

Machine-based classification of ADHD and nonADHD participants using time/frequency features of event-related neuroelectric activity



Hüseyin Öztoprak^{a,*}, Mehmet Toycan^a, Yaşar Kemal Alp^b, Orhan Arıkan^c, Elvin Doğutepe^{d,e}, Sirel Karakaş^{d,e}

^a Electric Electronic Engineering Department, Cyprus International University, Lefkosa, Turkish Republic of Northern, Cyprus

^b Radar, Electronic Warfare and Intelligence Systems Division, Aselsan, Ankara, Turkey

^c Bilkent University, Department of Electrical Engineering, 06533 Bilkent, Ankara, Turkey

^d Neurometrika Medical Technologies LLC, 06800 Ankara, Turkey

^e Doğuş University, Department of Psychology, 34722 Kadıköy, İstanbul, Turkey

ARTICLE INFO

Article history:

Accepted 2 September 2017

Available online 30 September 2017

Keywords:

Attention-deficit/hyperactivity disorder (ADHD)

Time-frequency Hermite atomizer

Machine learning

Classification

Feature selection

Support vector machine-recursive feature elimination (SVM-RFE)

HIGHLIGHTS

- TFHA extracts key ADHD biomarkers from ERP signals recorded during Stroop task.
- Feature selection with SVM-RFE leads to excellent classification performance.
- Patients with ADHD were best discriminated via the delta oscillations.

ABSTRACT

Objective: Attention-deficit/hyperactivity disorder (ADHD) is the most frequent diagnosis among children who are referred to psychiatry departments. Although ADHD was discovered at the beginning of the 20th century, its diagnosis is still confronted with many problems.

Method: A novel classification approach that discriminates ADHD and nonADHD groups over the time-frequency domain features of event-related potential (ERP) recordings that are taken during Stroop task is presented. Time-Frequency Hermite-Atomizer (TFHA) technique is used for the extraction of high resolution time-frequency domain features that are highly localized in time-frequency domain. Based on an extensive investigation, Support Vector Machine-Recursive Feature Elimination (SVM-RFE) was used to obtain the best discriminating features.

Results: When the best three features were used, the classification accuracy for the training dataset reached 98%, and the use of five features further improved the accuracy to 99.5%. The accuracy was 100% for the testing dataset. Based on extensive experiments, the delta band emerged as the most contributing frequency band and statistical parameters emerged as the most contributing feature group.

Conclusion: The classification performance of this study suggests that TFHA can be employed as an auxiliary component of the diagnostic and prognostic procedures for ADHD.

Significance: The features obtained in this study can potentially contribute to the neuroelectrical understanding and clinical diagnosis of ADHD.

© 2017 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Biological signals are a challenge to analyze. Among these signals, the electrophysiological activity of the human brain during cognitive processing is undoubtedly the most challenging. For decades, neuroscientists have been studying the brain-mind relation-

ship based on the electrophysiological responses of the brain. Most studies have been performed on signals in the time domain in the form of event-related potentials (ERPs). The starting point of a second approach was based on the principle that complex signals are composed of oscillatory responses of different frequencies (for a review, see Başar, 2011; Karakaş and Barry, 2017; Karakaş and Başar, 2004). This approach gained significant momentum with the introduction of new techniques that decompose complex sig-

* Corresponding author.

E-mail address: hoytoprak@ciu.edu.tr (H. Öztoprak).

nals into their oscillatory components within the time and frequency domains (Karakaş and Arıkan, 2006).

Attention-deficit/hyperactivity disorder (ADHD) has attracted significant attention because of its high incidence (0.2–12.2%) in children and because it is the most frequent diagnosis among children who are referred to psychiatry departments (Durukan et al., 2011; Rowland et al., 2002). However, interest in ADHD extends beyond its high incidence. Although this disorder was discovered at the beginning of the 20th century, it still remains under medical, psychological and social investigation (Cooper, 2001). Due to conflicting findings in the literature and the numerous theories concerning ADHD, our understanding of this condition remains unclear, and its diagnosis is confronted with many problems (for a review, see Karakaş, 2008).

The diagnostic procedure for psychiatric disorders is primarily based on empirically observed behaviour. Supportive evidence is primarily derived from psychometric test scores, which also essentially involve the observation of behaviour. The neuroelectricity of ADHD has mainly been studied in the time domain via ERPs or through the spontaneous activity of the brain (Robaey et al., 1992; Smith et al., 2003).

The study by Barry et al. (2003) is among the few that have studied oscillatory activity in relation to ADHD diagnoses. Others include the studies of Berdakh and Jinung (2012), who found that the power ratios between the frequency bands, specifically that between the theta and beta powers (theta-beta ratio: TBR), may be used for ADHD diagnosis. Recently, the Food and Drug Administration (FDA) approved the TBR as a tool to help healthcare providers diagnose ADHD. However, several recent publications have questioned this approach. One particular criticism concerns the use of fixed cut-off frequencies between the frequency bands. Saad et al. (2015) recommended that these frequencies should be determined individually for each patient. Moreover, studies that include neuroelectricity in clinical protocols are steadily arriving as biomarkers for neuropsychiatric disorders such as Alzheimer's disease and bipolar disorder (Başar et al., 2013).

Our group has previously studied the ERPs in ADHD and decomposed them into their oscillatory components using the time-frequency component analyser (TFCA; Alp et al., 2008; Özdemir et al., 2005). Preliminary work that tested the TFCA in healthy adults during wakefulness and sleep found high-resolution time-frequency signals with negligible cross-term contamination (Karakaş et al., 2006a,b; Tüfekçi et al., 2006). Subsequently, the Time Frequency Hermite Atomizer (TFHA; Alp and Arıkan, 2012) technique was developed and has provided improved characterization of the oscillatory components of ERPs that form the foundation of the investigation presented in this work. The application of machine-learning techniques to oscillatory activities (e.g., Ahmadlou and Adeli, 2010) for the classification of ADHD and non-ADHD control participants is a research area that needs attention.

An approach for discriminating ADHD from nonADHD participants is the use of machine-learning techniques (Anuradha et al., 2010; Tenev et al., 2014). Such studies markedly differ from each other in terms of the presence (or type) of the psychometric tasks used for measuring cognitive processes, the signal-processing techniques, and classification techniques. Mueller et al. (2010) measured ERP signals during a go/no-go task and extracted components by applying independent components analysis (ICA) to the outputs of a Support Vector Machine (SVM) algorithm. The average accuracy was found to be 92%. When the SVM algorithm module utilized radial basis kernels to achieve an automatic diagnosis of ADHD, the accuracy was 88% (Anuradha et al., 2010). Ahmadlou and Adeli (2010) reported a classification rate of 95.6% between an ADHD and a normal group using wavelet transformation domain classification with neural networks. Tenev et al. (2014) used a logical expression to create four SVM classifiers that

were trained with signals measured under different experimental conditions. The model yielded a classification accuracy of 82.3%. Berdakh and Jinung (2012) proposed a decision support system that uses a maximal discrepancy criterion to select the most distinguishing features for ADHD diagnosis. In their work, an SVM classifier was trained in a semi-supervised fashion to identify robust markers of electroencephalogram (EEG) patterns for accurate discrimination. The maximum accuracy was 97%. The accuracy rates that are reported for the foregoing studies vary between 82.3% and 97%. This variation may be due to the experimental conditions, such as the size of the population, the characteristics of the participants and the technique used for the training and testing of the proposed classifiers.

This work aims to contribute to computer-based decision-making regarding ADHD diagnosis. Specifically, the clinical utility of the features extracted by TFHA from ERP measurements necessitates special interest. The goals of the present study were the following: (1) to develop an approach that uses the output of an advanced time-frequency analysis technique and (2) to identify the biomarkers that best discriminate ADHD participants from nonADHD participants.

As illustrated in Fig. 1, the present study was basically conducted in three phases: (1) data acquisition, (2) feature extraction, and (3) classification. The data acquisition and feature extraction phases of the present study were conducted in a previously completed project (Karakaş et al., 2006c), and the data have been published in various other studies (Alp and Arıkan, 2012; Alp et al., 2008; Karakaş and Arıkan, 2006; Karakaş and Başar, 2006; Karakaş et al., 2006a,b; Tüfekçi et al., 2006).

2. Material and methods

2.1. Participants

The present study used the neuroelectrical data from a multi-centre, large-scale project (HUA-F-BAB 2006K-120-640-06-08; Karakaş et al., 2006c) in which multiple technologies were used for recording and data analysis. The original data were obtained from 70 boys in an ADHD group and 38 boys in an age-matched control group. Across the different experimental conditions, the sample size of the present study varied between 37–44 boys in the ADHD group and 32–38 boys in the healthy control group. The boys were 6–12 years old and were attending grades 1–6.

The clinical group consisted of children with ADHD of the combined (ADHD-C) subtype. The diagnoses were performed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association, 2000). The patients were unmedicated first referrals. They had not been previously diagnosed with ADHD and were not on any drug therapy with possible cognitive effects. Comorbidity was examined using the Schedule for Affective Disorders and Schizophrenia for School Age Children, Present and Lifetime version (K-SADS-PL; Gökler et al., 2004; Kaufman et al., 1997). Children with psychiatric or neurological comorbidities (e.g., oppositional defiant disorder or specific learning disabilities), clinical levels of anxiety or depression, and uncorrected visual or hearing defects were not included in the research sample. The children in both groups were at ages that were typical for their grade levels. The intelligence levels of both groups were within normal limits (IQ range = 90–129).

The parents of both the clinical and control groups were informed about the nature of the study. Parents who accepted their child's participation in the study signed a standard informed consent form. The study also required the assent of the cases/participants. The study was approved by the Hacettepe University

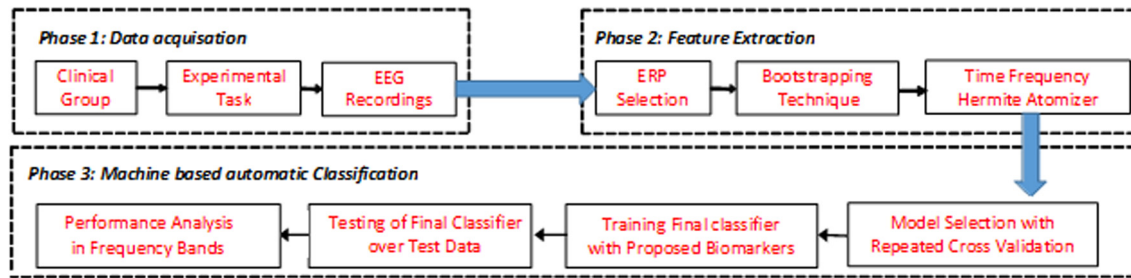


Fig. 1. Phases of the research protocol.

Ethics Committee and Gazi University Ethics Committee (HEK 05/13-32-Medical, Surgical and Drug Research Ethical Committee) and was conducted in compliance with the Declaration of Helsinki and the principles set forth by the Ministry of Health of the Turkish Republic for clinical studies.

2.2. Task procedures

Neuroelectrical responses were obtained as the participant performed the tasks required by the Stroop task of the TURCONS Neuropsychological Mapping Battery for Functional Magnetic Resonance Imaging (Karakas et al., 2013). The Stroop task measures complex attention, which is an executive function known to be affected in participants with ADHD (Barkley, 1997; MacLeod, 1991). The task blocks consisted of colour names that are congruently printed (e.g., the word “blue” printed in blue) or incongruently printed (e.g., the word “blue” printed in yellow). The stimuli were presented in 4 blocks. There were 15 words in each block (15 words \times 4 blocks = 60 stimuli). The task duration was approximately 4 min. The participants were asked to press the response button assigned to their index fingers when the colour word was congruent with the ink colour and to press the button assigned to their middle fingers when the colour and ink of the word were incongruent.

The independent variable was stimulus congruence (congruent/incongruent). The behaviourally dependent variable was response accuracy. Accuracy was classified as correct (types: hit/true positive, correct rejection/true negative) or incorrect (types: miss/false negative or false alarm/false positive).

2.3. Acquisition of the electrophysiological data

EEG recordings were collected in a Faraday cage (Lindgren) that attenuated the electrical artefacts and, due to its anechoic structure, prevented acoustic interference. The recordings were obtained through an electrode cap (Quik Cap 64) with 64 Ag-Ag/Cl sintered electrodes located according to the 10–10 system (reference: combined mastoids). As a preliminary approach, the present study analyzed data from only the midline electrodes and the bilateral F3 and F4 electrodes. No notch filter was activated. The EEG recordings were pre-amplified and filtered between dc and 100 Hz (fixed sampling rate: 1000 Hz). The impedance for all electrode sites was 3 Kohms or less. The trials in which the EEG exceeded $\pm 50 \mu\text{V}$ were automatically rejected online.

Bipolar recordings of the electro-ocular (EOG) activity and submental electromyographic (EMG) activity were used for artefact rejection (Compumedics Neuroscan-Scan 4-3). Continuous EEG recordings were studied for blink artefacts. The beginning and end of the blinks (approximate durations in the range of 200–400 s) were manually selected offline and marked as rejected regions. The averages were calculated for blinks, and the compo-

nents that formed the average blink were calculated using singular value decomposition. These components were subtracted from all instances of contaminated EEGs using a spatial filtering algorithm. The ERPs were then studied for muscular artefacts. The sections with muscular artefacts were selected offline and rejected from further analysis.

Artefact-free 1700-ms ERP epochs were manually selected from the EEG recordings. Each epoch consisted of a 500-ms pre-stimulus and a 1200-ms post-stimulus interval. The grand averages of the ERP recordings were calculated from the participant averages by the bootstrapping technique to control for and check the stability of the results. One great advantage of the bootstrap technique is its simplicity. This technique is a straightforward method for deriving estimates of the standard errors and confidence intervals for estimators of important statistical descriptors of the distribution, such as the percentile points, proportions, odds ratios, and correlation coefficients. Although, for most problems, it is impossible to know the true confidence interval, bootstrapping is asymptotically more accurate than the standard intervals obtained using the sample variance and assumptions of normality (Efron, 1987). For this process, a group of participants half the size of the total sample was randomly selected, and a grand average ERP was calculated. The participants were replaced, and another random selection and grand average calculation were performed. Attempts at obtaining summary statistics were repeated 20 times for each level of the independent variable (i.e., congruent vs. incongruent words).

2.4. Feature extraction via TFHA

The oscillatory components in the ERP signal were analyzed in the time and frequency domains using the Time-Frequency Hermite Atomizer (TFHA) technique (Alp and Arıkan, 2012). The TFHA technique uses Hermite-Gaussian functions as the basis of the decomposition of the signal components. Because the Hermite-Gaussian functions have optimal concentration properties, the identified ERP signal components in the delta, theta, beta, alpha, and gamma frequency bands are all high-resolution and highly localized in the time and frequency domains. The TFHA involves sequential extraction of the ERP signal components until the residuals of the reconstructed signal become negligible.

TFHA was applied to the different levels of stimulus congruence (congruent and incongruent), response accuracy (correct and incorrect), and recording sites (Fz, Cz, Pz, F3, and F4). Time-frequency components (delta, theta, alpha, beta and gamma) were studied in selected time windows (early, middle, and late).

In TFHA, the components are described according to 16 parameters that are grouped under four headings:

Wigner Distribution/Support Parameters (WD/Support): These are the temporal limits (start–end) and frequency limits (initial: freqi, final: freqf) of the extracted t–f components.

Wigner Distribution/Peak Parameters (WD/Peak): These are the time point at which the extracted t-f components reach their maximum values (tpeak), the frequency at this point (fpeak), the attained maximum value (value), and the total energy of the extracted t-f component.

Time Domain Signal/Peak Parameters (TD/Peak): These are the time points at which the ERP of the extracted component reaches its absolute peak voltage (TD-tpeak), the frequency at this point (TD-fpeak), the attained maximum value of the component in the frequency domain (TD-fvalue), and the energy of the extracted component.

Wigner Distribution/Statistical Parameters (WD/Statistical): These are the time and frequency centres of the component (t-c and f-c) and the time and frequency deviations of the component (t-std and f-std).

2.5. Machine-based automatic classification

The present paper attempted to classify the participants into ADHD and nonADHD groups using an automated approach that employed a computer algorithm based on the well-known SVM machine learning technique (Cristianini and Shawe-Taylor, 2000; Onton and Makeig, 2006).

2.5.1. Machine learning algorithms

The standard steps for classifying extracted components are as follows: (1) model selection, (2) training the site-specific classifiers using optimal feature lists, and (3) predicting the unknown class labels (ADHD and nonADHD). To establish an unbiased experimental setup, (1) and (2) should be exclusively performed with the training dataset, and (3) should be exclusively performed with the test dataset. As illustrated in Fig. 1, the classifier in this work was developed in accordance with this separation. In our setup, the training and testing datasets contain 30 and 10 samples, respectively. The sets are stratified with respect to the illness condition.

Several different approaches involving standard machine learning algorithms were used to demonstrate the selective effect of the feature selection method on classification performance. The ultimate goal was to arrive at a robust classification of the clinical and nonclinical control participants. Simplicity is obviously another important classifier property because a simple classifier is likely to be more comprehensible even via visual inspection and more generalizable because of the reduced risk of over-adaptation.

Based on its classification performance and straightforward implementation, SVM was identified as the classifier of choice. The important features of SVM are the absence of local solutions, a well-controlled capacity of the solution, and the ability to efficiently handle multidimensional input data (Cristianini and Shawe-Taylor, 2000; Onton and Makeig, 2006). In this study, the input to the SVM was an n -sized feature set that was extracted from the ERP data using TFHA. The feature set of each participant can be thought of as a point in an n -dimensional space. The current SVM was trained to classify the data points into two classes, i.e., ADHD and nonADHD, which will be referred to as the positive and negative classes, respectively. SVM chooses the hyperplane that provides the plane with the maximum margin between the positive and negative classes. The separating hyperplane is optimized when the distance between the closest data points, called the support vectors, is maximized. For the sake of simplicity, this study chose an SVM classifier with a linear hyperplane.

In the significant majority of the machine learning problems, the classification performance can be improved by employing an appropriate feature selection technique before the classification process. In this study, SVM-Recursive Feature Elimination (SVM-

RFE; Guyon et al., 2002), an algorithm that is embedded in SVM, was chosen as the main feature selection technique. SVM-RFE is a well-studied method that has been demonstrated to be successful in various applications including automated medical diagnosis (Duan et al., 2005; Li et al., 2012).

In selecting the appropriate set of features, the predictive power of the classifier that is trained with the specific features is of paramount importance. High predictive power can be achieved by using unbiased and powerful sampling techniques for splitting the training dataset into training and validation subsets. A widely used technique for assessing how the results of a statistical analysis will generalize to an independent dataset is cross validation (CV; Rosenblatt, 1958). In CV, the data are split k times to estimate the performance of the classifier. Specifically, $k - 1$ splits of the data are used for training the classifier, and the remaining data are used to validate its predictive power. In this work, we set $k = 5$, which corresponds to a testing/training ratio of 25%. This is a frequently applied configuration (Arlot and Celisse, 2010; Pollastri and McIysaght, 2005) because it utilizes most of the dataset for effectively training the classifier yet allocates a significant section of the dataset for precisely testing the trained classifier. A drawback of the classical CV is the restriction of the number of validation samples compared to the total sample size in the training dataset, which is 30 in this study. However, repeating the CV procedure n times with different repartitions (repeated cross validation: RCV) of the splits increases the number of the validation samples by the factor of n (Kim, 2009). In this study, we repeated the CV procedure 5 times.

Employing RCV provides a statistically powerful method of comparing the performances of the feature selection methods. However, the selected features and the consequent classifier might be different for each training RCV split. Therefore, to ensure an impartial setting, our experimental setup will be concluded by applying a final classifier on a separate test dataset. As detailed in Section 3.1, SVM-RFE performed well and consistently in this study; the top features that SVM-RFE produced in each RCV split are globally ranked by an ensemble feature selection technique to help identify the top biomarkers and employ them when designing a final classifier that can be tested with the participants in the testing dataset, which has been totally isolated from the training samples. The classifier to be applied to the independent test set is trained by the features that are obtained by employing RCV and Stability Selection to the internal training dataset.

2.5.2. Feature selection

Feature selection is designed to choose or to combine the features that preserve most of the relevant information and remove the redundant or irrelevant information. This approach is used to improve the efficiency and the robustness of the classifiers. Feature selection is usually difficult for any dataset and becomes even more complex when classification is performed on extracted data in which the features themselves are represented in different spaces and can vary in number over many orders of magnitude.

Feature selection techniques can differ significantly with respect to the assumptions of the technique, the size of the dataset and the size of the feature set. This study employed SVM-Recursive Feature Extraction (SVM-RFE), which is an embedded feature selection method that uses SVM weights to iteratively rank the set of features (Guyon et al., 2002). SVM-RFE uses the potential collaboration between the features because it evaluates the discriminative power of the feature subsets and not the power of the individual features. During the SVM training process, this technique calculates the rank of each feature in the feature subset. Stability Selection, which is an ensemble technique that was introduced by Meinshausen and Bühlmann (2010), performs well in conjunction with SVM-RFE in terms of the stability of the selected features

and the resulting error rate even when the feature set is multidimensional, and the sample size is small (Dernoncourt et al., 2014).

Two alternative univariate techniques, i.e., information gain and chi-square, were used in the present study to test the performance of the SVM-RFE. Overall, the univariate methods compute a specific ranking criterion and sort features independently. Information gain measures the amount of information (in bits) about class prediction if the only information available pertains to the presence of a feature and to the corresponding class distribution. Chi-square evaluates individual independence with respect to classes (Jin et al., 2006). Chi-square compares the number of cases in a class with the expected frequency in that class.

2.5.3. Criteria for testing the classification efficiency

The criteria for classifier efficiency were accuracy rate, error rate, sensitivity and specificity. The accuracy rate corresponds to the ratio of correctly classified participants to the total number of participants:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

where *TP*, *TN*, *FP* and *FN* correspond to the numbers of true positives (hits), true negatives (correct rejections), false positives (false alarms) and false negatives (misses), respectively. When accuracy approaches unity, the results can be interpreted more precisely by the error rate, which is given by the following:

$$\text{ErrorRate} = 1 - \text{Accuracy}.$$

Sensitivity is the performance in terms of correctly classifying the ADHD participants to the ADHD group and is given by the following:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Specificity is the performance in terms of correctly classifying the nonADHD participants to the control group and is given by the following:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

In the next section, the efficiency of the proposed classification technique will be presented according to the above-defined criteria.

3. Results and discussion

This section presents the findings regarding the efficiency of the SVM-RFE technique in classifying the participants as ADHD and nonADHD and provides a methodological discussion of the results. In this section, the performance of SVM-RFE is compared to the performances of the information gain and chi-square techniques, and the performance with only a subset is compared to that of the entire feature set. All analyses pertaining to data-mining were performed with the Waikato Environment for Knowledge Analysis (WEKA) software (Hall et al., 2009).

3.1. Classification efficiency of SVM-RFE

Fig. 2 illustrates the classification efficiency of SVM-RFE and the two univariate methods. The error rates of the two univariate methods were generally similar and were within the 0.05 and 0.30 range. The error rate of SVM-RFE was considerably lower than those of the two univariate methods for all the studied range of features.

Table 1 demonstrates the classification accuracy of the three methods in terms of sensitivity and specificity. The sensitivity (the accuracy in classifying the ADHD cases to the ADHD group)

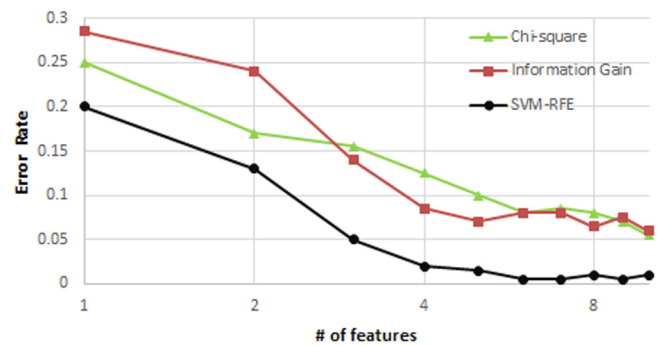


Fig. 2. Performances of the feature selection methods. The entire feature set was composed of 552 features. The number of features ranged from 1 to 10 and are presented in logarithmic base 2 to emphasize the lower range. The results are the averages of 25 RCV splits.

of SVM-RFE was higher than those of the two univariate methods for all numbers of features. When the number of features was ≥ 4 , the sensitivity values were comparable to that of the SVM-RFE. When the number was ≤ 3 , the sensitivity of SVM-RFE became markedly higher. For ≤ 3 features, the sensitivity of SVM-RFE varied between very high (98% for three features) and medium (84% for one feature).

Univariate methods compute a specific ranking criterion and sort the features independently. Accordingly, these techniques cannot evaluate the relationships (redundancy and complementarity) between the features. Thus, the feature subset is not, overall, optimized for a specific classification problem. In such a case, the selected features might not be the optimal feature subset for the classification of, for example, ADHD. The superior performance of SVM-RFE might be a result of its capacity to account for multivariate interactions between the features.

The specificity (the accuracy in the classification of the healthy participants to the nonADHD control group) of SVM-RFE was higher than those of the two univariate methods for all numbers of features. When the number of features was between 4 and 9, the specificity values of SVM-RFE were 100%. Specifically, when the number was ≤ 3 , the difference between the specificity values of the SVM-RFE and the univariate methods became markedly greater. For ≤ 3 features, the specificity of SVM-RFE varied between very high (93% for three features) and medium (74% for one feature). Notably (Table 1), the efficiency of the correct placement of the ADHD participants into the ADHD group (sensitivity) was higher than the correct placement of healthy control participants into the nonADHD group (specificity). These findings point to the necessity of the use of multivariate feature selection methods, such as SVM-RFE, for the accurate classification of ADHD.

The superior classification performance that is indexed by a decreased overall error rate (Fig. 2) and increased sensitivity and specificity (Table 1) demonstrates that SVM-RFE is capable of finding the discriminatory set of features and is able to place the participants in their relevant groups. A number of methodological factors may account for these findings. One contributory factor is the capability of SVM-RFE to account for multivariate interactions between the features. Another factor is the feature extraction method (TFHA) that the present study employed. TFHA offers a signal processing advancement over the presently existing methods. By fitting the most complex structure to a specific time-frequency area, TFHA vertically represents the signal under consideration. Consequently, TFHA provides a basis for the precise extraction of signal components (Alp and Arıkan, 2012). Third, the Stroop task by which the ERPs were triggered is highly relevant for measuring the attention deficit and impulsivity symptoms of ADHD (Bekçi, 2009; MacLeod, 1991). Thus, the task may have contributed

Table 1

Sensitivity and specificity of the feature selection methods.

| # of features | Chi-square | | Information gain | | SVM-RFE | |
|---------------|-------------|-------------|------------------|-------------|-------------|-------------|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| 1 | 78 | 69 | 73 | 74 | 84 | 74 |
| 2 | 87 | 78 | 82 | 72 | 92 | 80 |
| 3 | 92 | 78 | 90 | 82 | 98 | 93 |
| 4 | 95 | 81 | 96 | 87 | 98 | 98 |
| 5 | 95 | 85 | 97 | 91 | 99 | 100 |
| 6 | 95 | 89 | 97 | 87 | 99 | 100 |
| 7 | 96 | 88 | 95 | 89 | 99 | 100 |
| 8 | 96 | 88 | 95 | 90 | 99 | 100 |
| 9 | 97 | 89 | 95 | 90 | 99 | 100 |
| 10 | 97 | 91 | 95 | 92 | 99 | 99 |

Table 2

Feature ranking using stability selection.

| Ranking | Score | Feature |
|---------|-------|---------------------------------|
| 1 | 22 | CZ_11Incorrect_DeltaL_tstd |
| 2 | 19 | F3_22Correct_AlphaE_tc |
| 3 | 17 | FZ_11Correct_BetaE_start |
| 4 | 12 | CZ_22Correct_AlphaE_fpeak |
| 5 | 7 | CZ_22Correct_AlphaE_fc |
| 6 | 5 | F3_22Incorrect_DeltaE_TD-tvalue |
| 7 | 5 | FZ_22Correct_DeltaE_fc |
| 8 | 5 | F3_11Correct_DeltaE_fstd |
| 9 | 5 | F3_11Correct_DeltaE_end |
| 10 | 4 | F3_11Correct_DeltaE_tc |

to the deviation of the signal components collected from the ADHD and nonADHD groups and, consequently, boosted the classification performance.

3.2. Feature ranking using stability selection

Stability selection first discards all but the top d features in each of the 25 feature lists generated by SVM-RFE. The score of each feature in the dataset is then calculated by counting its incidence among all the reduced lists. Because the error rates of the top 5 features for each trial are almost zero, we chose $d = 5$. Table 2 tabulates the 10 features that had the highest overall scores.

The top-ranking feature of the SVM-RFE technique was a WD-statistics (tstd) representation of the delta band response that was obtained as the participants were responding incorrectly to the congruent stimulus. Fig. 3 illustrates this particular component for the ADHD (Fig. 3a) and the nonADHD (Fig. 3b) participants. The classification parameter of TFHA for the top-ranking feature was the standard deviation of the delta component over the time axis. The classification efficiency is demonstrated in Fig. 3 by the time distribution of the delta component: The TFHA component is distributed over the time axis in the ADHD group (Fig. 3a), while it is highly localized in the nonADHD group in the late part of the epoch (Fig. 3b).

The top 10 features presented in Table 2 can be categorized according to the experimental conditions and parameters given in Section 2.4. The feature lists contain the delta, alpha and beta bands. The alpha and beta features have already been identified as biomarkers for ADHD diagnosis. To our knowledge, the delta-band features have not been identified as indicators of ADHD. Table 2 indicates that, of the top features, 6 contained the delta band. In the existing literature (Saad et al., 2015), the frequency bands are generally studied with respect to their band power or the power ratios between the bands. Notably, the powerful delta-band features in this study belonged to the TFHA technique.

Among the other characteristics of the top 10 features are the time intervals at which the TFHA components were located; of

the 10 components, 9 occurred at the early time period. The featured electrodes (3 of the 5) had a left-lateralized frontocentral distribution. The TFHA parameters from all four groups occurred in the top 10 feature list. Five of these parameters belonged to the statistical parameters group, and the remaining features were distributed in the other parameter subgroups. Finally, 8 of the 10 features were obtained as the participants were responding correctly.

As seen from the presented results, there is not much room for improvement via the use of a more sophisticated classifier, such as SVM with non-linear kernels, which can generate non-linear classification boundaries. Using the simple SVM classifier, a more robust generalization can be achieved. The linear SVM also produces a more comprehensible decision criterion.

3.3. Performance of the proposed classifier on the test dataset

The final classifier was trained with the top features listed in Table 2. The number of features should be kept low because a complex model is likely to induce overfitting and consequently cause a decrease in the test performance. Therefore, as the internal validation performance becomes saturated around five features, the final classifier was trained with only the top five features. The normalized coefficients of the hyperplane of the output classifier are given below:

$$\text{SVM_hyperplane} = -1.6 \times f^1 + 1.42 \times f^2 + 1.41 \times f^3 + 0.75 \times f^4 + 0.75 \times f^5 - 1.91$$

where the features (f^1, f^2, f^3, f^4 , and f^5) correspond to the top five features listed in Table 2.

The performance of the classifier designed with RCV and stability selection resulted in 10 correct classifications, which is equivalent to an error rate of zero. This result demonstrates that the feature selection method that combined RCV with stability selection was successful in all the samples in the testing dataset. However, this result should be treated with caution because the number of tested participants was limited to ten.

3.4. Contributions of the top 3 features to ADHD classification

In previous studies, the ratio between two pre-determined features, namely, the theta and beta powers, was explicitly studied as a sign of ADHD. However, in the present study, we did not impose predefined ratios; rather, we used SVM-RFE to determine the features with ratios that indexed ADHD. Using SVM-RFE, it was also possible to study the ratios between different types of TFHA features, such as the starting point and the means of the frequency components.

Fig. 4 contains scatterplots that illustrate the distribution of the ADHD and nonADHD group participants to binary combinations (1,

2; 1, 3; and 2, 3) of the top 3 features in Table 1. In each of these figures, straight lines separate the ADHD group from the nonADHD group. These lines were found via a linear SVM and can be simply expressed using the generic formula, $y = mx + c$, where m is the ratio between the features, and c is a constant.

Feature pairs 1–2 (Fig. 1a) and 2–3 (Fig. 1c) were sufficient to discriminate the groups. Feature pair 1–3 (Fig. 1b) was nearly sufficient to discriminate between the nonADHD and ADHD groups,

excluding a single nonADHD participant whose corresponding features lied within the ADHD cluster. Fig. 4d is a scatterplot of the top 3 features. When all three features were used (Fig. 4d), the ADHD and nonADHD participants were separated with a linear plane with the widest margin among all the studied combinations (Fig. 4a–c).

For comparison purposes, scatterplots for the univariate techniques are also provided. Fig. 5a represents the scatterplot for the information gain, and Fig. 5b presents that for the chi-square.

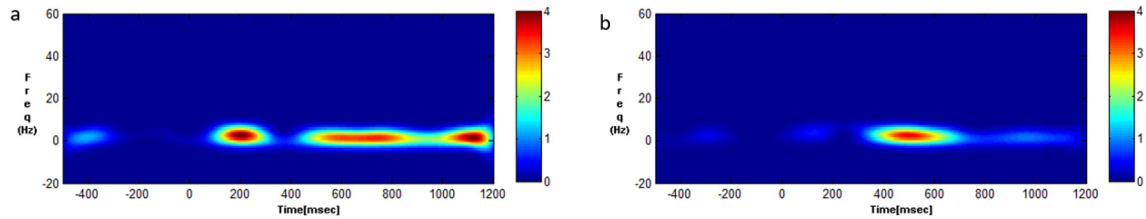


Fig. 3. TFHA output for feature 1 (CZ_11 Incorrect-DeltaL-tstd) in the feature rank list (Table 2). a: ADHD, b: nonADHD.

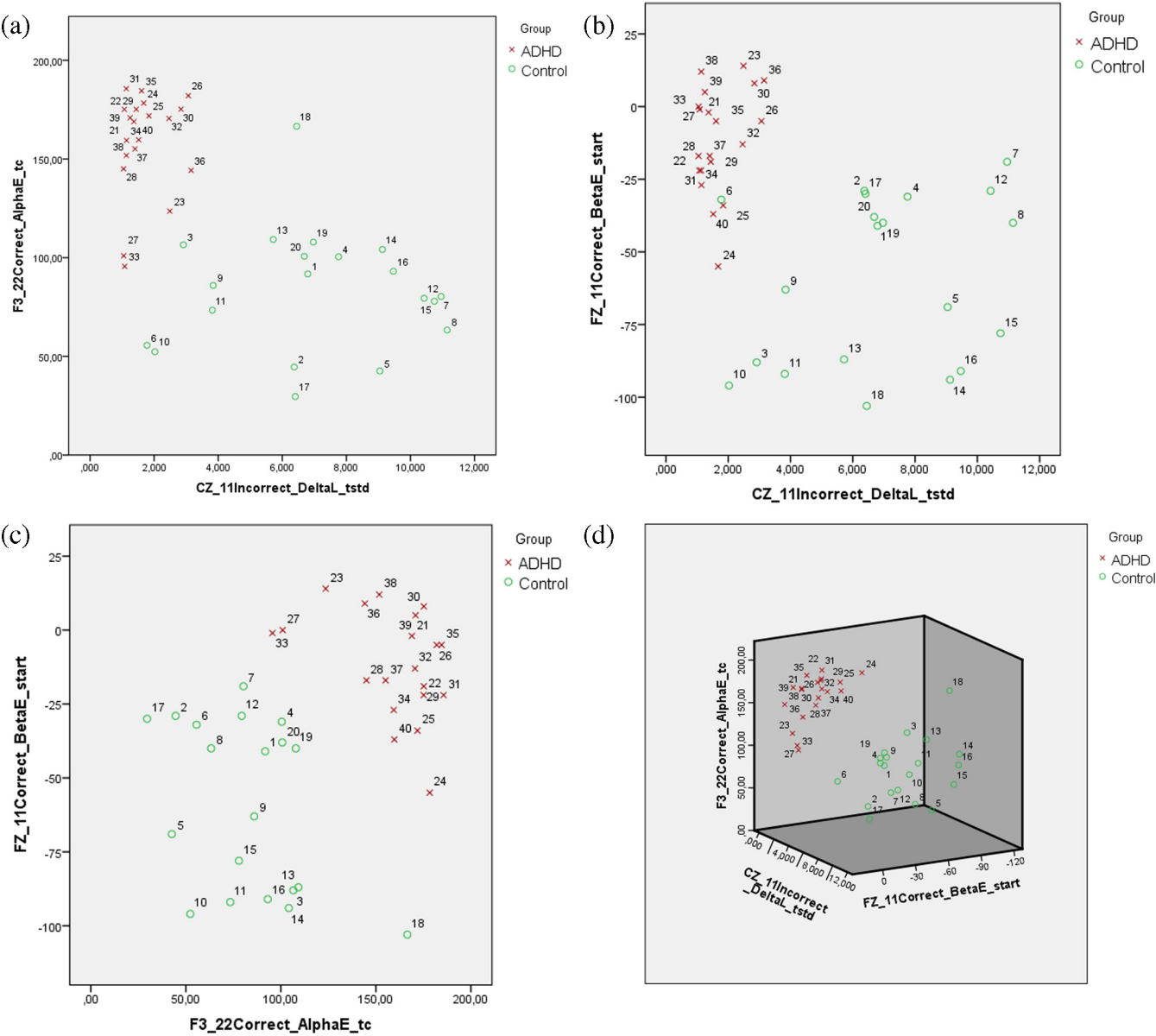


Fig. 4. Scatterplots for the participants for the combinations of the top 3 features (Table 2). (a) Feature pair 1 and 2, (b) Feature pair 1 and 3, (c) Feature pair 2 and 3, (d) All three Features (1, 2, and 3).

In accordance with the poorer sensitivities and the specificities of these techniques (Table 1), the margin between the ADHD and nonADHD samples was not clear; no visual indication of a linear plane that separated the ADHD group from the nonADHD group existed. Moreover, both techniques classified participant 7 as ADHD when he was in fact nonADHD. These graphical representations support the numerically inferior performance of the univariate feature selection methods presented in Table 1.

3.5. Contribution of the frequency bands to ADHD classification

In this subsection, we aim to compare the performance of the frequency bands in ADHD diagnosis and to obtain the best performing features. The classification setup was the same as the configuration explained in Section 2.5 (i.e., a classifier for each band was obtained by an internal RCV and applied on the separate test set). Fig. 6 illustrates the effect of the number of features on the error rates for each frequency band for the internal RCV. The error rates for the individual bands were within the 0.00 and 0.30 range. Although the powers varied, each of the bands could serve as ADHD markers. However, all the bands together were found to consistently outperform the performances of the individual bands.

Table 3 demonstrates the classification performances of the frequency bands based on sensitivity and specificity. The values indicate that the classifications based on the features derived from the beta and theta bands performed better in the classification of the participants with ADHD. These findings were consistent throughout all the tested features. In contrast, the alpha and delta features were better at classifying the nonADHD participants to their relevant (control) group.

Table 4 lists the top five features for each band that were found using SVM-RFE. All the features were obtained as the participants were responding correctly. Topographically, the alpha and beta features were mainly frontally distributed. The theta feature was posteriorly distributed. In contrast, the delta feature was distributed over the studied recording sites (Fz, Cz, Pz, F3, and F4).

The TFHA parameters from all four headings (Section 2.4) described the frequency bands. It might be expected that the top feature in Table 2, i.e., CZ-11-Incorrect-Delta-L-tstd, would be superior to the others; however, this was not the case because this feature was present only in 14 of the 25 lists. This result can be explained by the presence of similar information within the top delta features, i.e., the information contained in CZ-11-Incorrect-Delta-L-tstd might not be essential when all delta features are pre-

sent in the feature set. However, when the complete feature set is provided to SVM-RFE, it tends to select only a few delta features and further enriches the feature set with features from other bands because these features are likely to encompass information that differs from the information present in the delta bands. Similarly, the alpha and delta features are present in Table 2 but not in Table 4.

To test the performances of the frequency bands on the separate test dataset, a final classifier for each band was obtained by utilizing the respective features listed in Table 4. In accordance with the internal RCV results, the alpha, beta, delta and theta bands scored 9, 9, 10 and 8 of 10 participants correctly, respectively.

4. Conclusions

This paper presented a novel classification approach to discriminate between ADHD and nonADHD groups using an SVM classifier that was trained over a set of features extracted by the TFHA technique from ERP recordings. The best performing feature selection technique was identified using extensive experiments. We observed that the performance of SVM-RFE was superior to its alternatives. The RCV accuracy of SVM was 98% when 3 features were used. This performance increased to 99.5% when six features were used. Thus, we employed stability selection to assemble the top features that were obtained by applying SVM-RFE to 25 RCV

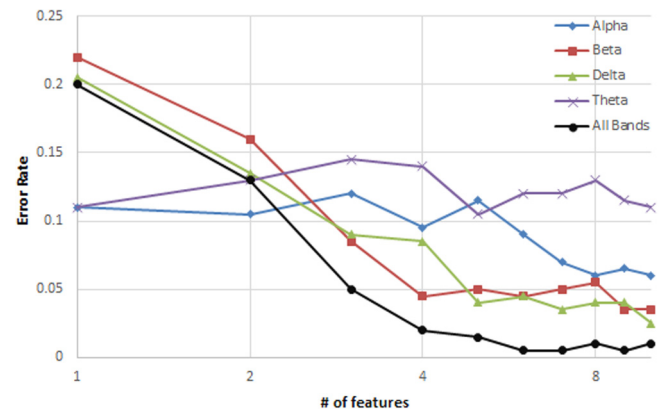


Fig. 6. Performances of the frequency bands. The alpha, beta, delta, gamma and theta bands were composed of 78, 89, 289, 36 and 60 features, respectively.

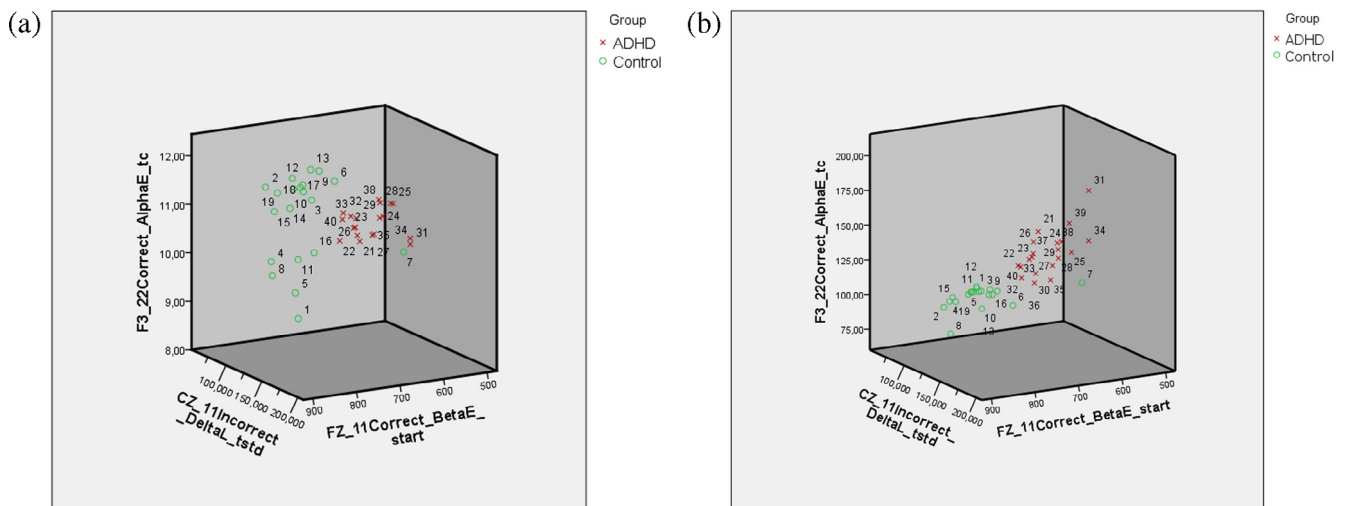


Fig. 5. Scatterplots for the participants for the top 3 features (Table 2) using the univariate methods. (a) Information gain, (b) chi-square.

Table 3
Sensitivities and specificities of the frequency bands.

| # of features | Alpha | | Beta | | Delta | | Theta | |
|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| 1 | 86 | 92 | 92 | 64 | 85 | 74 | 95 | 79 |
| 2 | 87 | 92 | 96 | 72 | 88 | 85 | 97 | 78 |
| 3 | 84 | 92 | 97 | 86 | 90 | 92 | 94 | 78 |
| 4 | 88 | 93 | 99 | 92 | 90 | 93 | 94 | 79 |
| 5 | 86 | 91 | 99 | 91 | 95 | 97 | 95 | 81 |
| 6 | 86 | 96 | 100 | 91 | 95 | 96 | 95 | 80 |
| 7 | 89 | 97 | 100 | 90 | 96 | 97 | 93 | 81 |
| 8 | 92 | 96 | 99 | 90 | 95 | 97 | 93 | 81 |
| 9 | 90 | 97 | 100 | 93 | 95 | 97 | 93 | 82 |
| 10 | 92 | 96 | 100 | 92 | 97 | 98 | 94 | 82 |

Table 4
SVM-RFE-based feature rankings of the frequency bands.

| Ranking | Alpha | | Beta | |
|---------|-------|------------------------------------|-------|--------------------------------------|
| | Score | Feature | Score | Feature |
| 1 | 25 | <i>F3_22Correct_AlphaE_tc</i> | 24 | <i>FZ_11Correct_BetaE_start</i> |
| 2 | 22 | <i>CZ_22Correct_AlphaE_tpeak</i> | 20 | <i>CZ_22Correct_BetaE_end</i> |
| 3 | 19 | <i>CZ_22Correct_AlphaE_end</i> | 18 | <i>FZ_11Correct_BetaE_freqf</i> |
| 4 | 15 | <i>CZ_22Correct_AlphaE_fpeak</i> | 15 | <i>FZ_11Correct_BetaE_wd_tpeak</i> |
| 5 | 5 | <i>FZ_22Correct_AlphaE_freqi</i> | 10 | <i>FZ_11Correct_BetaE_tpeak</i> |
| Ranking | Delta | | Theta | |
| | Score | Feature | Score | Feature |
| 1 | 17 | <i>PZ_11Incorrect_DeltaM_end</i> | 25 | <i>PZ_22Correct_ThetaM_end</i> |
| 2 | 16 | <i>CZ_11Incorrect_DeltaM_tstd</i> | 24 | <i>PZ_22Correct_ThetaM_tc</i> |
| 3 | 14 | <i>FZ_22Correct_DeltaE_fc</i> | 23 | <i>PZ_22Correct_ThetaM_tpeak</i> |
| 4 | 9 | <i>CZ_11Incorrect_DeltaM_start</i> | 22 | <i>PZ_22Correct_ThetaM_TD-fvalue</i> |
| 5 | 8 | <i>F3_11Correct_DeltaE_tc</i> | 18 | <i>PZ_22Correct_ThetaM_tstd</i> |

splits. Consequently, the subset of features that were consistently effective in discriminating the nonADHD group from the ADHD group was determined.

In the presented testing performance over a random but stratified separation of 40 samples to 30 samples of training and 10 samples of totally isolated testing datasets, the proposed classifier produced no misclassifications. Although the observed superior training performance on the validation data hints at a highly successful test performance, the 100% classification performance could be the result of the limited size of the testing dataset.

The top 3 features, i.e., CZ-11-Incorrect-Delta-L-tstd, FZ-11-Correct-Beta-E-start and F3-22-Correct-Alpha-E-tc, were from the delta, beta and alpha bands, respectively. Contrary to studies in which the effectiveness of the predefined power ratios are investigated using univariate feature selection techniques, SVM-RFE is inherently able to find the feature sets whose total discrimination power is the highest for discriminating children with and without ADHD.

Brain responses to stimuli can be decomposed into frequency components, and variations in these components represent cognitive and affective processes (for reviews, see [Karakaş and Başar, 2004, 2006](#)). Longstanding research specifically on neurofeedback training for the treatment of ADHD has used the ratio between the theta and beta bands as a biomarker for ADHD diagnosis and disease progression. This ratio no longer has scientific credibility for ADHD diagnosis (for a meta-analysis, see [Arn et al., 2013](#)). Indeed, the theta response was not among the top 10 features ([Table 2](#)) that discriminated children with ADHD from the non-ADHD controls in the present study.

In previous studies, the features associated with the delta band were often found to be ineffective ([Berdakh and Jinung, 2012](#)). An important finding in this study is the contribution of the delta band features to ADHD diagnosis. The cases with ADHD were best dis-

criminated via the delta oscillations. Of the 10 features that contributed to the 99% accuracy rate of the SVM classifier, 60% involved the delta band. The findings of the present study can be explained by the higher energy of the delta band, which makes it less vulnerable to measurement error. This fact, together with the versatile feature types provided by the TFHA, led to the success of the delta band features.

As demonstrated by its contribution to the P300, the delta response is an index of attention and decision-making; thus, it is obtained under conditions of high cognitive load ([Polich, 2007](#)). In accordance with this cognitive aspect, delta activity served as an important feature of the present study when it was recorded in response to the Stroop task within the frontocentral area between the Fz, Cz and F3 recording sites.

The findings of the present study also fit with the functional role of the delta band in neuropsychiatric disorders ([Başar and Güntekin](#)). A strong association has been found between delta anomalies and other neuropsychiatric disorders, which has led to the conclusion that delta response anomalies can be used as biomarkers for diagnoses and disease progression ([Güntekin and Başar, 2016](#)). The relevance of the delta oscillation to ADHD is demonstrated in research on the neurodevelopmental basis of ADHD. Consistent with the maturational lag hypothesis, the oscillatory activity in children with ADHD looks like that of younger children. Consistent with the maturational deviance hypothesis, there is a characteristic topographical distribution of the oscillatory activity regardless of age ([Burke and Edge, 2013](#)). Overall, there are increased relative and/or absolute delta and theta activities (slow activity) specifically in the frontal and midline sites (for a review, see [Barry et al., 2004; Mann et al., 1992](#)). There is decreased alpha (fast wave) activity in the parietal and temporal sites and decreased beta (fast wave) activity in the frontal, parietal and temporal sites

(Bresnahan and Barry, 2002; Hermens et al., 2005; Hobbs et al., 2007).

The second prominent feature of the present study was the alpha band. Variations in alpha oscillations represent mental processes that range from sensory to motor (for a review, see Başar and Güntekin, 2012). When confronted with a task that requires a high cognitive load, prolonged alpha oscillations are recorded over the frontal cortex (for a review, see Öñiz and Başar, 2009). As with the delta response, recent research has demonstrated that anomalies in the alpha response can act as biomarkers of neuropsychiatric disorders (for a review, see Başar and Güntekin, 2012). Serving the functions necessary for Stroop task performance, alpha oscillations were a feature of group differentiation. Among the features found in the present study, theta activity was the least paramount. However, the discriminatory role of the theta band for other neuropsychiatric disorders has also been documented in the frontocentral recording sites (Arnfred et al., 2011).

The aforementioned findings suggest that the features obtained by the SVM classifier of the present study can potentially contribute to an understanding of the associations between ADHD and the resulting neuroelectrical activity of the brain. These findings also suggest that frequency bands have the potential to act as biomarkers for ADHD diagnosis and disease progression.

In our study, all feature subgroups provided information about the presence of ADHD; however, the WD_Statistical features were more informative than the other feature subgroups. One important reason for this finding is the fact that statistical features provide more complete and robust information about the frequency components than instantaneous features, such as the time and frequency onset of the component and the highest power values of the bands. The success of the statistical features in this study suggests that higher-order moments (i.e., the skew and kurtosis) should be examined as potential biomarkers in the diagnosis of ADHD.

4.1. Limitations of the study

Following the conventions of the related literature, the pre-processing for eye movements was performed for blinks only in the present study. Blinks were removed using a spatial filtering algorithm, the software of which was provided as a part of the data acquisition system. Performance on the Stroop test produces lateral eye movements. In the present study, baseline correction was applied to the epochs using the algorithms supplied in the data acquisition system (Compumedics Neuroscan, Scan 3-4). Other than that, the data were not corrected for lateral eye movements, and this is a limitation of the study.

This study was conducted only on boys. The literature does not provide conclusive evidence on the effects of gender on ADHD. Some studies have reported such effects (Faraone et al., 2001), and other have not (Faraone et al., 2000). Gender differences in hereditary risk factors and the histories of comorbidity in families have not been reported (Faraone et al., 2001). Given these findings, gender was controlled at one level, i.e., boys. This choice was due to the high proportion of ADHD among boys compared girls (between 2:1 and 6:1 (American Psychiatric Association, 2000; Bhatia et al., 1991; Goodman and Stevenson, 1989)). The present study needs to be replicated with girls.

The cases in the present study were from the predominantly attention deficit and combined subtypes of ADHD. A replication of the study of only the predominantly attention subtype will be helpful for studying the state of attention deficit in relative isolation without the contamination of the hyperactivity/impulsivity symptoms that are represented for the combined subtype.

The study was conducted using a cross-sectional design. The disadvantage of the design was partially compensated for by form-

ing equivalent groups with respect to the critical variables (e.g., age, sex, intelligence quotient, comorbidity, medication, and health status) that would possibly have confounding effects on the findings. A future replication study should test the maturational delay and maturational deviance hypotheses with a longitudinal research design. Such a design would allow for the direct detection of the within-subject changes in ADHD symptoms and related cognitive processes.

The classification performance of this preliminary study demonstrated that TFHA produces features that are capable of modelling neuroelectric signals. This finding indicates that TFHA can be employed among the core components of the diagnostic and prognostic procedures of ADHD. This finding also demonstrates that the features and the classification technique that we used to arrive at these features can be used as auxiliary tools for ADHD diagnosis. Future aims will be to test the TFHA-based technique on not only the representative midline electrodes but also all the electrodes of the 10–10 system, to test the efficiency of the classification technique in the diagnosis of neuropsychiatric disorders other than ADHD and to test the concurrent validity of the technique by correlating the data with other data obtained from brain imaging and genetic studies.

Acknowledgements

This study was partially supported by the Scientific Research Unit of Hacettepe University (HUAF-BAB 2006K-120-640-06-08).

Conflicts of interest

None.

References

- Ahmadlou M, Adeli H. Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD. *Clin EEG Neurosci* 2010;41:1–10.
- Alp YK, Arıkan O. Time-frequency analysis of signals using support adaptive Hermite-Gaussian expansions. *Digit Signal Process* 2012;22:1010–23.
- Alp YK, Arıkan O, Karakaş S. Improving accuracy of source localization by a new time-frequency preprocessing technique. 14th World congress of psychophysiology. *The Olympics of the brain. Int J Psychophysiol* 2008;69:171.
- American Psychiatric Association. Attention deficit and disruptive behavior disorder. In: *Attention-deficit/hyperactivity disorder. Diagnostic and statistical manual of mental disorders*. 4th ed. Washington DC: American Psychiatric Association; 2000. p. 134–35.
- Anuradha J, Ramachandran V, Arulalan KV, Tripathy BK. Diagnosis of ADHD using SVM algorithm. In: *Proceedings of the third annual ACM Bangalore conference*. ACM; 2010.
- Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv* 2010;4:40–79.
- Arn M, Conners CK, Kraemer HC. A decade of EEG theta-beta ratio research in ADHD: a meta-analysis. *J Atten Disord* 2013;17:373–83.
- Arnfred SM, Mørup M, Thalbitzer J, Jansson L, Parnas J. Attenuation of beta and gamma oscillations in schizophrenia spectrum patients following hand posture perturbation. *Psychiat Res* 2011;185:215–24.
- Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol Bull* 1997;121:65–94.
- Barry RJ, Clarke AR, Johnstone SJ. A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography. *Clin Neurophysiol* 2003;114:171–83.
- Barry RJ, Clarke AR, McCarthy R, Selikowitz M, Rushby JA, Ploskova E. EEG differences in children as a function of resting-state arousal level. *Clin Neurophysiol* 2004;115:402–8.
- Başar E. Brain-body-mind in the nebulous Cartesian system: a holistic approach by oscillations. New York: Springer; 2011. p. 107–45.
- Başar E, Başar-Eroğlu C, Özderdem A, Rossini PM, Yener GG. Application of brain oscillations in neuropsychiatric diseases. *Suppl Clin Neurophysiol* 2013;62:343–65.
- Başar E, Güntekin B. A short review of alpha activity in cognitive processes and in cognitive impairment. *Int J Psychophysiol* 2012;86:25–38.
- Bekçi B, Karakaş S. Perceptual conflict and response competition: event-related potentials of the stroop effect. *Turk Psikiyatr Derg* 2009;20:127–37.
- Bhatia MS, Nigam VR, Bohra N, Malik SC. Attention deficit disorder with hyperactivity among paediatric outpatients. *J Child Psychol Psychiat* 1991;32:297–306.

- Bresnahan SM, Barry RJ. Specificity of quantitative EEG analysis in adults with attention deficit hyperactivity disorder. *Psychiat Res* 2002;112:133–44.
- Burke A, Edge A. Neurodevelopmental pathways of childhood ADHD into adulthood: maturational lag, deviation or both. In: Banerjee S, editor. *Attention deficit hyperactivity disorder in children and adolescents*. Intech; 2013. <https://doi.org/10.5772/53865>.
- Cooper P. Understanding AD/HD: a brief critical review of literature. *Child Soc* 2001;15:387–95.
- Cristianini N, Shawe-Taylor J. Support vector machines and other kernel based learning methods. Cambridge: Cambridge University Press; 2000.
- Dernoncourt D, Hanczar B, Zucker JD. Analysis of feature selection stability on high dimension and small sample data. *Comput Stat Data Anal* 2014;71:681–93.
- Duan KB, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans Nanobiosci* 2005;4:228–34.
- Durukan İ, Karaman D, Kara K, Türker T, Tufan AE, Yalçın Ö, Karabekiroğlu K. Çocuk ve ergen psikiyatrisi polikliniğine başvuran hastalarda tanı dağılımı. *Düşünen Adam* 2011;24:113–20.
- Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc* 1987;82:171–85.
- Faraone SV, Biederman J, Mick E, Williamson S, Wilens T, Spencer T, et al. Family study of girls with attention deficit hyperactivity disorder. *Am J Psychiat* 2000;157:1077–83.
- Faraone SV, Biederman J, Mick E, Doyle AE, Wilens T, Spencer T, et al. A family study of psychiatric comorbidity in girls and boys with attention-deficit/hyperactivity disorder. *Biol Psychiat* 2001;50:586–92.
- Gökler B, Ünal F, Pehlivanlıtürk B, Kültür ÇE, Akdemir D, Taner Y. Okul çağı çocukları için duygulanım bozuklukları ve şizofreni görüşme çizelgesi-şimdi ve yaşam boyu şekli- türkçe uyarlamasının geçerlik ve güvenilirliği [Reliability and validity of the Turkish adaptation of the schedule for affective disorders and schizophrenia for school-aged children-present and lifetime version]. *Türk J Child Adolesc Ment Health* 2004;11:109–16.
- Goodman R, Stevenson J. A twin study of hyperactivity – II. The aetiological role of genes, family relationships and perinatal adversity. *J Child Psychol Psychiat* 1989;30:691–709.
- Güntekin B, Başar E. Review of evoked and event-related delta responses in the human brain. *Int J Psychophysiol* 2016;103:43–52.
- Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389–422.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The Weka data mining software: an update. *SIGKDD Explor* 2009;11:10–8.
- Hermens DF, Soei EXC, Clarke SD, Kohn MR, Gordon E, Williams LM. Resting EEG theta activity predicts cognitive performance in attention-deficit hyperactivity disorder. *Pediatr Neurol* 2005;32:248–56.
- Hobbs MJ, Clarke AR, Barry RJ, McCarthy R, Selikowitz M. EEG abnormalities in adolescent males with AD/HD. *Clin Neurophysiol* 2007;118:363–71.
- Jin X, Xu A, Bie R, Guo P. Data mining for biomedical applications. In: PAKDD 2006 workshop, BioDM 2006, Singapore, April 9, 2006. Lecture notes in computer science, vol. 3916. Berlin, Germany: Springer; 2006. p. 106–15.
- Karakaş S. Dikkat eksikliği hiperaktivite bozukluğu: Kuram ve modeller. In: Karakaş S, editor. *Kognitif nörobilimler*. Ankara: MN Medikal & Nobel; 2008. p. 303–22.
- Karakaş S, Arıkan O. Hypothesis testing for gamma response generation using alternative signal analysis techniques. 13th World congress of psychophysiology. The Olympics of the brain. *Int J Psychophysiol* 2006;61:324.
- Karakaş S, Baran Z, Ceylan AO, Tileylioglu E, Tali T, Karakaş HM. A comprehensive neuropsychological mapping battery for functional magnetic resonance imaging. *Int J Psychophysiol* 2013;90:215–34.
- Karakaş S, Barry RJ. A brief historical perspective on the advent of brain oscillations in the biological and psychological disciplines. *Neurosci Biobehav Rev* 2017;75:335–47. <https://doi.org/10.1016/j.neubiorev.2016.12.009>.
- Karakaş S, Başar E. Oscillatory responses of the brain and their cognitive correlates. In: Adelman G, Smith BH, editors. *Encyclopedia of neuroscience*. 3rd ed. San Diego (CA): Elsevier; 2004.
- Karakaş S, Başar E. Models and theories of brain function in cognition within a framework of behavioral cognitive psychology. *Int J Psychophysiol* 2006;60:186–93.
- Karakaş S, Doğutepe E, Çakmak ED, Baran Z, Özkan A, Tüfekçi İ, et al. Time-frequency analysis of neuroelectric responses obtained under the standard paradigms of psychophysiology. 13th World congress of psychophysiology. The Olympics of the brain. *Int J Psychophysiol* 2006a;61:370.
- Karakaş S, Doğutepe E, Tüfekçi Dİ, Bekçi B, Çakmak ED, Arıkan O. Gamma response in sleep and wakefulness. 13th World congress of psychophysiology. The Olympics of the brain. *Int J Psychophysiol* 2006b;61:372.
- Karakaş S, Gücüyener K, Tali T, Topçu M, Arıkan O, Karakaş M, et al. Diagnosis of attention deficit hyperactivity disorder and its subtypes: a multidisciplinary and Multicenter approach. Project no: DPT-HÜAF 2006K120-640-06-08; 2006c.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc* 1997;36:980–8.
- Kim JH. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 2009;3735–3745.
- Li X, Peng S, Chen J, Lü B, Zhang H, Lai M. SVM-T-RFE: a novel gene selection algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles. *Biochem Biophys Res Commun* 2012;419:148–53.
- MacLeod CM. Half a century of research on the Stroop effect: an integrative review. *Psychol Bull* 1991;109:163–203.
- Mann CA, Lubar JF, Zimmerman AW, Miller CA, Muenchen RA. Quantitative analysis of EEG in boys with attention-deficit-hyperactivity disorder: controlled study with clinical implications. *Pediatr Neurol* 1992;8:30–6.
- Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol* 2010;72:417–73.
- Mueller A, Candrian G, Kropotov JD, Ponomarev V, Baschera GM. Classification of ADHD patients using a machine learning system. *Nonlin Biomed Phys* 2010;4:1.
- Öniz A, Başar E. Prolongation of alpha oscillations in auditory oddball paradigm. *Int J Psychophysiol* 2009;71:235–41.
- Onton J, Makeig S. Information-based modelling of event-related brain dynamics. In: Neuper, C, Klimesch, W, editors. *Progress in brain dynamics*. Amsterdam: Elsevier, vol. 159; 2006. p. 99–120.
- Özdemir AK, Karakaş S, Çakmak ED, Tüfekçi Dİ, Arıkan O. Time-frequency component analysis and its application to brain oscillatory activity. *J Neurosci Methods* 2005;145:107–25.
- Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 2007;118:2128–48.
- Pollastri G, Mclysaght A. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 2005;21:1719–20.
- Robaey P, Breton F, Dugas M, Renault B. An event-related potential study of controlled and automatic processes in 6–8-year-old boys with attention deficit hyperactivity disorder. *Electroencephalogr Clin Neurophysiol* 1992;82:330–40.
- Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386–408.
- Rowland AS, Lesesne CA, Abramowitz AJ. The epidemiology of attention-deficit/hyperactivity disorder (ADHD): a public health view. *Ment Retard Dev Disabil Res Rev* 2002;8:162–70.
- Saad JF, Kohn MR, Clarke S, Lagopoulos J, Hermens DF. Is the theta/beta EEG marker for ADHD inherently flawed? *J Atten Disord* 2015. <https://doi.org/10.1177/1087054715578270>.
- Smith JL, Johnstone SJ, Barry RJ. Aiding diagnosis of attention-deficit/hyperactivity disorder and its subtypes: discriminant function analysis of event-related potential data. *J Child Psychol Psychiat* 2003;44:1067–75.
- 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.
- Tüfekçi Dİ, Karakaş S, Arıkan O. Comparison of two different early gamma response detectors in the REM stage of sleep. 13th World congress of psychophysiology. The Olympics of the brain. *Int J Psychophysiol* 2006;61:372.