K-Complex Identification In Sleep EEG Using MELM-GRBF Classifier

Seyed Mohammad Reza Noori,Amin Hekmatmanesh, Mohammad Mikaeili
Department of Engineering
Biomedical Engineering Group, Shahed University
Tehran, Iran
smrezanoori@gmail.com

Khosro Sadeghniiat-Haghighi
Occupational Sleep Research Center
Tehran University of Medical Science
Tehran, Iran

***Abstract*— K-complexes like spindles are hallmark patterns of stage 2 sleep. Due to correlation between these patterns and some diseases, it is necessary to develop algorithms to detect them. In this study, a new method is used to detect K-complexes automatically. 10 time-series and chaotic features were used in order to extract the K-complex waves from stage 2 sleep. To use the most effective features, feature space dimension is reduced with Sequential Forward Selection method. The reduced feature space is classified using Generalized Radial Basis Function Extreme Learning Machine (MELM-GRBF) algorithm. GRBFs make the modification of the RBF possible by adjusting a new parameter τ. We're applied this methodology to K-complex classification for the first time. The classifier gives noticeably better results compared to ELM-RBF method for sensitivity and accuracy 61.00 ± 6.6 and 96.15 ± 3.7, respectively.**

Keywords-K-Complex; EEG; Sleep; Classifier, Extreme Learning Machine, Radial Basis Function

I. INTRODUCTION

One of the well-known transient patterns of sleep EEG signals, which can occur in response to auditory stimuli is K-Complex [1, 2]. According to American Academy of Sleep and Medicine (AASM), K-Complex is defined as a transient pattern with the following properties; a sharp negative transition which is immediately followed by a positive component. K-Complex duration should last more than 0.5 sec. The considered frequency is at 8-16 Hz which scored and elicited from the Central channels. Some researchers used slightly different definitions. For instance some researchers used the AASM definition along with minimum amplitude of 75 μV [3] as threshold or some used 100 μV pick to pick amplitude criterion [4]. Also, there are variations in time duration range for K-Complex pattern definition, some investigators used 1 to 5 second [5]. Another time limitation which is used is 1 to 3 second [4,6]. High variation in K-complex time duration in different subjects has caused different definitions.

Different methods are also used to detect K-complexes. Results are usually reported using three well-known statistical explanations including; accuracy, sensitivity and specificity. In short, an electronic system based on digital logic circuit was implemented by Bremer et al. [4], and 68% sensitivity for K-complex detection was achieved. Another researcher used matched filter method with specific limitations for k-complexes, but according to their report the results were not satisfying [7,8]. Rosa and Paiva et al. used stochastic algorithms based on feedback loops of rhythms that is driven by white noise to simulate K-complex waves [5]. Jansen et.al [6] used Neural Networks as a classifier to identify K-complexes from non K-complex waves. Bankman et al. [9] used both feature extraction and Neural Networks as a classifier, the reported results showed better performance. Due to different datasets used on each study, the reported results are not comparable.

In this study, considering the acceptable results of Bankmen et al. [9], we used feature extraction with a MELM-GRBF classifier. Navarro et al. [10] applied this classifier to 15 datasets taken from the UCI repository [11]. This is the first time that such a classifier is utilized for K-complex classification. Features are a combination of some fractal dimensions as chaotic features and some time series features that are acquired based on the K-complex definition.

II. DATASET

12 healthy subjects (average age: 22.4 (22-26 years)) were used to record sleep EEG signals using 10-20 system at Baharloo Hospital, Sleep Clinic, Tehran. Ten channels were used to record the sleep EEG signals which included; C4-M1, C3-M2, F4-M1, F3-M2, Cz-M1, Fz-M1, T3-M2, T4-M1 O2-M1, and O1-M2. Central channels (C3-M2, C4-M1) were used for K-complex analysis. The sampling frequency was set 200Hz. Subjects slept for afternoon naps on two consecutive days which the second day sleep was used to score the K-complexes manually by an expert. The K-complex patterns were scored according to the AASM standard. Finally 581 K-complexes data were extracted from stage 2 sleep. Fig. 1 shows a K-complex pattern extracted from recorded EEG manually.

III. PREPROCESSING

Signals were filtered by a 50Hz Notch filter and a 0.1- 40Hz band-pass filter which was provided by the device, then a 16th order type two 8-16Hz Bandpass Chebyshev filter was applied. IIR filters cause phase distortion which was omitted by applying zero-phase digital filtering by processing the input signals in both forward and reverse directions [12]. Assuming EEG signals to be stationary, signals were segmented using a sliding one second window.

978-1-4799-7418-4/14/$31.00 ©2014 IEEE

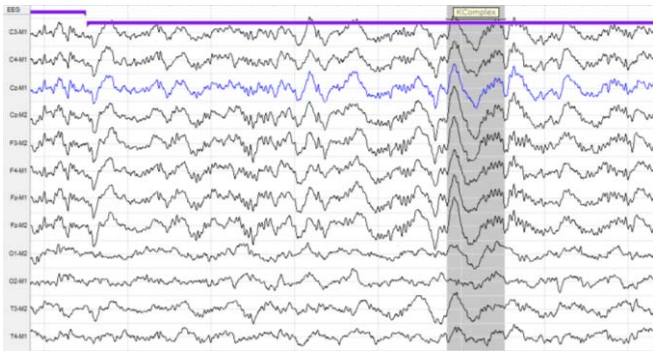

Figure 1. K-complex pattern extracted from recorded EEG manually.

IV. METHODS

A. Time Series Features

Time series traits are the ones which are elicited from K-complex definition. Time series features consist of; minimum amplitude, maximum amplitude, average, 75μv threshold, standard deviation (SD) and energy of signal.

B. Fractal Dimension and nonlinear Features

- Sevcik Fractal Dimension

. Sevcik method is a fractal dimension that is approximated with N samples and is based on time axes normalization and EEG signals [13,14]. To compute the fractal dimension, it is suggested to normalize the metric space as [15]:

$$i' = \frac{i}{N}, q'(i') = (q(i) - q_{min})/(q_{max} - q_{min}) \qquad (1)$$

Where $q'(i')$ is the normalized EEG sample, q(i) is the i^{th} EEG signal sample, q_{max} and q_{min} are maximum and minimum values in EEG signal, respectively. Sevcik fractal dimension is formulated as:

$$FD = 1 + \frac{\ln(L)}{\ln(2(N-1))} \qquad (2)$$

Where L is the normalization length of the signal.

- Higuchi Fractal Dimension

Higuchi's model is based on a different measure for the length of a signal which is slided. q_m^k is a new time series which is formulated as below [15]:

$$q_m^n = \left\{q(m), q(m+n), q(m+2n), \dots, q\left(m + \left[\frac{N-m}{n}\right]n\right)\right\} m = 1,2,3,\dots n \qquad (3)$$

Where n is the step parameter, m is an initial value and [.] is the bracket operator. Finally, the length of the new time series (Lm(n)) is formulated as:

$$L_m(n) = \frac{\{\sum_{i=1}^{\lfloor\frac{N-m}{n}\rfloor} |q(m+in) - q(m+(i-1)n)|.\frac{N-1}{n}\}}{\left[\frac{N-m}{n}\right]n} \qquad (4)$$

$\frac{(N-1)}{\left[\frac{N-m}{n}\right].n}$ is employed to normalize the values and $L_m(n)$ is used to calculate the average length of the time series which means, that $L(n) = \frac{1}{n}\sum_{m=1}^{n} L_m(n)$. $L(n)$ is equivalent to n-FD, where FD is fractal dimension value. Finally, the Higuchi fractal dimension is obtained by using the slop of $Log(L(n))$ versus $Log(1/n)$ [13].

C. Entropy

A nonlinear chaotic parameter which is based on phase space and considered as a nonlinear quantification is Entropy. This quantifier depicts the complexity of a system and quantifies the system tendency to chaos. Principals of this method is based on the trend of trajectory variations. More regularity means system is driving to less complexity and vice versa. In the other hand, higher complexity means system's more tendency to chaos. For analyzing the entropy property, phase space should be reconstructed and let the trajectory grow in the reconstructed phase space as much as it can. This phase space slides in to cells that are marked as a(n). After a specific lag the trajectory will migrate from a(0) cell to the next which is called a(1). All the cells are signed to construct a time series, a(0), a(1),…, a(N). This procedure is an ongoing process on the other trajectories. The initial values are not identical; hence various time series are constructed. At last, the Entropy is calculated as [16]:

$$z_n = -\sum_i S(i)\ln S(i) \qquad (5)$$

$S(i)$ is the number of appearance of the trajectory in the i-th cell.

V. FEATURE SELECTION

Among 10 features, Sequential Forward Selection (SFS) method was employed to select the best features. SFS method is a greedy procedure which starts with empty set as the first level. In this process, a new feature is added to the feature space and combined with the other features to optimize the objective function in every stage sequentially. As an evaluation function in SFS method, the difference of total Mahalanobis distances between first class data and the center of the second class is calculated. On the contrary, the difference of total Mahalanobis distances between second class data and the center of the first class is calculated. The higher distance results in higher weight assigned to the feature. Therefore, the features which can maximize the objective function, will be chosen as the most valuable features. By employing SFS method and considering the results, five features are selected as the most qualified features. These are; Sevcik fractal dimension, Entropy, Katz fractal dimension [17, 18], mean and Higuchi fractal dimension, respectively.

VI. CLASSIFIER

Our approach is focused on a new kernel which is based on probability distribution function of data in feature space. The kernel parameters are calculated in this section.

RBF kernels are among the most popular kernels used in kernel based classifiers. But they have certain limitations [19-21]. In order to avoid those limitations, a more generalized parametric model should be defined to have better description of

the statistical behavior of distributions. To this end, the modified density function is presented as follow:

$$p(X;C,r,\tau)=\frac{\tau}{2r\Gamma(\frac{1}{\tau})}\exp(-\frac{X-C}{r^{\tau}}) \quad (6)$$

Where C is the mean, $r>0$ and $\tau>0$ are width and shape parameters of the distribution, respectively and, Γ is the Gamma function.

Parameter r controls the descend rate of the density function and is proportional to the standard deviation, expressed with Eq. 7.

$$r=\sigma\sqrt{\frac{\Gamma(\frac{1}{\tau})}{\Gamma(\frac{3}{\tau})}} \quad (7)$$

The described Generalized Gaussian Distribution (GGD) has more extensive capability for modeling and representing statistical behavior than Gaussian Distribution (GD). In GGD, parameter r models the peak width. The new model is stable and is able to describe various statistical behaviors. The new kernel is presented in Eq. 8

$$\emptyset_j(\mathbf{X};\mathbf{C_j},r_j,\tau_j)=\exp(-\frac{\|\mathbf{X}-\boldsymbol{C_j}\|^{\tau_j}}{r_j^{\tau_j}}) \quad (8)$$

Where $\boldsymbol{X_i}=(x_{i1},\dots,x_{ik})^T$ is the input vector, K is the number of inputs, r_j is the GRBF width, $\boldsymbol{C_j}=(c_{j1},\dots c_{jk})$ is the center and τ_j is the i-th GRBF Shape parameter.

Fig. 2 shows the variations of the basis function shape with parameter τ. This basis function makes a better match between the kernel and data distributions in the feature space. By adjusting parameter τ, the concavity and convexity of the kernel could be controlled, around its center [22].

A new classifier based on the new kernel is designed with following properties: the hidden layer is one Single Layer Feed Forward Neural Network (SLFN) with m neurons. The output layer neuron is set a linear function to guarantee linear training for parameters in the output which is explained in Eq. (11).

Where β is the training weight matrix, T is the desired output matrix, and H is the output of the hidden layer which is explained in the equations (9), (10) and (11). This training procedure is developed for SLFNs by Huang et al. [20-22] and was named as Extreme Learning Machine (ELM). In this study, the ELM kernel is modified to be able to choose input weights randomly, this classifier is named MELM-GRBF [10].

$$H=(h_1,h_2,\dots,h_m)=\begin{pmatrix}\emptyset_1(\boldsymbol{X}_1;\boldsymbol{C}_1,r_1,\tau_1) & \dots & \emptyset_m(\boldsymbol{X}_1;\boldsymbol{C}_m,r_m,\tau_m)\\ \dots & \dots & \dots\\ \emptyset_1(\boldsymbol{X}_n;\boldsymbol{C}_1,r_1,\tau_1) & \dots & \emptyset_m(\boldsymbol{X}_n;\boldsymbol{C}_m,r_m,\tau_m)\end{pmatrix}_{n\times m} \quad (9)$$

$$\boldsymbol{T}=(t_1,t_2,\dots,t_n)_{n\times J}^{T} \quad (10)$$

$$\boldsymbol{\beta}=(\beta_1,\beta_2,\dots,\beta_m)_{m\times 1} \quad (11)$$

In this system, training parameters are modifiedto minimize a squared error function, according to Eq. 12 [23].

$$SE=\sum_{i=1}^{n}(o_i-t_i)^2 \quad (12)$$

The SLFN configuration with generalized radial basis function is depicted in Fig. 3. Since our classification problem is a two class (K-complex and non K-complex) problem, a single neuron would be sufficient in the output layer and the classification can be done using a threshold on the output.

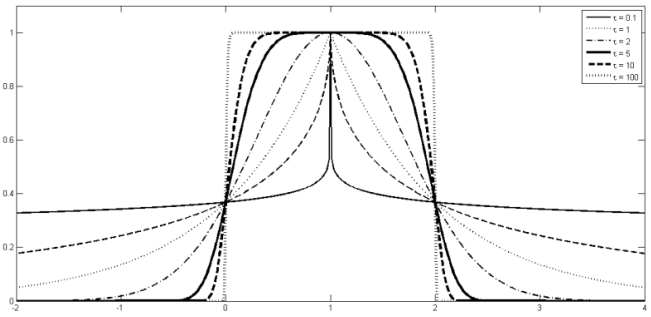

Figure 2. Effect of the parameter τ on the kernel behavior with c =1 and r =1

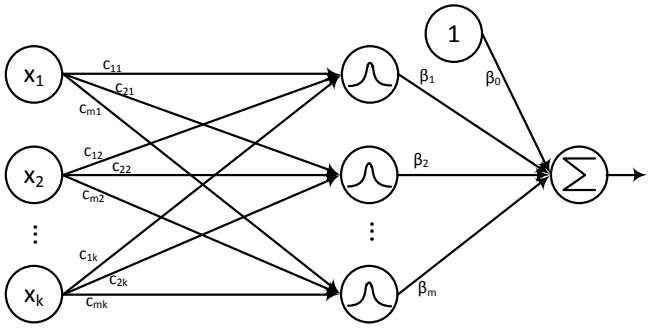

Figure 3. Scheme of generalized radial basis function used in SLFN

VII. CLASSIFIER PARAMETERS

In the proposed SLFN, the single neuron in the output layer, makes a weighted sum of the outputs of the hidden neurons (Fig. 1). In this network, centers of the hidden layer neurons are selected from the input patterns, randomly. For choosing r in ELM-RBF, different methods based on maximum distance among input neurons, number of hidden layer neurons, K-nearest neighbors and… are used. In MELM-GRBF method, width (r) and shape (τ) parameters are sensitive to the distribution of distances.

In order to calculate kernel parameters, minimum and maximum values of the hidden layer should be determined. To this end, the minimum (S) and the maximum distance (L) in the feature space are used to produce the lowest and highest value of the kernel, respectively. These two parameters are related to λ, which are defied in Eq. 13 and Eq. 14.

$$\exp\left(-\left(\frac{L}{\mathrm{r}}\right)^{\tau}\right)=\lambda \qquad (13)$$

$$\exp\left(-\left(\frac{S}{\mathrm{r}}\right)^{\tau}\right)=1-\lambda \qquad (14)$$

For calculating (L) for each hidden layer neuron, according to Eq. 15, select minimum distance between the neuron and all its neighbor neurons. Since parameter τdetermines the tail of the kernel, so S values for all neurons could be selected equally and be calculated from Eq. 16.

$$S_i=\|c_i-c_j\| \qquad (15)$$

Which j-th index is the nearest hidden layer neuron to the i-th neuron.

$$S=\sqrt{\delta^2\times k} \qquad (16)$$

δ is a very small distance in each dimension and k is the number of the observations. In the employed classifier the δ is chosen as 0.1, experimentally. According to the distribution of distances in the feature space λ is obtained as 0.05, experimentally. By choosing the free parameter, proper values for τ and r in each kernel of hidden layer neuron is calculated as in Eq. 17 and Eq. 18.

$$\tau=\frac{\ln\left(\frac{\ln(\lambda)}{\ln(1-\lambda)}\right)}{\ln(\frac{L}{S})} \qquad (17)$$

$$\mathrm{r}=\frac{L}{(-\ln(\lambda))^{1/\tau}} \qquad (18)$$

In short, in order to train the presented classifier, the number of hidden layer neurons (m) and λ values should first be determined. Then, neuron centers should be selected among the input learning patterns, randomly. Using Eq. 15-18, S, L, τ and r parameters are then calculated, respectively. Therefore, the kernels are configured and the outputs of the hidden layer (H) is calculated. Using H and T matrixes, the training parameters can be obtained as follows (Eq. 19).

$$\hat{\beta}=[(H^TH)^{-1}H^T]\times T \qquad (19)$$

VIII. Results

In this part the results of the classifier for classifying EEG signals with K-Complex patterns from signals without K-Complex patterns using ELM-RBF and MELM-GRBF methods is reported. We used Matlab 2013, on a Core i5 Intel CPU, 2.53GHz PC with 4GB Ram. Since the number of free parameters in MELM-GRBF is more than ELM-RBF, the new method is more time consuming and costly.

According to the training procedure, first step is to select m and γ. After evaluating several values for m including: 10, 15, 20, … and 60, we selected 40 as the best one. In addition, several different values for γ such as 0.01, 0.02,… and 0.2, was tested and 0.05 was the best. Our algorithm calculates parameters τ and r in the extent of 1.4 to 2.8 and 0.5 to 4.1, respectively.

In this classifier, 70% of the data (814 observations) was allocated to the training set and the rest of the dataset (30% or 348 observations) was used for testing set. The results of MELM-GRBF and ELM-RBF are presented in tables (1) and (2), respectively. Accuracy, sensitivity and their Standard Deviation (SD) are reported in percentage. These results are after 15 times train and test and then averaging. It's clear that in MELM-GRBF accuracy and sensitivity are improved for 2.53% and 7.04%, respectively. In addition, standard deviation of the accuracy and sensitivity have decreased 1.2% and increased 1.2%, respectively. These results proved MELM-GRBF matched better to K-complex patterns in comparison to ELM-RBF methodology.

TABLE I EVERY COLUMN IS THE MELM- GRBF RESULTS AVERAGE (AVE%) AFTER 15 TIMES TRAIN AND TEST.

Statistical Measurements	Value in percentage
Mean Accuracy	96.15
Mean Accuracy SD	3.7
Mean Sensitivity	61.00
Mean Sensitivity SD	6.6

TABLE II EVERY COLUMN IS THE ELM- RBF RESULTS AVERAGE (AVE%) AFTER 15 TIMES TRAIN AND TEST.

Statistical Measurements	Value in percentage
Mean Accuracy	93.62
Mean Accuracy SD	4.9
Mean Sensitivity	52.94
Mean Sensitivity SD	5.4

IX. Discussion

Sequential Forward Selection (SFS) method indicated that the information extracted from the nonlinear and fractal dimension features gives more valuable information than the time features. This can be due to the nonlinear behavior of brain as a signal generator. Since visual distinguishing of the nonlinear and dimensional information of K-complexes is impossible, the importance of this automatic K-Complex detection algorithm is doubled.

Presented method is used in order to compensate some of the RBF limitations using a free parameter (τ). The parameter τ would adjust the convexity and concavity of the kernel tail. Hence, τ can match betterto the distribution of distances in the feature space. Obviously, better matching and more compatibility with the interested K-complex patterns drives the classifier to be more powerful and enhances the statistical evaluations. For training, modified ELM method is used which is proposed by Haung et al. [28] for single hidden layer feed forward neural networks. According to the applied modification, the method is named MELM-GRBF. This kind of classifier has not been used for K-complex classification, before. In this new method only two initial values namely the number of hidden layer neurons and γ must be selected. As shown in the table. 1 and 2, MELM-GRBF has improved the performance of the K-Complex detection results.

REFERENCES

[1] M. Kryger, T. Roth and W. Dement, Principles and practice of sleep medicine, 4th edition, Elsevier Saunders, ISBN: 1416003207, 2005.

[2] A. Rechtschaffen ,and A. Kales, "A manual of standardized terminology and scoring system for sleep stages in human subjects," U.S. Government Printing Office, Washinghton, DC; 1968.

[3] A. Rodenbeck, R. Binder, P. Geisler, H. Danker-Hopfe, R. Lund, F. Raschke, HG. Wee and H.Schulz, "A review of sleep EEG patterns. Part I: A compilation of amended rules for their visual recognition according to Rechtschaffen and Kales," Somnologie, Vol 10(4), pp 159-175, 2006

[4] G. Bremer, JR. Smith and I. Karacan, "Automatic detection of the Kcomplex in sleep electroencephalograms," IEEE Trans. Biomed. Eng., vol 17(4), pp 314-23, 1970.

[5] AC. Da Rosa, B. Kemp, T. Paiva Lopes, FH.Da Silva and HA. Kamphuisen, "A model-based detector of vertex waves and K complexes in sleep electroencephalogram ," Electroencephalogr. Clin. Neurophysiol.; vol 78(1), pp71-79 1991

[6] Jansen BH., "Artificial neural nets for K-complex detection.," IEEE Eng Med Biol Mag., vol 9(3),pp 50-52, 1990.

[7] O. Sheriff, B. Pagnrek, S. Mamouhd and R. Broughton, "Automatic detection of K-complex in the sleep EEG," Int. Electrical and Electronic Conf. and Exp., vol 81,1977

[8] M. Woertz, T. Miazhynskaia, P. Anderer and G. Dorffner, "Automatic K-complex detection: comparison of two different approaches," Abstracts of the ESRS, JSR 13(Suppl.1),1, Prague 2004.

[9] N. Bankman, VG. Sigillito, R. A. Wise and P. L. Smith, "Featurebased detection of the K-complex wave in the human electroencephalogram using neural networks," IEEE transactions on biomedical engineering, vol. 39, no12, pp. 1305-1310, 1992.

[10] F. Fernández-Navarro, F. Hervás-Martínez, C. Sanchez-Monedero, J. Gutiérrez, P. Antonio, MELM-GRBF: A modified version of the extreme learning machine for generalized radial basis function neural networks, Neurocomputing 74 (2011) 2502-2510.

[11] A. Asuncion, D. Newman, UCI Machine Learning Repository, 2007. URL </http://www.ics.uci.edu/~mlearn/MLRepository.html>.

[12] Oppenheim, A.V., and R.W. Schafer, "Discrete-Time Signal Processing," Prentice-Hall, 1989, pp.284–285.

[13] Sevcik C (1998) A procedure to estimate the fractal dimension of waveforms. Complexity nt'l. 5.
http://www.complexity.org.au/ci/vol05/sevcik/sevcik.html

[14] B. S. Raghavendra, and D. D. Narayana, "Computing fractal dimension of signals using multiresolution box-counting method," International Journal of Information and Mathematical Sciences 6, no. 1, 2010, 50-65.

[15] A. Allahverdy, Ali Moti Nasrabadi, Mohammad Reza Mohammadi, "Detecting ADHD Children Using Symbolic Dynamic of Nonlinear Features of EEG," Electrical Engineering (ICEE), 19th Iranian Conference, 2011

[16] R. Hilborn, "Chaos and nonlinear dynamics: an introduction for scientists and engineers," Oxford university press, 2000.

[17] Katz MJ (1988) Fractals and the analysis of waveforms. Comp Biol Med 18:145–156.

[18] Esteller, Rosana, George Vachtsevanos, Javier Echauz, and Brian Litt. "A comparison of waveform fractal dimension algorithms." Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on 48, no. 2 (2001): 177-183.

[19] F. Scarselli, A. Chung Tsoi, Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results, Neural Networks 11 (1) (1998) 15–37.

[20] F. Fernandez-Navarro, C. Hervas-Martınez, P.A. Gutierrez, M. Carboreno, Evolutionary q-gaussian radial basis functions neural networks for multi-classification, Neural Networks, in press. URL /http://dx.doi.org/10.1016/ j.neunet.2011.03.014S.

[21] F. Fernandez-Navarro, C. Hervas-Martınez, M. Cruz, P.A. Gutierrez, A. Valero, Evolutionary q-Gaussian radial basis function neural network to determine the microbial grow/thno growth interface of Staphylococcus aureus, Applied Soft Computing 11 (3) (2011) 3012–3020.

[22] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: IEEE International Conference on Neural Networks—Conference Proceedings, vol. 2, 2004, pp. 985–990.

[23] G. Feng, G.B. Huang, Q. Lin, R. Gay, Error minimized extreme learning machine with growth of hidden nodes and incremental learning, IEEE Transactions on Neural Networks 20 (8) (2009) 1352–1357.