Ensemble learning in detecting ADHD children by utilizing the non-linear features of EEG signal*

Pham Thi Viet Huong[1], Nguyen Anh Tu[2], Tran Anh Vu[2]**

[1] International School, Vietnam National University, Hanoi, Vietnam
[2] Hanoi University of Science and Technology, Hanoi, Vietnam
**vu.trananh@hust.edu.vn

Abstract. Electroencephalogram (EEG) has play a critical role in the assessment of Attention-Deficit Hyperactivity Disorder (ADHD) in patients. In this paper, we proposed a novel method, which utilizes the non-linear features of EEG signal in discriminating EEG children with healthy group. Since most of the previous research focused on linear feature of EEG, this paper opens a new aspect on analyzing EEG in the task of detecting ADHD in humans. Our dataset is recently published in 2020 in ieee-dataport.org. We use the Fractal Dimensions (FD) as non-linear features with different method of feature selection. Finally, we use ensemble learning as a classifier to discriminate ADHD children with healthy group. Our result confirmed our methodology as it has higher accuracy when compared with state-of-the-art studies..

Keywords: Attention-Deficit Hyperactivity Disorder (ADHD), Electroencephalogram (EEG), Fractal Dimension (FD), Ensemble learning.

1 Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a mental disorder that is characterized by an ongoing pattern of inattention and/or hyperactivity impulsivity that interferes with functioning or development [1]. According to recent studies, around 5% of children are affected by the ADHD, with boys having a higher risk than girls [1] [2]. Normally, ADHD symptoms appear in preschool age and become critical in primary shool age. The main problem of ADHD in children is the lack of concentration and weak regulation of their behaviors, so they do not show appropriate react to the surrounding environment [3] [4] [5]. Therefore, early diagnosis of ADHD is extremely important in preventing later complications such as negative impacts on children's social interactions.

** Corresponding Author.

Usually, the diagnosis of ADHD is mainly based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD) [1] [6]. This diagnosis is highly dependent on a parent or teacher's perception of the psychologist's questions and the truthfulness of their answers. To minimize this subjective factor, objective ways have been developed to identify children with symptoms of ADHD. One way is to use electroencephalogram (EEG) in the diagnosis [7] [8] [9] [10], which is a recording of brain activity. In order to get EEG, small sensors are attached to the scalp to catch the electrical signal produced when brain cells send message to each other.

EEG processing has become one of the most widely used techniques for ADHD diagnosis due to its accessibility and non-expensive characteristics. Researchers have been developed several methods to deal with EEG in differentiating ADHD group and healthy group. The very first research in developing a rationale for the diagnosis of ADHD was taken in [11] for 15 years. He found that in ADHD people, the theta activity increased, and beta power dramatically reduced. In [12], 30 ADHD children and 30 healthy children were studied and results showed that ADHD group had greater absolute power in delta and theta oscillations in all regions of their brain. ADHD adults and healthy groups were classified using support vector machine based on power spectra in [13].

The most commonly used machine learning algorithms for classification of ADHD patterns using EEG are Logistic Regression [14], Linear Discriminant Analysis (LDA) [15], K-Nearest Neighbor [16], Support Vector Machine (SVM) [17], Principal Component Analysis (ICA) [18], Fast Fourier and Wavelet Transform [19] and Neural Networks [20] [21]. Deep learning methods are also utilized to perform the task, for example, convolution neural networks (CNN) [22] [23].

The non-linear features of EEG signal such as entropy and Lyapunov exponent were taken advantage in differentiating the ADHD group in [24]. In order to improve the classification results, the double input symmetrical relevance (DISR) and minimum Redundancy Maximum Relevance (mRMR) methods were used to choose the best features to put into the neural network. Results showed that the extracted non-linear features revealed that non-linear indices were greater in different regions of the brain of the ADHD children compared to healthy children. As expected, ADHD children have more delays and less accurate in cognitive tasks.

Our proposed method also utilized from the non-linear features of the EEG signal. We use fractal dimension (FD) based metrics such as Higuchi, Katz and Petrosian fractal dimensions to define the chaotic pattern in EEG signal. Instead of using some given tools in Matlab to select the features, such as DISR and mRMR [24], we perform different methods: filter method, Correlation-based Feature Selection (CFS), Lasso method, logistic method, wrapper method, recursive feature elimination (RFE), which dig more into the physics of the EEG signal. After feature selection, we use ensemble learning to perform the task. Our achieved results are better than current research for the same purpose.

Our paper is organized as follow. Section I is the introduction. Section II presents the dataset and methodology we use to perform the task. Section III shows the experiment and results. Section IV concludes the paper.

2 Data and Methodology

2.1 Dataset

Our dataset is taken from ieee-dataport.org, which is IEEE's dataset storage and dataset search platform. The dataset is the EEG signal from 61 children with ADHD and 60 healthy controls (boys and girls, age 7-12). The ADHD group was diagnosed using DSM-IV criteria by a qualified psychiatrist and this group was given Ritalin for up to 6 months. DSM-IV criteria is the official guide of the American Psychiatric Association, which is intended to offer a framework for categorizing disorders and defining diagnostic criteria for the disorders listed. None of the children in the control group had a history of psychiatric disorders, epilepsy, or any report of high-risk behavior. EEG recording was performed based on 10-20 standard by 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) at 128 Hz sampling frequency. The A1 and A2 electrodes were the references located on earlobes.

The EEG recording methodology was based on visual attention tasks, since visual attention is one of the impairments in in ADHD children. A series of cartoon character photos were given to the children, and they were instructed to count the figures. The number of characters in each image was chosen at random between 5 and 16, and the images were large enough for children to be easily see and count. To have a continuous stimulation during the signal recording, each image was presented immediately and without interruption after the child's reaction. As a result, the length of EEG recording during this cognitive visual task was determined by the child's performance (i.e. response speed).

2.2 Methodology

Data preprocessing

EEG recording was performed based on 19 channels at 128Hz sampling frequency. Our obtained signal was in the range 0-64Hz as in **오류! 참조 원본을 찾을 수 없습니다.**. We process the signal using Fast Fourier Transform (FFT) filter and remove the noise at 50Hz, we obtain the clean signal as in **오류! 참조 원본을 찾을 수 없습니다.**.

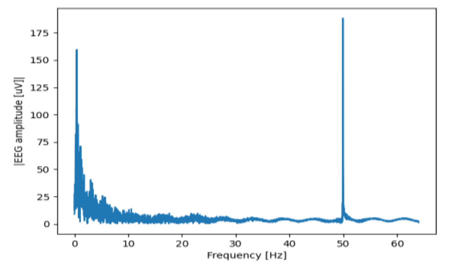

Fig. 1. Original EEG signal at Fp1

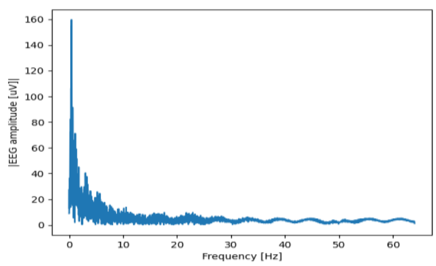

Fig. 2. Processed EEG signal at Fp1

Feature extraction

We utilized the fractal dimension (FD), which is non-linear and represents the chaotic pattern of the EEG signal. FD is a ratio giving a statistical index of complexity in terms of details in the pattern variations with the scale [25] [26]. In our paper, we calculate three FD: Higuchi, Katz and Petrosian. All these features are computed for 19 channels.

Katz Fractal Dimension is calculated as follows [25]

$$FD = \frac{\ln (N-1)}{\ln(N-1)-\ln (\frac{d}{L})} \tag{1}$$

where L is the sum of distances between consecutive points, N is the length of data sequence and d is the diameter of data sequence.

Higuchi Fractal Dimension is calculated based on a time series $x(1), x(2), \dots, x(N)$ as an input then a new time series is obtained [26]

$$Fx_m^k = \left\{x(m), x(m+k), x(m+2k), \dots, x(m+\left\lfloor\frac{N-m}{k}\right\rfloor k\right\} \tag{2}$$

for $m = 1, 2, 3, \dots, k$

where m is the first sample and $\lfloor . \rfloor$ indicates the integer part of series. Length $L_m(k)$ for x_m^k is given by

$$L_m(k) = \frac{\sum_{i=1} |x(m+ik)-x(m+(i-1)k|(N-1)}{\lfloor\frac{N-m}{k}\rfloor k} \tag{3}$$

$$d[x_m(i), x_m(j)] = max_{k=1,2,\dots,m}(|s(i+k-1)-(j+k-1)|) \tag{4}$$

$$x_m(i) = \{s(i), s(i+1), \dots, s(i+m-1)\}; 1 \leq i \leq N-m+1 \tag{5}$$

where m and r_f are positive real integers and indicate data length and filtering level, respectively. N is the number of samples and d is the distance between $x_m(i)$ and $x_m(j)$

Petrosian Fractal Dimension was introduced in [27]. In this calculation, samples of a time series are subtracted consecutively, and a new time series is produced. Then, positive and negative samples are allocated to 1 and -1. Hence, the number of sign changes in the produced time series is equal to the number of local extrema in the primary time series. The Petrosian FD is calculated as

$$D = \frac{\log_{10} n}{\log_{10} n + \log_{10}\left(\frac{n}{n+0.4N_\Delta}\right)} \tag{6}$$

where n and N_Δ are the number of samples and number of sign changes in the binary time series, respectively. In this algorithm, the N_Δ is important, while in the Katz FD calculation, the amplitude differences are important. Hence, the Petrosian method is faster and more sensitive to noise.

Feature selection

At first, using all of the extracted feature appears to be logical, however this will result in the inclusion of irrelevant or duplicate data, reducing classification accuracy. In our proposed method, we use several methods to select the appropriate features and figure out which method works best for our dataset. Following are those method that we apply to select feature in our dataset.

+) *The Filter* approach rates each feature based on a uni-variate metric and then selects the features with the highest ranking. The following are some examples of uni-variate metrics [28]:

- Variance: eliminating features that are constant or quasi-constant
- Chi-square: a categorization tool. It is a statistical test of independence used to detect if two variables are dependent on each other.
- Correlation coefficients: duplicate features are removed
- Information gain or mutual information: Examine the independent variable's role in predicting the target variable.

+) *The Correlation Feature Selection (CFS) method*, which is a simple approach that uses a correlation-based heuristic evaluation function to rank feature subsets. The feature subset evaluation function in CFS is defined as follows [29] [16]:

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \tag{7}$$

where M_s is the evaluation of a subset of S consisting of k features, $\overline{r_{cf}}$ is the average correlation value between features and class labels, and $\overline{r_{ff}}$ is the average correlation value between two features.

+) *The Lasso method* imposes a limit on the total of the absolute values of the model parameters: it must be smaller than a predetermined value (upper bound). To do so, the method uses a shrinkage (regularization) procedure in which the coefficients of the regression variables are penalized, with some of them being reduced to zero. The variables with a non-zero coefficient following the shrinking procedure are chosen to be part of the model during the feature selection procedure. The purpose of this procedure is to reduce the prediction error as much as possible [30].

+) *The Logistic method* includes a set of diagnostic tools that allow us to quantify the proposed model's goodness-of-fit and choose features accordingly. The maximum value of the log likelihood (LL) reached for each feature is used to evaluate the model's performance. D is a type of deviation that is defined as [31] [32]:

$$D = -2(\textit{LL of the current model} - \textit{LL of the saturated model}) \tag{8}$$

The saturated model has the same number of parameters as the sample size and has a probability of one. Low deviance values suggest a strong match or, in other words, a strong predictive value for the features. When comparing the two models, the deviation is useful.

+) *The Recursive Feature Elimination (RFE) method* is a feature selection algorithm with a wrapper. The method works by looking for a subset of features in the training dataset, starting with all of them and successfully deleting them until just the target number remains.

+) *Wrapper method:* To forecast the target variable, the wrapper approach looks for the optimal subset of input information. It chooses the features that give the model the best accuracy. Wrapper approaches employ past model inferences to determine if a new feature should be included or eliminated [28].

Ensemble learning for classification

Ensemble learning is a method of solving a computational intelligence problem by intentionally generating and combining many models, such as classifiers or experts. Ensemble learning is primarily used to improve a model's performance (classification, prediction, function approximation, etc).

The ensemble learning includes:

- *Boosted Trees:* The method is with the training parameters based on the Weighted Majority voting rule and the AdaBoost ensemble approach in this study. The learner type is Decision tree, with a maximum of 20 splits, 30 learners, and a 0.1 learning rate.
- *Bagged Trees:* The weight average rule employs the bag ensemble method with 30 learners and a Decision tree learner type.
- *Subspace KNN:* The training parameters in this work are based on the simple Majority Vote rule, and the proposed method uses the Subspace ensemble approach.
- *Subspace Discriminant:* The majority voting rule was utilized to create the subspace discriminant ensemble, which used the random subspace ensemble approach with 30 linear discriminant learners and two subspace dimensions.
- *RUS Boosted Trees:* It is employing Combined RUS and normal boosting technique of AdaBoost with RUSBoost ensemble approach as training parameters in this study. The decision tree is the learner type, with a maximum of 20 splits, 30 learners, and a learning rate of 0.1.

3 Experiment setup and results

After pre-processing signal, we apply the method of feature extraction for each of the 19 channels [24]. Then, feature selection algorithms are applied. As a result, we get 58 feature sets from 3 methods of calculating FD. Then we implement feature selection methods to reduce the number of features as in Table 1.

Table 1. Results of feature selection

Model Selection	Feature Set (58)
Filter Method	12 features
CFS Method	7 features
Lasso Method	38 features
Logistic Method	15 features
RFE Method	20 features
Wrapper Method	25 features

For a more detail result of feature selection, see Table 2

Table 2. Detail results of feature selection

Logistic Method	1. Fp1_Kat 2. F3_Hig 3. C3_Kat 4. C3_Hig	5. P3_Pet 6. O1_Hig 7. F7_Pet 8. F8_Pet	9. P3_Hig 10. P4_Pet 11. F8_Hig 12. T7_Pet	13. T7_Hig 14. P7_Pet 15. P8_Pet 16. P8_Hig
Lasso Method	1. F4_Hig 2. P7_Hig 3. P4_Hig 4. F7_Hig 5. Pz_Hig 6. C3_Hig 7. Fp1_Kat 8. P8_Hig 9. O1_Kat 10. Fp2_Hig	11. Cz_Kat 12. P3_Kat 13. P8_Kat 14. Fz_Kat 15. Fp2_Kat 16. C4_Kat 17. F4_Kat 18. T7_Kat 19. O2_Kat 20. F7_Kat	21. C3_Kat 22. T8_Hig 23. P7_Kat 24. P4_Kat 25. Pz_Kat 26. F8_Pet 27. T8_Kat 28. F3_Pet 29. Cz_Hig 30. O1_Hig	31. C4_Hig 32. F3_Hig 33. O2_Hig 34. Fp1_Hig 35. F8_Hig 36. Fz_Hig 37. P3_Hig 38. C4_Pet
Wapper Method	1. Fp1_Kat 2. Fp2_Pet 3. F3_Hig 4. F4_Hig 5. C3_Kat 6. C3_Hig 7. P3_Pet	8. P3_Kat 9. P3_Hig 10. P4_Pet 11. O2_Kat 12. F8_Pet 13. F8_Hig 14. T7_Pet	15. T7_Hig 16. O1_Hig 17. O2_Pet 18. T8_Pet 19. T8_Hig 20. P7_Pet 21. P8_Pet	22. P8_Hig 23. Fz_Hig 24. Cz_Kat 25. Pz_Pet
Filter Method	1. Fp1_Kat 2. F3_Kat 3. F4_Kat	4. C3_Kat 5. P3_Kat 6. O1_Kat	7. Fz_Kat 8. Cz_Kat 9. Pz_Kat	10. P7_Hig 11. P8_Kat 12. P8_Hig
RFE Method	1. Fp1_Pet 2. Fp2_Pet 3. F3_Pet 4. F3_Hig 5. F4_Pet	6. F4_Hig 7. C3_Pet 8. C3_Hig 9. C4_Pet 10. P3_Pet	11. P4_Pet 12. O1_Pet 13. O2_Pet 14. F8_Pet 15. T7_Pet	16. P7_Pet 17. P8_Pet 18. Fz_Pet 19. Cz_Pet 20. Pz_Pet
CFS Method	1. Fp1_kat 2. P3_Kat	3. P7_Hig 4. P8_Hig	5. Cz_Kat 6. Pz_Kat	7. Pz_Hig

The extracted features are input to the ensemble learning. Training and testing set are divided with ratio 80:20. We set the labels of ADHD children and Control Children by 1 and -1, respectively. The accuracy of the classification is given in Table 3. We see that with the subspace KNN and RUS boosted trees, the best results are obtained. We also present the confusion matrix and the RoC for those cases.

Table 3. The accuracy of training data

Feature Selection	Ensemble learning				
	Boosted Tress	Bagged Tress	Subspace KNN	Subspace Discriminant	RUS Boosted Trees
Filter Method	77.7	86.8	**90.9**	71.9	90.1
CFS Method	87.2	90.9	**91.7**	74.8	90.9
Lasso Method	79.8	91.3	91.3	79.8	**94.6**
Logistic Method	90.5	89.7	89.7	76.4	**92.6**
RFE Method	75.2	88	88.8	80.2	**91.3**
Wrapper Method	75.2	91.3	91.3	81.0	**94.6**

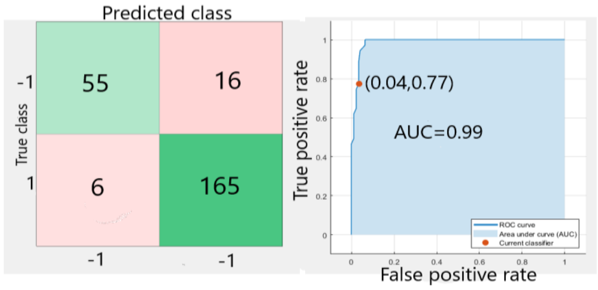

Fig. 3. The Confusion matrix and ROC of Subspace KNN (Filter Method)

Fig. 4. The Confusion matrix and ROC of Subspace KNN (CFS Method)

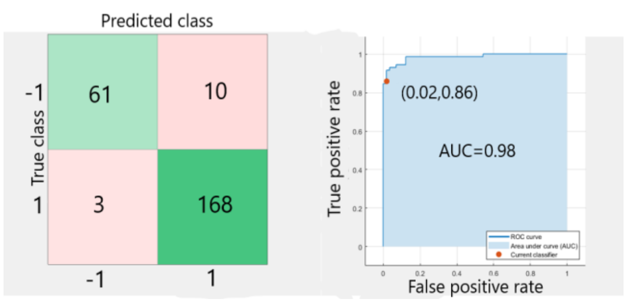

Fig. 5. The Confusion matrix and ROC of RUS Boosted Trees (Lasso Method)

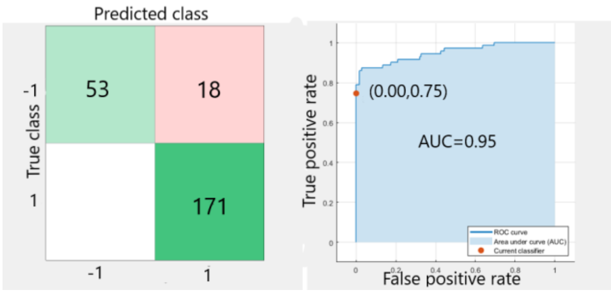

Fig. 6. The Confusion matrix and ROC of RUS Boosted Trees (Logistic Method)

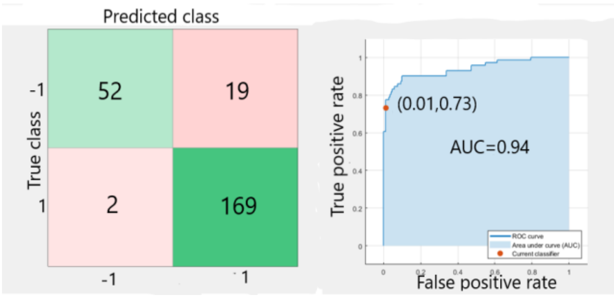

Fig. 7. The Confusion matrix and ROC of RUS Boosted Trees (RFE Method)

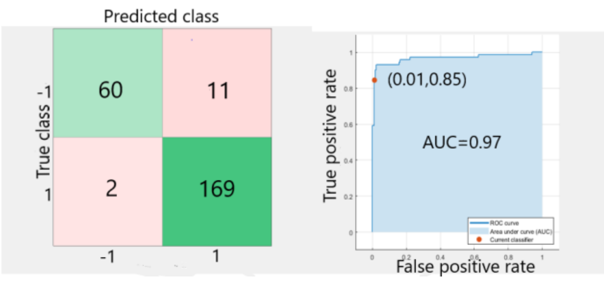

Fig. 8. The Confusion matrix and ROC of RUS Boosted Trees (Wrapper Method)

The confusion matrix results showing the true positive rates/false negative rates and the positive predictive values/false discovery rates are illustrated in Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8. In addition, the ROC curves are all normal.

The accuracy on testing data is given in Table 4. The highest accuracy 98.33% is obtained with logistic method feature selection and RUS boosted trees.

Table 4. The accuracy of training and testing data

		Train (80%)	Test (20%)
Filter Method	Subspace KNN	90.9	80.0
CFS Method	Subspace KNN	91.7	81.6
Lasso Method	RUS Boosted Trees	94.6	95
Logistic Method	RUS Boosted Trees	92.6	98.33
RFE Method	RUS Boosted Trees	91.3	88.33
Wrapper Method	RUS Boosted Trees	94.6	83.33

Table 5. Comparison of the model accuracy with some state-of-the art studies in this field

Study	Year	Dataset	Feature selection	Classifier	Accuracy
This study	2021	61 ADHD children, 60 healthy children	Katz FD, Higuchi FD, Petrosian FD	Ensemble learning	98.33%
[21]	2016	31 ADHD children, 30 healthy children	Lyapunov Exponent, Katz FD, Higuchi FD, Petrosian FD	MLP NN	93.65%
[33]	2019	50 ADHD children, 51 healthy children	Mutual information Connectivity matrix	Deep CNN	94.67%
[34]	2019	50 ADHD children, 57 healthy children	Filter Bank Common Spatial Patterns Gradient-weighted Class Activation Mapping	Deep CNN	90.29%
[35]	2019	47 ADHD children, 50 healthy children	Phase space reconstruction of EEG, CFS and PSO feature selection	SVM, NN k-NN, and naive-Bayes classifier	93.3%

Table 5 show how our study outperforms the state-of-the art studies in accuracy for the same purpose.

4 Conclusion

In general, ADHD is a disorder that is common in children and it affects to children's reaction to the environment. Hence, early diagnosis of these symtoms is very important in the child's development. In our paper, we use the non-linear features of EEG signals to differentiate between ADHD children and healthy children. Our dataset is published in 2020 in ieee-dataport.org. So far, most studies have used linear features (spectral, time, spatial or time-frequency features) to categorized ADHD patients. Although some of these studies have provided promising results, new advanced methods are still in need to analyze EEG signals. Non-linear features of EEG signal in children's brain has only reported in [21] with the dataset of 31 ADHD children and 30 healthy children. They used the same set of non-linear features but different feature selection methods by using the given tools in Matlab. In our study, instead of using tools in Matlab, we used some modified feature selection method, which focuses more on the physics and the structure of the EEG signals. For classifier, we use ensemble learning, which is more simple method than neural network [21]. We get better results of 98.33% accuracy with a larger and more updated dataset of 61 ADHD children and 60 healthy control. Our results show that the non-linear features are appropriate features to analyze and characterize the EEG signals. The application of non-linear analysis to EEG has opened a new door in analyzing EEG signals in order to discriminate ADHD patients from the healthy group.

References

1. A. P. Association: American Psychiatric Association. DSM-5 Task Force. Diagnostic and statistical manual of mental disorders: DSM-5. In 5th American Psychiatric Association, Washington DC (2013).
2. A. Meysamie, M. D. Fard and M.-R. Mohammadi: Prevalence of attention-deficit/hyperactivity disorder symtoms in preschoolaged Iranian children. Iran J Pediatr. 21(4), (2011).
3. J. A. King, M. Colla, M. Brass, I. Heuser and D. v. Cramon: Inefficient cognitive control in adult ADHD: evidence from trial-by-trial Stroop test and cued task switching performance. Behav Brain Funct, 3(42), (2007).
4. F. Aboitiz, T. Ossandón, F. Zamorano, B. Palma and X. Carrasco: Irrelevant stimulus processing in ADHD: catecholamine dynamics and attentional networks. Front Psychol. (2014).
5. M. R. Mohammadi, N. Malmir, A. Khaleghi and M. Aminiorani: Comparison of Sensorimotor Rhythm (SMR) and Beta Training on Selective Attention and Symptoms in Children with Attention Deficit/Hyperactivity Disorder (ADHD): A Trend Report. Iran J Psychiatry, 10(3), (2015).
6. O. WH: The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. WHO. Geneva (1992).
7. J. F. Lubar: Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. Biofeedback Self Regul. 16(3), (1991).
8. A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller and G. Candrian: Machine learning approach for classification of ADHD adults. Int J Psychophysiol. 93(1), 162-6, (2014).
9. S.-S. Poil, S. Bollmann, C. Ghisleni and R. L. O'Gorman: Age dependent electroencephalographic changes in Attention Deficit/Hyperactivity Disorder (ADHD). Clinical Neurophysiology, 125(8), 1626-1638, (2014).
10. A. Mazaheri, C. Fassbender, S. Coffey-Corina, T. A. Hartanto, J. B. Schweitzer and G. R. Mangun: Differential oscillatory electroencephalogram between attention-deficit/hyperactivity disorder subtypes and typically developing adolescents. Biol Psychiatry. 76(5), 422-429, (2014).

11. J. F. Lubar: Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. Applied Psychophysiology and Biofeedback 16(3), 201-225, (1991).
12. L. C. Fonseca, G. M. A. S. Tedrus, C. d. Moraes, A. d. V. Machado, M. P. d. Almeida and D. O. F. d. Oliveira :Epileptiform abnormalities and quantitative EEG in children with attention-deficit/hyperactivity disorder. Arq Neuropsiquiatr 66(3A), 462-7, (2008).
13. A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller and G. Candrian: Machine learning approach for classification of ADHD adults. Int J Psychophysiol. 93(1) 162-166, (2014).
14. InezBuyck and J. R.Wiersema: Resting electroencephalogram in attention deficit hyperactivity disorder: Developmental course and diagnostic value Author links open overlay panel. Psychiatry Research 216(3), 391-397, (2014).
15. R. M. N. H. &. D. P. W. M Duda: Use of machine learning for behavioral distinction of autism and ADHD. Translational Psychiatry, vol. 6, (2016).
16. S. Kaur et. Al: Phase Space Reconstruction of EEG Signals for Classification of ADHD and Control Adults. Clinical EEG and Neuroscience, (2020).
17. A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin and M. Elsharnouby: Detecting ADHD patients by an EEG-based serious game. IEEE Transactions on Instrumentation and Measurement, (2018).
18. J. R. Wessel: Testing Multiple Psychological Processes for Common Neural Mechanisms Using EEG and Independent Component Analysis. Brain Topography, vol. 31, 90-100, (2016).
19. K. Katoab, K. Takahashia, N. Mizuguchiac and J. Ushiba: Online detection of amplitude modulation of motor-related EEG desynchronization using a lock-in amplifier: Comparison with a fast Fourier transform, a continuous wavelet transform, and an autoregressive algorithm. Journal of Neuroscience Methods , vol. 293, 289-298, (2018).
20. K. M. A. Allahverdy Armin, M. R. Mohammadi and M. N. Ali: Detecting ADHD Children using the Attention Continuity as Nonlinear Feature of EEG. Frontiers Biomed Technol, 3(1-2), 28-33, (2016).
21. A. K. Mohammad Reza Mohammadi, A. M. Nasrabadi, S. Rafieivand, M. Begol and H. Zarafshan: EEG classification of ADHD and normal children using non-linear features and neural network. Biomedical Engineering Letters, vol. 6, 66-73, (2106).
22. A. Vahid, A. Bluschke, V. Roessner and S. S. a. C. Beste: Deep Learning Based on Event-Related EEG Differentiates Children with ADHD from Healthy Controls. Journal of Clinical Medicine, 8(7), (2019).
23. Z. Li and C. Weike: Application of Deep Convolutional Neural Networks in Attention-Deficit/Hyperactivity Disorder Classification: Data Augmentation and Convolutional Neural Network Transfer Learning. Journal of Medical Imaging and Health Informatics, 9(8), 1717-1724, (2019).
24. M. R. Mohammadi, A. M. N. Ali Khaleghi, S. Rafieivand, M. Begol and H. Zarafshan: EEG classification of ADHD and normal children using non-linear features and neural network. Biomedical Engineering Letters , 66-73, (2016).
25. T. Higuchi: Approach to an irregular time series on the basis of the fractal theory. Physica D, Nonlinear Phenomena, (1988).
26. A. Petrosian: Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns. In Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems, Lubbock, TX, USA, (1995).
27. R. L. M. Petre Stoica: Introduction to Spectral Analysis, the University of Michigan: Prentice Hall (1997).

28. [Online]. Available: https://towardsdatascience.com/feature-selection-identifying-the-best-input-features-2ba9c95b5cab.
29. A.-B. A. T.-S. M. Sánchez-Maroño N: Filter Methods for Feature Selection – A Comparative Study," Intelligent Data Engineering and Automated Learning - IDEAL 2007. IDEAL 2007. Lecture Notes in Computer Science, vol. 4881, Springer, Berlin, Heidelberg. (2007).
30. [Online]. Available: https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf.
31. O. S. Qasima and Z. Y. Algamalb: Feature selection using particle swarm optimization-based logistic regression model. Chemometrics and Intelligent Laboratory Systems, 182(15), 41-46, 2018.
32. Q. Cheng, P. Varshney and M. Arora: Logistic Regression for Feature Selection and Soft Classification of Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters, 3(4), 491-494, (2006).
33. C. He, S. Yan and L. Xiaoli: A deep learning framework for identifying children with ADHD using an EEG-based brain network. Neurocomputing, 356(3), 83-96, (2019).
34. H. Chen, Y. Song and X. Li: Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. Journal of Neural Engineering, 16(6), (2019).
35. S. S. Simranjit Kaur1, P. Arun, D. Kaur and M. Bajaj: Phase Space Reconstruction of EEG Signals for Classification of ADHD and Control Adults. Clinical EEG and Neuroscience, (2019).