

Automated and ERP-Based Diagnosis of Attention-Deficit Hyperactivity Disorder in Children

Abstract

Event-related potential (ERP) is one of the most informative and dynamic methods of monitoring cognitive processes, which is widely used in clinical research to deal with a variety of psychiatric and neurological disorders such as attention-deficit/hyperactivity disorder (ADHD). In this study, there were 60 participants including 30 patients with ADHD and 30 subjects as a control group. Their ERP signals were recorded by three electrodes in two modalities. After a preprocessing step, several features such as band power, fractal dimension, autoregressive (AR) model coefficients and wavelet coefficients were extracted from recorded signals. The aim of this study is to achieve a high classification rate. The results show that the fractal dimension–wavelet combination features provided a good discriminative capability; it should be noted that this improvement was achieved by combining all sets of features and applying a feature selection algorithm, which resulted in a maximum accuracy rate of 88.77 and 95.39% in support vector machine (SVM) and ν -SVM classification algorithms using a 10-fold cross-validation approach, respectively. ERP has been widely used for clinical diagnosis and cognitive processing deficits in children with ADHD. To increase the accuracy of the diagnostic process of ADHD, ERP signals were recorded to extract some specific ERP features related to this disease for classifying the two groups. The results show that the Fra-wave characterization produced the best average accuracy with an efficiency of 99.43% for ν -SVM classifier, compared with 97.65% efficiency for the wavelet features and the other features.

Keywords: Algorithms, attention, attention deficit disorder with hyperactivity child, cognition, cognition disorders, control groups, electrodes, evoked potentials, fractals, humans, nervous system diseases, support vector machine

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common psychiatric disorders in school-aged children, which is characterized by a persistent pattern of impaired attention, impulsive behavior, and excessive motor hyperactivity. These manifestations, related to a child's conduct, directly affect academic activities, familiar dynamics, and performance in the social environment, which can negatively influence the personality development of a child.^[1] The clinical diagnosis of this disease requires several medical specialists as well as high-cost expenditure, which makes it difficult to access this benefit among the low-income population. Event-related potentials (ERPs) are defined as changes in the ongoing electroencephalogram (EEG) because of a sensory stimulus.^[2] Because these potentials

are physiologically correlated with neuro-cognitive functions, they have been widely used for clinical diagnoses, brain computer interface, and especially, in investigations of perceptual and cognitive processing deficits in children with ADHD. The most popularly assessed ERP features for interpretation of cognitive processes are the areas and the peaks of ERP components, defined by the mean and peak-to-peak voltages, respectively.^[3] The analyses of these parameters are usually performed in the time domain, whereby amplitudes and latencies of prominent peaks in the averaged potentials are measured and correlated to information processing mechanisms.^[4] Although the quantification of ERP components by areas and peaks is the standard procedure in fundamental ERP research, the conventional approach has two drawbacks. First, ERPs are time-varying signals, which reflect the sum of

Hossein R. Jahanshahloo, Mousa Shamsi, Elham Ghasemi¹, Abolfazl Kouhi

Department of Electrical Engineering, Sahand University of Technology, ¹Department of Agriculture Engineering, Tabriz University, Tabriz, Iran

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work noncommercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Jahanshahloo HR, Shamsi M, Ghasemi E, Kouhi A. Automated and ERP-based diagnosis of attention-deficit hyperactivity disorder in children. *J Med Sign Sence* 2017;7:26-32.

Address for correspondence:

Mr. Hossein Reza Jahanshahloo, Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran.
E-mail: h.jahanshahlo518@gmail.co

Website: www.jmss.mui.ac.ir

underlying neural events during stimulus processing, that operate in different time scales ranging from milliseconds to seconds. Various procedures such as ERP subtraction or statistical methods have been employed to separate functionally meaningful events that partly or completely overlap in time. However, the reliable identification of these components in ERP waveforms still remains a problem. Second, analyses in the frequency domain have revealed that EEG/ERP components in different bands (delta, theta, alpha, beta, and gamma) are functionally related to information processing. However, the Fourier transform (FT) of ERP lacks the time localization information of transient neural events. Therefore, using efficient algorithms for signal analysis in time–frequency domain is very important to extract and relate distinct functional components. These limitations, as well as the problems related with time-invariant methods, can be solved by using the wavelet formalism. The wavelet transform (WT) is a time frequency representation that has an optimal resolution both in the time and frequency domains and has been successfully applied to the study of EEG–ERP signals.^[5] Although ERP feature extraction from the time frequency domain based on the discrete WT (DWT) has been growing increasingly popular, this approach gives better results for pathology detection purposes, particularly in ADHD identification. Fractal dimension is an important parameter that has significant applications in various fields including signal processing. Signal analysis is a high-level signal processing technique for identifying the signal features such as roughness, smoothness, and solidity.

Likewise, because of lack of a suitable test or biomarker for diagnosis of ADHD in children, their identification requires a complete evaluation of children’s behavior both at home and at school. As a result, their diagnosis would be somewhat less accurate and time consuming. By taking into account the impacts of ADHD on the future lives of the children suffering from it, there is an absolute need for a biological marker that is capable enough to diagnose ADHD in children at the beginning by observing its symptoms in childhood. To introduce working biomarkers, some significant research has been conducted during the past few decades about EEG signals and ERP for diagnosis of ADHD in children.

The main goal of this study is to present a supporting diagnostic tool that uses signal processing for feature selection and machine learning algorithms for diagnosis. Particularly, for feature selection, information theoretic method is proposed, which is based on fractal dimension Higuchi algorithm and wavelet coefficients measure. In this method, a maximal discrepancy criterion is introduced for selecting distinct (most distinguishing) features of two groups as well as a semi-supervised formulation for efficiently updating the training set. Furthermore, support vector machine (SVM) classifier is trained and tested for

identification of robust marker of ERP signal for accurate diagnosis of ADHD group. The results show that the proposed approach provides higher accuracy in the diagnostic process of ADHD in comparison to the few currently available methods.

The result of the paper is organized as follows: “Materials and Methods” section gives a description of the method and data acquisition condition. “Feature Extraction” section presents the details of the features extraction containing band power, fractal dimension Higuchi algorithm, autoregressive (AR) coefficients, and wavelet coefficients. A brief description of SVM and ν -SVM classifiers is presented in “Classifiers” section. Performance analysis of SVM and ν -SVM classifiers using a 10-fold cross-validation approach on the extracted features is reported in “Results” section. Conclusions from this study are summarized in “Conclusion and Discussion” section.

Materials and Methods

Subjects

The participants in the research were children belonging to educational institutions of the metropolitan area of Manizales, aged between 4 and 15 years, and whose medical diagnosis had been determined by neurophysiological evaluation based on the clinical criteria of Diagnostic and Statistical Manual of Mental Disorder (DSM-IV).^[6] The children were tested in a room under the same lighting and noise conditions, and with uniformity in the following considerations: nonabnormalities in physical examination, normal visual and hearing ability, and intelligence quotient (IQ) greater than 80. Patients with pharmacologic management (methylphenidate, 20 mg) did not take their medicine for up to 24 h before the test.

In addition, the comorbidities in the children (oppositional defiant disorder, specific phobia, and learning problems) were taken into account. In this study, 60 children participated, which included 30 children with ADHD and 30 children as a control group. Their ERP signals were recorded by three electrodes located in the midline of the head (Pz, Cz, and Fz) according to 10–20 international system in two modalities, that is, auditory and visual, at sampling rate of 640 samples per second.

Data acquisition

The examination protocol was conducted according to the criteria of Oddball paradigm in the auditory and visual modalities. The first test involved the emission of 80 dB tone lasting 50 ms, with a frequency of 1000 Hz for frequent stimulus and 3000 Hz for infrequent stimulus, which were applied randomly at every 1.5 s. In the visual modality of the test, the children were asked to watch a monitor placed 1 m away, which showed an image with a consistent pattern (a checkerboard of 16 squares), which was the frequent stimulus. The rare stimulus was the presentation of a target in the center of the screen with the same common pattern in the background; the child must press a button each

time the unusual stimulus appeared. The experiment consisted of 200 stimuli, of which 80% were frequent and 20% were infrequent stimuli.

Preprocessing

Each of the classifiers was trained with the corresponding dataset. Because of the large set of features obtained from ERP measurements (values from all three electrodes in all bands), it was a good idea to reduce this set and feed the classifier with the most appropriate subset of these features. In addition, the complexity and dimensionality were reduced. For each SVM classifier [Figure 1], the forward selection scheme was applied to choose the best attributes or features that corresponded to the selected classifier. Forward selection scheme was an algorithm that iteratively selected subset of features from a set of features, such that the chosen subset of features was most relevant to the discrimination of the data. By applying this technique to each classifier that corresponded to a different condition, the features were obtained, which represented the most relevant discrimination of the data in the corresponding condition.

Feature extraction

During signal acquisition, the recordings were filtered from 0.3 to 100 Hz. This was necessary to improve the signal/noise ratio and to adapt them to later stages. ERP data was first band pass filtered (0.3–100 Hz). Then bad channels were identified and replaced, and an average reference was computed. The signals were analyzed for artifact rejection and power analysis. Blinks and other artifacts were automatically identified with a threshold of $\pm 100 \mu\text{V}$ and excluded from analyses. The clean data was then submitted to a fast Fourier transform (FFT) with a 1-s Hanning window and 50% overlap. Next, independent component analysis (ICA) was applied to eliminate dependence among the input signals. Later, features were estimated from dataset to form the initial space of characteristics. Finally, training of the classifier and assessment of the classification performance were done with the selected features [Figure 1].

In this section, the extracted features from ERP signals are explained. A group of these features consisted of

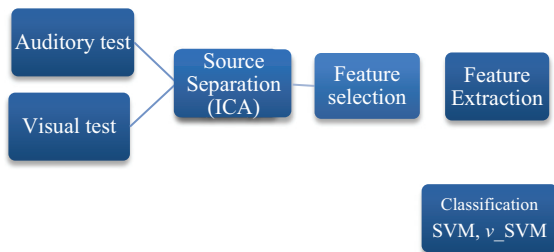


Figure 1: Block diagram of the overall procedure for features extraction and classification using raw ERP signals

fractal-based features,^[7-9] AR coefficients,^[10] band power,^[11] and wavelet coefficients.^[12]

Higuchi fractal dimension

If we consider a time series as $x(1), x(2), \dots, x(n)$, k new subsequences, x_m^k can be constructed as: $x_m^k = \{x(m), x(m+k), x(m+2k), \dots, x(m+\frac{N-m}{k}k)\}$ for $m = 1, 2, \dots, k$, where m indicates the index of first sample and k indicates time interval or delay between points in the k embedding sequences. Average length for each time series x_m^k is defined as:

$$L_m(k) = \frac{(N-1) \sum_{i=1}^{(N-m)/k} |x(m+ik) - x(m+(i-1)k)| (n-1)}{N - m/kk} \tag{1}$$

where N is the length of data sequence x . Total average length, $L(k)$, for scale k is computed as follows:

$$L(k) = \sum_{m=1}^k L_m(k) \tag{2}$$

where $L(k)$ is proportional to k^{-D} , and D is the fractal dimension that is defined using Higuchi's method.^[8] In fact, Higuchi dimension is tightly related with Takens' theory,^[13] in which a signal is considered as the output of a complex system. To characterize this system, signal samples were assigned to state variables activity. Any subsequent activity in Higuchi method showed the activity of the state variable through time. Average length of the subsequent activities was interpreted as the average length of the signals produced by the state variables.

Autoregressive coefficients

The AR models are used to describe a time series. This model estimates each sample as a weighted sum of previous samples by a recursive linear filter. The integer parameter p is called the order of the AR model. The AR model of a signal $x(t)$ in discrete time t is defined as follows:

$$x(t) = \sum_{i=1}^p a_i x(t-1) + \epsilon(t) \tag{3}$$

where a_1, a_2, \dots, a_p are the coefficients of recursive filter, p is order of the model, and $\epsilon(t)$ is uncorrelated noise. The Burg method^[10] is applied to fit a P th order AR model to the input signal, x , by minimizing (least squares) the forward and backward prediction errors while constraining the AR parameters to satisfy the Levinson–Durbin recursion. These coefficients can model the whole variation of a signal but they are sensitive to additive noise.

Band power

ERP can reflect the brain activity in different frequencies.^[11] The raw ERP is generally described in terms of five different frequency bands: gamma (>30 Hz), beta (13–30 Hz), alpha (8–12 Hz), theta (4–8 Hz), and delta (<4 Hz). To determine the band power in the mentioned frequency intervals, raw ERP signal was filtered through band pass filters (elliptic order six) to represent ERP content in the five successive frequency bands. In the output of each band pass filter, each sample was squared and averaged over several consecutive samples providing an estimation of band power in a window with the length of 1 s.

Wavelet

The DWT is an efficient tool in signal representation.^[12] Each mother wavelet can be compressed to provide a higher resolution in a dyadic form and also shifted in each resolution to model different locations of a signal through the time domain. A wavelet function in the scale a and time shift b is expressed as:

$$\psi_{(a,b)}(t) = 2^{a/2} \psi(2^{a/2}(t - b)) \quad (4)$$

where $\psi(t)$ is a wavelet function. Compressed and stretched derivations of wavelet function model are the high and low frequency components of a signal. Hence, by projecting an original signal into different resolutions, details of the signal can be obtained.^[12] By using DWT, signals are decomposed into two low and high bands at each level that are called approximations and details. Approximations represent low frequency components of the signal while details represent high frequency components. In this study, signals were decomposed into subspaces using the mother wavelet “db4.” Then energies corresponding to approximations and details were considered as signal features in each time frame.^[14]

Feature selection

Feature selection is a kind of commonly used dimensionality reduction method, as opposed to feature extraction such as principle component analysis (PCA), in which new low-dimensional embedding is produced using the original features. PCA is one of the most popular appearance-based methods used mainly for dimensionality reduction in compression and recognition problems. It can be used to reduce the dimensionality of the feature vectors and eliminates all the statistical covariance in the transformed feature vectors.

Classifiers

SVM

After selecting the relevant features for each dataset, the datasets were fed to SVM classifiers and the models were built. For generalization of each classifier model, 10-fold cross-validations were used.

Cross-validation is a technique that clarifies how the chosen classifier model will generalize to an independent dataset that is different from the one that has trained the model. It partitions the data into n complementary subsets and uses $n-1$ of the subsets for training the model, and the remaining set for testing the model. This procedure is repeated n times so that each of the subsets is used exactly once as a testing set. The results are then averaged over the rounds to get the final estimation.

ν -SVM

SVM is a method for supervised learning that is used for classification or regression analysis. That means, given an input data sample, in which each data point is marked as belonging to one of the two possible groups or classes, SVM builds a model that is then used for classifying new data points to one of the two classes.^[15]

For the independent and identically distributed training data,

$$(x_i, y_i) \in R^N \times \{+1, -1\}, i = 1, 2, \dots, l \quad (5)$$

and the original optimization problem in ν -SVM classification algorithm is as follows:

$$\min \tau(w, \xi, \rho) = \frac{1}{2} w - \nu \rho + \frac{1}{l} \sum_i \xi_i \quad (6)$$

$$y_i [(x_i \cdot w) + b] \geq \rho - \xi_i \quad (7)$$

$$\rho \geq 0, \xi_i \geq 0 \quad (8)$$

Applying Lagrangian optimization theory and incorporating kernels for dot products leaves the following dual quadratic optimization problem:

$$\max w(a) = -\frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K(x_i, x_j) \quad (9)$$

$$0 \leq a_i \leq \frac{1}{l} \quad (10)$$

$$\sum_i a_i y_i = 0 \quad (11)$$

$$\sum_i a_i \geq \nu \quad (12)$$

The parameters in Eqs. (5)–(12) are defined as follows: ν is an upper bound on the fraction of margin errors and lower bound on

the fraction of support vectors (SVs) and $0 \leq \nu \leq 1$. To understand the role of ρ , note that for $\xi = 0$, the constraint of Eq. (7) simply states that the two classes are separated by the margin $2\rho/w$ and other parameters have the same meanings as those in the original method.^[16]

To compute b and ρ , two sets S_{\pm} are considered, of identical size $s > 0$, containing SVs $0 < a_i < 1/l$ and $y_i = \pm$, respectively. Then, as a result of Karush Kuhn Tucker (KKT) conditions, constraint Eq. (7) becomes an equality with $\xi_i = 0$. Hence, in terms of kernels,

$$b = -\frac{1}{2s} \sum_{x \in S_+} \sum_j a_j y_j K(x, x_j) \quad (13)$$

$$\rho = \frac{1}{2s} \left(\sum_{x \in S_+} \sum_j a_j y_j K(x, x_j) - \sum_{x \in S_-} \sum_j a_j y_j K(x, x_j) \right) \quad (14)$$

the resulting decision function can be shown to take the following form:

$$f(x) = \text{sgn} \left(\sum_i a_i y_i K(x, x_i) + b \right) \quad (15)$$

Results

The recording acquisition per patient was made by using three electrodes (Fz, Cz, and Pz), with two different types of stimulus, auditory and visual; so, a total of six recordings were acquired per patient. To avoid correlations between electrodes, ICA was applied to generate three uncorrelated sources (for each electrode), on each one of the modalities.

Tables 1–4 show classification rate of power band, fractal dimension, AR coefficient, wavelet coefficient features, and fractal–wavelet (Fra-wave) in delta, theta, alpha, and beta bands frequency with SVM and ν _SVM classifiers, respectively.

This study presents an implementation of ADHD automatic detection system, focusing on the discriminative capability analysis of four different sets of the features and Fra-wave to get the best classification performance.

Table 1 presents the performance results of SVM and ν _SVM classifiers considering different features in delta band frequency. It is presented that the Fra-wave characterization produced the best mean accuracy of 92.93% utilizing SVM and 96.07% using ν _SVM, compared with other features. As it is demonstrated in Table 1, ν _SVM accuracy is better than SVM accuracy considering the whole set of features.

The theta band features are classified using SVM and ν -SVM methods and the results are presented in Table 2. On the basis of this table, the best average accuracy is obtained for Fra-wave characterization with an efficiency of 92.62% for

Table 1: Classification rate in delta band frequency

Feature	Class type	Auditory modality			Visual modality		
		Fz	Cz	Pz	Fz	Cz	Pz
Power band	SVM	88.1	97.4	92.8	92.4	96.8	96.6
	ν _SVM	92.3	100	95.8	100	77.5	91.4
Fractal	SVM	91.2	77.8	81.2	98.7	87.1	84.1
	ν _SVM	88.5	98.3	85.8	100	94.5	85.3
AR coefficient	SVM	97.4	92.3	96.2	85.4	85.4	94.1
	ν _SVM	95.6	98.9	82.5	100	96.4	95.8
Wa coefficient	SVM	97.4	88.9	85.6	78.4	75.4	82.5
	ν _SVM	100	94.1	96.5	84.1	100	86.5
Fra-wave	SVM	96.9	92.2	92.5	90.1	94.5	91.4
	ν _SVM	96.8	98.8	96.5	94.5	99.4	90.4

Table 2: Classification rate in theta band frequency

Feature	Class type	Auditory modality			Visual modality		
		Fz	Cz	Pz	Fz	Cz	Pz
Power band	SVM	96.5	92.1	95.4	92.4	77.5	78.5
	ν _SVM	97.7	98.7	96.0	100	89.8	93.5
Fractal	SVM	89.7	88.1	96.6	78.8	85.8	99.5
	ν _SVM	95.4	88.7	95.4	89.9	97.8	97.4
AR coefficient	SVM	96.8	88.1	96.4	82.5	89.4	86.9
	ν _SVM	97.8	92.5	100	100	89.5	91.4
Wa coefficient	SVM	82.1	84.5	89.7	74.1	78.5	92.5
	ν _SVM	97.4	98.5	98.7	88.4	100	97.1
Fra-wave	SVM	93.4	90.9	92.4	95.4	90.1	93.5
	ν _SVM	98.4	99.1	97.5	98.8	100	98.8

Table 3: Classification rate in alpha band frequency

Feature	Class type	Auditory modality			Visual modality		
		Fz	Cz	Pz	Fz	Cz	Pz
Power band	SVM	95.6	88.4	92.3	96.6	75.4	85.4
	ν _SVM	96.8	92.2	96.6	98.7	90.5	92.2
Fractal	SVM	75.4	70.9	88.4	78.8	98.1	78.8
	ν _SVM	90.8	95.8	100	94.4	98.8	88.9
AR coefficient	SVM	88.6	88.9	84.7	91.5	92.1	81.1
	ν _SVM	92.8	95.8	88.9	84.5	99.8	97.8
Wa coefficient	SVM	93.6	90.4	88.8	74.8	76.9	90.8
	ν _SVM	100	98.8	100	98.9	90.4	97.8
Fra-wave	SVM	100	96.8	94.5	96.8	90.8	92.5
	ν _SVM	99.2	99.8	100	97.8	99.8	100

ν _SVM classifier, whereas it is 98.77% for the wavelet feature and other features.

Generally, on the basis of the information provided in Tables 1–4, the auditory case has produced a better accuracy compared with visual modality.

Table 4: Classification rate in beta frequency

Feature	Class type	Auditory modality			Visual modality		
		Fz	Cz	Pz	Fz	Cz	Pz
Power band	SVM	96.8	88.9	92.1	96.8	75.5	90.1
	v_SVM	100	94.5	92.5	100	92	93.3
Fractal	SVM	82.5	71.4	78.7	79.8	84.5	89.4
	v_SVM	86.5	100	95.8	81.4	92.5	97.8
AR coefficient	SVM	88.9	96.8	91.4	90	86.5	92.5
	v_SVM	94.8	100	95	96.4	100	98.7
Wa coefficient	SVM	92.5	86.5	90.1	88.1	74.8	75.8
	v_SVM	90.2	97.8	99.8	95.6	90.8	93.5
Fra-wave	SVM	96.5	92.5	93.5	90.5	96.9	95.4
	v_SVM	100	98.8	100	94.5	100	96.8

Conclusion and Discussion

Many studies have tried to identify behavioral disorders among children by using different diagnostic methods and tools. Some researchers have used multiple sources to gather the comprehensive information.^[17,18] Some of these sources were questionnaires, interviews with families, parents and child, and clinical observation. However, they had considered much information simultaneously, which took a long time and also caused diagnostic errors. In some studies, teacher’s ratings were focused,^[19] so the risk of misdiagnosis was high. In some studies, information obtained from both parents and teacher’s ratings were used for diagnosing behavioral disorders;^[20] however, using this amount of information for diagnosis takes a long time, with high cost and high risk of errors. ERP and EEG were applied for accurate diagnosis.

ERP has been widely used for clinical diagnosis and cognitive processing deficits in children with ADHD. To increase the accuracy of diagnosis processes for ADHD, ERP signals were recorded to extract some specific ERP features related to this disease and the participants were classified in two groups. In this way, after preprocessing of signals, several features were extracted from the recorded signals, some of which reflected the complexity and roughness of the signals (fractal-based features), and the others modeled the signals in the time (AR coefficients), spectral (band power), and time–frequency domains (wavelet coefficients). Next, these extracted features were applied to the classifiers such as SVM and v_SVM. Results of this experiment showed the supremacy of v_SVM classifier.

This study presents an implementation of an ADHD automatic detection system, focused on the discriminative capability analysis of four different sets of features to get the best classification performance.

The example, Table 2 presents the performances results, showing that the Fra-wave characterization produced the best average accuracy with an efficiency of 98.77% for v_SVM classifier, compared with the 96.68% efficiency for the wavelet features and the other features. However,

the combination of them produced a better accuracy for v_SVM classifier. The fact can also be seen in Tables 1–4 that auditory modality produced a better accuracy compared with visual modality.

The results can be compared with some previous research. As the first study, Jiao and colleagues used channel 128, K-nearest neighbors (K-NN), and SVM to classify ADHD. The results showed that the best classification accuracy of 83.33% was achieved by K-NN classifier.^[21]

Mueller et al. used two groups of age-matched adults (75 ADHD and 75 controls) and performed a visual two stimulus go/no-go task. ERP responses were decomposed into independent components, and a selected set of independent ERP component features were used for SVM classification. Using a 10-fold cross-validation approach, classification accuracy was 91%. Predictive power of SVM classifier was verified on the basis of the independent ADHD sample (17 ADHD patients), resulting in a classification accuracy of 94%.^[22] On the other hand, in our research, we have used visual and auditory stimulus to achieve better classification accuracy.

Zeynab et al. have obtained 215 wavelet coefficients and 20 recurrence quantification analysis (RQA) features that could discriminate 31 ADHD and 37 normal groups significantly. They have used K-fold (5-fold) cross-validation in which all data have been randomly divided into five groups. Classification results have demonstrated that combination of two kinds, Morlet and Mexican hat kernel classification of features, represents the signal in a better way.

Results show that there is approximately no difference between accuracy of Morlet and Mexican hat kernel classification approaches, and the accuracy of both methods was 98%.^[23]

Our study can be enhanced by considering the following two main factors: using more signal recording channels and increasing the number of participants in the test.

This study is the first attempt to classify ADHD patients by means of SVM and independent ERP signal with combined feature Fra-wave. The results demonstrate the efficiency of the utilized approach. As a future study, independent component decomposition and feature extraction procedures can be considered. Moreover, this promising approach can easily be applied to other clinical problems.

Acknowledgements

We thank Dr. P. Castro-Cabrera University of Universidad Nacional de Colombia, sede Manizales for providing the ERP dataset.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Lazzaro I, Gordon E, Li W, Lim CL, Plahn M, Whitmont S, *et al.* Simultaneous EEG and EDA measures in adolescent attention deficit hyperactivity disorder. *Int J Psychophysiol* 1999;34:123-34.
- Mavrogiorgou P, Juckel G, Frodl T, Gallinat J, Hauke W, Zaudig M, *et al.* P300 subcomponents in obsessive-compulsive disorder. *J Psychiatr Res* 2002;36:399-406.
- Bostanov V. Data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans Biomed Eng* 2004;51:1057-61.
- Demiralp T, Istefanopulos Y, Ademoglu A, Yordanova J, Kolev V. Analysis of functional components of P300 by wavelet transform. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol 20, 1998. p. 4.
- Quiroga RQ. Obtaining single stimulus evoked potentials with wavelet denoise. *Physica D* 2000;145:278-92.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorder*. 4th ed. Washington, DC: American Psychiatric Press; 1994.
- Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D* 1998;31:277-83.
- Petrosian A. Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns. *Proceedings of the IEEE Symposium Computer Based Medical Systems*, 1995. p. 212-7.
- Stoica P, Moses RL. *Introduction to Spectral Analysis*. Upper Saddle River, New Jersey: Prentice-Hall; 1997.
- Moretti DV, Babiloni C, Binetti G, Cassetta E, Dal Forno G, Ferreric F, *et al.* Individual analysis of EEG frequency and band power in mild Alzheimer's disease. *Clin Neurophysiol* 2004;115:299-308.
- Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Appl Comput Harmon Anal* 2011;30:129-150.
- Li Y, Zhang S. Apply wavelet transform to analyses EEG signal. *Engineering in Medicine and Biology Society, Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, vol 3, 1996. p. 1007-8.
- Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D* 1998;31:277-83.
- Galka A. *Topics in Nonlinear Time Series Analysis with Implication for EEG Analysis*. Singapore by Fulstrand Offset Printing: World Scientific; 2000.
- Tenev A, Markovska-Simoska S, Kocarev L, Pop-Jordanov J, Müller A, Candrian G. Machine learning approach for classification of ADHD adults. *Int J Psychophysiol* 2013;93:162-6.
- Qi-Hua XU, Jun SH. Fault diagnosis for aero-engine applying a new multi-class support vector algorithm. *Chin J Aeronaut* 2006;19: 175-82.
- Harrington R. Assessment of psychiatric disorders in children. *Psychiatry* 2005;4:19-22.
- Bruce H, Evans N. Assessment of child psychiatric disorders. *Psychiatry* 2008;7:242-5.
- Musisi S, Kinyanda E, Nakasujja N, Nakigudde J. A comparison of the behavioral and emotional disorders of primary school-going orphans and non-orphans in Uganda. *Afr Health Sci* 2007;7:202-13.
- Langberg JM, Froehlich TE, Loren RE, Martin JE, Epstein JN. Assessing children with ADHD in primary care settings. *Expert Rev Neurother* 2008;8:627-41.
- Yang J, Li W, Wang S, Lu J, Zou L. Classification of children with attention deficit hyperactivity disorder using PCA and K-nearest neighbors during interference control task. *Adv Cogn Neurodyn* 2016;447-53.
- Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera G-M. Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: A validation study. *Nonlinear Biomed Phys* 2011;5:5. doi: 10.1186/1753-4631-5-5.
- Zeynab E, Ali M, Saeed M. An auditory brainstem response-based expert system for ADHD diagnosis using recurrence qualification analysis and wavelet support vector machine. *23rd Iranian Conference on Electrical Engineering*, 2015.