Sleep Spindle Detection Using Modified Extreme Learning Machine Generalized Radial Basis Function Method

Amin Hekmatmanesh, Seyed Mohammad Reza Noori, Mohammad Mikaili
Department of Engineering, Biomedical Engineering Group, Shahed University, Tehran, Iran
amin_hm2002@yahoo.com

Abstract—**Spindles are the main indicators of stage two sleep. Manual detection of spindles is tedious and time-consuming, hence many attempts have been performed to automatically detect them. Scientists have found that spindles are related to diseases such as Alzheimer's and Schizophrenia. Therefore, designing algorithms to extract certain spindle features to diagnose them with high accuracy is valuable. In this study, new chaotic features, along with a new training and classification algorithm has been used to detect spindles: an algorithm called MELM-GRBF (Modified Extreme Learning Machine). GBRF (Generalized Radial Basis Function) can update the radial basis function (RBF) using a parameter (τ), and the training algorithm is MELM. GRBF centers are selected randomly from the training data. The width and the parameter τ of GRBF are achieved considering two limitations: locality and coverage.**

In order to have the most similarity between spindle and non-spindle segments and to demonstrate the capability of these (chaotic features and classifier) algorithms, non-spindles are extracted immediately before spindles occur for one second in sleep stage two. After presenting chaotic (Higuchi Fractal Dimension, Katz Fractal Dimension, Sevcik Fractal Dimension, Largest Lyapunov and Entropy) and time series features, the ELM-RBF and MELM-GRBF statistical results are compared. The Best average results of MELM-GRBF classifier after 15 training trials were, 93.10%, 90.34% and 95.90% for accuracy, sensitivity and specificity, respectively and the variances were 0.78, 1.61 and 1.16, respectively. While the ELM-RBF results for accuracy, sensitivity and specificity were 91.06%, 85.83% and 96.32% with 1.64, 2.39 and 1.18 variance, respectively.

Keywords-Spindle; Generalized basis function classifier; Radial basis function; Modified extreme learning machine; Chaotic features, Fractal dimensions

I. INTRODUCTION

Loomis et al. first described spindles as waves driven and distributed from thalamus to cortex (with a frequency of 12 to 14Hz) [1]. Spindles became an indicator for some diseases like Alzheimer's [2], Rett Syndrome [3] and Schizophrenia [4]. Researchers are interested in finding the causes of EEG changes when spindles are generated. First step in considering spindle changes is detecting the pattern automatically to get rid of the tiresome task of manual detection.

Many methods and features are used to detect spindles automatically. Researchers mostly use the visual characteristics of spindles in their manual detection. In short, Barros et al. [5], Görür et al. [6], Huuppone et al. [7], and Ahmed et al. [8] used various methods to extract specific features from EEG signals in order to detect spindles, other researchers like Babadi et al. [9] used creative methods to detect spindles. Artificial neural networks (ANN) have a vast application for classifying feature spaces. In this study we used a modified ANN i.e. Generalized Radial Basis Function (GRBF) [10] to detect spindles. This method is based on the Radial Basis Function Neural Network (RBFNN) method. Generally, RBFNNs employ hyper-ellipses to classify the feature space [11]. RBFs use distance to estimate the result. Gaussian RBFs are based on the Gaussian density function characterized by a center position and a variance parameter; these functions have the highest output when the input is closest to the center position [10]. In kernel based RBFNNs, there must be a counterbalance between coverage and locality. Coverage means, the superposition of RBFs in all patterns, it must have a high value (close to one) in the data set. Locality is a high value (near one) for patterns that are close to the RBF and low values (close to zero) for patterns that are not located in the region where the RBF is centered [10]. Two methods of learning can be used in RBFNNs; the first is calculation of controlling parameters and weights of hidden layer (in a two-step process, clustering in an unsupervised mode [12] or vector quantization method [13]). The second method is based on gradient descent that optimizes all of the parameters in a supervised model [14-16]. The gradient descend algorithm has limitations such as learning rate dependency, local minimums and number of iterative steps. Due to these limitations Haung et al. [17,18] proposed an algorithm, called Extreme Learning Machine (ELM), in which ELM is an effective training algorithm for a single hidden layer feed forward neural network [17,18].

II. DATASET

Sleep EEG signals of 12 healthy (6 men, 6 women, average age: 22.4 (22-25 years)) subjects were recorded according to 10-20 standard at Baharloo Hospital, Sleep Clinic, Tehran. The recording device was EMBLA (N7000 version) with 10 channels C3-M2, C4-M1, Cz-M1, F4-M1, F3-M2, Fz-M1, T3-M2, T4-M1, O2-M1, and O1-M2 with reference to both mastoids. Only C3-M2, C4-M1 channels were used for processing to detect spindles. The sampling frequency was adjusted to 200Hz. Subjects were advised to avoid caffeine and nicotine at least four hour prior to the test. Subjects took

978-1-4799-4409-5/14/$31.00 ©2014 IEEE

afternoon naps on two consecutive days. The first day nap was for adaptation and on the second day EEG signals were recorded. These signals were manually scored by an expert at the sleep clinic. Stages were scored according to the American Academy Sleep Medicine (AASM) standard. The expert scored 720 spindles in stage two sleep.

III. PROCEDURE

Signals were passed through, a 50Hz Notch filter, a 40Hz low-pass filter and a 0.3Hz high-pass filter, then an IIR Chebyshev type two filter of order 16 was employed as band pass filter between 8 and 16Hz. The high phase distortion resulted from this filter was omitted by performing zero-phase digital filtering by processing the input data in both forward and reverse directions [19]. Considering EEG signals as stationary, signals were segmented using a sliding one second window. Spindle signals were selected from stage two and non-spindle signals were chosen one second before each spindle. The reason we selected non-spindles at this positions was primarily the most similarity between spindle and non-spindle signals to show the abilities of chaotic features and secondly to ensure that the non-spindle signals were also in stage two. After extracting features from the segments the feature space was classified by MELM-GRBF classifier.

IV. METHODS

A. Time Series Features

Time series features are ordinary features that can be perceived by watching through the signal. They consist of; minimum, maximum, standard deviation (SD), power, zero crossing and slope changes of the signals.

B. Chaotic Features

1) Higuchi Fractal Dimension

Higuchi's method is based on a different measure of length of a signal to be analyzed. z_m^k is a new time series to be constructed as below [20]:

$$z_m^k = \{z(m), z(m+k), z(m+2k), \dots, z(m+[\frac{N-m}{k}]k)\} \quad (1)$$

$$m = 1,2,3,\dots,k$$

Where m is an initial point, k is the step parameter and [] sign is the bracket function. The length of the new time series ($L_m(k)$) is defined as:

$$L_m(k) = \frac{\{\sum_{i=1}^{[\frac{N-m}{k}]} |z(m+ik) - z(m+(i-1)k)| . \frac{(\mathrm{N}-1)}{[\frac{\mathrm{N}-\mathrm{m}}{\mathrm{k}}].\mathrm{k}}\}}{k} \quad (2)$$

Where $\frac{(N-1)}{[N-m].k}$ is the normalization factor. The average length of time series is measured by averaging $L_m(k)$ which means, $L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k)$. L(k) is equivalent to k^{-FD}, where FD is fractal dimension value. Higuchi fractal dimension is calculated by estimating the slop of Log(L(k)) versus Log(1/k) line a line drawn [21].

2) Sevcik Fractal Dimension

Sevcik [21] proved that fractal dimension can be approximated with N signal samples. Sevcik method is based on EEG signals and time axes normalization [21,22]. To calculate Sevcik fractal dimension, it is proposed to normalize the metric space by [20]:

$$i' = \frac{i}{N}, Z'(i') = (Z(i) - Z_{\min}) / (z_{\max} - z_{\min}) \quad (3)$$

Where $Z(i)$ is the EEG signal and $Z'(i')$ is the normalized EEG signal of the i_{th} value in the signal, $Z_{\max}$ and $Z_{\min}$ maximum and minimum values in EEG signal, respectively. Sevcik fractal dimension is formulated as:

$$FD = 1 + \frac{\ln(L)}{\ln(2(N-1))} \quad (4)$$

Where L is the length of normalized signal.

3) Largest Lyapunove Exponent

Wolf proposed a method to calculate the Largest Lyapunove Exponent (LLE) [20]. In this method, first phase space reconstruction has to be calculated then the nearest neighbor for the first embedding vector. Distance between two trajectories has to be calculated by growing the trajectory, the distance will be calculated again with a specific time delay for several times. The LLE is calculated as bellow [20]:

$$L_1 = \frac{1}{t_k - t_0}\sum_{i}^{k} \ln \frac{L'(t_i)}{L(t_i - 1)} \quad (5)$$

Where L_0 is the initial and L_i is the new distance. K is the number of steps in which distances will be measured.

4) Entropy

Entropy is the other method of chaotic quantification which is based on phase space. This quantifier demonstrates the level of system complexity and system tendency to chaos. Complexity is based on the trajectory regularity, if the regularity increased, the complexity will decrease and vice versa. High complexity shows high system tendency towards chaos.

After reconstructing the phase space, let the trajectory grows as much as it wants. Then this phase space is divided in to cells that are labeled as b(n). b(0) is the first label. After delay d the trajectory will be in the next cell witch is named b(1), all the cells are labeled to construct a time series, b(0), b(1),…, b(N). All this process is repeated on the other trajectory. The initial conditions are not the same, so various time series are constructed. Entropy is calculated as S_n [23]:

$$S_n = -\sum_i q(i)\ln q(i) \qquad (6)$$

Where q(i) shows the number of times that trajectory appears in a call.

V. MELM-GRBF CLASSIFIER

This method has proposed by Francisco et al. in 2011. Generalization Gaussian Density (GGD) is the GRBF kernel that is calculated by (7) and (8) as bellow [10]:

$$p(X,c,r,\tau) = \frac{\tau}{2r\Gamma(1/\tau)}\exp\left(-\frac{\|X-c\|}{r^{\tau}}\right) \qquad (7)$$

where c, r>0 and τ >0 are center, variance and the shape of distribution function, respectively. The variance r demonstrates the width of distribution and calculated as (8):

$$r = \sigma\sqrt{\frac{\Gamma(1/\tau)}{\Gamma(3/\tau)}} \qquad (8)$$

Where σ is the normal SD and $\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$, for z>0.

Value of τ shows the descent rate of density function. $\tau = 2$ achieves Gaussian density and $\tau = 1$ results is a Laplacian distribution density. If $\tau \to 0$ distribution tends to an impulse and if $\tau \to \infty$ the distribution becomes uniform (Fig.1). GGD is more stable and flexible than GD because fewer parameters are needed to calculate it. GGD Model is calculated for a k dimensional input space as bellow:

$$\varnothing_j\left(X;c_j,r_j,\tau_j\right) = \exp\left(-\frac{X-c_j^{\tau_j}}{r_j^{\tau_j}}\right) \qquad (9)$$

Where $X_i = \left(x_{i1},\ldots,x_{ik}\right)^T$ is the input vector, K is the number of inputs, r_j is the GRBF width, $c_j = \left(c_{j1},\ldots,c_{jk}\right)^T$ is the center and τ_j is the i_{th} GRBF. The GGD function gives better matching between the shape of the kernel and the distribution of the distances, because of the τ parameter trigger convexity around the point where the distance is the radial of the kernel (Fig.2) [10]. It is considerable that in this method constant value of λ , that determines the minimum output and the optimum value for it experimentally achieved 0.05 [10].

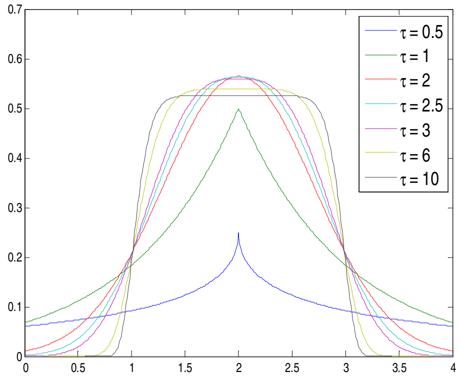

Fig.1 Probability density function of the GGD for different values of τ (c=2, r=1).

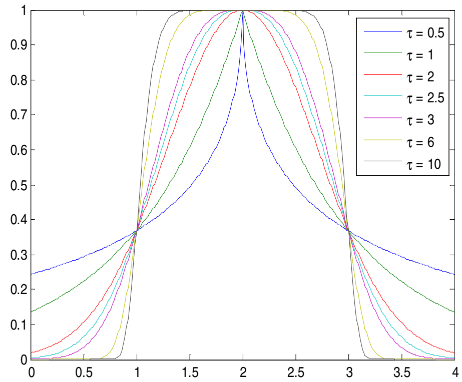

Fig.2 Demonstrates the radial unit activation for various values of τ in the GRBF model (c=2 , r=1).

VI. STATISTICAL MEASURES

We report our results based on three parameters; Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe) that are proposed in previous papers [24, 25].

VII. RESULTS

In this study we have repeated the MELM-GRBF, Training-test sequence for 15 times. and the results are recorded as seen in Table II. The input neurons were 15, the RBF and MELM-GRBF kernels were GD and GGD, respectively. In the ELM-RBF the parameters were τ=2 and r=2.5, while the parameters τ and r achieved in MELM-GRBF were 1.6 and 1.9, respectively. The average GRBF accuracy value and SDs for 15 times training in its best state was 93.10%, sensitivity and specificity were 90.34% and 95.90%, respectively (Table I). Average results for conventional RBF with the similar inputs, were 91.06% for accuracy and sensitivity and specificity were 85.83% and 96.32%, respectively (Table 2). The data was processed by MELM-GRBF algorithm with MATLAB 2012a software.

I. DISCUSSION

Usually two methods are recommended to improve spindle detection; using specific features and using new classifications. Because of intrinsic complexity and nonlinearity of EEG signals we expected the chaotic features to help improve spindle detection rates. Using new methods to classify the data is another useful method to improve our accuracy. Single hidden layer feed forward neural network can learn distinct observations with small errors by choosing the activation function properly [26-28]. We used a powerful training algorithm for single hidden layer feed forward neural network named ELM which is proposed by Haung et al. [17,18]. The Modified ELM algorithm, named MELM, does not have the training gradient descend problems [10]. MELM-GRBF is the GRBF classifier trained by MELM. MELM-GRBF is more robust and accurate in comparison to the ELM-RBF method. As exhibited in the tables the MELM-GRBF method, the SDs average variances are more reliable than the RBF (compare table I and table II).

Table I. Every column is the MELM- GRBF results average (ave%) after 15 times train and test.

Statistical Measurements	**Value in percentage**			
Acc ave	92.94	93.04	93.17	**93.10**
Acc SD ave	1.01	1.02	1.08	**0.78**
Sen ave	90.53	90.13	90.05	**90.34**
Sen SD ave	2.01	1.79	1.65	**1.61**
Spe ave	95.37	95.85	96.25	**95.90**
Spe SD	1.36	0.85	1.21	**1.16**

Table II. Every column is the ELM- RBF results average (ave%) after 15 times train and test.

Statistical Measurements	**Value in percentage**			
Acc ave	89.92	90.15	**91.06**	90.92
Acc SD ave	1.46	1.70	**1.64**	1.42
Sen ave	84.03	84.40	**85.83**	85.1
Sen SD	2.34	3.02	**2.39**	3.04
Spe ave	95.86	96.03	**96.32**	96.63
Spe SD ave	1.16	1.05	**1.18**	1.72

In short, as illustrated in the tables, GRBF is more robust and gives better statistical results with lower variance. In terms of mathematics, centers and radius in ELM-RBF algorithm are selected randomly among training data, while, in the MELM-GRBF only centers are selected among them. Also, radius and τ values in GRBF are calculated using equations [10] that fulfills the coverage and locality limitations. GD in GRBF is parameterized by τ and results in GGD. Since, GGD has various shapes with various τ's, this generates different distribution forms, Fig.1 (like impulse, uniform, Laplasian, Gaussian). The other features that is improved in ELM-RBF are: 1. Flexibility in modeling the vast variety of statistical behaviors 2. Fewer parameter to calculate 3. Less time needed to train and test the data [10]. In brief, GGD needs only one more parameter in comparison to the GD, and it makes the prediction of various statistical distributions possible. According to the described causes, in GRBF some ELM-RBF limitations are eliminated and the detection gives us better results with low variances in comparison to the ELM-RBF with GD kernel in spindle detection.

REFERENCES

[1] L. De Gennaro and M. Ferrara, "Sleep spindles: an overview," Sleep Medicine Reviews, 2003,Vol. 7, No. 5, pp. 423-440.

[2] J. Dauwels, F. Vialatte, and A. Cichocki. "Diagnosis of alzheimers disease from EEG signals: Where are we standing?," Current Alzheimer Research 7, no. 6, 2010, 487-505.

[3] E. A. Garofalo, I. Drury, and G. W. Goldstein. "EEG abnormalities aid diagnosis of Rett syndrome," Pediatric neurology 4, no. 6 ,1988, 350-353.

[4] E. J. Wamsley, M. A. Tucker, A. K. Shinn, K. E. Ono, S. K. McKinley, A. V. Ely, D. C. Goff, R. Stickgold, and D. S. Manoach, "Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation?," Biological psychiatry 71, no. 2, 2012, 154-161.

[5] A. K. Barros, R. Rosipal, M. Girolami, G. Dorffner, and N. Ohnishi, "Extraction of sleep-spindles from the electroencephalogram (EEG)," in Proc. Artif. Neural Netw. Med. Biol. (ANNIMAB-1), vol. 7, 2000, pp. 125–130.

[6] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen, and K. Leblebicioglu, "Sleep spindles detection using short time Fourier transform and neural networks," in Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on, 2002, pp. 1631-1636. (Pubitemid 34647874).

[7] E. Huupponen, G. Gomez-Herrero, A. Saastamoinen, A. Varri, J. Hasan, S. L. Himanen, "Development and comparison of four sleep spindle detection methods, Artificial Intelligent In Medicine," Artificial Intelligence in Medicine (2007) 40, 157-170.

[8] B. Ahmed, A. Redissi, and R. Tafreshi, "An automatic sleep spindle detector based on wavelets and the teager energy operator," in Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, 2009, pp. 2596-2599.

[9] B. Babadi, S. M. McKinney, V. Tarokh, and J. M. Ellenbogen. "DiBa: A Data-Driven Bayesian Algorithm for Sleep Spindle Detection," Biomedical Engineering, IEEE Transactions on 59, no. 2, 2012, 483-493.

[10] F. Fernandez-Navarro, C. Hervas-Martinez, and J. Sanchez-Monedero, "MELM-GRBF: A modified version of the extreme learning machine for generalized radial basis function neural networks", Neurocomputing 74, 2011, 2502 –2510

[11] P.A. Gutierrez, C. H. Martı´nez, M. Carbonero, and J.C. Fernandez, "Combined projection and kernel basis functions for classification in evolutionary neural networks," Neurocomputing 72, 13–15, 2009, 2731–2742

[12] Z. Uykan, and C. Guzelis, "Input–output clustering for determining centers of radial basis function network," Proceedings of the 1997 European Con-ference on Circuit Theory and Design, vol. 2, Technical University of Budapest, European Circuit Society, Hungary, 1997, pp. 435–439.

[13] M. Vogt, "Combination of radial basis function neural networks with optimized learning vector quantization," in: Proceedings of IEEE International Conference on Neural Networks (ICNN), vol. 3, San Francisco, CA, 1993, pp. 1841–1846.

[14] A. Staiano, R. Tagliaferri, and W. Pedrycz, "Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering," Neurocomputing 69 (13–15), 2006, 1570–1581.

[15] L. Xu, "RBF nets, mixture experts, and Bayesian Ying-Yang learning," Neuro-computing 19 (1–3), 1998, 223–257.

[16] G. Bugmann, "Normalized Gaussian radial basis function networks," Neuro-computing 20 (1–3), 1998, 97–110.

[17] G. Feng, G.B. Huang, Q. Lin, and 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.

[18] J. Cao, Z. Lin, and G.B. Huang, “Composite function wavelet neural networks with extreme learning machine,” Neurocomputing 73 (7–9) (2010) 1405–1416

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

[20] 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

[21] 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

[22] 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.

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

[24] J. C. da Costa, M. D. Ortigueira, A. Batista, and T. S. Paiva. “Threshold choice for automatic spindle detection,” In Systems, Signals and Image Processing (IWSSIP), 19th International Conference on, pp. 197-200. IEEE, 2012.

[25] S. Devuyst, T. Dutoit, J. F. Didier, F. Meers, E. Stanus, P. Stenuit, and M. Kerkhofs. “Automatic sleep spindle detection in patients with sleep disorders,” In Engineering in Medicine and Biology Society, 28th Annual International Conference of the IEEE, pp. 3883-3886. IEEE, 2006.

[26] G.B. Huang, Q. Zhu, and C. Siew, “Extreme learning machine: theory and applications,” Neurocomputing 70 (1–3), 2006, 489–501.

[27] G.B. Huang, and L. Chen, “Enhanced random search based incremental extreme learning machine,” Neurocomputing 71 (16–18), 2008, 3460–3468.

[28] Y. Lan, Y.C. Soh, and G. Huang, “Ensemble of online sequential extreme learning machine,” Neurocomputing 72 (13–15), 2009, 3391–3395.