A NEW APPROACH FOR EPILEPTIC SEIZURE DETECTION USING EXTREME LEARNING MACHINE

Yuedong Song
Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, U.K.

Sarita Azad
Department of Applied Mathematics and Theoretical Physics, University of Cambridge
Wilberforce Road, Cambridge CB3 0WA, U.K.

Pietro Lio
Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, U.K.

Keywords: Epileptic seizure detection, Electroencephalogram (EEG), Discrete Wavelet Transform, Extreme Learning Machine (ELM).

Abstract: In this paper, we investigate the potential of discrete wavelet transform (DWT), together with a recently-developed machine learning algorithm referred to as Extreme Learning Machine (ELM), to the task of classifying EEG signals and detecting epileptic seizures. EEG signals are decomposed into frequency sub-bands using DWT, and then these sub-bands are passed to an ELM classifier. A comparative study on system performance is conducted between ELM and back-propagation neural networks (BPNN). Results show that the ELM classifier not only achieves better classification accuracy, but also needs much less learning time compared to the BPNN classifier. It is also found that the length of the EEG segment used affects the prediction performance of classifiers.

1 INTRODUCTION

Epilepsy, the second most common serious neurological disorder in human beings after stroke, is a chronic condition of the nervous system and is characterized by recurrent unprovoked seizures. Approximately one in every 100 persons (about 50 million people) worldwide are suffering from epilepsy(Iasemidis et al., 2003). Electroencephalography (EEG) is an important clinical tool for monitoring, diagnosing and managing neurological disorders related to epilepsy.

In recent years, there has been an increasing interest in the application of pattern recognition (PR) methods for automatic epileptic seizure detection. Several methods have been developed for handling EEG signals classification, and among these methods, Multi-layer Perceptron Neural Network (MLPNN)(Acir et al., 2005; Kalayci and Ozdamar, 1995; Jahankhani et al., 2006; Wang et al., 2005; Übeyli, 2009; Ghosh-Dastidar et al., 2007) and Support Vector Machine (SVM)(Guler and Übeyli, 2007; Chandaka et al., 2009; Lima et al., 2009) are two widely used classification paradigms.

The general trend in automatic epileptic seizure detection has focused on high accuracy but has not considered the time taken to train the classification models, which should be an important factor of developing an EEG-based automatic detection device for epileptic seizures because the device will need to update its training during use. Therefore some classification models with high classification accuracy may not be satisfactory when considering the trade-off between the classification accuracy and the time for training the classification models. In our study, to obtain a balance between high classification accuracy and short training time, we investigate the use of a novel paradigm of learning machine called Extreme Learning Machine (ELM)(Huang et al., 2004). In recent years, Extreme Learning Machine has been increasingly popular in classification task due to its high generation ability and fast learning speed. In (Wang and Huang, 2005), a classification system is built using ELM to classify protein sequences with ten classes of super-families obtained from a domain database, and its performance is compared with that of Backpropagation Neural Networks. The results

Song Y., Azad S. and Lio P. (2010).
A NEW APPROACH FOR EPILEPTIC SEIZURE DETECTION USING EXTREME LEARNING MACHINE.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 436-441
DOI: 10.5220/0002745904360441

show ELM greatly outperforms BPNN in terms of both training time and classification accuracy.(Zhang et al., 2007) developed an ELM for multi-category classification in three Cancer Microarray Gene Expression datasets, and the results reveal that ELM can not only obtain high classification accuracy but also avoid problems such as overfitting, local minima, and improper learning rate. In addition to the field of Bioinformatics, Extreme Learning Machine has also been successfully applied to the field of Biosignal Processing. (Kim et al., 2007) proposed an ELM based classification scheme for arrhythmia classification, and finally achieved 97.5% in average accuracy, 97.44% in average sensitivity, 98.46% in average specificity, and 2.423 seconds in learning time.

Up to now, to the best of our knowledge, there is no study in the literature related to the assessment of ELM classification performance when applied specifically to the epileptic/non-epileptic discrimination problem.

The present study is organized as follows: Section 2 describes the EEG signals benchmark dataset used in the experiments. Section 3 presents the mechanisms of discrete wavelet transform (DWT), and provides a description of backpropogation neural network (BPNN) and extreme learning machine (ELM) classifiers. The feature extraction from DWT and performance comparison of ELM with BPNN is reported in Section 4. Finally, Section 5 concludes the paper.

2 DATA ANALYZED

In this study, a publicly-available database introduced in (Andrzejak et al., 2001) has been used. The whole data is composed of 5 sets (denoted A-E), each containing 100 single-channel EEG data of 23.6s duration. Sets A and B were taken from surface EEG recordings of 5 healthy volunteers with eyes open and closed, respectively. Sets C, D, and E originated from the EEG archive of presurgical diagnosis. Signals in Set C were recorded from the hippocampal formation of the opposite hemisphere of the brain, and signals in Set D were recorded from within the epileptogenic zone. While Sets C and D contain only brain activity measured during seizure free intervals, Set E contains only seizure activity. All EEG signals were recorded with the same 128-channel amplifier. The data were digitized at 173.6 samples per second at 12-bit resolution. Band pass filter was set to 0.53–40 Hz. In the present work, we classified only two (A and E) of the complete datasets. The total number of EEG signals is 200 (100 normal signals and 100 seizure signals). Each data set has 4096 sampling points.

3 THEORY AND MODELS

3.1 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) can be thought of as an extended version of the classic Fourier Transform and instead of working on a single scale, it works on a multi-scale basis where each signal is decomposed into several scales, each scale providing a particular coarseness of the signal. Each phase of decomposition of a signal is composed of two down samplers by 2 and two digital filters. Figure 1 gives the process of Discrete Wavelet Transform. Here in each phase $h(.)$ is the high-pass filter of that phase and serves as the discrete mother wavelet; and $g(.)$ gives the low-pass filter, which acts as the mirror version of the corresponding $h(.)$. The down-sampled outputs of the first low-pass and high-pass filters supply the approximation $A1$, and the detail $D1$, respectively. The first approximation, $A1$, is further decomposed and the procedure is continued. Each low-pass filter g satisfies the quadrature mirror condition given in (Guler and Übeyli, 2007):

$$G(Z)G(Z^-1)+G(-Z)G(-Z^{-1})=1 \qquad (1)$$

where $G(Z)$ stands for the $Z-$transform of the filter g. Its corresponding high-pass filter can be obtained as:

$$H(Z)=ZG(-Z^{-1}) \qquad (2)$$

Therefore, a sequence of filters, in increasing subscript of i can be obtained:

$$G_i+1(Z)=G(Z^{2^i})G_i(Z), \qquad (3)$$

$$H_i+1(Z)=H(Z^{2^i})H_i(Z), i=0,...,I-1 \qquad (4)$$

Where initially $G_0(Z)=1$.

For classification purpose two models, namely Backpropagation neural network and Extreme learning machine, are discussed in this work.

3.2 Extreme Learning Machine (ELM)

Recently a new learning algorithm called Extreme Learning Machine (ELM) has been proposed for single-hidden layer feedforward networks in(Huang et al., 2004). In comparison with some well-known classification paradigms such as Support Vector Machine (SVM) and Backpropagation Neural Network (BPNN), ELM has some prominent features such as higher generalization ability, faster learning speed,

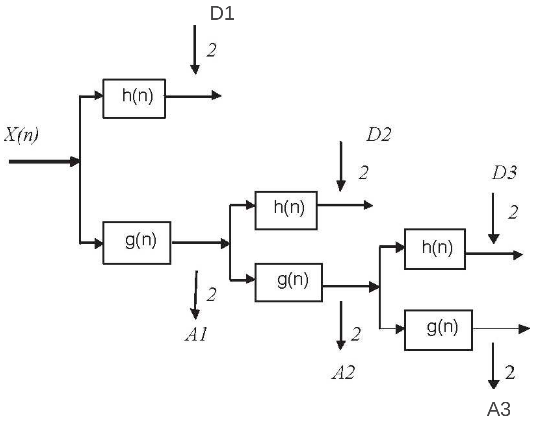


Figure 1: Wavelet decomposition of a signal.

and suitability for almost all nonlinear activation functions, and it can avoid problems like local minima and improper learning rate, which are usually faced by traditional learning methods. Figure 2 shows the structure of ELM. Here, we briefly present the idea behind ELM.

Suppose learning N arbitrary different instances (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in R^n$, and $t_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in R^m$, standard single-layer feedforward networks with N_h hidden neurons and activation function $g(x)$ are mathematically modelled as a linear system

$$\sum_{i=1}^{N_h} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, ..., N, \tag{5}$$

where $w_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ denotes the weight vector connecting the ith hidden neuron and the input neuron, $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ denotes the weight vector connecting the ith hidden neuron and output neurons, and b_i represents the threshold of the ith hidden neuron. $w_i * x_j$ represents the inner product of w_i and x_j. If the single-layer feedforward network with N_h hidden neurons with activation function $g(x)$ is able to approximate N distinct instances (x_i, t_i) with zero error means that

$$\mathbf{H}\beta = \mathbf{T}, \tag{6}$$

where

$$\mathbf{H}(w_1, ..., w_{Nh}, b_1, ..., b_{Nh}, x_1, ..., x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \ldots & g(w_{N_h} \cdot x_1 + b_{N_h}) \\ . & . & . \\ . & . & . \\ . & . & . \\ g(w_1 \cdot x_N + b_1) & \ldots & g(w_{N_h} \cdot x_N + b_{N_h}) \end{bmatrix}$$

$$w = \begin{bmatrix} w_1^T \\ . \\ . \\ . \\ w_{N_h}^T \end{bmatrix}_{N_h \times m} \qquad \mathbf{T} = \begin{bmatrix} t_1^T \\ . \\ . \\ . \\ t_N^T \end{bmatrix}_{N \times m}$$

$\mathbf{H}$ is the hidden layer output matrix of the SLFN. Hence for fixed arbitrary input weights w_i and the hidden layer bias b_i, training a single-layer feedforward network equals to discovering a least-squares solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$, $\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$ is the best weights, where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse. In terms of(Huang et al., 2006), Extreme Learning Machine utilizes such Moore-Penrose inverse approach for obtaining good generalization performance with extremely fast learning speed. Unlike some conventional methods, for example Backpropagation algorithm, ELM is able to avoid problems in tuning control parameters (learning epochs, learning rate, and so on) and keeping to local minima.

The procedure of ELM for single-layer feedforward networks is expressed as follows:

1. Choose arbitrary value for input weights w_i and biases b_i of hidden neurons.
2. Calculate hidden layer output matrix $\mathbf{H}$.
3. Obtain the optimal $\hat{\beta}$ using $\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$.

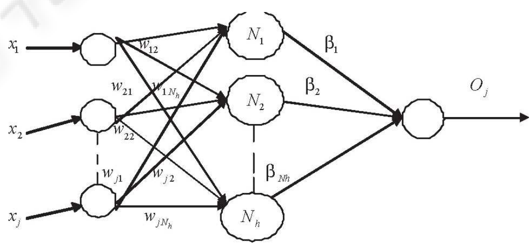
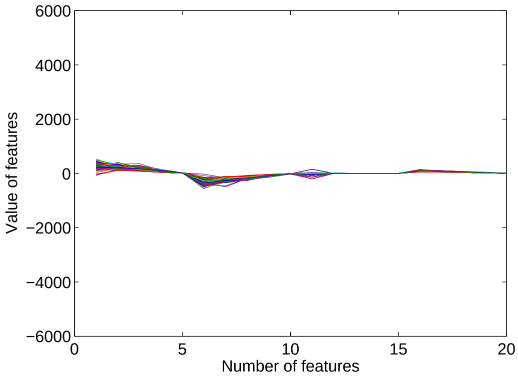


Figure 2: The structure of ELM.

4 EXPERIMENTS AND RESULTS

As previously mentioned, the whole dataset is divided into 5 categories (Sets A-E), each containing 100 EEG signals of 23.6s. In the present work, following the work of(Subasi, 2007) and (Chandaka et al., 2009), we use only Set A (normal EEG signals) and Set E (epileptic seizure signals) to conduct the computational simulation with ELMs. Therefore, totally 200 EEG signals are obtained, and each EEG signal contains 4096 sampling points. The discrete wavelet



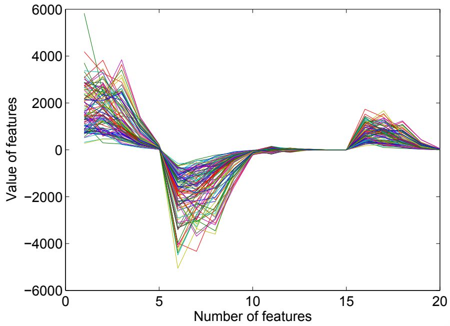


Figure 3: EEG data after feature extraction (Left: Set A. Right: Set E.).

transform (DWT) was used to extract features from these two EEG data sets, since DWT has ability to capture transient characteristics in EEG signals and to localize them in both time and frequency content accurately according to(Subasi, 2007). After the process of multi-resolution decomposition, the EEG signals were decomposed into details D1-D4 and one approximation A4. These approximation and details were reconstructed from the Daubechies 4-wavelet filter. In order to further reduce the dimensionality of these extracted features, statistics over these wavelet coefficients were utilized and four statistical features was used: (1) Maximum of the wavelet coefficients in each sub-band; (2) Minimum of the wavelet coefficients in each sub-band; (3) Mean of the wavelet coefficients in each sub-band; (4) Standard deviation of the wavelet coefficients in each sub-band. Hence, each EEG data finally consists of 20 features. From Figure 3, we can see the difference between these two EEG datasets clearly. The values of features (in amplitude) extracted from EEG signals in Set E are much larger than the values of features extracted from EEG signals in Set A. Hence, these statistical features which represent the significant properties of the two data sets can be utilized for evaluating the performance of classifiers.

All the simulations are based on a 2.27 GHz 2-core CPU with 2 GB memory. In order to compare the performance of ELM classifiers, we also implemented a backpropagation neural network (BPNN) based on a Levenberg-Marquardt backpropagation (LMBP) learning algorithm which is thought of as the fastest method for training moderate-sized feedforward neural networks according to(Hagan and Menhaj, 1994). For the BPNN and ELM, all of the input value were normalized in the range of [-1,1]. The weight vector w_i and the threshold b_i of ELM were randomly generated in the range of [-1,1]. The number of hidden neurons in BPNN was set to 25 according to (Subasi, 2007) and the number of hidden neurons in ELM was set to 20. The performance of the BPNN and ELM algorithms was evaluated by the following measures:

- Sensitivity: number of true positive detections/number of actually positive subjects.
  $Sensitivity = TP/(TP+FN)$.
- Specificity: number of true negative detections/number of actually negative subjects.
  $Specificity = TN/(TN+FP)$.
- Classification Accuracy: number of correct detections/total number of subjects.
  $Classification \quad Accuracy = (TP+TN)/(TP+TN+FP+FN)$.
- Learning time: A measure of the time spent in training classifiers.

Among these 200 EEG data (100 normal signals and 100 epileptic seizure signals), half of the normal EEG signals and half of the epileptic EEG signals were used for training, and the rest for testing.

After training BPNN and ELM, the test data was used to evaluate the performance of these two classification algorithms. Classification results of the BPNN and the ELM classifiers are displayed by confusion matrices. The confusion matrices demonstrating classification results of two classifiers are given in Table 1. In terms of the confusion matrices, 5 epileptic subjects were classified incorrectly by the BPNN classifier as normal subjects, however only 2 epileptic subjects were classified incorrectly using the ELM classifier. Both classifiers identified correctly all normal subjects. In order to calculate the average test performance of two classifiers, the classification experiment was repeated for ten times and all the simulation results were averaged over 10 trials. The results are given in Table 2.

Table 1: Confusion matrices of the classifiers.

Classifiers	Desired result	Output result: Epileptic	Output result: Normal
BPNN	Epileptic	45	5
	Normal	0	50
ELM	Epileptic	48	2
	Normal	0	50

Table 2: Performance comparison of ELM and BPNN.

Classifiers	Average Learning Time(Seconds)	Average Sensitivity (%)	Average Specificity (%)	Average Accuracy (%)
ELM	0.0187	93.6	100	96.8
BPNN	52.9969	90.6	100	95.3

As quoted in Table 2, it is clear that the ELM classifier outperforms the BPNN classifier in its average classification accuracy of the EEG signals. It is shown that performance rate with ELM is 96.8% whereas with BPNN it is 95.3%. The ELM classifier recognized precisely all the normal and epileptic subjects with sensitivity 93.6% and specificity 100%, followed by the BPNN classifier with sensitivity 90.6% and specificity 100%. Thus the ELM classifier can obtain better generalization than the BPNN classifier. In addition, the learning time of the ELM classifier is 0.0187s while the learning time of the BPNN classifier is 52.9969s. The ELM classifier can run 2834 times faster than the BPNN classifier. Hence, in the case of real-time implementation of epilepsy diagnosis support system, ELM classifiers are more appropriate than BPNN classifiers.

In Table 3, we present a comparison in classification performance achieved by different methods. we have quoted results from our present proposed method (Wavelet-ELM) and also from recently reported in(Subasi, 2007) and (Chandaka et al., 2009). The datasets used in these experiments are the same (Set A and Set E). It is shown in the table that our method generates better classification accuracy compared with those of others.

Table 3: Comparison of statistical parameters of various classification algorithms.

Type of Classifier	ME model (Subasi2007) (%)	Cross-correlation SVM(Chandaka2009) (%)	Wavelet ELM (%)
Specificity	94	100	100
Sensitivity	95	92	93.6
Accuracy	94.5	95.96	96.8

Table 4: Effect of change in window size.

Window Size	Average Test Accuracy (%)
4096	96.80
2048	96.35
1024	96.15
512	95.76
256	94.18

In the next step of the experiment, we further investigate the generalization of our Wavelet-ELM classification scheme by varying the length of EEG signals. Four rectangular windows are formed, using 2048, 1024, 512 and 256 sampling points respectively, such that each EEG signal is divided into 2, 4, 8, 16 small segments. Hence, in Case Study 2, there are 200 EEG signals with 2048 sampling points in each set; in Case Study 3, there are 400 EEG signals with 1024 sampling points in each set; in Case Study 4, there are 800 EEG signals with 512 sampling points in each set; in Case Study 5, there are 1600 EEG signals with 256 sampling points in each set. In all case studies, we still use 50% EEG data from each set for training and the remaining 50% for testing. Table 4 describes the average test accuracy over 10 trials using ELM classifiers for every case study. It is obvious from the results that the larger the window size, the greater the average test accuracy. Figure 4 summarises the obtained results.

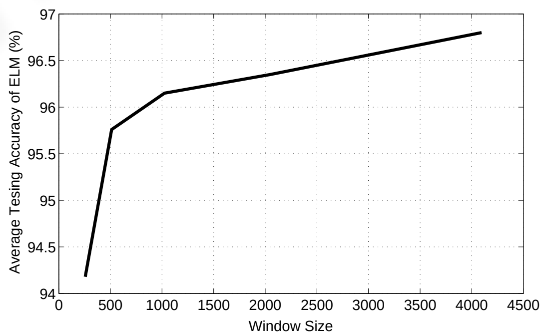


Figure 4: The relationship between the generalization performance and window size.

5 CONCLUSIONS

In this paper, we have investigated the Wavelet-Extreme Learning Machine classification scheme for identifying epileptic seizures. Using statistical features extracted from the DWT sub-bands of EEG signals, ELM and BPNN classifiers are built and compared according to their test accuracy and learning time. The proposed system using the ELM classifier can achieve test accuracy as high as 96.8%, as compared to BPNN classifier and two recently-proposed methods, where the test accuracies are 95.3%, 94.5% and 95.96% respectively. In addition, this study also shows that the ELM classifier needs much less learning time compared to the stand-alone BPNN classifier for the task of epileptic seizure detection, which demonstrates its potential for real-time implementation in an epilepsy diagnosis support system.

ACKNOWLEDGEMENTS

This work is financed by the EU 6 Framework Programme Project: Measuring and Modelling Relativistic-Like Effects in Brain and NCSs.

REFERENCES

Acir, N., Oztura, I., Kuntalp, M., Baklan, B., and Guzelis, C. (2005). Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Transactions on Biomedical Engineering*, 52(1):30–40.

Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):61907.

Chandaka, S., Chatterjee, A., and Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems With Applications*, 36(2P1):1329–1336.

Übeyli, E. (2009). Statistics over features: EEG signals analysis. *Computers in Biology and Medicine*, 39(8):733–741.

Ghosh-Dastidar, S., Adeli, H., and Dadmehr, N. (2007). Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE Transactions on Biomedical Engineering*, 54(9):1545–1551.

Guler, I. and Übeyli, E. (2007). Multiclass support vector machines for EEG-signals classification. *IEEE Transactions on Information Technology in Biomedicine*, 11(2):117–126.

Hagan, M. and Menhaj, M. (1994). Training feedforward networks with the marquardt algorithm. *IEEE transactions on Neural Networks*, 5(6):989–993.

Huang, G., Zhu, Q., and Siew, C. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE International Joint Conference on Neural Networks, 2004. Proceedings*, volume 2.

Huang, G., Zhu, Q., and Siew, C. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501.

Iasemidis, L., Shiau, D., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., Carney, P., Prasad, A., Veeramani, B., and Tsakalis, K. (2003). Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, 50(5):616–627.

Jahankhani, P., Kodogiannis, V., and Revett, K. (2006). EEG signal classification using wavelet feature extraction and neural networks. In *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing, 2006. JVA'06*, pages 120–124.

Kalayci, T. and Ozdamar, O. (1995). Wavelet preprocessing for automated neural network detection of EEG spikes. *IEEE Engineering in Medicine and Biology Magazine*, 14(2):160–166.

Kim, J., Shin, H., Lee, Y., and Lee, M. (2007). Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 3257–3260.

Lima, C., Coelho, A., and Chagas, S. (2009). Automatic EEG signal classification for epilepsy diagnosis with relevance vector machines. *Expert Systems With Applications*, 36(6):10054–10059.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4):1084–1093.

Wang, B., Jun, L., Bai, J., Peng, L., Li, Y., and Li, G. (2005). EEG recognition based on multiple types of information by using wavelet packet transform and neural networks. In *Proc. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China*, pages 5377–5380.

Wang, D. and Huang, G. (2005). Protein sequence classification using extreme learning machine. In *2005 IEEE International Joint Conference on Neural Networks, 2005. IJCNN'05. Proceedings*, volume 3.

Zhang, R., Huang, G., Sundararajan, N., and Saratchandran, P. (2007). Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 485–495.