Identification of Attention Deficit Hyperactivity Disorder with Deep Learning Model

Ömer KASIM (omerksm@gmail.com)
Kutahya Dumlupinar University: Kutahya Dumlupinar Universitesi https://orcid.org/0000-0003-4021-5412

Research Article

Keywords: Attention Deficit Hyperactivity Disorder, Signal Decomposition, Power Spectral Density, Neighborhood Component Analysis, Deep Learning Model

Posted Date: April 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1227039/v1

Identification of Attention Deficit Hyperactivity Disorder with Deep Learning Model

Ömer KASIM[a*]

[a]Department of Electrical and Electronics Engineering, Simav Technology Faculty, Kutahya Dumlupinar University, Kutahya 43500, Turkey

∗Corresponding author.

E-mail address: omer.kasim@dpu.edu.tr (Ö.Kasım).

Abstract

This article explores the detection of Attention Deficit Hyperactivity Disorder (ADHD), a neurobehavioral disorder, from electroencephalography (EEG) signals. Due to the unstable behavior of EEG signals caused by complex neuronal activity in the brain, frequency analysis methods are required to extract the hidden patterns. In this study, the feature extraction was performed with the Multitaper and Multivariate Variational Mode Decomposition methods. Then, these features were analyzed with the neighborhood component analysis and the features that contribute effectively to the classification were selected. The deep learning model including the convolution, pooling, and bidirectional long short term cell and fully connected layer was trained with the selected features. The trained model could effectively classify the subjects with ADHD with a deep learning model, support vector machines and linear discriminant analysis. The proposed approach was validated with an ADHD open access dataset (doi:10.21227/rzfh-zn36). Experimental results showed that the proposed approach can innovatively classify ADHD subjects from Control group effectively. The proposed method was able to classify 1210 test samples in 0.1 seconds with an accuracy of 95.54%. The proposed method is promising for distinguishing subjects with attention deficit hyperactivity disorder effectively.

Keywords: Attention Deficit Hyperactivity Disorder, Signal Decomposition, Power Spectral Density, Neighborhood Component Analysis, Deep Learning Model.

1. INTRODUCTION

Classification of EEG signals is an important step in the design of the brain-computer interface (BCI) [1]. One of the BCI applications is Attention Deficit Hyperactivity Disorder (ADHD) detection. ADHD, a neurodevelopmental disorder, is characterized by executive functions and attention deficit. It affects approximately 5% of adults and 10% of children worldwide [2]. Also, it varies according to the population and the disease can be up to 20% of the population [3]. For the diagnosis of ADHD, experts use neuropsychological assessments and heterogeneous cognitive profiles. However, wide cognitive profiles complicates the diagnosis [4]. One of the methods used to support the diagnosis is the evaluation of Electroencephalography (EEG) signals. The diagnosis of ADHD can be made more safely by examining the signal changes in the response of the patients to different stimuli. A clear diagnosis of ADHD is important in solving individuals' social and psychiatric problems [5].

Various methods such as ERP method [1], statistical analysis of the signal [6], and observation of the Power Spectral Density (PSD) of the signal [7], application of photic stimuli [8] were proposed by using EEG in detecting ADHD. In these studies, wavelet transform, frequency space transformation, welch power spectrum transform were used and it was stated that it caused significant changes in the alpha band in the PSD of individuals with ADHD [9]. It was stated in studies that ADHD classification can be performed when power of the signal changes are applied to artificial intelligence algorithms [10]. When applying frequency transformation to extract spectral information from a signal, it is assumed to be a reliable representation of the relative phase of power coefficients obtained versus frequency. However, this assumption is not always valid. The average of the signal is used to solve this problem. Taking averaging weakens signal components. It is also unreliable in small data sets [11]. Instead of the mean process, the Multitaper method is the motivation of the study as it creates PSD and reduces the prediction bias by obtaining more than one independent estimation from the same sample. Frequency powers were also obtained in 3 different forms, and properties of outputs at different power levels were obtained with the Multivariate Variational Mode Decomposition (MVMD) method. These features allow the evaluation of the signal in 3 different bands.

Different methods have been used in the literature for the diagnosis of ADHD. These studies generally are performed with the artificial intelligence algorithm. The PSD changes in EEG signals with stimuli, features obtained with the help of wavelet transform and chaotic analysis are classified by machine learning algorithms, neural networks and deep learning models. Yang et al. processed the data obtained from the 128 channels with PCA algorithm for ADHD classification. The authors applied the processed data to the K-nearest neighbors (KNN) and SVM classifier as a set of features. In the experimental results, 83.33% cross-validation success was obtained with the K-NN algorithm with the highest success rate [12]. Khoshnoud et al. used 19-channel EEG signals in their study. Data recording was performed while resting with eyes closed. Approximate Entropy, Lyapunov Exponent and Multifractal Singularity Spectrum were used for feature extraction in EEG signals. These features applied to the radial basis function network (RBFN) and SVM to classify them. In the study, classification based on frequency band power was evaluated using the same type classifiers. An accuracy of 83.33% was achieved with SVM under four-fold cross-validation test. As a result, it has been observed that nonlinear features provide better separation between ADHD and control than band power characteristics [13]. Chen et al. obtained features from the power spectrum for ADHD detection from EEG signals. These properties applied to SVM are divided into 4 groups as relative spectral power, spectral power ratio, complexity and dual phase. An accuracy of 84.59% was obtained in the SVM method used in the classification performed with these features [14]. Jahanshahloo et al. obtained the feature vector of the study using the fractal dimension, band power, and wavelet and Autoregressive (AR) coefficients. ADHD classification with this feature vector was performed by SVM method. In the experimental results, it has been observed that the combination of fractal dimension and wavelet transform features achieve well discrimination ability. In the classification made using these features, 88.77% success was achieved with the SVM method as a result of the 10-fold cross-validation approach [15]. Mueller et al. used two age-matched groups of adults in their study. Two visual stimuli were applied to the 2 classified groups in their study. The ERP responses in EEG recordings were separated into independent component analysis (ICA) and ADHD classification was performed by SVM method. The classification accuracy was obtained as 91% by using the 10-fold cross-validation [16]. Dea et al. obtained the PSD properties of the EEG signal. The authors used principal component analysis (PCA) to reduce the data size of PSD features. Reduced features are classified with Support Vector Machine (SVM). As a result of the experiments, healthy individuals were classified with a success rate of 94.1% with ADHD samples [17]. Altınkaynak et al. used morphological features and wavelet coefficients properties for ADHD classification. ADHD classification has been performed by using various machine learning techniques with the obtained feature vectors. The highest success was obtained from the Multilayer Perceptron (MLP) method with 91.3% among Naive Bayes, Support Vector Machines, Multilayer Perceptron, Random Forest and Logistic Regression methods [18]. Khaleghi et al. aimed to find the most distinguishing features for ADHD detection from EEG signals. Double input symmetrical relevance (DISR), Mutual Information Maximization (MIM), Fast Correlation-Based Filter (FCBF) and Conditional Mutual Information Maximization criterion (CMIM) algorithms have been tested respectively. As a result, the highest accuracy (91.83%) was obtained with the combination of DISR and MLP [19]. Boroujeni et al. used features based on a combination of nonlinear features in the ADHD classification. In the calculation of the features, the chaotic time series of the EEG obtained from FP1, FP2, F3, F4 and Fz were analyzed. It provided an accuracy of 96.05% in the experiments performed [20].

Studies involving deep learning methods have also been suggested in the literature. Chen et al. proposed a Convolutional Neural Network (CNN) -based method for detecting ADHD from EEG signals in their study. The feature extraction was performed by arranging the order of the channels belonging to the EEG signals. In addition to these features, a success of 94.67% was achieved with the feature matrix obtained by calculating 13 features [21]. Dubreuil et al. detected ADHD with a CNN model trained using the stacked multi-channel EEG time-frequency separations of ERP. With its model trained with 2800 feature vectors, higher success (88%) was obtained than RNN [22]. Marcano et al. used 5 EEG channels selected for the ratio of theta and beta power values measured during an attention task. A probability ratio detector is designed in the study. The area under curve (AUC) at resting and one excitation was achieved as 73%. It was also obtained as a false positive rate (FPR) of 0.32 [23].

In this study, the Multitaper method and MVMD were used in an innovative way together with the Neighborhood component analysis (NCA) and deep learning model (DLM) with Bidirectional Long Short Term Memory (BLSTM) cell. The frequency-power values generated by the data obtained from each EEG channel with the Multitaper method and MVMD were used to obtain the components of the EEG signal. The 2% tolerated 408 features were selected with the NCA algorithm. Support vector machines (SVM), Linear Discriminant Analysis (LDA) and DLM classifiers were trained with 2/3 of these features. Then the 1/3 holdout validation result of the experiments, the highest success was obtained by using NCA and DLM methods together with 95.54%. Since the NCA method enables the classifier to classify with fewer features, the processing time is shortened.

The main contributions of this study are as follows.

1. The MVMD contributes to noise immunity and mode alignment, as well as the removal of negativity in the EEG signal. While the Multitaper method creates power spectral density (PSD), data loss is prevented by taking averages because it reduces the prediction bias by obtaining more than one independent estimation from the same sample.

2. The proposed method containing Multitaper-NCA and DLM was tested by separating from the 33.3% holdout validation group data set using EEG signals from patients diagnosed with ADHD. The method suggested as a result of the test reached a classification accuracy of 95.54%, which is much higher than the LDA SVM and DLM. When effective features are used in classification; the high ADHD classification is clearly obtained and processing time is substantially decreased.

The article is organized as follows. Section 2 covers the methodology of detection of ADHD disorder in obtained dataset from hospital. Experimental results are shown in Section 3. A detailed discussion is given the proposed approach based on DLM with ADHD patients EEG data analysis in Section 4. Finally, a brief conclusion is given in Section 5.

2. MATERIAL and METHODS

2.1. Architecture of the Proposed ADHD Identification Model

In the method developed in the proposed study, Multitaper, MVMD, NCA and DLM were used together. There are 3630 data in the data set in which the model is evaluated. 1830 of these data are subjects with ADHD and the rest of it consists of a control group. In the recording of the data, RAW EEG records belonging to 8 of 19 channels that included potential differences between electrodes were obtained using the international 10-20 system. These channels are C3, C4, P3, P4, T5, T6, O1, and O2. These channels are especially preferred because they are regions where eye-blink artifacts have few effects. As shown in Figure 1, power values of 1-49 Hz frequencies were obtained by applying the Multitaper method and the EEG signals of each one of the 8 channels are divided into 3 components. 407 features were selected that achieve optimal classification by using the NCA feature selector algorithm. The subjects with ADHD were classified into the *malignant* class and the control group were labeled as *benign* class by the DLM.

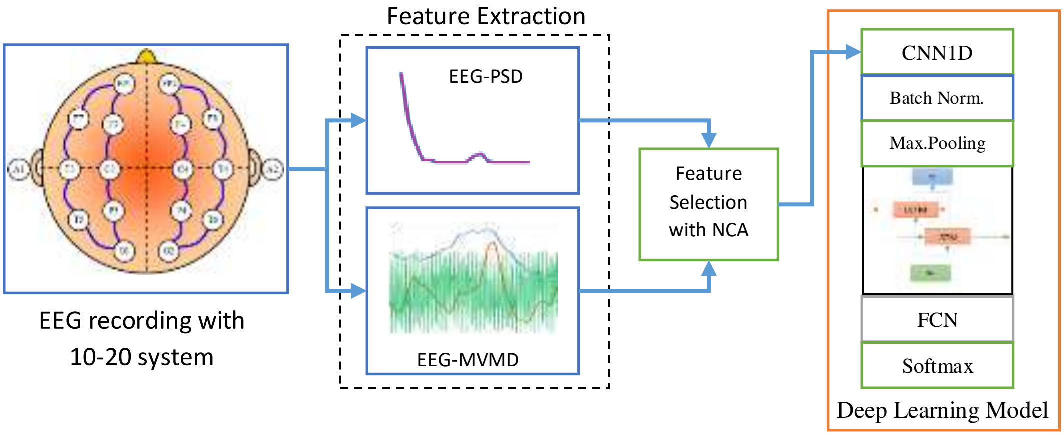

Figure 1. The overall framework of proposed system

2.2. ADHD Dataset

The EEG data used in the study were obtained using the potential differences between the electrodes placed according to the international 10-20 system. Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2 channels, data received from 19 channels with 128 Hz sampling frequency are included. The A1 and A2 electrodes were the references with the earlobes. Visual attention tasks are included in the EEG recording protocol. With the continuous stimulation applied on each task, the subjects were asked to count certain visuals. EEG recordings corresponding to these stimuli were obtained. Records in the data set belong to 121 subjects, aged 7–12 years [24]. Of these, 61 subjects were evaluated in ADHD and 60 were in the control group. There were no reports of psychiatric disorders, epilepsy, or any high-risk behaviors in the control group. EEG signals recorded from 121 subjects presented in the data set were used to obtain the data used in the study. The 30 segments were obtained from each of these signals in 10-second intervals. These segments were obtained randomly from different locations in time without overlapping each other. Experiments were conducted with a total of 3630 data.

2.3. Feature Extraction with Multivariate Variational Mode Decomposition

MVMD makes it possible to use one-dimensional Variational Mode Decomposition as multi-dimensional. EEG recordings consisting of multi-channel signals can be processed with MVMD. In addition, this method ensures consistency of multi-channel component frequencies [34]. MVMD involves extending the signal to multivariate data instead of parsing a one-dimensional signal $u_k(t)$ into k mode $x(t) = \sum u_k(t)$. Hilbert-Huang Transform is used to obtain the one-sided spectrum. Then u(t) is used to determine the center frequency. It is then modulated to the fundamental frequency corresponding to the frequency spectrum of each mode, multiplied by the exponential term to determine the corresponding center frequency w(t). In MVMD, the multivariate modulated oscillations of K are calculated by obtaining $u_k(t) = [u_1(t) + u_2(t) + u_3(t)]$ using the signal $u_k(t)$. The optimization function in this case is obtained by Equation 1 [25].

$$minimize(\sum_k \sum_c \left\| \partial_t \left[u_{k,c}(t). e^{-jw_k t} \right] \right\|^2) \tag{1}$$

In Equation 1, the term $u_{k,c}(t)$ is a complex valued signal with a single frequency w_k component in each channel. The channel number c and the mode number k indicate the analytically modulated signal. For the optimization specified in Equation 1, firstly, the constrained optimization problem is transformed into an unconstrained optimization problem. With this transformation, the problem is obtained as an augmented Lagrangian function by adding two penalty terms. The Lagrangian function is expressed by Equation 2.

$$\mathrm{L}(\mathrm{u}_{k,c}, \omega_k, \lambda_c) = \alpha \sum_k \sum_c \left(\partial_t [u_{k,c}(t). e^{-jw_k t}]\right)^2 + \sum_c x_c(t) - \sum_k u_{k,c}(t)^2 + \sum_c \lambda_c, x_c(t) - \sum_k u_{k,c}/t) \quad (2)$$

In Equation 2, α denotes the balance parameter used to provide the necessary data accuracy constraint. λ is the Lagrange multiplier. The problem that turns into an unconstrained optimization problem is solved by using the alternative direction method (ADMM) algorithm of the multipliers and the components of different channels and frequency bands are obtained.

2.4. Feature extraction using Multitaper

The Multitaper method is used to obtain the power spectral density by moving the information contained in a signal to the frequency space. The power spectrum is formed by distributing the average power of a signal to certain frequency values in the signal [26]. The average power of the $x[n]$ that is the discrete time signal in the range n_1 and n_2 is obtained. The total energy of the signal in N finite time is also finite. This situation is shown by Equation 3.

$$\frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N.F_s} \int_0^{F_s} |T_x(f)|^2 df \quad (3)$$

In Equation 3, F_s is the sampling frequency and T is the Fourier transform of f. The power spectrum of $x[n]$ and the discrete time Fourier transform of $x[n]$ are limited by finite time N is defined by Equation 4.

$$S_x^N(f) = \frac{1}{N.F_s} |T_x(f)|^2, f \epsilon [0, F_s] Hz \quad (4)$$

The average power of $x[n]$ from 0 to $N-1$ is obtained, when integrating $S_x^N(f)$from 0 to F_s in Equation 2. Thus, $S_x^N(f)$ is calculated as the distribution of the finite average power over the frequency bands of x[n]. In the Multitaper method, the contracting window function is used instead of rectangular data windows to reduce the periodogram-deviation problem.

A tapering window function w is defined by k and the sub-sequence $x_j[n]$ periodogram corresponding to column j of a single-conical spectrogram is calculated by Equation 5.

$$S_{x_j}^{L,k}(f) = \frac{1}{F_s} \left| T_{w_k^L x_j}(f) \right|^2, f \epsilon [0, F_s] Hz \quad (5)$$

In Equation 5, T function is the Fourier transform of the w function. The estimator S is obtained by multiplying this transformation by the sampling frequency. Although the obtained S is an approximation of the long-term power spectrum, the periodogram-variance problem needs to be solved. In solving this problem, more than one tapering window functions are used to reduce the deviation and variance found in the periodogram. The windows are shown with $\mathrm{W} = \{\mathrm{w}_1^L, \mathrm{w}_2^L, \ldots, \mathrm{w}_k^L\}$. Each W window in the array is defined as K tapering window of length L. The multi-stage spectral estimator using W windows is obtained by Equation 6.

$$S_{x_j}^{w}(f) = \frac{\sum_{k=1}^{K} S_{x_j}^{L,k}(f)}{K}, f \epsilon [0, F_s]\ Hz \quad (6)$$

2.5. Feature selection applying dimensionality reduction Neighborhood component analysis (NCA) algorithm

NCA is a nonparametric feature selection method. It produces non-negative weights for all properties. NCA produces non-negative weights. Relief's negative weights mean an excess of features. The negatively weighted features are pruned. Then, the positive weighted features are selected by using the most distinctive features to generate weights. Before weights are produced, the properties are normalized using min - max normalization [27].

$$W = (\mathrm{minmax}(\mathrm{fr}), \mathrm{t\_v}) \tag{7}$$

In Equation 7, W denotes the weight vector of NCA. The feature vector is fr and t_v is the target vector. Weights are obtained by matching the properties normalized with min-max with the target vector.

2.6. Linear Discriminant Analysis

The LDA method separates the two classes using a linear boundary between properties. In separation, the argument is expressed as a linear variable. This argument appears as a label for a class [28]. First, models of probability density functions are obtained for data generated from each class. Then, a new data point is classified by determining the probability density function whose values are greater than the others. The separation function of the LDA classifier is a linear compound of X's complements. This calculation is expressed as in Equation 8 [28].

$$\mathrm{D} = \mathrm{wX} + \mathrm{m_0} \tag{8}$$

In Equation 7, w is the weight vector and m_0 is the bias value. The decision-making for classes is defined with D value. The X is a pXN_k matrix of N_k samples. These samples are p-dimensional data from class k. μ_k means the previous probabilities of each class and δ is the covariance matrix. Each x value is obtained with argmax as in Equation 9. The resulting LDA decision boundaries are linear across data classes.

$$argmax\, x^T \delta^{-1} \mu_k - \frac{1}{2} \mu_k^T \delta^{-1} \mu_k \tag{9}$$

Consequently, discrimination is a predominantly linear combination of predictors. Generally, estimators with large differences between class averages will have larger weights, also when the class averages are similar the weights will be small.

2.7. Support Vector Machines

SVM method transforms the input data vectors into a higher dimensional by passing it through a kernel process. The data in the area resulting from the transformation are classified by modeling complex decision boundaries with a hyperplane. In the classification process, the distance between the hyperplane and the nearest data point is maximized [29]. Generally, SVM can be formulated as seen in Equation 9.

$$\min_{(w,b)} \sum_{i=1}^{N} f\left(y_i, x_i^T . w + b\right) + \alpha . y(w) \tag{10}$$

In Equation 10, w is the weight vector and b is the bias value. The i'th input and output pair (x_i,y_i) is obtained, where x_i is the input and y_i is the output. The estimated output value of the i th sample is calculated with $x_i^T . w + b$. N is the number of samples and y(w) is the regularized term. α, on the other hand, is a non-negative parameter used to balance between the data fitting loss term and the regulator term.

2.8. DLM with Bidirectional Long Short Term Memory Cell

Long and Short Term Memory (LSTM) model is used in solving sequential classification problems. The LSTM unit determines whether the existing memory will be stored or updated with new information. Therefore, the LSTM-RNN is capable of modeling long-range dynamic dependencies, thus avoiding the problem of vanishing gradients problem during training [30]. LSTM architecture has an input gate, forgetting gate and output gate. A single LSTM unit is defined by Equation 11.

$$i_t^j = \sigma(U_i x_t + W_i h_{t-1} + b_i)^j \tag{11}$$

In Equation 10, W is the weight matrix and b is the deviation variable. i is the entrance gate of the j th LSTM unit at time t. σ is expressed as the sigmoid function. The input data at time t is expressed as x_t and the output of the previous LSTM unit is h_t.

$$f_t^j = \sigma(U_f x_t + W_f h_{t-1} + b_f)^j \tag{12}$$

In Equation 12, f_t describes the forgetting gate. In the forget gate, the importance of information is calculated and unnecessary information is discarded.

$$c_t^{\sim j} = \tanh(U_c x_t + W_c h_{t-1} + b_c) \tag{13}$$

In Equation 13, $c_t^{\sim j}$ represents the new memory gate unit and the memory content of the previous unit is expressed as b. The new memory content is calculated by forget gate unit. This represents updated memory content.

$$c_t^j = f_t^j c_{t-1}^j + i_t^j c_t^{\sim j} \tag{14}$$

The update in LSTM block is performed and c_t^j is obtained by using Equation 14. The o_t^j expressed as the output unit that controls the final output state. The output of the LSTM cell h_t^j is calculated by Equation 15 at the last LSTM output unit that enabled at time t.

$$h_t^j = o_t^j.\tanh\left(c_t^j\right) = \sigma(U_0 x_t + W_0 h_{t-1} + b_0)^j.\tanh\left(c_t^j\right) \tag{15}$$

BLSTM model has the ability to access content in both forward and backward directions. The demonstration of the BLSTM model is presented in Figure 2.

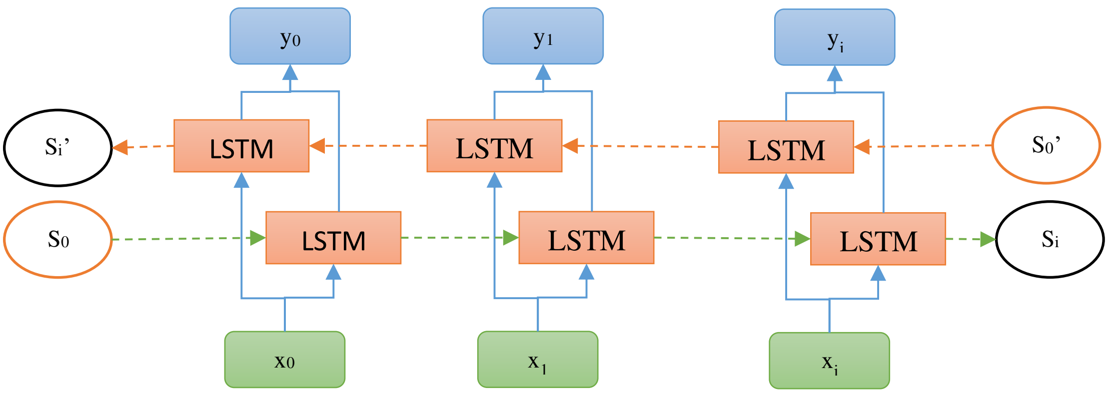

Figure 2. Architecture of Bidirectional Long Short Term Memory Networks

For the hyper-parameters of DLM, the intermediate layer number was set as 100, the initial learning rate was 0.05, the gradient threshold was 1, and the mini batch size was 384 with the best performance. The parts in the architecture of the layers belonging to the DLM model created in this study are shown in Table 1.

Table 1. The DLM structure of the Proposed Classifier

Layer (type)	**Output Shape**
conv1d_1 (Conv1D)	(1,407)
batch_normalization_1	(1,407)
activation_1(Activation)	(1,407)
Blstm_1 (BLSTM)	(100)
dense_1(Dense)	(50)
dense_2(Dense)	(2)

3. EXPERIMENTAL RESULTS

This section covers the application and evaluation of the combined use of Multitaper, MVMD, NCA and DLM modules proposed in this study. First, the experimental setup, performance criteria and dataset are expressed. Then, the results of the experiments performed in the data set of the study are presented to validate the approach. Finally, a comparison is made between the approach applied and the ADHD classification methods suggested in the literature. The proposed method has been implemented in the Python programming environment. Experiments with the proposed method were carried out on an i7 9900 Intel processor running at 2.40GHZ, 32GB of RAM and an NVidia GPU.

The accuracy, precision, recall and f1 score metrics are used to measure the performance of the proposed approach. The true positive (TP), true negative (TN), false positive (FP) and false negative (FN) expressions are used in the calculation of these metrics. These expressions are derived from the confusion matrix. TP refers to ADHD subjects who are correctly classified. FP shows the subjects with ADHD but included in the control group. Shows subjects in the FN control group but classified as ADHD. TN represents the correct classification in the control group. The parameters obtained with these parameters are obtained by Equations 16, 17, 18 and 19 respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$Prn = \frac{TP}{TP + FP} \tag{17}$$

$$Rcl = \frac{TP}{TP + FN} \tag{18}$$

$$f_1 = \frac{2. Prn. Rcl}{Prn + Rcl} \tag{19}$$

EEG signal segments were obtained in 10 seconds using raw data. Power distribution at frequencies between 1-49Hz and features of 8 channels divided into 3 components were obtained by applying the Multitaper transformation and MVMD of these segments. The data set of the study, which consists of 407 selected features by NCA, includes 3630 feature vectors. It contains the PSD graph of the individual with ADHD. There is the PSD graphic of the normal individual. In the study, *nw* parameter was chosen as 1.25. This parameter mostly reflects the power change in the graphics marked in red. As can be seen in the graphs, more power changes occur in subjects with ADHD. An example of the signal separated into its components by MVMD is shown in Figure 3.

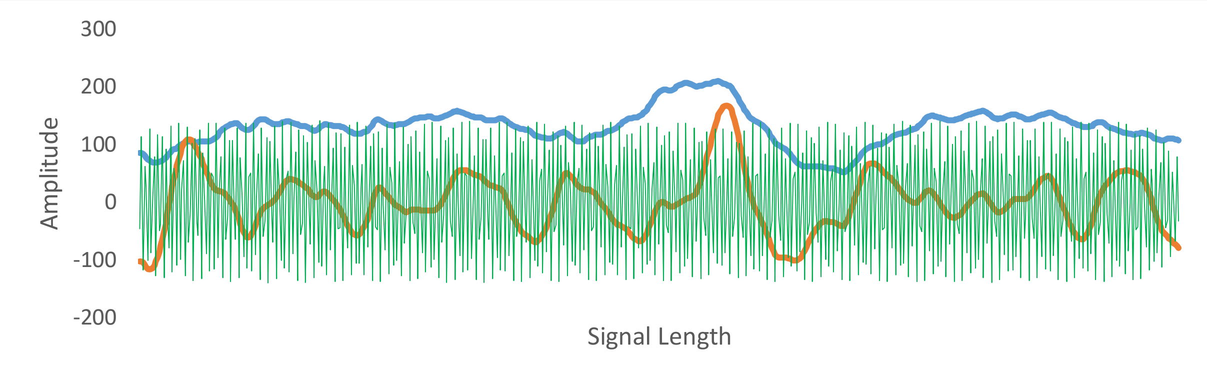

Figure 3. Sample of signal separated into components with MVMD (Blue: Low frequency part of the signal, orange: medium frequency part of the signal, Green: High frequency part of the signal)

4. DISCUSSION

2420 of 3630 data were used in the training of 3 separate classifiers. The classifiers trained in these data were applied to holdout validation with 1210 data, and their superiority to each other was revealed. While 76.38% accuracy was obtained with LDA classifier, 81.69% accuracy was obtained as a result of experiments with SVM classifier. The accuracy was achieved as 95.54% with the DLM designed in this study.

Table 2. Confusion Matrix of Proposed Multitaper-NCA-BLSTM Method

	Actual Normal	Actual With-ADHD
Predicted Normal	568	32
Predicted With-ADHD	22	588

The NCA feature selection algorithm was applied to the data set of the study to increase the effectiveness of it. The method, which was optimized by selecting the best 407 features of NCA, both increased the success of DLM and an effective system that works faster has emerged as a result of experiments. Finally, the performance criteria of the proposed approach are compared with different studies developed in the literature.

The confusion matrix of the best model is presented in Table 2. Of the 600 subjects in the control group, 568 were correctly classified. Among the group with ADHD, 588 out of 610 subjects were classified correctly. The FP rate was only 22 (3.6%). Precision, recall and f1 score were calculated as 0.95, 0.96 and 0.95, respectively. The ROC curve for the success of the method is presented in Figure 4. The area under the curve was obtained as 0.96.

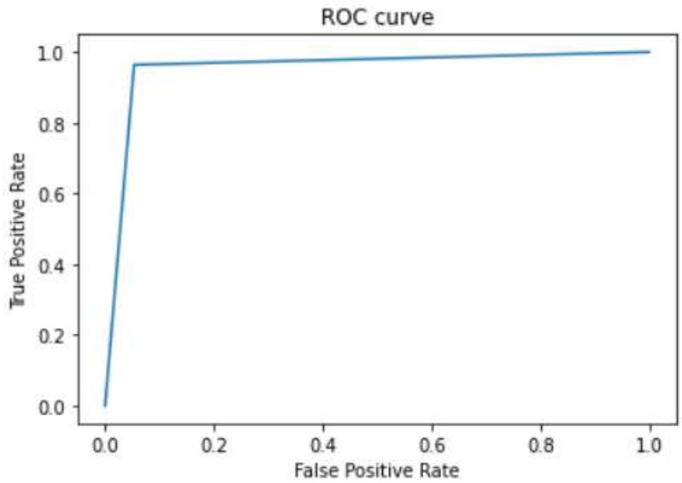

Figure 4. ROC Curve of holdout validation results of the proposed model

Table 3. Training and Testing Time comparison for single DLM, SVM and LDA with proposed method

Methods	Training Time (sec) (60% of data)	Testing Time (sec) (40% of data)
Pure SVM	72.223	0.0406
SVM with NCA	72.068	0.0142
Pure LDA	0.2340	0.0268
LDA with NCA	0.0427	0.0057
DLM	124.71	0.1770
DLM with NCA	117.63	0.1069

The performance analysis of the study was made by comparing the performance of other known classification methods. The experimental results obtained regarding the identifying of ADHD times are shown in Table 3. As a result, Multitaper-NCA-DLM with 407 features reduced training and test times were compared with SVM, LDA and BLSTM. The proposed method allows very low processing time compared to BLSTM with a similar number of features (407 features).

The detection accuracy of the model proposed in the relevant studies in the literature in ADHD classification and the accuracy rates obtained from other classification algorithms are compared in Table 4. Different methods have been used for ADHD detection. Feature extraction has been performed using methods such as multifractal singularity spectrum, approximate entropy, PSD, largest Lyapunov exponent, wavelet coefficients, chaotic time series analysis and Spectrogram. The feature vector is given to the classifier directly or by processing with feature reduction algorithms. Classification is made with machine learning, neural network or deep learning models by processing features with algorithms such as PCA and DISR.

Among the studies using SVM machine learning, Dea et al. achieved 94.1% success by using PSD, PCA and SVM methods together [17]. The highest accuracy with MLP was obtained as 91.83% by using DISR and MLP together in Khaleghi et al.'s study [19]. With the combination of Spectrogram and CNN deep learning method, 88% success was achieved. Fouladvand et al. stated that they made the detection of ADHD with LSTM with 84% accuracy [31].

In the experiments conducted with the same data set, it was shown that the entropy measurements of especially the recordings in the C3 channel were effective in detecting ADHD. In the study, subjects with ADHD were classified with an accuracy of 93.65% with holdout validation 30% data slice [32]. When the trained model was tested with a 1/3 of the whole dataset, a test accuracy of 95.54% was achieved. These results show that the proposed method is more successful and effective than the methods suggested in the literature. While using the Multitaper method (PSD) used in the proposed method, data loss was prevented by taking an average because it reduces the prediction bias by obtaining more than one independent estimate from the same sample. Abbas et al found distinctive features in the Beta power band in their experiments with the same data set. In the experimental results, the AUC success metric for ADHD detection in this band was 0.7585 [33].

5. CONCLUSION

In this study, a method that performs ADHD diagnosis from EEG signals in which Multitaper, MVMD, NCA and DLM are used together in an innovative way is proposed. The dataset obtained from the EEG data obtained from 121 subjects. In addition, the results of previous studies are compared with the performance metrics obtained. PSD values of 8 channels and 3 signal components were obtained with Multitaper and MVMD, which were least affected by eye-blink artifact and significantly increased in stimulation in ADHD subjects. The feature vector reduction was implemented with NCA to improve performance metrics. Many DLM variants were also checked for false positives to achieve the best data generalization. It was found that the best 407 feature selection and hyper-parameters presented in the study were improved. 1210 data could be classified with 95.54% holdout validation accuracy. Furthermore, the classification performance obtained with deep learning was obtained more successfully than SVM and LDA classifiers. It shows that the proposed method for accuracy in experiments deals with False-positives less than other ADHD classification methods. Faster training and higher success level of DLM used with NCA provided an advantage over deep learning methods. The training takes about 117 seconds and checks the ADHD in 1210 data about 0.1 seconds.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This work was not funded by any organization.

REFERENCES

1. Homri, I., & Yacoub, S. (2019). A hybrid cascade method for EEG classification. Pattern Analysis and Applications, 22(4), 1505-1516.
2. Dubreuil-Vall, L., Ruffini, G., & Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. Frontiers in neuroscience, 14.
3. Taylor, E. (1994). Syndromes of attention deficit and overactivity. *Child and adolescent psychiatry: Modern approaches*.
4. Amado-Caballero, P., Casaseca-de-la-Higuera, P., Alberola-Lopez, S., Andres-de-Llano, J. M., Lopez-Villalobos, J. A., Garmendia-Leiza, J. R., & Alberola-Lopez, C. (2020). Objective

ADHD diagnosis using Convolutional Neural Networks over Daily-Life Activity Records. *IEEE Journal of Biomedical and Health Informatics*.
5. Alba, G., Pereda, E., Mañas, S., Méndez, L. D., González, A., & González, J. J. (2015). Electroencephalography signatures of attention-deficit/hyperactivity disorder: clinical utility. *Neuropsychiatric disease and treatment*, *11*, 2755.
6. Kompatsiari, K., Candrian, G., & Mueller, A. (2016). Test-retest reliability of ERP components: a short-term replication of a visual Go/NoGo task in ADHD subjects. *Neuroscience Letters*, *617*, 166-172.
7. Marcano, J. L., Bell, M. A., & Beex, A. L. (2018). Classification of ADHD and non-ADHD subjects using a universal background model. *Biomedical signal processing and control*, *39*, 204-212.
8. Tosun, M. (2021). 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, 1-10.
9. McAuliffe, D., Hirabayashi, K., Adamek, J. H., Luo, Y., Crocetti, D., Pillai, A. S., ... & Ewen, J. B. (2020). Increased mirror overflow movements in ADHD are associated with altered EEG alpha/beta band desynchronization. *European Journal of Neuroscience*, *51*(8), 1815-1826.
10. Aydemir, E., Tuncer, T., & Dogan, S. (2020). A Tunable-Q wavelet transform and quadruple symmetric pattern based EEG signal classification method. *Medical Hypotheses*, *134*, 109519.
11. Oliveira, G. H., Coutinho, L. R., da Silva, J. C., Pinto, I. J., Ferreira, J. M., Silva, F. J., ... & Teles, A. S. (2020). Multitaper-based method for automatic k-complex detection in human sleep EEG. *Expert Systems with Applications*, 113331.
12. Yang, J., Li, W., Wang, S., Lu, J., & Zou, L. (2016). Classification of children with attention deficit hyperactivity disorder using PCA and K-nearest neighbors during interference control task. In *Advances in Cognitive Neurodynamics (V)* (pp. 447-453). Springer, Singapore.
13. Khoshnoud, S., Nazari, M. A., & Shamsi, M. (2018). Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals. *Journal of integrative neuroscience*, *17*(1), 17-30.
14. Chen, H., Chen, W., Song, Y., Sun, L., & Li, X. (2019). EEG characteristics of children with attention-deficit/hyperactivity disorder. *Neuroscience*, *406*, 444-456.
15. Jahanshahloo, H. R., Shamsi, M., Ghasemi, E., & Kouhi, A. (2017). Automated and ERP-based diagnosis of attention-deficit hyperactivity disorder in children. *Journal of medical signals and sensors*, *7*(1), 26.
16. Mueller, A., Candrian, G., Grane, V. A., Kropotov, J. D., Ponomarev, V. A., & Baschera, G. M. (2011). Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. *Nonlinear biomedical physics*, *5*(1), 5.
17. De Dea, F., Ajčević, M., Stecca, M., Zanus, C., Carrozzi, M., Cuzzocrea, A., & Accardo, A. (2019). A Big-Data-Analytics Framework for Supporting Classification of ADHD and Healthy Children via Principal Component Analysis of EEG Sleep Spindles Power Spectra. *Procedia Computer Science*, *159*, 1584-1590.
18. Altınkaynak, M., Dolu, N., Güven, A., Pektaş, F., Özmen, S., Demirci, E., & İzzetoğlu, M. (2020). Diagnosis of Attention Deficit Hyperactivity Disorder with combined time and frequency features. *Biocybernetics and Biomedical Engineering*.
19. Khaleghi, A., Sheikhani, A., Mohammadi, M. R., Nasrabadi, A. M., Vand, S. R., Zarafshan, H., & Moeini, M. (2015). EEG classification of adolescents with type I and type II of bipolar disorder. *Australasian physical & engineering sciences in medicine*, *38*(4), 551-559.
20. Boroujeni, Y. K., Rastegari, A. A., & Khodadadi, H. (2019). Diagnosis of attention deficit hyperactivity disorder using non-linear analysis of the EEG signal. *IET systems biology*, *13*(5), 260-266.
21. Chen, H., Song, Y., & Li, X. (2019). A deep learning framework for identifying children with ADHD using an EEG-based brain network. *Neurocomputing*, *356*, 83-96.

22. Dubreuil-Vall, L., Ruffini, G., & Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Frontiers in neuroscience*, *14*.
23. Marcano, J. L., Bell, M. A., & Beex, A. L. (2018). Classification of ADHD and non-ADHD subjects using a universal background model. *Biomedical signal processing and control*, *39*, 204-212.
24. Ali Motie Nasrabadi, Armin Allahverdy, Mehdi Samavati, Mohammad Reza Mohammadi, June 10, 2020, "EEG data for ADHD / Control children", IEEE Dataport, doi: https://dx.doi.org/10.21227/rzfh-zn36.
25. Gavas, R., Jaiswal, D., Chatterjee, D., Viraraghavan, V., & Ramakrishnan, R. K. (2020, March). Multivariate Variational Mode Decomposition based approach for Blink Removal from EEG Signal. In 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 1-6).
26. Wieczorek, M. A., & Simons, F. J. (2007). Minimum-variance multitaper spectral estimation on the sphere. *Journal of Fourier Analysis and Applications*, *13*(6), 665-692.
27. Goldberger, J., Hinton, G. E., Roweis, S. T., & Salakhutdinov, R. R. (2005). Neighbourhood components analysis. In *Advances in neural information processing systems* (pp. 513-520).
28. McLachlan, G. J. (2004). *Discriminant analysis and statistical pattern recognition* (Vol. 544). John Wiley & Sons.
29. Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of machine learning research*, *2*(Dec), 125-137.
30. Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*, *18*(5-6), 602-610.
31. Fouladvand, S., Hankosky, E. R., Henderson, D. W., Bush, H., Chen, J., Dwoskin, L. P., ... & Zhang, G. Q. (2018, June). Predicting Substance Use Disorder in ADHD Patients using Long-Short Term Memory Model. In *2018 IEEE International Conference on Healthcare Informatics Workshop (ICHI-W)* (pp. 49-50).
32. Mohammadi, M. R., Khaleghi, A., Nasrabadi, A. M., Rafieivand, S., Begol, M., & Zarafshan, H. (2016). EEG classification of ADHD and normal children using non-linear features and neural network. *Biomedical Engineering Letters*, *6*(2), 66-73.
33. Abbas, A. K., Azemi, G., Amiri, S., Ravanshadi, S., & Omidvarnia, A. (2021). Effective connectivity in brain networks estimated using EEG signals are altered in children with attention deficit hyperactivity disorder. Computers in Biology and Medicine, 104515.
34. Chang, L., Wang, R., & Zhang, Y. (2022). Decoding SSVEP patterns from EEG via multivariate variational mode decomposition-informed canonical correlation analysis. Biomedical Signal Processing and Control, 71, 103209.