji, S. Nagahiro, Akt2 and akt3 play a pivotal role in malignant gliomas, Neuro-oncology 12 (3) (2009) 221–232, http://dx.doi.org/10.1093/neuonc/nop026.
[41] G.S. Yochum, D.E. Ayer, Role for the mortality factors morf4, mrgx, and mrg15 in transcriptional repression via associations with pf1, msin3a, and transducin-like enhancer of split, Mol. Cell. Biol. 22 (22) (2002) 7868–7876.
[42] M. Deng, F. Li, B.A. Ballif, S. Li, X. Chen, L. Guo, X. Ye, Identification and functional analysis of a novel cyclin e/cdk2 substrate ankrd17, J. Biol. Chem. 284 (12) (2009) 7875–7888.
[43] K. Strub, P. Walter, Assembly of the alu domain of the signal recognition particle (srp): dimerization of the two protein components is required for efficient binding to srp rna, Mol. Cell. Biol. 10 (2) (1990) 777–784.

Supriya Patil is an Associate professor in Electronics and Telecommunication Engineering Department at Padre Conceicao College of Engineering, Verna, Goa. Presently, she is pursuing her Ph.D at Goa University. She has obtained B.E degree from DKTE Textile and Engineering Institute, Ichalkaranji, India and M.E degree from Walchand College of Engineering, Sangli, India. She has 20 years of teaching experience in the field of Electronics and communication. Her research interests are Digital Signal Processing and Artificial Neural Network.

t

Prof. Gourish M. Naik is presently working as Professor and Head of Electronics department at Goa University. He obtained his Ph.D from Indian Institute of Science, Bangalore in 1987. He has guided six Ph.D's. He is also coordinator of DEITI (an educational broadcast studio supported by Indian Space Research). His other commitments are regulating digitization Center at Goa University to support the various Digital repository projects like DIGITAP (Digital Repository for Fighter Aircrafts of Indian Navy) Million Book project of Ministry of Information Technology, New Delhi and Antarctica Study Center (NCAOR), Govt. of India. He has to his credit around 50 research papers, many invited talks and keynote addresses. He has authored two books on Embedded Systems published by Springer (Holland). He is a member of Goa State Rural Development Authority, member of advisory board, Goa Police Cyber Crimes and also advisor for Directorate of Technical Education. He was the chairman of Goa University Technical Advisory Committee.

Prof. K.R. Pai has retired as Professor and Head in Electronics and Telecommunication Engineering at P.C. College of Engineering, Goa. He received B.E degree in Electrical Engineering from NITK, Surathkal, M.Tech degree from IISC, Bangalore and Ph.D from IIT, Mumbai. He has 35 years of teaching experience with specialization in communication and signal processing. He has authored and coauthored more than 25 journal and conference papers in the area of digital signal processing. He has guided three Ph.Ds in the field of Signal Processing and Biomedical Engineering. He has been referee for Ph.D thesis and Examiner for many Ph.D comprehensive and defense viva-voce at various universities. He has chaired/co-chaired several technical sessions in National and International conferences.

Dr. Rajendra S. Gad is Associate Professor of Electronics at Goa University. He received B.Sc (Physics) and M.Sc (Electronics) degrees from Goa University in 1995 and 1997 respectively. He worked for his Ph.D in areas of non-invasive measurements to understand problem of human blood glucose measurement in the year 2009. His team was judge winner at Mentor Graphics University Design Contest to design LC3 Processor in the year 2010. He established MOU with AlTERA Inc., USA under University program to develop FPGA SoC laboratory. His research interest is multivariate signal processing having application in the areas of multispectral and 3D biometrics, MIMO communication systems and Autonomic Neural System.