Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Article

Diagnosis of Attention Deficit Hyperactivity Disorder with combined time and frequency features

Miray Altınkaynak [a,*], *Nazan Dolu* [b], *Aysegül Güven* [a], *Ferhat Pektas* [c], *Sevgi Özmen* [d], *Esra Demirci* [d], *Meltem Izzetoglu* [e]

[a] *Department of Biomedical Engineering, Engineering Faculty, Erciyes University, Turkey*
[b] *Department of Physiology, Medical Faculty, Baskent University, Turkey*
[c] *Department of Physiology, Medical Faculty, Kırsehir Ahi Evran University, Turkey*
[d] *Department of Child Psychiatry, Medical Faculty, Erciyes University, Turkey*
[e] *Electrical and Computer Engineering Department, Villanova University, USA*

ARTICLE INFO

Article history:
Received 13 November 2018
Received in revised form
17 February 2020
Accepted 10 April 2020
Available online 4 May 2020

Keywords:
Attention Deficit Hyperactivity Disorder
Machine learning
Classification
Auditory evoked potentials
Fractal dimension
Discrete Wavelet Transform

ABSTRACT

The aim of this study was to build a machine learning model to discriminate Attention Deficit Hyperactivity Disorder (ADHD) patients and healthy controls using information from both time and frequency analysis of Event Related Potentials (ERP) obtained from Electroencephalography (EEG) signals while participants performed an auditory oddball task. The study included 23 unmedicated ADHD patients and 23 healthy controls. The EEG signal was analyzed in time domain by nonlinear brain dynamics and morphological features, and in time-frequency domain with wavelet coefficients. Selected features were applied to various machine learning techniques including; Multilayer Perceptron, Naïve Bayes, Support Vector Machines, k-nearest neighbor, Adaptive Boosting, Logistic Regression and Random Forest to classify ADHD patients and healthy controls. Longer P300 latencies and smaller P300 amplitudes were observed in ADHD patients relative to controls. In fractal dimension calculation relative to the control group, the ADHD group demonstrated reduced complexity. In addition, certain wavelet coefficients provided significantly different values in both groups. Combining these extracted features, our results indicated that Multilayer Perceptron method provided the best classification with an accuracy rate of 91.3% and a high level of reliability of concurrence (Kappa = 0.82).

The results showed that combining time and frequency domain features can be a useful and discriminative for diagnostic purposes in ADHD. The study presents a supporting diagnostic tool that uses EEG signal processing and machine learning algorithms. The findings would be helpful in the objective diagnosis of ADHD.

* *Corresponding author at*: Department of Biomedical Engineering, Faculty of Engineering, Erciyes University, 38039 Kayseri, Turkey.
E-mail addresses: eem.miray@gmail.com, miray@erciyes.edu.tr (M. Altınkaynak).
https://doi.org/10.1016/j.bbe.2020.04.006

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most prevalent psychiatric disorder in school age children which is characterized by inattention, hyperactivity and impulsivity symptoms [1]. All these symptoms negatively affect the academic activities, social behavior, personality development of the child and can continue into adulthood [2].

Changes of brain structure and functions in ADHD patients were investigated by using various neuroimaging and neurophysiological techniques [3–9]. Numerous research revealed that Electroencephalography (EEG) is a discriminative tool for ADHD and healthy control particularly regarding frontal and central measurements. In time domain, EEG data of ADHD participants have been analyzed principally via ERP trial averaging. Morphological features, mainly peak amplitudes and latencies of many components (P1, N1, P2, N2, P3, etc.) were evaluated in many studies for the interpretation of cognitive processing in ADHD. Most consistently, P300 component, was found discriminative for ADHD and control [10] Results of other components were quite variable [11]. P300 emerges approximately 300 ms. following stimulus presentation and is correlated with attention and widely used in evaluating cognitive deficits in various brain disorders. In ERP studies generally, other than a few predetermined extremums all other information contained in the ERP time series is overlooked. Therefore, in this study, together with morphologic features of AEP, (Latency, amplitude and latency/amplitude ratio of P300 wave) the whole ERP signal was also analyzed both in time-frequency plane via Discrete Wavelet Transform (DWT).

In the frequency domain, power of different EEG frequency bands was evaluated in many studies to diagnose ADHD. Increased theta power and high theta/beta ratio in ADHD compared to controls are most consistent findings. Other frequency bands such as alpha and beta have been more variable among children with ADHD [12]. In some studies, ERPs were decomposed into different components and several frequency domain features were evaluated for classifying ADHD and control groups [13,14]. Wavelet transform (WT) is a widely used method for multi-resolution analysis of AEPs in ADHD [15]. WT decompose the signal into high and low frequency components and obtain significant features of original data. In this study, entropy of the DWT coefficients of AEPs were selected as one of the feature data set to be used in the discrimination between healthy controls and ADHD patients.

Recently, nonlinear approaches have been widely used to measure complexity of EEG signal in ADHD children. Most consistently, decreased EEG complexity was observed in ADHD subjects relative to controls [16]. Over the past decade, several complexity estimates such as; entropy, fractal dimension (FD), Lyapunov exponent, Hurst component and Lempel-Ziv complexity have been used to measure the complexity of EEG in ADHD patients [17]. Alba et al. [18] reviewed EEG based linear and nonlinear biomarkers and reported among nonlinear features LZC and approximate entropy are more discriminative features for ADHD and control. FD which is an important parameter that can reveal the self-similarity and complexity of a signal has emerged as being sensitive to neurological and psychiatric diseases [19]. Comparison of EEG fractal dimension between the ADHD patients and controls showed that complexity is reduced in correlation with cognitive impairment [20]. In this study, FD of EEG was calculated by Higuchi algorithm. This algorithm estimates complexity directly from time series so it is faster and simpler than the other estimators. So, in this study Higuchi Fractal Dimension (HFD) of full band EEG was selected as another feature for classifying ADHD patients and healthy controls. Finally, we combined morphological, wavelets, and nonlinear based features in order to discriminate ADHD and control group.

The extracted features were first evaluated in terms of their separation capability between ADHD and healthy controls. Then the features that provide statistical significance were selected to be used in machine learning algorithms. Previously, linear and nonlinear classifiers such as Linear Discriminant Analysis (LDA) [9], Support Vector Machines (SVM) [13] and Neural networks (NN) [16] were commonly utilized for classification between ADHD and control group. In this study various machine learning algorithms including; Multilayer Perceptron, k nearest neighbors, Naïve Bayes (NB), SVM, Random Forest (RF), Adaboost and Logistic Regression (LR) were applied and evaluated for this purpose. In Table 4 we summarized recent diagnostic classification studies in ADHD based on EEG measurements and compared with our results.

The objective of this study was discriminating ADHD patients and control groups that used information from both time and frequency analysis of ERP obtained from EEG signals while participants performed an auditory oddball task. To our knowledge, there has been no previous investigation on the use of machine learning techniques that evaluated multiple classifiers for classification with both time and frequency domain features of AEPs.

2. Material and methods

2.1. Participants

Participants comprised 23 unmedicated ADHD children and 23 healthy controls with age of 7–12 years. The healthy controls were selected according to the following criteria; neurological, endocrine or psychiatric illness absence and normal hearing function. The ADHD patients were referred from Child and Adolescent Psychiatry Department of Erciyes University, Medical Faculty Hospital. Diagnosis of patients was performed by professional psychiatrists according to DSM IV. ADHD children were totally drug-naïve. Exclusion criteria were evidence of neurological disorders and having hearing impairment. All participants were right-handed. The study was approved by the local ethics committee of Erciyes University. Child's assent and parental consent were obtained for all participants. Demographic characteristic of the participants is shown in Table 1.

2.2. Experimental design

EEG was recorded while participants performing an auditory 'oddball' paradigm. Experimental paradigm consisted of 128

Table 1 – Demographic characteristics of the participants.

	ADHD	Control	t; df; p
Participants	23	23	
Gender (M:F)	16:7	14:9	
Age in years	9.09 ± 1.621	9.13 ± 1.632	0.091; 43.9; 0.9288^{ns}
Education in years	3.13 ± 1.56	3.03 ± 1.71	0.179; 43.67; 0.859^{ns}
WISC-R	103.96 ± 22.56	107.57 ± 10.08	0.700; 30.44; 0.487^{ns}

Groups were compared with two-tailed unpaired Student's t tests. *Abbrevations*: WISC-R, The Wechsler Intelligence Scale for Children-Revised; t, t-value; df, degree of freedom; p, p value; ns, non-significance.

standard and 32 target trials. Infrequent target stimuli were presented as 2000 Hz tone bursts and standard stimuli were presented as 1500 Hz tone burst. The interstimulus intervals were randomized between 1250 and 2500 ms. Both auditory stimuli had durations of 50 ms. Before measurements, subjects were instructed to keep their eyes open, not to move, to listen carefully for the target stimuli, and press a response button when they hear target trials. The experiment was conducted in a sound and light attenuated, electrically shielded room.

Fig. 1 is the flow chart that summarize the study. Details of each step are presented in the following sections.

2.3. *EEG data acquisition and pre-processing*

EEG was obtained at Fz, Cz, Pz, and Oz positions using international 10–20 system with active Ag/AgCl electrodes. The ground electrode was placed on the right ear and the reference electrode was placed on the left ear. The sampling frequency was 2500 Hz. Vertical and horizontal Electrooculogram (EOG) were recorded differentially via Ag/AgCl electrodes. All electrode impedances were below 5 kΩ. EEG and EOG amplitudes greater than 50 μV were classified as artifacts and data from trials with artifacts were rejected. Preprocessing included band pass filtering between 0.05–100 Hz and band stop (Notch) filtering of 50 Hz to remove additional noise. AEPs were recorded with MP150 acquisition system (Biopac System Inc.).

2.4. *Feature extraction*

Previously, ERP signals were commonly analyzed in time or frequency domains in order to extract discriminative features for ADHD and controls. In this study we combined time and frequency information of ERPs to enable higher classification accuracy.

The initial feature space consisted of three sets of features that have shown optimum performance in similar studies. The first group comprised 3 morphological features latency, amplitude, latency/amplitude ratio of P300 component, which consisted of parameters measured over the ERP signal. The second group was WT–based features, which consisted of WT coefficients obtained through decomposition of ERP into three levels, using DWT. Hence, four sets of WT coefficients were generated: coefficients in the 0.05–12.5 Hz band (A3; approximation coefficients of the third level), 12.5–25 Hz band (D3; detail coefficients of the third level) 25–50 Hz band (D2; detail coefficients of the second level), and 50–100 Hz band (D1; detail coefficients of the first level). Then, the entropy was calculated for each WT coefficient set. Thus, the group was comprised of 4 WT coefficients entropy value. The last set was made up of a nonlinear dynamics based feature: the fractal dimension of EEG signal. Higuchi algorithm was used to calculate FD. Therefore, 3 morphological features, 4 WT-based features, and a nonlinear dynamics based feature consisted the initial feature space. These features were analyzed statistically by independent-sample t-test searching for significantly different ones in both groups, which can provide better classification accuracy. Accordingly, 2 WT-based features; detail coefficients of the first level and approximation coefficients of the third level were not included in the classification step. The remaining 6 features formed the feature set to discriminate normal controls and ADHD patients. A brief description of the used methods and features are presented below.

2.4.1. *Morphological features*

AEPs were obtained by averaging target trials from ongoing EEG [21] that decrease the noise level. Averaged AEP data were

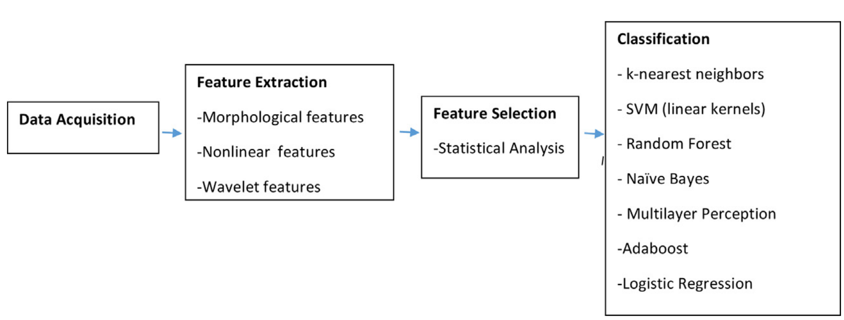

Fig. 1 – Proposed signal processing algorithm, which consist of four sequential steps: Data acquisition from 46 subjects,

baseline corrected. In this study we utilized P300 latency and amplitude parameters that give one of most consistent findings of EEG-ADHD diagnosis studies. The P300 was selected according to following criteria by specialists. It has the largest positive peak around 300 ms (although the range can vary between 250–500 ms) [22]. Latency and amplitude values of P300 component compared between groups in Fz, Cz, Pz, and Oz locations. In each location, P300 latencies were longer in ADHD group compared to controls (Fz = 336.59 ± 13.35; 296.66 ± 21.21; t = 7.64; $p < 0.05$; Cz = 338.8 ± 13.6; 294.2 ± 24.6; t = 7.587; $p < 0.05$, Pz = 333.9 ± 17.7; 295.94 ± 24.5; t = 6.01; $p < 0.05$, Oz = 328.17 ± 18.37; 296.08 ± 22.15; t = 5.34; $p < 0.05$). However, no significant difference in terms of amplitude for Cz, Pz, Oz sensors was observed (Fz = 19.63 ± 9.10; 30.23 ± 11.85; t = 3.23; $p = 0.002$, Cz = 22.7 ± 8.7; 25.54 ± 9.47; t = 0.925; $p = 0.361$, Pz = 20.13 ± 11.24; 24.49 ± 10.2; t = 1.37; $p = 0.175$, Oz = 18.36 ± 9.08; 19.01 ± 8.91; t = 0.24; $p = 0.806$). The descriptive values given (mean ± std) results belong to ADHD group and controls, respectively. Since ADHD is associated with frontal dysfunctions [23] and we observed that Fz measurements were more discriminative between groups, further analysis in Fz channel. Latency, amplitude and latency/amplitude values of Fz sensor were the first three biomarkers in the feature set.

2.4.2. *Nonlinear features*

Recently, nonlinear dynamics have been widely used to estimate complexity of EEG signal in ADHD children. In this study, we investigated complexity of EEG data in both groups via FD. The utility of FD in the discrimination of ADHD and control had been reported previously [17]. There are various algorithms to calculate FD such as Higuchi's, Katz's, and Petrosian's. Several neurophysiological research concerning fractal analysis of EEG analysis showed the superiority of Higuchi's algorithm over other algorithms [24]. Higuchi proposed an algorithm that estimates FD directly from time series, hence it is a time-consuming and simple method. Thus, Higuchi's algorithm [25] (Higuchi 1988), was chosen to calculate FD values in this study.

Higuchi algorithm

Given the time sequence $x(1)+x(2)+\ldots+x(N)$ the algorithm constructs k new self-similar time series x_m^k as:

$$x_m^k = \left\{x(m)+x(m+k)+x(m+2k),\ldots,x\left(m+\lfloor\frac{N-m}{k}\rfloor k\right)\right\}, \quad m = 1,2,3,\ldots,k \tag{1}$$

where m represents the initial time and k indicates the discrete interval time.

For each curve x_m^k, the average length L_m^k is computed as:

$$L_m^k = \frac{\sum_{i=1}^{\lfloor (N-m)/k \rfloor} |x(m+ik) - x(m+(i-1)k|}{\lfloor \frac{N-m}{k} \rfloor k}(N-1) \tag{2}$$

Where N is the total length of the data and $\frac{N-1}{\lfloor (N-m)/k \rfloor}k$ is a normalization factor. Then for all time series, an average length is computed as:

$$L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m^k \tag{3}$$

Finally, the slope of the line that fits the pairs $\ln[L(k), \ln(1/k)]$ for $k = 1, 2 \ldots k_{max}$ in a least-squares sense gives the Higuchi dimension. The saturation point of FD is chosen as the appropriate value of the parameter k_max [26]. According to this approach k_max was chosen as 30 for our study.

2.4.3. *Wavelet features*

Wavelet transform gives information both in time and frequency domain. WT decomposes the signal into weighted set of scaled functions (Eq. (4)) [27]:

$$F(a,b) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \tag{4}$$

where a defines the scale, b defines the shift.

DWT is employed in this study for multi-resolution decomposition of the original AEP signal. This method uses discrete scales and shift factors to decompose the signal into low and high frequency bands at each level that are called approximations for low frequency components ($g[n]$) and details for high frequency components ($h[n]$) (Fig. 2). In many studies, the features extracted from the detailed coefficients revealed the characteristics of the time series and were used in the discrimination between ADHD and normal controls. In this study, signals were decomposed into 3 level sub-spaces using the "Daubechies (db2)" because it is an appropriate mother wavelet for multi-resolution analysis of AEPs [28] (Fig. 3). Then, entropy values of approximations and details were investigated and discriminative features for ADHD and healthy controls were selected for classification.

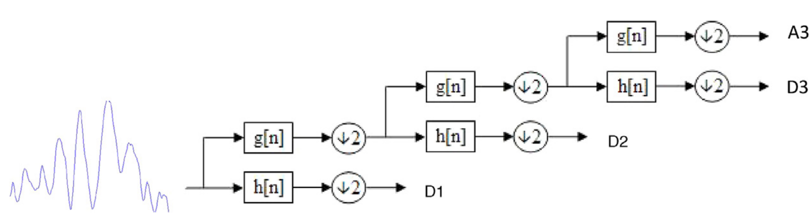

Fig. 2 – Three-level wavelet decomposition tree. *h[n]* indicates high pass filter and *g[n]* indicates low pass filter. The high pass filter generates detail information *D[n]* at each level.

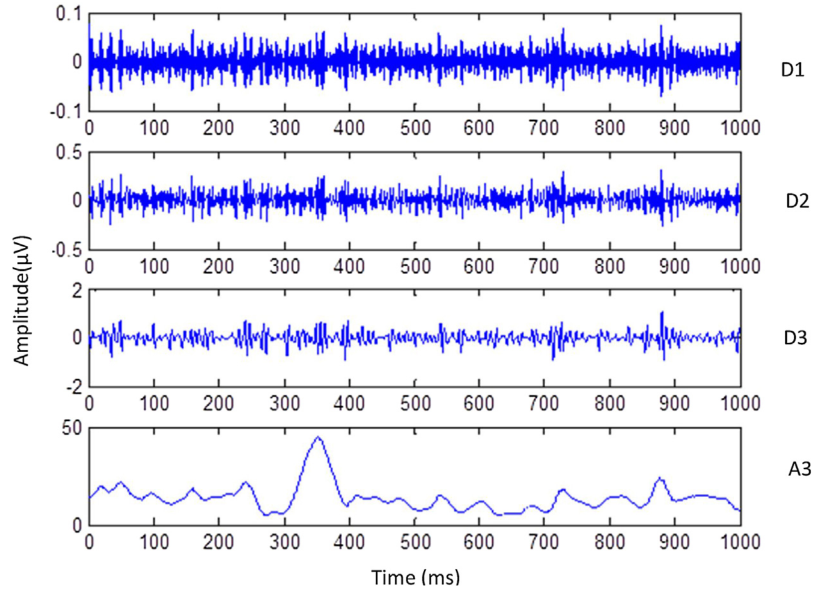

Fig. 3 – Wavelet coefficients of a sample healthy AEP data. Signals were decomposed into 3 level sub-spaces using the "Daubechies (db2)".

2.5. *Feature selection*

After extracting features using the aforementioned methods, these three different sets of features (morphological, wavelets, and nonlinear based) were first analyzed statistically by independent-sample *t*-test, searching for significantly different ones in both groups, which can provide better classification accuracy. Shapiro–Wilk normality test was used to confirm the normal distribution of variables. Each feature was evaluated separately. Statistical threshold defined as 0.05.

2.6. *Classification*

After extracting and selecting the features using the aforementioned methods, various machine learning algorithms including LR, MLP, k-NN, NB, SVM, RF, and Adaboost were applied and compared in terms of their performance in classifying ADHD and healthy controls. A brief description of the utilized algorithms are described below.

Support Vector Machine: The linear SVM classification is based on a dot product between vectors. If every data point is mapped into a higher dimensional space, the dot product is replaced by kernel functions [29]. If the mapping function is denoted as Φ, Φ: $\vec{x} \rightarrow \phi(\vec{x})$, kernel function can be written as:

$$F(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (5)$$

k Nearest Neighbors Algorithm: The method is based on feature similarity which is defined according to a distance metric between two data points. For a given feature vector given $x \in X$, first Euclidean distance between x' and each of the feature vectors in training set is computed. The class labels of the *k* nearest vectors from training set are used to predict the class of x' according *k* closest neighboring samples in the training set [30].

Naïve Bayes: The method assumes the values of features given in a class are independent of any other feature. When vector $X = (x_1, x_2, \ldots, x_n)$ represents some *n* features are to be classified into y, NB classifier can be formally defined as [31]:

$$P(x|y) \approx \prod_{k=1}^{n} P(x_k|y)^{w_i} \quad (6)$$

where $P(x|y)$ denotes the conditional probability distribution of attribute x_k conditioned by the given class y and wi refers to the weight of the attribute x_k.

Multilayer Perceptron Neural Network; MLP uses a supervised learning technique called back-propagation [32]. This algorithm iteratively calculates the error function in weight space until the prediction error reaches minimum. Given a set of features $X = \{x_1, x_2, \ldots, x_n\}$ and a target y, MLP mathematically can be written as:

$$y = \varphi \sum_{i=1}^{m} w_i x_i + b \quad (7)$$

where φ is the activation function, *w* represents weights vectors, *b* denotes the bias.

Random Forest: RF consists of decision trees from randomly selected set of features and splits them at every node. For classification, the input vector is fed to each decision tree in the RF and each tree votes for a class. Finally, the RF chooses the most voted prediction as the result [33].

Adaboost: This method combines different simple learning algorithms (weak learners) and generate a strong learner [34]. An Adaboost classifier can be defined as:

$$H_f = \sum_{t=1}^{T} \alpha_t C_t(x) \tag{8}$$

where $C_t(x)$ denotes the weak learner, α_t refers to the weight of the tth weak learner and T is the total number of iterations.

Logistic Regression; Logistic regression estimates probabilities of the possible outcomes using a logistic function [35]. The logistic function is a sigmoid function, which maps any real value into another value between 0 and 1. The standard logistic function σ: R → (0,1) is defined as:

$$\sigma(t) = \frac{1}{1+e^{-t}} \tag{9}$$

2.7. Training procedure

In this study, leave-one subject out cross validation (LOSO) method was used in order to evaluate classifiers' performance. It is a common technique for the assessment of the accuracy when the sample size is small. This method makes unbiased comparation with the selection of each input as validation data uniquely where other inputs are used as training data [36]. Subset selection effect is neutralized with this approach. Classification results of classifiers were obtained using WEKA (Waikato Environment for Knowledge Analysis) Version 3.8.2. Grid search algorithm was used to tune hyperparameters of classifiers.

For SVM; we used a linear kernel and the C-SVC formulation. Gamma was 0 and cost was 1. For LR; max iteration defined as 100 with grid search in which we obtained highest accuracy. The MLP structure with the sigmoid activation function has an input layer with 6 neurons, one hidden layer with 4 neurons, and one output layer with 2 neurons. Learning rate and momentum were varied from 0.1 to 0.9 during the learning rate tuning step using a step size of 0.1. Both parameters were chosen as 0.2 with grid search algorithm. For k-NN; k value of k-NN classifier varied from 1 to 10 using a step size of 1. k Value was chosen as 3 and distance metric was Euclidean. In RF, the number of trees is optimized in the range 100–1000 and step of size of 100. The number of trees was 100, the seed was 1 and dept was 2. For Adaboost; we used Decision Tree Classifier as weak learner for training purpose. The weight threshold for weight tuning was 100, number of iterations was 10 and learning rate was 1.

We compare performance of the methods in terms of classification error, sensitivity, specificity, accuracy, and receiver operating characteristic (ROC) area. In addition, Cohen's Kappa coefficient that is used to test inter-rater reliability is reported for each classifier. It measures the agreement between two raters [37]. Cohen's kappa value is defined as:

$$\text{Kappa} = \frac{AAC_O - ACC_E}{1 - ACC_E} \tag{10}$$

where ACC_O is the observed accuracy and ACC_E is the expected accuracy. Generally, Kappa ranges between 0 to 1 where larger numbers corresponds more reliability. In addition, F-statistics that is a value showing the degree of significant difference between two populations [38] were also reported for each classifier.

3. Results

Different sets of features based on morphological, wavelet, and nonlinear based methods were first analyzed statistically with independent t-test to evaluate their discriminatory capabilities between ADHD and healthy controls, and then, combined to enable higher classification accuracy. Table 2 shows morphological, wavelets, and nonlinear based features' student t-test results. Longer P300 latencies and smaller P300 amplitudes were observed in ADHD patients relative to controls. In fractal dimension calculation relative to the control group, the ADHD group demonstrated reduced complexity. The entropy value of D_1 and A_3 coefficients of the wavelet transform were eliminated from the feature set because they were not significantly different between groups ($p > 0.05$). The features that were significantly different in both groups ($p < 0.05$) were selected to be input vectors of the classifiers.

Sensitivity (true positive factor), specificity (1-False positive factor), accuracy, Receiver operating characteristic (ROC) area, F and Kappa-statistics and mean absolute error (MAE) were calculated to measure performance of the classifiers. The results are shown in Table 3.

The results suggested that use of appropriate features and selected classification methods can improve the discrimination capability between ADHD patients and control groups. ROC curves of classifiers presented in Fig. 4.

Table 2 – Mean, standard deviation and statistical analysis of selected features.

	ADHD	Healthy control	t; p
Latency	336.59 ± 13.35	296.66 ± 21.21	7.64; 0.000^s
Amplitude	19.63 ± 9.10	30.23 ± 11.85	3.23; 0.002^s
Latency/Amplitude	22.68 ± 15.29	11.34 ± 4.87	3.32; 0.002^s
Fractal Dimension	1.0528 ± 0.0109	1.0721 ± 0.0304	2.85; 0.007^s
Entropy of D2	2.3479 ± 0.2414	2.5485 ± 0.3575	2.22; 0.031^s
Entropy of D3	3.4547 ± 0.2178	3.6523 ± 0.3239	2.42; 0.019^s

Table 3 – The summary of performance scores for different classification methods.

	SVM	NB	k-NN	MLP	RF	Adaboost	LR
Sensitivity	0.95	0.86	0.91	0.91	0.82	0.95	0.91
Specificity	0.82	0.86	0.82	0.91	0.82	0.82	0.86
Accuracy (%)	89.13	86.95	86.95	91.3	82.6	89.13	89.13
ROC Area	0.89	0.94	0.89	0.91	0.91	0.91	0.86
Kappa	0.78	0.73	0.73	0.82	0.65	0.78	0.78
F-Statistic	0.89	0.87	0.86	0.91	0.84	0.89	0.89
MAE	0.108	0.120	0.15	0.103	0.217	0.13	0.12

4. Discussion

The aim of the present study was to discriminate ADHD and control groups automatically by using time-frequency features of frontal EEG signals in a machine learning model. Various structural and functional brain imaging studies indicated that ADHD is associated with cognitive impairment especially in frontal cortex [39]. Our results indicated that Fz measurements were more discriminative between groups. Thus, we focused on Fz channel to evaluate ADHDs' brain function. We obtained ERPs from ongoing EEG signal by grand average method. Recently, Johnstone et al. [11] reviewed ERP literature in relation to ADHD. Accordingly, reaction time to stimuli, N100, P200, P300 latency and P300 amplitude contributed most to classification. In this current study AEPs were analyzed both in time and frequency domains. In time domain, the features were generated from latency, amplitude and latency/amplitude values of AEPs' P300 component and FD values of original EEG signal via HFD algorithm. For AEP targets, we observed smaller P300 amplitudes and longer latencies in ADHD group. We included Latency/Amplitude value as a feature due to its high capability in the discrimination of both groups. In FD calculation relative to the control group, the ADHD group demonstrated reduced complexity. These findings are consistent with previous research and confirm the theory of neurological diseases reduce complexity of brain functions [40]. In time-frequency domain entropy of DWT coefficients were selected as another feature data set for the discrimination of healthy control and ADHD patients. Finally, we combined statistically significant features in order to explore a robust classification.

Various machine learning techniques were introduced for the classification of ADHD and healthy controls. In literature SVM and k-NN are two classifiers that are commonly used for classifying ADHD and healthy control as seen in Table 4. In this study in addition to these classifiers, LR, NB, RF, Adaboost and MLP classifiers were also implemented and their performances were compared with each other.

In Table 4 we compared our results with diagnostic classification studies in ADHD that used EEG/ERP signals.

As seen in Table 4, Tenev et al. [39] (2014) used power spectra of EEG measurements in different conditions and, Mueller et al. [13] (2010) decomposed ERP signals into independent components and extracted features were used for SVM classification. In our study, time-frequency features were used for classification and the success rate was higher than both studies. This is maybe due to the fact that we have used smaller sample size compared to these studies. In order to make a robust comparison we need to test our data with larger data sets. Nazhvani et al. [41] (2013) used wavelet coefficients, Ghassemi et al. [42] (2010) used several frequency features that were extracted from different independent components. Their success rate was high (92%) but the number of participants for both groups was quite different. In our study the number of participants for both groups was equal. In another study by Ghassemi and Moradi [43] (2012), wavelet and FD results were used as inputs to k-NN. Mohammadi et al. [16] (2016) used non-linear features of EEG signal and employed MLP for classification. They obtained a high accuracy. Sadatnezhad et al. [44] (2011) could classify these two groups with 86% success rate. Ahmadlou and Adeli [47] (2010) obtained higher accuracy (96%) but their ADHD and control sample size are quite different in order to make a robust evaluation of classifier success. Recently, deep learning (DL) techniques has been utilized for ADHD detection. Chen et al. [50] combined EEG based brain network with convolutional neural network (CNN) to classify ADHD and obtained 94.67% accuracy on test data. They trained three different models (MLP, SVM and CNN) with three dataset and compared the performance of the models appropriately. CNN models had better performance than MLP and SVM models. In another study, Vahid et al. applied deep learning methods on event-related EEG data and classified ADHD and control group with 83% accuracy [51]. There are also deep learning studies conducted with fMRI for the discrimination of ADHD [51,52]. These studies show deep learning approaches are promising to detect ADHD.

In this study we have applied various machine learning techniques to discriminate ADHD and healthy controls. The most common problem in machine learning is overfitting that reduces the generalizability of the model. In order to avoid overfitting; we reduced feature set before classification, parameters of classifiers were tuned appropriately with grid search algorithm, we included ensemble learning methods (RF, Adaboost) that reduces variance and we prefered simple models since we had a small sample size. We evaluated classifiers' performance with LOSO. During each LOSO iteration, features from all but one subject constituted the training set and the left-out features of one subject constituted the validation set. Among the classifiers we obtained better results with MLP as shown in Table 3. To avoid overfitting we structured a simple MLP with one hidden layer with 4 neurons. A higher number of hidden layers increases the model complexity and training time and as known complex ANNs are more prone to overfitting.

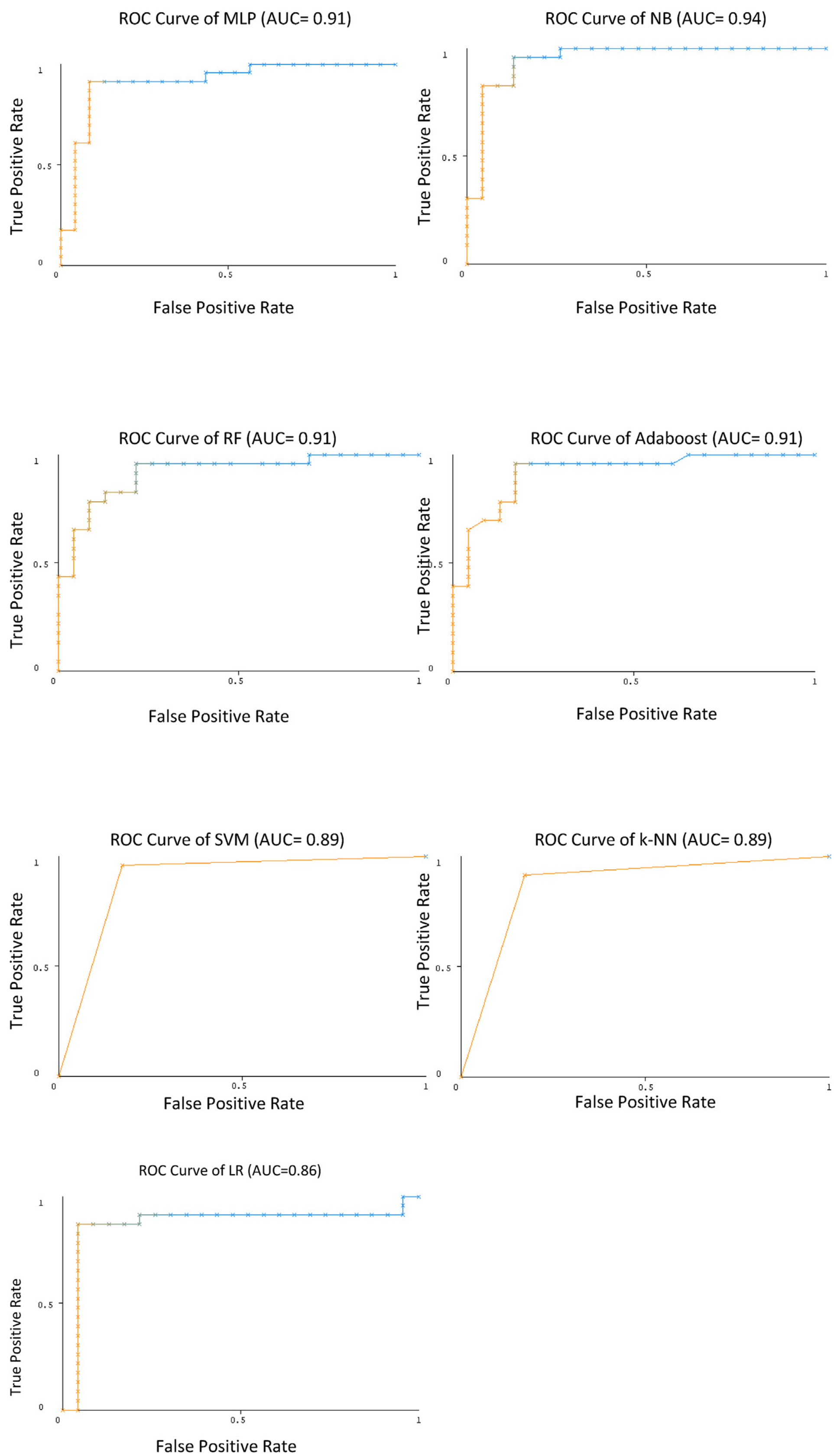

Fig. 4 – ROC curves of classifiers.

Table 4 – Comparison of our results to the similar methods in literature.

Ref.	Participants	Source	Methods	Classifier	Accuracy
[39]	117 (67 ADHD, 50 HC)	VCP, ECP	PSD	SVM	82.3%
[13]	148 (74 ADHD,74 HC)	ERP	ICA	SVM	90%
[40]	36 (12 ADHD, 12 HC, 12 BMD)	VEP	WT	k-NN	92.5%
[41]	50 (10 ADHD,40 HC)	ERP (VCP)	ICA	k-NN	92%
[42]	66 (35 ADHD, 31 HC)	AEP	WT, FD	k-NN	88%, 96%
[43]	60 (30 ADHD, 30 HC)	EEG	NA	MLP	92.28%
[44]	43 (21 ADHD, 22 BMD)	EEG	FD, WT	XCSF XCSF-LDA	78.55%, 86.44%
[45]	30 (ADHD, HC)	EEG	PCA	k-NN	83.33%
[46]	99 (52 ADHD, 47 HC)	EEG	PSD	SVM	70%
[47]	54 (47ADHD, 7 HC)	EEG	WT	RBF-NN	96%
[48]	38 (28 ADHD, 10 HC)	EEG	PSD	NN	70%
[49]	144 (52 ADD, 48 ADHD, 44 HC)	ERP	NM	CNN	83%
[50]	101 (50 ADHD, 51 HC)	EEG	DF,Brain Network Measures	CNN MLP SVM	94.67%, 84.53%, 84.17%
Our Study	46 (23 ADHD, 23 HC)	EEG/ERP	Morphologic Features, Non-linear Features, Wavelet	SVM k-NN RF Adaboost MLP NB LR	89.13%, 86.95%, 82.6%, 89.13%, 91.3%, 86.95%, 89.13%

HC: healthy control; VCP: visual continuous performance test; ECP: emotional continuous performance test; CSD: Current Source Density; NA: nonlinear analysis; NN: neural networks; BMD: bipolar mood disorder; LDA: Linear Discriminant Analysis; PCA: Principal Component Analysis, units, WT: wavelet transform; CSD: Current Source Density; DF: Deep Features; CNN: convolutional neural network; NM: Neurophysiological Measures.

Compared with traditional ADHD diagnosis methods, our method has the following advantages; 1. The source was EEG signal recorded during an auditory oddball task which is noninvasive, economical, objective, and used routinely. 2. AEPs were analyzed both in time and frequency domains rather than studying only in the time domain or in frequency domain provided appropriate, physiologically relevant, and information rich biomarkers. 3. A good discrimination of ADHD and healthy control with up to 91.3% accuracy was obtained with the use of only 6 features. 4. Many machine learning techniques were introduced for classifying ADHD and healthy control. Some classifiers were used first time to distinguish these group. 5. The method had extremely fast discrimination speed (0.03 s) and satisfactory high classification accuracy (91.3%). 6. We used an easy, short task that can be appropriate for ADHD children who have difficulty in sustaining attention to effortful, long cognitive tasks. 7. MLP could classify groups with 78.26% accuracy with only AEP time domain features (latency, amplitude, latency/amplitude). Using FD and wavelet features in addition to these features increased the accuracy by 13.04%.

Finally, to evalute our models, we did not do any preprocessing feature selection using data from the test-set and did not tuned the parameters. 80.4%, 82.6%, 78.26, 82.6,80.4, 76.08, 78.26 accuracies were observed for SVM, NB, k-NN, MLP, RF, Adaboost and LR models respectively. The results show that the models accuracies increase with feature selection by 2.2–13.05% range.

4.1. *Potential clinical significance*

ADHD is one of the most prevalent developmental disorders in children. The disorder negatively affects social and academic skills and often persist into adulthood. Currently ADHD is diagnosed by clinicians via interview-based (rated by parents or teachers) evaluation of the DSM-IV or DSM V criterion. Thus, the clinical diagnosis of ADHD often entails subjective evaluation [53].

Considering its high prevalence, large economic and societal costs, and the negative effects on the future lives of the children, more objective approaches, preferably based on biomarkers would be extremely valuable. Neuroimaging studies have presented new approaches for objective diagnosis of the ADHD. Numerous research mainly focused on EEG and ERP signals to differentiate ADHD patients from controls. Loo et al. [12] (2005) reported the clinical utility of EEG in ADHD in details.

ADHD demonstrated cognitive impairments during a wide range of tasks. Monden et al. [54] (2015) claimed alternating tasks such as oddball paradigms or go/no-go blocks are more appropriate for ADHD participants. Tasks including rest period is not appropriate for ADHD children because they may exhibit unexpected movements and hyperactive behaviors during resting period. Appropriately, we used auditory oddball paradigm and did not adopt rest period. We selected an easy task that can be appropriate for ADHD children who have trouble performing mental arithmetic or motor imagery. Because long attentional tasks may not be appropriate for ADHD children, we chose shorter auditory task that contains only 160 stimuli. In short, this EEG based method offer clinical advantage for the diagnosis of ADHD children: It is simple, robust and applicable to ADHD children as young as 7 years old.

4.2. *Limitations*

There are several limitations in this present study. First, small sample size limits the generalizability of our results. In machine learning small-sample models tend to report higher classification accuracy. Although we applied seven different machine learning algorithm and obtained high accuracies, a larger sample size would be more robust and preferable in feature exploration. Second, we could not evaluate difference between subtypes of ADHD because our sample size was small. With larger sample sizes it would be possible to test if they differ in terms of neurological measures. Third we evaluated only four locations in EEG including Fz, Cz, Pz and Oz. To assess regional differences more channel measurements (Fp1, Fp2) are needed. Additionally, although we tested hearing functions with Rinne and Weber test, there could be difference in hearing threshold between subjects. Despite of these limitation the results of this study are highly promising.

5. Conclusion

This study presents a machine learning model to support the early and objective diagnosis of ADHD that is frequently encountered childhood neurodevelopmental disorder. Our findings suggested that combining time and frequency features obtained from EEG and AEPs might be a useful and discriminative tool for diagnostic purposes of ADHD. In addition, this approach might bring a new perspective to the diagnosis of ADHD and the findings would be helpful in the regulation of therapy.

As future work, we plan to continue acquiring more data to study the electrical brain activity of healthy controls and ADHD patients. In addition, we will also expand feature vectors that can increase the accuracy of the classification and develop the software that will be readily available to the physicians which can further be used in the evaluation of the efficacy of treatment.

Author statement

Conceptualization: Nazan Dolu, Miray Altınkaynak
Data curation: Ferhat Pektas, Nazan Dolu
Formalanalysis: Nazan Dolu, Esra Demirci, Sevgi Özmen
Fundingacquisition: Nazan Dolu
Investigation: Miray Altınkaynak, Aysegül Güven
Methodology: Miray Altınkaynak, Aysegül Güven
Project administration: Nazan Dolu
Resources: Nazan Dolu, Miray Altınkaynak
Software: Miray Altınkaynak, Aysegül Güven
Supervision: Aysegül Güven, Meltem Izzetoglu, Sevgi Özmen, Esra Demirci
Validation: Esra Demirci, Ferhat Pektas, Aysegül Güven
Visualization: Ferhat Pektas, Meltem Izzetoglu
Roles/Writing – originaldraft: Miray Altınkaynak
Writing – review & editing: Miray Altınkaynak, Aysegül Güven, Meltem Izzetoglu, Nazan Dolu, Sevgi Özmen

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgement

This study was supported by the TUBITAK under project number 114S470.

REFERENCES

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edition. 1994 (DSM-IV).
[2] Lazzaro I, Gordon E, Li W, Lim CL, Plahn M, Whitmont. et al. Simultaneous EEG and EDA measures in adolescent attention deficit hyperactivity disorder. Int J Psychophysiol 1999;34:123–34.
[3] Peng X, Lin P, Zhang T, Wang J. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLOS ONE 2013;8(11):e79476.
[4] Paloyelis Y, Mehta MA, Kuntsi J, Asherson P. Functional magnetic resonance imaging in attention deficit hyperactivity disorder (ADHD): a systematic literature review. Expert Rev Neurother 2007;7(10):1337–56.
[5] Sato JR, Hoexter MQ, Castellanos XF, Rohde LA. Abnormal brain connectivity patterns in adults with ADHD: a coherence study. PLoS ONE 2012;7(9):e45671.
[6] Monden Y, Dan H, Nagashima M, Dan I, Kyutoku Y, Okamoto M, et al. Clinically-oriented monitoring of acute effects of methylphenidate on cerebral hemodynamics in ADHD children using fNIRS. Clin Neurophysiol 2012;123:1147–57.
[7] Silk TJ, Vance A, Rinehart N, Bradshaw JL, Cunnington R. White matter abnormalities in attention-deficit/hyperactivity disorder: a diffusion tensor imaging study in adult patients. Hum Brain Mapp 2009;30(9):2757–65.
[8] Román AC, González AC, Hernández EP, Unturbe FM, Alonso TO, Marquéz JG. The attentional effect in attention deficit/hyperactivity disorder (ADHD) by magnetoencephalography (MEG). Clin Neurophysiol 2008;119(9):e153–4.
[9] Gonzalez JJ, Méndez LD, Mañas S, Duque MR, Pereda E, De Vera L. Performance analysis of univariate and multivariate EEG measurements in the diagnosis of ADHD. Clin Neurophysiol 2013;124(6):1139–50.
[10] Szuromi B, Czobor P, Komlosi S, Bitter I. P300 deficits in adults with attention deficit hyperactivity disorder: a meta-analysis. Psychol Med 2011;41(7):1529–38.
[11] Johnstone SJ, Barry RJ, Clarke AR. Ten years on: a follow-up review of ERP research in attention-deficit/hyperactivity disorder. Clin Neurophysiol 2013;124:644–57.
[12] Loo SK, Makeig S. Clinical utility of EEG in attention-deficit/hyperactivity disorder: a research update. Neurotherapeutics 2012;9(3):569–87.
[13] Mueller A, Candrian G, Kropotov JD, Ponomarev VA, Baschera GM. Classification of ADHD patients on the basis of independent ERP components using a machine learning system. Nonlinear Biomed Phys 2010;4(Suppl. 1):S1.
[14] Ponomarev VA, Mueller A, Candrian G, Grin-Yatsenko VA, Kropotov JD. Group Independent Component Analysis (gICA) and Current Source Density (CSD) in the study of EEG in ADHD adults. Clin Neurophysiol 2014;125:183–97.
[15] Heinrich H, Dickhaus H, Rothenberger A, Heinrich V, Moll GH. Single-sweep analysis of event-related potentials by wavelet networks methodological basis and clinical application. IEEE Trans Biomed Eng 1999;46(7):867–79.
[16] Mohammadi MR, Khaleghi A, Nasrabadi A, Rafieivand S, Begol M, Zarafshan H. EEG classification of ADHD and normal children using non-linear features and neural network. Biomed Eng Lett 2016;6:66–73.
[17] Sohn H, Kim I, Lee W, Peterson BS, Hong H, Chae JH, et al. Linear and non-linear EEG analysis of adolescents with attention-deficit/hyperactivity disorder during a cognitive task. Clin Neurophysiol 2010;121:1863–70.
[18] Alba G, Pereda E, Mañas S, Méndez LD, González A, González JJ. Electroencephalography signatures of

attention-deficit/hyperactivity disorder: clinical utility. Neuropsychiatr Dis Treat 2015;11:2755–69.
[19] Esteban FJ, Beltrán LD, Di Ieva A. The fractal geometry of the brain. New York: Springer-Verlag; 2016.
[20] Sebastián MV, Navascués MA. A relation between fractal dimension and Fourier transform–electroencephalographic study using spectral and fractal parameters. Int J Comput Math 2008;85:657–65.
[21] Davila CE, Srebro R. Subspace averaging of steady-state visual evoked potentials. IEEE Trans Biomed Eng 2000;47 (6):720–8.
[22] Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007;118(10):2128–48.
[23] Bush G, Valera EM, Seidman LJ. Functional neuroimaging of attention-deficit/hyperactivity disorder: a review and suggested future directions. Biol Psychiatry 2005;57:1273–84.
[24] Kesic S, Spasic SZ. Application of Higuchi's fractal dimension from basic to clinical neurophysiology: a review. Comput Methods Programs Biomed 2016;133:55–70.
[25] Higuchi T. Approach to an irregular time series on the basis of the fractal theory. Physica D 1988;31:277–83.
[26] Accardo A, Affinito M, Carrozzi M, Bouquet F. Use of the fractal dimension for the analysis of electroencephalographic time series. Biol Cybern 1997;77:339–50.
[27] Quiroga RQ, Sakowitz OW, Basar E, Schürmann M. Wavelet transform in the analysis of the frequency composition of evoked potentials. Brain Res Protoc 2001;8:16–24.
[28] Brandley AP, Wilson WJ. On wavelet analysis of auditory evoked potentials. Clin Neurophysiol 2004;115:1114–28.
[29] Boser BE, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory; 1992. p. 144–52.
[30] Horton P, Nakai K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. ISMB-97 Proceedings; 1997. p. 147–52.
[31] Rish I. An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3(22). 2001. pp. 41–6.
[32] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation parallel distributed processing: explorations in the microstructure of cognition, Volume 1: Foundation. MIT Press; 1986.
[33] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7.
[34] Kuncheva LI. Combining pattern classifiers: methods and algorithms. John Wiley & Sons; 2014.
[35] Menard S. Applied logistic regression. 2nd ed. SAGE; 2002, ISBN 978-0761922087.
[36] Cawley GC, Talbot NLC. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. Neural Netw 2004;17:1467–75.
[37] Smeeton NC. Early history of the kappa statistic. Biometrics 1985;41:795.
[38] Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol 2011;1:37–63.
[39] 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 2014;93:162–6.
[40] Goldberger AL, Peng CK, Lipsitz LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging 2002;23:23–6.
[41] Nazhvani AD, Boostani R, Afrasiabi S, Sadatnezhad K. Classification of ADHD and BMD patients using visual evoked potential. Clin Neurol Neurosurg 2013;115:2329–35.
[42] Ghassemi F, Moradi MM, Doost MT, Abootalebi V. Classification of ADHD normal participants using frequency features of ERP's independent components. Proceedings of the 17th Iranian Conference of Biomedical Engineering (ICBME2010); 2010.
[43] Ghassemi F, Moradi MM. Using non-linear features of EEG for ADHD/normal participants classification. Proc Soc Behav Sci 2012;32:148–52.
[44] Sadatnezhad K, Boostani R, Ghanizadeh A. Classification of BMD and ADHD patients using their EEG signals. Expert Syst Appl 2011;38(3):1956–63.
[45] 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.
[46] Lenartowicz A, Delorme A, Walshaw PD, Cho AL, Bilder RM, McGough JJ, et al. Electroencephalography correlates of spatial working memory deficits in attention-deficit/hyperactivity disorder: vigilance, encoding, and maintenance. J Neurosci 2014;34:1171–211.
[47] Ahmadlou M, Adeli H. Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD. Clin EEG Neurosci 2010;41:1–10.
[48] Alba-Sanchez F, Yanez-Suarez O, Brust-Carmona H. Assisted diagnosis of attentiondeficit hyperactivity disorder through EEG bandpower clustering with self-organizing maps. Conf Proc IEEE Eng Med Biol Soc 2010;2447–50.
[49] Vahid A, Bluschke A, Roessner V, Stober S, Beste C. Deep learning based on event-related EEG differentiates children with ADHD from healthy controls. J Clin Med 2019;8(7):1055.
[50] Chen H, Song Y, Li X. A deep learning framework for identifying children with ADHD using an EEG-based brain network. Neurocomputing 2019;356:83–96.
[51] Kuang D, He L. Classification on ADHD with deep learning. International Conference on Cloud Computing and Big Data. 2014. pp. 27–32.
[52] Riaz A, Asad M, Alonso E, Slabaug G. DeepfMRI: end-to-end deep learning for functional connectivity and classification of ADHD using fMRI. J Neurosci Methods 2020;335.
[53] Thomas CG, Johnson LG. ADHD: is objective diagnosis possible? Psychiatry (Edgmont) 2005;2(11):44–53.
[54] Monden Y, Dan I, Nagashima M, Dan H, Uga M, Ikeda T, et al. Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS. Neuroimage Clin 2015;9:1–12.