DOI: 10.22067/cke.2022.70922.1014

Exploring Effective Features in ADHD Diagnosis among Children through EEG/Evoked Potentials using Machine Learning Techniques*

Research Article

Faezeh Rohani[1] Kamrad Khoshhal Roudposhti[2] Hamidreza Taheri[3] Ali Mashhadi[4] Andreas Mueller[5]

Abstract: With the aid of intelligent system approaches, the present study aimed at extracting and investigating effective features for detecting Attention-Deficit/Hyperactivity Disorder (ADHD) in children. With this end in view, 103 children, aged from 6 to 10, were recruited for this study, among which 49 cases were assigned to the treatment group (ADHD children) and the remaining 54 cases to the control group (healthy children). The disorder diagnosis was performed using the well-known, relevant psychological questionnaires and clinical interviews with expert psychologists. Data collection consisted of EEG signals in eyes open and eyes closed states, as well as GO/NOGO task for about 3 hours for every participant. The extracted features consisted of the amplitudes and latency in Event-Related Potential (ERP) and the power spectrum in the sleep mode signals. Approximately 826 features of 19 channels were extracted in the standard 10-20 system and different task conditions. A set of features were selected with the aid of the feature selection methods, and then the selected features were analyzed by neuroscientists, and the irrelevant ones were removed. Next, the classification methods and their performance evaluation were applied. Finally, the best results in terms of the corresponding feature vector and classification method were presented. The healthy and ADHD groups were classified with 75.8% accuracy using the Support Vector Machine (SVM) method. The results showed that the use of selection of effective features with the aid of intelligent system techniques under the supervision of experts leads us to reach robust biomarkers in the detection of disorders.

Keywords: Attention Deficit Hyperactivity Disorder (ADHD), EEG/Evoked Potentials, Feature Extraction, Feature Selection

1. Introduction

Psychiatric disorders are complex because psychological, biological, and genetic factors influence cognition, emotions, and behavior in certain areas [1]. With questionnaires and clinical interviews, it has been found that the diagnosis of disorders relies on mental descriptions and external observations. Therefore, such diagnoses are prone to error due to the complexity of psychiatric disorders, intrinsic mentality, and even the use of the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition: DSM-5 [2] diagnostic guide. Accordingly, researchers have made significant efforts to obtain biological markers of mental disorders [3-10]. Most of these markers are genetic, biochemical, blood epigenetic, and blood plasmatic [11, 12]. However, some of these markers are electroencephalographic letters, induced potentials, and magnetic resonance imaging [13]. Unhealthy groups and healthy individuals have complex characteristics and are difficult to detect using individual markers. Henceforth, the symptoms of the diagnosis can be obtained by different neurobiological pathways [14]. Attention-Deficit Hyperactivity Disorder (ADHD), a neurological disorder, affects an estimated 4% to 12% of school-aged children worldwide [15]. Based on DSM-5, this disorder consists of three types, namely hyperactive and impulsive, inattentive, and combined [2].

The present study investigated and extracted the Electroencephalography (EEG) and Event-Related Potential (ERP) features that have been studied concerning the EEG and ERP indicators and brain function of ADHDs [16-19]. The principal advantage of using ERP includes the possibility of nonaggressive cognitive processes in milliseconds [20]. In recent years, machine learning methods have been widely used in the medicine and health realms [21-23]. Nevertheless, in psychiatry, due to limitations such as the absence of data, fear of distancing from diagnostic measures, and inadequate knowledge, this technique has been applied less frequently. However, the needs suggest that the combinatorial biomarkers have better performance compared with individual values [24].

Extensive research at the Switzerland Brain and Trauma Foundation has shown that biological boundaries can be traced through the stimulated potential to create biological markers (a measurable indicator for biological conditions) [25]. Moreover, in this research center, psychological neuroscience is used as an indicator to identify a specific disorder in the brain. The foundation also believes that none of the markers can help the diagnosis alone but that the diagnosis must be made through the proper usage of a set of these markers [6]. In this view, researchers using machine learning methods for the separation of ADHD and control groups in adults (74 cases in the ADHD group, 74 cases between 18-50 years old in the control group) observed that with GO/NOGO task, the accuracy of the Support Vector Machine (SVM) method was 92 % [6].

In another study, researchers using machine learning

* Manuscript received: 12 June 2021, Revised, 26 June 2021, Accepted, 10 September 2021.
1. PhD Candidate, Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran.
2. Corresponding author. Assistant Professor, Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran. **Email:** kamrad@liau.ac.ir.
3. Professor, Department of Motor Behavior, Ferdowsi University of Mashhad, Mashhad, Iran.
4. Professor, Department of Clinical Psychology, Ferdowsi University of Mashhad, Mashhad, Iran.
5. Professor in Brain and Trauma Foundation, Grisons/Switzerland, Chur, Switzerland.

methods on 117 adults (67 particiants in the ADHD group and 50 participants in the control group) showed that the classification accuracy for separating groups was about 69.2% in Visual Continuous Performance Test (VCPT) mode and 72.6 and 70.9% in eyes closed and eyes open states. However, in the form of scoring, the results showed up to 82.3 % change [26].

Oztoprak et al., using the time-frequency amplitude characteristics of EPR with strop test, classified the ADHD and control groups with 100% accuracy using the SVM method. This accuracy was for 3 to 5 features in the delta frequency band. In their study, all participants were male and in the age range of 6 to 12 years old, and the sample included 44 cases in the ADHD group and 38 cases in the control group [27].

Helgadotter et al. had 310 participants in the ADHD group and 351 participants in the control group, aged from 5.8 to 14. Their method accuracy rate was about 81% when analyzed by age and 73% the other way round (i.e., not based on age) [3].

Heinrich et al. investigated the neural mechanisms of motor control using the potentials in combination with MRI, obtaining a classification rate of 90% in a linear analysis. The study suggested that both cognitive and motor inhibition should be regarded as fundamental problems in children with ADHD [28].

Meuller et al. used machine learning techniques to separate ADHD from healthy participants. Their experimental EEG and ERP data were collected from 181 ADHD and 147 healthy participants. Spectral power, ERP amplitude, and latency measures were extracted and used as a feature vector for the input of their machine-learning framework. ADHD patients and healthy participants were classified by logistic regression model with accuracy values between 72% and 76%, while their specificity values slightly decreased over time (between 64% and 67%) [29].

During the review of the related literature, various studies have reported good EEG classification capability and ERP. These methods had different accuracy rates according to the selection of different effective features, their numbers of features, and the applied classification technique. Therefore, the number of features and the type of features are effective in obtaining accuracy. With this end in view, this study aims at extracting effective features to diagnose ADHD in children under the supervision of neuroscientists. Figure 1 shows the workflow of the current study.

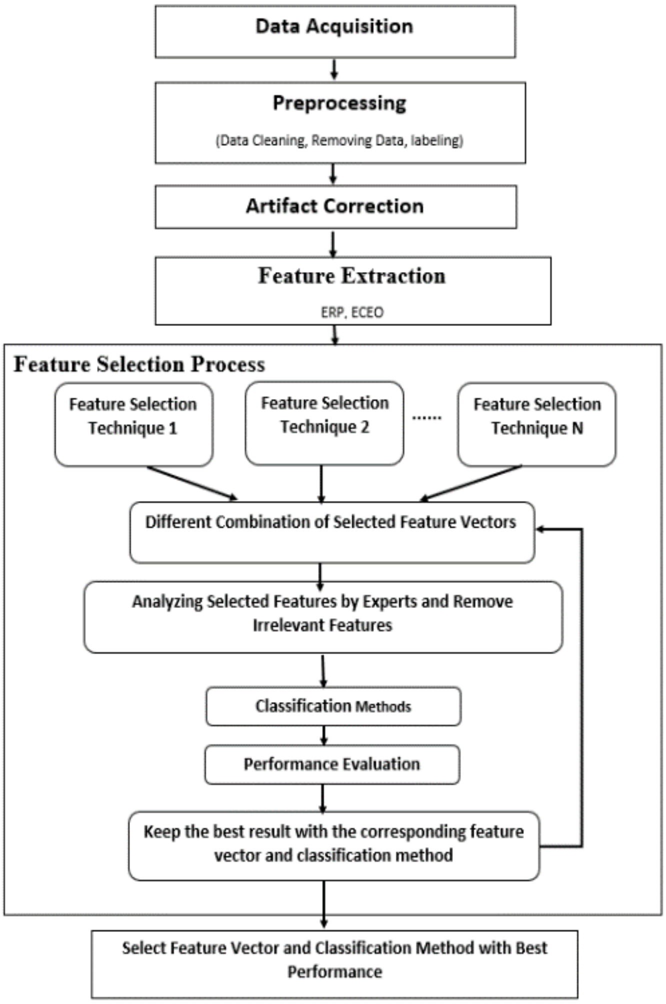

Figure 1. Workflow of the research framework. ECEO denotes EEG signals from eyes closed and eyes open states

2. Data collection

2.1. Participants

The participants consisted of 103 participants from 7 to 10 years old. According to the DSM-5, 49 participants were diagnosed with ADHD (22 females, 27 males), and the remaining 54 participants were healthy participants (24 females,30 males). The ADHD participants were recruited from clinics, and the members of the control group were selected from summer leisure classes of Ferdowsi University of Mashhad, Iran. Deprivation criteria in this study were an IQ scoring below 75, epilepsy, and comorbidities disorder with ADHD. Control patients who consumed a drug were not included in the study. The ADHD patients who had medication under the supervision of their doctors did not take drugs before testing. Therefore, all the participants did not receive any medication at the time of testing.

2.2. Procedure

Data was collected in the motor behavior lab at Ferdowsi University from July 2019 to February 2020. All ADHD participants were screened medically by medical doctors. As the first step in this project, parents filled out a set of such questionnaires as Child Behavior Checklist (CBCL), AMEN, ADHD, Cognitive Change Index (CCI), and the Swanson, Nolan, and Pelham (SNAP). For the IQ test, the Riven test was applied [30]. Participants were tested in a single session for about 3 hours, including recording their EEGs/ERPs and taking the IQ tests. The parents were aware of this study and agreed to use clinical data for research purposes. They had signed consent forms before the start of the study.

2.3. EEG and ERP task

EEG was recorded for 10 minutes (5 minutes with eyes closed and 5 minutes with eyes opened), and ERP was recorded for 20 minutes. The ERP test was Go/NOGO task that contained 400 trials. This task had four conditions, namely A-A (animal-animal), A-P (animal-plant), P-H (Plant-Human), and P-P (plant-plant). Each condition involved 100 trials. The task had novel sounds along with human images in the P-H state. The details of this task are provided in [5].

2.4. Data recording and pre-processing

The EEG was recorded with the aid of the “NeuroAmp® x23” and “ERPrec software” (BEE Medic GmbH, Switzerland). The Raw EEG was analyzed by Matlab. The sampling rate of the input signals was 500 HZ, and it was referenced with linked-earlobes and filtered by band-pass between 0.5 and 50 HZ with a 45-55 Hz notch filter. The Electro-Cap electrode application system (19channel, Electro-Cap, International Inc, USA) that worked with the international 10-20 system was used in the present study. The impedance for all electrodes was not more than five kOhm. Neuronal activity of 19 brain channels including Fp1, Fp2, F3, F4, F7, F8, F8, Fz, C3, C4, Cz, T3, T4, T5, T6, P6, P3, P4, Pz, O1, and O2 and linked earlobes and such frequency bands as Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30), and Gamma (30-50 Hz) were recorded.

For artifact removing, the starting raw EEGs were first removed. Then eye-blink and horizontal eye movements were detected, with the aid of independent component analysis (ICA) decomposition removed from the EEGs. The remaining artifacts were removed from the slow (e.g., sweat artifact)/fast (e.g., muscle artifacts) wave correction (i.e., excessive activity in the 0-3 Hz and 20-50 Hz frequency bands). Finally, the amplitudes range of more than 100 µV were removed.

3. Method

3.1. Feature extraction

In signal processing, features are generally divided into the time, frequency, and time-frequency domains. The time-domain characteristics refer to directly extracted features from the signal itself without altering such signal spaces as mean, standard deviation, energy and power, entropy, skewness, kurtosis, auto-regressive coefficient, zero-crossing percentile, and Hjorth parameters [31-39].

The purpose of applying a mathematical transformation to a signal is to obtain additional information that is not available in the original raw signal. However, time domain-based analysis of the signals is popular, but in many cases, the useful information of the signal lies in its frequency content, which is called the signal spectrum. Simply, the spectrum of a signal represents the frequencies’ amplitude in that signal. Examples of approaches for extracting frequency range features are the Fourier transform, Short-Term Fourier Transform (STFT), spectral entropy, spectral centroid, spectral spread, spectral roll-off, harmonic parameters, and power spectral density [40- 43].

According to the description of the extraction feature, the features extracted in this study included the density spectrum of 5 frequency bands and 17 channels of EEGs in eyes closed and eyes opened states. The spectral power density was a description of power distributed over the frequencies in the limited data set signal, so the power spectrum density unit was the power in each frequency unit (watts per Hz). The density spectrum indicates at what frequencies the signal strength changes are weaker and at what frequencies they are stronger.

Amplitude and latency peaks were extracted for ERP in eight task conditions for the 17 channels [5]. The conditions included four main states (A-A, A-P, P-P, and P-H) and four mixture conditions amid all states (A-A/P, A-P-A-A, P-P/H, P-H-P-P). For ERP, usually, the first, second, and third peaks from the curves would be extracted.

The VCPT has two stimuli, and usually, the features should be extracted on the second stimulus, and the events and peaks are examined after the second stimulus appearance. In this case, the peaks will be considered after the second stimulus appearance and are positive or negative. The first positive peak is called P100, the second P200, and the third P300. The first negative peak is called N100, and the second negative peak is called N200, and this cycle, as shown in Figure 2 [44], continues. Therefore, knowing that the second stimulus appears in 1,400 milliseconds, the signal analysis time interval can be from 1,300 to 2,400 milliseconds, and in cases where it is necessary to check the events of the first stimulus, the time interval is between 300 to 1,100. Besides, to align all the signals, a baseline is set in the range of 1,300 to 1,400 milliseconds.

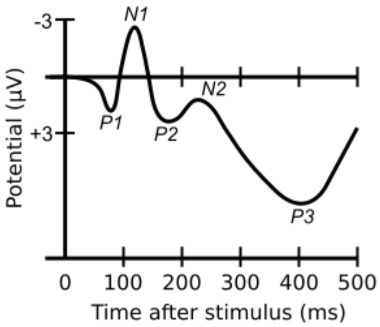

Figure 2. A waveform showing several ERP components, including the N100 (labeled N1) and P300 (labeled P3). Note that the ERP is plotted with negative voltages upward, a common but not universal practice in ERP research.

In ERP, to obtain the appropriate peaks, the average ERP diagrams were considered for all participants. Moreover, to obtain the lowest and highest points along with the signals, curves of the time window, which are one of the features of ERP components, were considered. The size of the time window was fixed at 45% of the time interval from the highest peak to the adjacent peak in the average main ERP curve. To reach the main peak in this time window, different methods such as measuring the area under the ERP curves in the time window range or measuring the curve in the specified time window are applied. In this study, the curve range method has been used. Another list of features, including arousal index, reaction time, theta/beta ratio, C3/C4 index, and omission and commission error, was also extracted. Features related to reaction time, commission, and omission are behavioral parameters compared with other features that are characteristic of the brain.

One of the major points in extracting features is to identify the important frequency bands for specific disorders. Based on the past studies, it has been found that the significant frequency bands in the diagnosis of ADHD are F3, F4, F8, Fz, C3, C4, Cz, Cz, T5, T6, P2, O1, and O2. However, since the purpose of the study was to obtain more variant characteristics, all frequency bands except FP1 and FP2 (due to artifact in the data and meanness in ADHD) were examined. The importance of the features is described in the feature selection part below.

3.2. Feature selection

A set of features has been extracted from the EEG/ERP signals, and it is evident that all of these features did not relate to ADHD. Thus, it was necessary to reduce features to achieve effective features, prevent over-fitting, and reduce computational efforts [45]. Therefore, in this study, to limit the number of features, a combinational approach using intelligent feature selection methods with a neuroscientist's supervision was proposed. Based on this approach, several feature selection methods have been used to select different sets of effective features. Then neuroscientists examined the selected features and selected a set of effective features.

One of the feature selection methods that was used in the present study was the combined Hybrid Structured sparse learning method [46]. This method is the same as the regression of Least-squares, which contains two regulating modes, L1-norm and L2.1-norm, as follows:

$$\min_{w} J(W) = \|X^T W - Y\|^2 + \gamma_1 \|W\|_{1.1} + \gamma_2 \|W\|_{2.1} \qquad (1)$$

Equation 1 is a target function in which $X = [x_1 . x_2 \dots . x_n] \in R^{d\times n}$ where n training samples and d features are applied, and $Y = [y_1 . y_2 \dots . y_c] \in R^{n\times c}$ where c is the number of classes for each x_i training data. By finding the optimal values of the parameters γ_1 and γ_2, the optimal coefficient matrix for each feature of x_i can be obtained. To get the best k features, the features would be sorted based on their effectiveness, and then the k feature is selected with the highest rank.

The sequential floating forward selection (SFFS) [47] is another implemented feature selection method in the present study. This algorithm finds an optimal subset of features by addition (adding a new feature to the subset of previously selected features) and subtraction (removing a feature from the subset of previously selected features).

Therefore, amongst all the features selected by automatic methods, after being analyzed by an expert, a set of features were finally selected. Table 1 shows the group of features.

Table 1. The group of features

Group	Features name
EC/EO/VCPT	Arousal index
EC/EO/VCPT	Theta/beta ratio
EC/EO	frequency spectra (coherence)
Behavioral in VCPT	Omission errors
	Commission errors
	Reaction time
ERP	Min amplitudes
	Max amplitudes
	Min latency
	Max latency

3.3. Classification

Supervised machine learning methods work in such a way that in them, a set of input vectors such as $X = \{x_n\}$ and the corresponding output vector $T = \{t_n\}$ are given. The goal for the machine, using those training data for the new x input, is to be able to predict t [48]. In this regard, two distinct modes can be considered. Regression, in which t is a continuous variable and classification and belongs to a discrete set. In the learning process, the system first needs to be trained, and then in the testing process, the trained system is used to predict the output concerning the new input values. Support Vector Machine (SVM) is a well-known supervised machine learning method and one of the simplest types of SVMs (i.e., linear SVM), which finds a hyperplane that separates sets of positive and negative samples with the maximum distance. A couple of the most accurate approaches, SVM and ensemble classification models, were used and reported in this study.

3.4. Cross-validation and evaluation

In the supervised learning methods, there are two sets of data (i.e., train data set and the test data set), which are managed in different ways for validation. Here, the K-fold method was used for validation. K-fold cross-validation is one of the most common methods of validating machine learning systems. In this method, the whole set of data is divided into K equal parts. Form the *K* parts, *K-1* parts are used as a set of training data, based on which the model is constructed, and with the remaining part, the testing process is performed. The number of repetitions of this process will be *K* times such that each *K* part is used only once for evaluation, and the accuracy for the model is calculated each time. In this evaluation method, the final accuracy of the system will be equal to the average of all obtained *K* accuracies [49].

Confusion matrix: This matrix shows how the classification technique works. This is according to the separate input datasets for different class categories [50]. In what follows, TP, TN, FN, and FP and their relationships in the present study are explained.

- True Negative (TN) = correctly rejected. This rate indicates the number of records whose true category has been negative, and the classifier has identified them as negative. In this study, it is the correct diagnosis of the control group, the participants who have been correctly diagnosed as healthy ones.
- False Positive (FP) = incorrectly identified. The misdiagnosis with ADHD, meaning control group participants who have been misdiagnosed with ADHD.
- False Negative (FN) = incorrectly rejected. The misdiagnosis of the control group. That is the participants who were ADHD but were misdiagnosed as healthy ones.
- True Positive (TP) = correctly identified. Correct diagnosis of ADHD, participants who were in the ADHD group and were diagnosed with ADHD.

Accuracy: The most important criterion for determining the performance of the classification technique is the accuracy criterion. This measure computes the total accuracy of a classification and illustrates that the designed classification correctly classifies a few percent of the entire set of experimental records. The accuracy of the classification based on the concepts expressed in the confusion matrix is calculated by the following equation:

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

Scoring: The main scoring criterion is to evaluate the performance of the Receiver Operating Characteristic (ROC) area under the receiver operating characteristic curve (AUC). This criterion shows the overall performance of a model by combining the actual-positive rate (sensitivity) and the false positive rate (1-specificity). For binary classifiers, the AUC value varies from 0.5 to 1, in which 1 indicates the full performance of a classifier [51].

4. Results

The effectiveness of the proposed method in this paper has been investigated with the aid of data collected from control group children and children with ADHD. In all classification processes, the 5-fold cross-validation approach was applied to validate the model, and for evaluation, accuracy criteria from the confusion matrix of each classifier were calculated. To stabilize the final output of the classifiers and provide a reliable answer based on the evaluation criteria, the results were an average of 10-trial classification.

In the first step, the data was presented directly to the classifiers without selecting the subset of features. In the second step, the data was first presented to the feature selection algorithms and then to the classifiers. After obtaining their accuracies, the features were checked by the neuroscience specialist, and then the features were given to the classifiers again. The final output is shown in Table 2. The total number of features was 826, the number of features in each section was 30, 5, and 37, and finally, the number of effective features that have been obtained in combination methods was about 113 features.

Based on the results, all of the selected methods and features were not approved by the specialist, so according to the expert's opinion and previous studies, combining the features was necessary to obtain the appropriate accuracy to separate the control group from the ADHD group. Moreover, based on the results, 37 features were approved by experts [9, 52] for the data of this study that had an accuracy of 61.9%, which slightly showed the specific characteristics of this research data.

Table 2. The performance of different feature selection techniques and classifier models

o Features	Feature Selection	Model	TP	TN	FP	FN	ACC	AUC	Expert Approved
826	No Feature Selection	Tree	80	67	20	33	73.8	0.74	–
826	No Feature Selection	Ensemble RUS Boosted tree	85	61	15	39	73.8	0.68	–
113	Combine	SVM-Linear	83	67	17	23	75.8	0.75	Yes
37	Neuroscience	Ensemble Subspace Discriminant	69	54	31	46	61.9	0.58	Yes
30	Hybrid Structured Sparse Learning (HSSL)	Logistic Regression	81	88	19	12	84.5	0.90	–
5	Sequential Floating Forward Selection (sffsAB)	Cosine KNN	96	59	4	41	78.6	0.83	–
64	Sequential Floating Forward Selection Standard (SffsSt)	Ensemble Subspace Discriminant	70	76	30	24	70.9	0.75	–

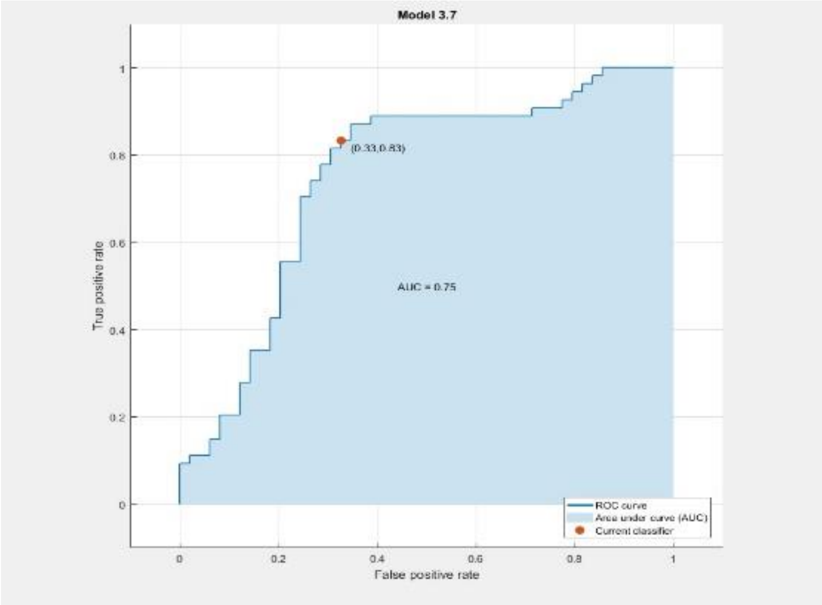

Figure 3. ROC score of the selected method

The methods which used the feature selection method of HSSL and SFFS with 84.5% and 78.6% accuracy were not approved by the neuroscientist, and to the best of neuroscientist's knowledge, most of the selected features were not relevant to the diagnosis of ADHD. Therefore, under the supervision of the neuroscientist, a small number of significant features were selected as effective features.

By combining the features obtained from the selection methods that have been approved by the specialists and the proposed and approved features of the neuroscientist concerning the significance of ADHD and behavioral features, 113 features were obtained with a 75.8% accuracy rate. As shown in Table 2, using the SVM method, the correct detection rate of ADHD (TP) and control (TN) were 83% and 67%, respectively. Accordingly, the misdiagnosis of ADHD (FP) and control (FN) groups were 17% and 33%, respectively. Figure 3 shows the ROC diagram of the classifier result.

5. Discussion

In this paper, all the mentioned features were extracted from the raw signal in the closed and open eye modes, as well as ERP and behavioral features. To select the best features, we used the methods of selecting the feature of the HSSL and SFFS. The method of extracting and selecting the feature vector from raw signals significantly impacts the obtained results. Consequently, we tried to use brain signal processing and extract the best features in diagnosing ADHD in the first stage. Then those features were approved by a specialist.

In the present study, features included the theta, beta, and alpha frequency bands of Pz, O1, O2, T5, T6, C6, Cz, Fz, C3, C4, F3, F4, and F8 electrodes, the maximum and minimum latencies, and the highest and lowest domains in ERP. The effective features were obtained through feature selection methods with the approval of neuroscientists, and finally, for classification, the linear SVM was used. The feature vector with 113 features, which was obtained with a combination strategy, was used for the classification process by the SVM method. The obtained result showed that the accuracy of the proposed approach was 75.8%.

Due to changes in brain functionality and the instability of their brain signals, the diagnosis of ADHD in children aged 6 to 10 is very limited in the literature. Therefore, to compare with previous studies, the same research method and executive protocol must be applied to record data. This is a research constraint that limits comparison with accessible studies. TableIII summarizes the studies conducted on the diagnosis of ADHD in children.

As shown in Table 3, different methods have been used in different studies for data collection. Moreover, the applied tests and the data registration conditions were different. One of the advantages of the present study is using all conditions in one setting: raw signal and ERP signal.

Some studies like [3], have only used closed-eye data for diagnosis and analysis, in which case the type of data and the number of participants examined affected the results. In [3], due to the large number of participants, one of the prominent features was the age of the participants, while the number of participants of the present study was fewer, and all conditions, that is, raw signal (eyes closed and eyes opened) and the event-dependent potential were used.

In some studies like [27], only male participants were recruited, and ERP was also performed by color strop test. In such studies, with about 3 to 5 behavioral features (omission and commission error), an accuracy of 99.5% was achieved. With respect to what experts claim, this number of features is not acceptable and comparable with the present study. In this study, with a few features, the observed accuracy was above 80%. However, some of the features were approved by the experts as criteria for ADHD detection.

In [56], to diagnose ADHD through the pre-forehead cortex, NIRS data, strop test, and behavioral data were collected where with the aid of SVM, the accuracy rate was 86%. The difference between this method and the one in the current study is the type of data collection procedure followed.

Table 3. Studies conducted on the diagnosis of ADHD in children. SVM-RFE denotes support vector machine recursive feature.

Ref.	Number of participants	Age/ gender	Accuracy	Classification method	Feature selection method	Selected feature(s)	Device and system	Data collection method
[27]	70 ADHD 37 Control	6 to 12 Boy	99.5%.	SVM	(SVM-RFE)	Three features (behavioral features) include omission, commission, errors	64 electrode 10-10 system	ERP with Stroop task
[54]	62 ADHD 39 Control	7 to 16 Boy/ girl	58-63% 85%	Statistical analysis with Ancona	No	Theta/beta ratio Theta at Cz Beta at Cz Omission errors	19 channel 10-20 system Mitsar 201	ERP Go/No Go
[55]	7 ADHD 7 Control	8 to 12 Boy/girl	-	statistical analysis	No	ERP Spectral perturbation Inter-trial coherence Time locked on each stimulus Omission, commission errors reaction time	14 channels (Fz, F3, F4, Cz, C3, C4, Pz, P3, P4, Oz, O3, O4, and M1-M2 for the left and right mastoids), 10-20 system Ant company	ERP Go/No Go
[56]	108 ADHD 108 Control	~10 Boy/girl	86%	SVM	No	Reaction time Behavioral	10-20 system NIRS[i] system	Reverse Stroop task
[3]	310 ADHD 350 Control	5.8 to 14 Boy/girl	76%	SVM	No	20 features Age dependent Coherence Power, Relative power	17 electrodes: Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, T5, T6, Pz, P3, P4, O1 and O2 NicoletOne	Eyes Closed EEG
[28]	19 ADHD 21 Control	9 to 14 Boy/girl	90%	Statistical analysis	statistical analysis with Ancona	ERP	BrainAmp 10-20 system	ERP Go/No Go and TMS data

6. Conclusion

In this study, with the aid of intelligent techniques under a neuroscientist's supervision for diagnosing ADHD, a new strategy was proposed to select effective EEG/ERP-based features. A new dataset was also collected for applying and evaluating the proposed method. The limitations of previous researches were discussed it was tried to improve them. The automatic feature selection techniques usually try to find a set of features that increase the accuracy measurement. Since the number of samples is limited, the automatic techniques can be affected by the experimental-based artifacts and can find some irrelevant features that can increase the system's accuracy for that specific dataset but might not work in others. Thus, we have proposed an expert's supervision-based feature selection technique to achieve an acceptable result with the expert's approval. In this study, due to the characteristics of the data, the effective feature was confirmed by experts. As experts stated, integrating all dimensions (including lifestyle, questionnaire, interview, and psychiatric examination) is essential in the diagnostic process [57]. In short, the results are promising and can be expended by taking into account such factors as the effects of age on more data samples. By increasing the number of features, the feature selection techniques show a weak performance or will be a time-consuming task. Thus, using optimization methods for the mentioned purpose can be a proper solution for future related works.

Declaration of competing interest

The authors have no conflict of interest to disclose.

7. Acknowledgements

A special thanks goes to the Brain and Trauma Foundation in Switzerland, headed by Dr. Andreas Mueller and his coworker Gian Candrian, who supported the study of Switzerland Opportunity, the HBI Foundation, and the BioMed Institute, which provided the hardware and software needed to record the signal. We also appreciate the Ferdowsi University School of Sports Science and Soroush Psychological Clinic for their help in collecting data and selecting samples. Without the support of these groups, the current research would not be possible at its best quality. We also appreciate all the children and their parents who patiently accompanied us.

7. References

[1] Guntern, G. "Auto- organization in human systems", Behavioral Science, Vol. 27(4), pp. 323-337, 1982.

[2] Association, A.P., "Diagnostic and statistical manual of mental disorders (DSM-5®)", American Psychiatric Pub, 2013.

[3] Helgadóttir, H., Gudmundsson, Ó. Ó., Baldursson, G., Magnússon, P., Blin, N., Brynjólfsdóttir, B., ... & Johnsen, K., "Electroencephalography as a clinical tool

for diagnosing and monitoring attention deficit hyperactivity disorder: a cross-sectional study", BMJ open, Vol. 5(1), 2015.

[4] Müller, A., Candrian, G., Kropotov, J., "ADHS-Neurodiagnostik in der Praxis", Springer-Verlag, 2011.

[5] Mueller, A., Candrian, G., Kropotov, J. D., Ponomarev, V. A., & Baschera, G. M., "Classification of ADHD patients on the basis of independent ERP components using a machine learning system", In Nonlinear biomedical physics, Vol. 4, No. 1, pp. 1-12, BioMed Central, June, 2010.

[6] Mueller, A., Candrian, G., Grane, V. A., Kropotov, J. D., Ponomarev, V. A., & Baschera, G. M., "Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study", Nonlinear biomedical physics, Vol. 5(1), pp. 1-18, 2011.

[7] Dubreuil-Vall L, Ruffini G, Camprodon J. A., "Deep learning convolutional neural networks discriminate adult adhd from healthy individuals on the basis of event-related spectral eeg", Frontiers in neuroscience. Apr 9;14:251, 2020.

[8] Furlong S, Cohen J. R, Hopfinger, J., Snyder, J., Robertson, M. M., Sheridan, M. A., "Resting-state EEG Connectivity in Young Children with ADHD", *Journal of Clinical Child & Adolescent Psychology*, Aug 18:1-7, 2020.

[9] Kaiser A, Aggensteiner PM, Holtmann M, Fallgatter A, Romanos M, Abenova K, Alm B, Becker K, Döpfner M, Ethofer T, Freitag CM. "EEG Data Quality: Determinants and Impact in a Multicenter Study of Children, Adolescents, and Adults with Attention-Deficit/Hyperactivity Disorder (ADHD)", Brain Sciences, Feb, Vol. 11(2), pp. 214, 2021.

[10] Tosun M. Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning. Physical and Engineering Sciences in Medicine. May 27:1-0, 2021.

[11] Cubero-Millán, I., Ruiz-Ramos, M. J., Molina-Carballo, A., Martínez-Serrano, S., Fernández-López, L., Machado-Casas, I., ... & Muñoz-Hoyos, A., BDNF concentrations and daily fluctuations differ among ADHD children and respond differently to methylphenidate with no relationship with depressive symptomatology. Psychopharmacology, Vol. 234(2), pp. 267-279, 2017.

[12] Wang, L. J., Li, S. C., Lee, M. J., Chou, M. C., Chou, W. J., Lee, S. Y., & Kuo, H. C., "Blood-bourne MicroRNA biomarker evaluation in attention-deficit/hyperactivity disorder of Han Chinese individuals: an exploratory study", Frontiers in psychiatry, 9, 2018.

[13] Kropotov, J. D., Grin-Yatsenko, V. A., Ponomarev, V. A., Chutko, L. S., Yakovenko, E. A., & Nikishena, I. S., "ERPs correlates of EEG relative beta training in ADHD children", *International journal of psychophysiology*, Vol. 55(1), pp. 23-34, 2005.

[14] Insel, T. R., & Cuthbert, B. N., "Brain disorders? Precisely", Science, Vol. 348(6234), pp. 499-500, 2015.

[15] Krieger, V., & Amador-Campos, J. A., Assessment of executive function in ADHD adolescents: contribution of performance tests and rating scales. Child Neuropsychology, Vol. 24(8), pp. 1063-1087, 2018.

[16] Yang, M. T., Hsu, C. H., Yeh, P. W., Lee, W. T., Liang, J. S., Fu, W. M., & Lee, C. Y., "Attention deficits revealed by passive auditory change detection for pure tones and lexical tones in ADHD children", Frontiers in human neuroscience, Vol. 9, pp. 470, 2015.

[17] Lenartowicz, A., & Loo, S. K., "Use of EEG to diagnose ADHD", Current psychiatry reports, Vol. 16(11), pp. 498, 2014.

[18] Kakuszi, B., Tombor, L., Papp, S., Bitter, I., & Czobor, P., "Altered response-preparation in patients with adult ADHD: A high-density ERP study", Psychiatry Research: Neuroimaging, Vol. 249, pp. 57-66, 2016.

[19] Snyder, S. M., Rugino, T. A., Hornig, M., & Stein, M. A., "Integration of an EEG biomarker with a clinician's ADHD evaluation", Brain and behavior, Vol. 5(4), e00330, 2015.

[20] Banaschewski, T., & Brandeis, D., "Annotation: what electrical brain activity tells us about brain function that other techniques cannot tell us–a child psychiatric perspective", *Journal of child Psychology and Psychiatry*, Vol. 48(5), pp. 415-435, 2007.

[21] Khalifa, M., "Health Analytics Types, Functions and Levels: A Review of Literature", ICIMTH, pp. 137-140, 2007.

[22] Meskó, B., Hetényi, G., & Győrffy, Z., "Will artificial intelligence solve the human resource crisis in healthcare?", *BMC health services research*, Vol. 18(1), pp. 1-4, 2018.

[23] Islam, M. S., Hasan, M. M., Wang, X., & Germack, H. D., "A systematic review on healthcare analytics: application and theoretical perspective of data mining", In Healthcare, Vol. 6, No. 2, pp. 54, *Multidisciplinary Digital Publishing Institute*, june, 2018.

[24] Jollans, L., & Whelan, R., "Neuromarkers for mental disorders: harnessing population neuroscience", *Frontiers in psychiatry*, Vol. 9, pp. 242, 2018.

[25] Mandal, A. https://www.news-medical.net/health/What-is-a-Biomarker.aspx.

[26] Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A., & Candrian, G., "Machine learning approach for classification of ADHD adults", *International Journal of Psychophysiology*, Vol. 93(1), pp. 162-166, 2014.

[27] Öztoprak, H., Toycan, M., Alp, Y. K., Arıkan, O., Doğutepe, E., & Karakaş, S., "Machine-based classification of ADHD and nonADHD participants using time/frequency features of event-related neuroelectric activity", *Clinical Neurophysiology*, Vol. 128(12), pp. 2400-2410, 2017.

[28] Heinrich, H., Hoegl, T., Moll, G. H., & Kratz, O., "A bimodal neurophysiological study of motor control in attention-deficit hyperactivity disorder: a step towards core mechanisms?", *Brain*, Vol. 137(4), pp. 1156-1166, 2014.

[29] Müller A, Vetsch S, Pershin I, Candrian G, Baschera GM, Kropotov JD, Kasper J, Rehim HA, Eich D., "EEG/ERP-based biomarker/neuroalgorithms in adults with ADHD: Development, reliability, and application in clinical practice", *The World Journal of Biological Psychiatry*, May 7, 2019.

[30] Raven, J., Court, J. H., "Manual for Raven's progressive matrices and vocabulary Scales", 1991, San Antonio, TX: Harcourt Assessment, 2003, updated 2004.

[31] Zoubek, L., Charbonnier, S., Lesecq, S., Buguet, A., & Chapotot, F., "Feature selection for sleep/wake stages classification using data driven methods", *Biomedical Signal Processing and Control,* Vol. 2(3), pp. 171-179, 2007.

[32] Tzanetakis, G., & Cook, P., "Musical genre classification of audio signals", *IEEE Transactions on speech and audio processing,* Vol. 10(5), pp. 293-302, 2002.

[33] Tsallis, C., Mendes, R., & Plastino, A. R., "The role of constraints within generalized nonextensive statistics", *Physica A: Statistical Mechanics and its Applications,* Vol. 261(3-4), pp. 534-554, 1998.

[34] Mormann, F., Andrzejak, R. G., Elger, C. E., & Lehnertz, K., "Seizure prediction: the long and winding road", *Brain,* Vol. 130(2), pp. 314-333, 2007.

[35] Shannon, C. E., "A mathematical theory of communication", *ACM SIGMOBILE mobile computing and communications review,* Vol. 5(1), pp. 3-55, 2001.

[36] Rényi, A., "On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability", Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California, 1961.

[37] Nai-Jen, H., & Palaniappan, R., "Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals", *In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 1, pp. 507-510, IEEE,September, 2004.

[38] Bai, J., & Ng, S., "Tests for skewness, kurtosis, and normality for time series data", *Journal of Business & Economic Statistics,* Vol. 23(1), pp. 49-60, 2005.

[39] Ansari-Asl, K., Chanel, G., & Pun, T., "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood", *In 2007 15th European Signal Processing Conference*, pp. 1241-1245, IEEE, September, 2007.

[40] Tang, W. C., Lu, S. W., Tsai, C. M., Kao, C. Y., & Lee, H. H., "Harmonic parameters with HHT and wavelet transform for automatic sleep stages scoring", REM, Vol. 365, pp. 8-6, 2007.

[41] Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S., "The timbre toolbox: Extracting audio descriptors from musical signals", *The Journal of the Acoustical Society of America,* Vol. 130(5), pp. 2902-2916, 2011.

[42] Kıymık, M. K., Güler, İ., Dizibüyük, A., & Akın, M., "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application", *Computers in biology and medicine,* Vol. 35(7), pp. 603-616, 2005.

[43] Percival, D. B., Walden, A. T., "Wavelet methods for time series analysis", Cambridge university press; 2000.

[44] Kaiser, A., Aggensteiner, P. M., Baumeister, S., Holz, N. E., Banaschewski, T., & Brandeis, D., "Earlier versus later cognitive event-related potentials (ERPs) in attention-deficit/hyperactivity disorder (ADHD): a meta-analysis", *Neuroscience & Biobehavioral Reviews,* Vol. 112, pp. 117-134, 2020.

[45] Jenke, R., Peer, A., & Buss, M., "Feature extraction and selection for emotion recognition from EEG", *IEEE Transactions on Affective computing,* Vol. 5(3), pp. 327-339, 2014.

[46] Park, K. S., Choi, H., Lee, K. J., Lee, J. Y., An, K. O., & Kim, E. J., "Emotion recognition based on the asymmetric left and right activation", *International Journal of Medicine and Medical Sciences,* Vol. 3(6), pp. 201-209, 2011.

[47] Ververidis, D., & Kotropoulos, C., "Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition", *Signal processing,* Vol. 88(12), pp. 2956-2970, 2008.

[48] Bishop, C. M., & Tipping, M., "Variational relevance vector machines", *arXiv preprint arXiv,*1301.3838, 2013.

[49] Borra, S., & Di Ciaccio, A., "Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods", *Computational statistics & data analysis,* Vol. 54(12), pp. 2976-2989, 2010.

[50] Fawcett, T., "An introduction to ROC analysis", *Pattern recognition letters,* Vol. 27(8), pp. 861-874, 2006.

[51] Hajian-Tilaki, K., "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation", *Caspian journal of internal medicine,* Vol. 4(2), pp. 627, 2013.

[52] Kropotov JD. Quantitative EEG, event-related potentials and neurotherapy. Academic Press; 2010.

[53] Thome, J., Ehlis, A. C., Fallgatter, A. J., Krauel, K., Lange, K. W., Riederer, P., & Gerlach, M., "Biomarkers for attention-deficit/hyperactivity disorder (ADHD)", A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *The World Journal of Biological Psychiatry,* Vol. 13(5), pp. 379-400, 2012.

[54] Ogrim, G., Kropotov, J., & Hestad, K., "The QEEG theta/beta ratio in ADHD and normal controls: sensitivity, specificity, and behavioral correlates", *Psychiatry Research,* Vol. 198(3), pp. 482-488, 2012.

[55] Baijot, S., Cevallos, C., Zarka, D., Leroy, A., Slama, H., Colin, C., & Cheron, G., "EEG dynamics of a go/nogo task in children with ADHD", *Brain sciences,* Vol. 7(12), pp. 167, 2017.

[56] Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., ... & Inagaki, M., "Applied machine learning method to predict children with ADHD using prefrontal cortex activity: a multicenter study in Japan", *Journal of attention disorders,* Vol. 24(14), pp. 2012-2020, 2020.

[57] Bzdok, D., & Yeo, B. T., "Inference in the age of big data: Future perspectives on neuroscience", *Neuroimage,* Vol. 155, pp. 549-564, 2017.