

Recognition of Attention Deficit/Hyperactivity Disorder (ADHD) Based on Electroencephalographic Signals Using Convolutional Neural Networks (CNNs)

Sara Motamed^{1*}, Elham Askari¹

¹.Department of Computer Engineering , Fouman & Shaft Branch, Islamic Azad University, Fouman, Iran

Received: 25 Jun 2021/ Revised: 06 May 2022/ Accepted: 06 Jun 2022

Abstract

Impulsive / hyperactive disorder is a neuro-developmental disorder that usually occurs in childhood, and in most cases parents find that the child is more active than usual and have problems such as lack of attention and concentration control. Because this problem might interfere with your own learning, work, and communication with others, it could be controlled by early diagnosis and treatment. Because the automatic recognition and classification of electroencephalography (EEG) signals is challenging due to the large variation in time features and signal frequency, the present study attempts to provide an efficient method for diagnosing hyperactive patients. The proposed method is that first, the recorded brain signals of hyperactive subjects are read from the input and in order to the signals to be converted from time range to frequency range, Fast Fourier Transform (FFT) is used. Also, to select an effective feature to check hyperactive subjects from healthy ones, the peak frequency (PF) is applied. Then, to select the features, principal component analysis and without principal component analysis will be used. In the final step, convolutional neural networks (CNNs) will be utilized to calculate the recognition rate of individuals with hyperactivity. For model efficiency, this model is compared to the models of K- nearest neighbors (KNN), and multilayer perceptron (MLP). The results show that the best method is to use feature selection by principal component analysis and classification of CNNs and the recognition rate of individuals with ADHD from healthy ones is equal to 91%.

Keywords: Hyperactivity; Electroencephalographic Signals; Convolutional Neural Networks (CNN); Principal Component Analysis (PCA).

1- Introduction

Diagnosis of hyperactivity based on history and experiment remains essentially clinical and can be supported by neuropsychological assessments. But due to heterogeneous cognitive profiles in patients with hyperactivity, it is not clearly diagnosed. In general, there are various conditions that often complicate the diagnosis due to the irregularity, impulsivity, and range of natural cognitive profiles with variable strengths and weaknesses that are widespread in these areas. Hence, a biomarker will be of great value in reducing the intrinsic uncertainty of clinical diagnosis. Electroencephalography (EEG) signals contain rich information related to functional dynamics in the brain. The use of EEG in hyperactive subjects was begun more than 75 years ago with Jasper et al. (1938),

that reported the increasing of the power of EEG with low-frequencies in Front-central regions [1]. Studies on EEG abnormalities in hyperactive patients were first performed by Lubber in 1973. He concluded that theta activity increased in the brains of hyperactive individuals, and beta power is significantly reduced in these patients [2]. In other studies, some factors for hyperactivity diagnosis through electroencephalography signals were introduced to learn abnormalities [3]. Since then, human electrophysiological studies have been presented using EEG spectral analysis and Event-Related Potentials (ERPs) of functional performance in the hyperactive patients [4]. In contrast to EEG signals, ERPs reflect changes in the electrical activity of the brain that are blocked by the occurrence of a particular event, i.e., a response to a discrete external stimulus or an internal mental process [5]. ERPs also provide high-resolution

✉ Sara Motamed
Sara.Motamed@iau.ac.ir

non-invasive neurophysiological measurements. This allows the inefficient dynamics of the brain to be assessed and cognitive processes that may not be apparent at the behavioral level to be identified [6].

Artificial neural networks have recently been introduced as an encouraging application of artificial intelligence that is very effective in recognizing brain models. Machine learning, a subset of artificial intelligence and deep learning, a specialized sub-discipline of machine learning, have been increasingly used in clinical research with promising results. Machine learning can be described as the practice of using algorithms to train a system using a large amounts of data, with the goal of giving it the ability to learn how to do a particular task and then classify or accurately predict. Deep learning is a subset of machine learning algorithms that introduce tasks in smaller units that often provide higher levels of accuracy [7]. Neural networks are characterized by their network architecture which is defined by the anatomical arrangement of its connected processing units, i.e., artificial neurons with a loss or optimization function that determines the overall purpose of the learning process. Connections are trained or teach how to perform the desired task and by using of a training algorithm, change the parameters of the neural network experimentally. This is done in such a way the target function is eventually optimized based on the inputs received by the neural network. There are different types of neural networks with different designs and architectures from different principles and for various purposes [8, 9].

In this paper, convolutional neural network (CNN) method has been used to find the most efficient electrode to diagnose patients. The hierarchy of our proposed model is such that after reading the signals from the input, they are pre-processed by the filtering method and then a FFT and PF are applied to all normalized signals. The output of this step enters the next step, i.e., feature selection. In this step, Principal Component Analysis (PCA) is used. Finally, the CNN classification method with 8 convolutional layers and 2 fully connected layers will be applied to learn the obtained features and the results will be discussed. The structure of the current study is as follows: In section 2, the database used, is introduced. In Section 3, the methods used in this study are briefly described. In section 4, the main structure of the proposed model is introduced and sections 5 and 6, respectively, introduce the experiments are done and express the results. The conclusion of the present study is presented in Section 7.

2- Database

The present study uses the standard database introduced in [10], which includes 57 females and 39 males. The data used for this study has been processed by Alpha-Neuro Center that is a neuropsychology research laboratory. Sampling rate

is 2000 Hz / filter channel is below than 250 Hz. The received signals from 19 channels were recorded at rest for 5 minutes and subjects were instructed to look at a certain point on the wall and move as little as possible and also to prevent movement and or blinking. NeuroGuide / WinEEG software has also been used to remove artifacts [10].

In the section on the separation of effective electrodes in this dataset, five groups of electrodes named Frontal, Central, Temporal, Parietal, Occipital are introduced, and the spectrum of brain waves in this dataset named delta (4-0), theta (4-8), alpha (8-13) and beta (13-32) Hz (11) are divided into Hz(11).

3- Method

The dataset used in the current study includes pre-processed EEG signals on which filtering operations have been performed. On all signals the gap filter of 55-65 Hz, the low cut filter of 0.3 Hz, and the high cut filter of 30 Hz have been applied (the reason for such a low filter is that this data is used to create the neuro feedback protocol and does not use gamma wave bands for neuro feedback). Therefore, after reading the recorded brain signals of hyperactive individuals and in order to convert the signals from time range to frequency range, FFT is used and to select the effective feature for examining hyperactive subjects from those of healthy, PF are utilized. Important features of EEG signals are then extracted and PCA is applied to all features. Finally, the outputs obtained from the previous stage are sent to the classification to first determine the most effective electrode and second to determine the recognition rate of hyperactive subjects from those are healthy (Fig. 1).

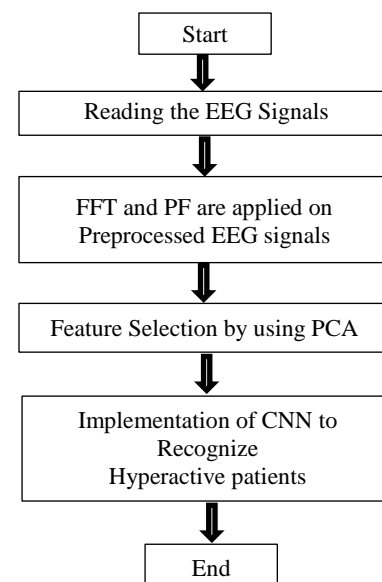


Fig. 1 General overview of the proposed method

3-1- Fast Fourier Transform (FFT)

FFT is one of the most important algorithms used in signal processing and data analysis. Fourier analysis can transform a signal from the main domain, which is usually time or space, into a frequency domain display and vice versa.

It is assumed that the discrete version is represented at the time of the audio signal by length N and the sampling rate f_s with $x[j]$. The frequency content of the $x[j]$ signal over a given period of time can be expressed using discrete Fourier transforms (FT) over time as a function of frequency and by using of the FT coefficients $X[k]$. The parameters transform between the time domain and the frequency begins using the Perceval's theorem; Perceval's theorem states that the sum of the squares of a function with the sum of its transformed squared is equal to the Eq. (1):

$$\sum_{j=1}^{N-1} |x[j]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]X^*[k]| = \sum_{k=0}^{N-1} p[k] \quad (1)$$

Where $P[k]$ is the power spectrum without phase and k is the frequency index. Usually the content of the frequency resulting from the FT is symmetric with respect to the zero frequency, so when using the power spectrum, the whole or only a part of it can be considered [12].

Since FFT converts a signal from a time or space range to a frequency range, it facilitates the analysis of a given signal, which is why this method is used in the present study.

3-2- Peak Frequency (PF)

The PF is defined as the maximum amount of power in the EEG frequency spectrum between the range of 7.5 and 12.5 Hz. According to researches accomplished on PF, several important interpersonal and intrapersonal differences have been identified. Interpersonal differences are attributed to genetic factors. Low values of this feature indicate brain damage such as chronic fatigue syndrome (CFS), Alzheimer's disease, hyperactivity, etc. [10]. Based on the researches, it can be noted that the amount of PF is different in subjects with ADHD and healthy subjects [17]. The PF also varies with age and gender. In healthy adults, for example, the PF is hidden between 9.5 to 11.5 Hz. At PF, the PF location within the alpha band increases with age in childhood, culminates in early adulthood, and then decreases in older adulthood [17].

3-3- Feature Selection

The performance of a classifier depends on the relationship between the number of samples, the number of attributes, and the complexity of the classifier. Therefore, by having appropriate features, classifier performance and recognition rate could be increased. On the other hand, it is observed practically that if the number of training samples compared to the number of features is

relatively small, additional features could reduce the classifier performance. Therefore, in the present study, principal component analysis is used.

3-3-1- Principal Component Analysis (PCA)

In PCA, the principal data space is described based on the special vectors of the covariance matrix, and the specific values corresponding to the special vectors express the attributes' energy in line with these vectors. When the correlation between the variables of the problem is linear, linear PCA will be the first choice. However, in situations where the problem has a nonlinear correlation, taking benefit of nonlinear versions could improve the function [13].

Technically, PCA removes the least important variables, while the most valuable parts of all variables are remained. That is why, this method is used in the present study to reduce the complexity of the calculations and keep the best features.

3-4- Convolutional Neural Networks (CNN)

The CNN is a kind of artificial neural network that is inspired by the function of the human and animals' visual cortex of the brain and is applied for functions such as image and video recognition, speech recognition, recommendation systems, natural language processing and other cases. The basic assumption of CNN's architecture is that operations are performed on input data to preserve spatial and neighborhood information in the data, and ultimately a vector of encoded attributes is obtained. In general, a CNN network consists of three main layers: the convolutional layer, the pooling layer, and the fully connected layer. Different layers perform different tasks. There are two phases of training in each CNN. Progressive phase and Back propagation phase [7]. During the training process, the common weights in the convolutional layers as well as the weights among the fully connected convolutional layers significantly reduce the number of free trainable network parameters and thus increase generalizability. The CNN used in this study is generally consists of the following layers:

Convolutional layer: This layer is the main core of the CNN. The convolutional layer parameters include a set of learnable filters. In these layers, the CNN uses various filters to convolute the input data as well as the mapping of intermediate features, and such mapping of different features has several main advantages. First, the weight-shared mechanism in each feature mapping drastically reduces the number of parameters, and the local connection learns the relationship among neighboring pixels. It also causes the invariability and stability of the object's displacement, and the ratio of the freedom degree of the system and the number of samples required for learning is remarkably increased, which makes the generalizability of the system stronger. As mentioned

above, this layer performs convolution on the input EEG signal using the kernel.

ReLU layer: This layer introduces a nonlinear method to the network, which is the most common activator function (Fig. 2):

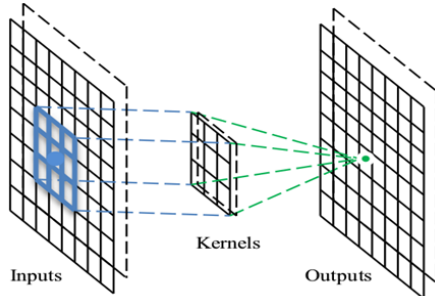


Fig. 2. Convolutional operation

Fig. 2 shows the ReLU layer, in which it introduces a nonlinear method to the network that is the most common activator function.

Pooling Layers: A pooling layer is usually placed after a convolutional layer and can be used to reduce the mapping size of network attributes and parameters. Like convolutional layers, pooling layers are remained unchanged toward displacement considering of the neighboring pixels in their own calculations. Pooling layer implementation using the maximum and average functions are the most common implementations (Fig. 3).

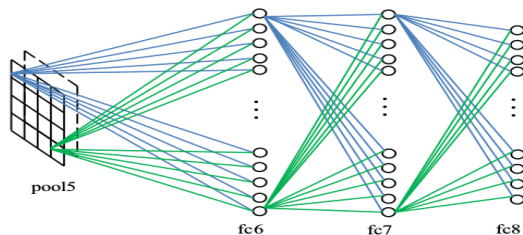


Fig. 3. Pooling operation on feature mapping.

Fully connected layer: After the last pooling layer, as shown in Fig. 4, there are fully connected layers. Fully connected layers perform like their counterparts in traditional artificial neural networks. The fully connected layer allows the network result to be displayed in the form of a specific size vector. This vector can be used for following further processing.

4- The Proposed Method

The purpose of the current study is to provide an efficient method based on reducing the dimensions of selected features and appropriate classification in order to achieve

the best recognition rate for diagnosing subjects with ADHD at the right time. The proposed method operates in such a way that first the recorded brain signals of hyperactive individuals are read from the input, then, FFT is applied for the signals to be transformed from time range to frequency range. Also, to select an effective feature to check hyperactive individuals from those of healthy, PF is used. Therefore, the reason for selecting FFT is the easier analysis of pre-processed signals, and the reason for choosing the PF is that the amount of this feature varies in individuals with ADHD and healthy subjects, and depends on gender and age factors. So, it can be claimed that these two features could help increase the diagnosis rate of ADHD patients. PCA and No-PCA will then be applied to all features in the feature selection section. The reason for giving importance to this section is that the feature selection and extraction stage are very important. Because the more correctly the features are selected, the better the results will be in the classification stage. Finally, CNN classification is applied to diagnose and evaluate subjects with ADHD from healthy ones [16]. The reason for selecting CNN classification is that this classification is able to store data throughout the network and the ability to work with incomplete knowledge, as well as the ability of high error tolerance. Therefore, the classification is expected to show desirable results. The architecture of the deep neural network is that first a convolutional layer with a nonlinear ReLU function along with Dropout and BN, and then a Max-pooling layer is added. Over several times of repeating, a two-dimensional matrix will be obtained and will produce a total of 78,432 parameters. In the architecture, the first layer of large size filter (128×1) and in the next layers, smaller size filters (16×1) are used, and finally, the feature vectors selected with two fully connected layers with nonlinear function ReLU and Softmax are used to automatically recognition of different stages of ADHD. Also, in the network training section, and to determine the network meta-parameters, the Trial-error method and the Cross-entropy function and the Adam optimizer with a learning rate of 0.002 have been used. The total number of epochs applied in the proposed model is 150 and the 10 Fold method is used for Cross Validation of data.

5- Performance Analysis

This section presents an efficient method based on PCA, No-PCA and CNN to optimally identify hyperactive subjects using electroencephalography signals and also to determine the best and most effective electrode for better diagnosis of the disorder. Also, to display the performance of the proposed model, it will be compared with Multi-Layer Perceptron (MLP) and K-Nearest Neighbors (KNN). MLP classification with back-propagation learning

algorithm consists of three layers: input, hidden and output. In input layers, the number of neurons is equal to the length of the input vector or the number of features. The most important parameters are the number of hidden layers, the number of neurons in each layer, the amount of learning and the learning time in data training and testing. Here, for data classification, a hidden layer with five neurons is considered [18].

In KNN classification, the number of neighborhood is considered 2, and Euclidean distance is used to calculate the distance between neighbors [10]. The findings in the following section are resulted from the features obtained from PCA and No-PCA and CNN, MLP and KNN classifications on central, temporal, occipital, frontal and parietal electrodes and include Accuracy, Recall, and F1-score of the classification [14, 15].

6- Discussion of the Results

The experiments performed in this study are divided into several sections. The first section is related to the investigation of the most functional and effective electrode in the diagnosis of hyperactive subjects, which will be calculated using PCA and NO-PCA feature selection. The second section deals with the recognition rate obtained of the most efficient electrode introduced by CNN classification. The third section of the tests is allocated to diagnosing the total rate of hyperactive patients who have been normalized in the pre-processing stage, and the extent of their disorder by applying the standard deviation threshold obtained from the clinic that the amount of which is presented in the dataset. Then, the patients are divided into three groups including those with low hyperactivity, moderate hyperactivity and hyperactivity. In the last part, the experiments are allocated to comparing the proposed model with the competing models of KNN and MLP in the current study.

6-1- Results Obtained from PCA and No-PCA Features and CNN Classification

In the proposed model, first all normalized signals are read from the input and will be divided into three groups of patients with low hyperactivity, moderate hyperactivity and hyperactivity, and FFT and PF are applied on all read signals. In the feature selection stage, on all the selected signals, once PCA and once No-PCA are applied, and at the end, the output of the feature selection stage enters the CNN classification. The results of the experiments performed based on the 10-fold evaluation criterion are reported in Tables (1) and (2), respectively.

Table 1: Results of PCA and CNN classification on different electrodes

<i>Electrode's name</i>	<i>Accuracy</i>	<i>Recall</i>	<i>F1 score</i>
Central	54%	0.08	0.66
Temporal	43%	0.07	0.66
Occipital	33%	0.05	0.45
Frontal	45%	0.08	0.61
Parietal	66%	0.04	0.73

Table 2: Results of No-PCA and CNN classification on different electrodes

<i>Electrode's name</i>	<i>Accuracy</i>	<i>Recall</i>	<i>F1 score</i>
Central	60%	0.08	0.69
Temporal	54%	0.09	0.66
Occipital	60%	0.08	0.69
Frontal	44%	0.10	0.56
Parietal	63%	0.07	0.54

Tables (1) and (2) shows that the most effective electrode is related to the parietal which has a recognition accuracy of PCA and CNN classification of 66% and a recognition accuracy by No-PCA and CNN classification of 63%. This means that the use of PCA is effective in increasing the accuracy of the disorder diagnosis.

6-2- Experiment Results on the Proposed Model and Competing Models of KNN and MLP

In this part of the experiments, the purpose is to evaluate the performance of the proposed model. That is why, the introduced method will be compared with competing models of KNN and MLP. The results based on the 10-fold evaluation criterion are shown in Table (3).

Table 3: Review of the three classifiers of CNN, KNN, and MLP

<i>Classifier</i>	<i>FFT +PCA</i>	<i>FFT +No-PCA</i>
CNN applied on the whole subjects in the dataset	accuracy: 90%	accuracy: 85%
KNN applied on the whole subjects in the dataset	accuracy: 84%	accuracy: 83%
MLP applied on the whole subjects in the dataset	accuracy: 61%	accuracy: 71%
CNN applied On parietal electrode	accuracy: 55%	accuracy: 51%
KNN applied On parietal electrode	accuracy: 61.5%	accuracy: 44%
MLP applied On parietal electrode	accuracy: 42%	accuracy: 35.5%

Table (3) shows the results of the three classifiers of CNN, KNN and MLP on all electrodes as well as the most effective electrode (parietal electrode). As it was explained, the input of the proposed model is the signals of healthy and unhealthy subjects that first the pre-processed

data is read and then the FFT and PF will be applied to the signals. Next, the PCA feature selection is applied to the extracted features and the best features will be selected. Eventually, all features will enter the classification stage. Also to ensure the correct operation of the proposed model, the experiments with No-PCA are performed once again. The results reveal that the best method is to use PCA and CNN classification on the subjects in the database with a recognition rate of 91%. While the competing models, KNN, and MLP show recognition rate of 88% and 66%, respectively.

As shown in Fig. 5, for the accuracy of the PF performance in the feature extraction section, all of these experiments have been performed once by eliminating this feature. Investigating the results obtained from Fig. 5, it is observed that the use of FFT methods and PF facilitates the analysis of the results, and PCA causes raising the classification rate presented in this study by selecting the appropriate features.

7- Conclusion

The present study investigated the recognition of brain signals in hyperactive patients and the goal was to find the most effective model with the highest recognition rate to diagnose hyperactive subjects. Also, during the experiments, the most effective electrode with a high recognition rate in diagnosing hyperactive subjects has been identified. The proposed model hierarchy works in such a way that after reading the pre-processed signals from the data set introduced in the text, the FFT and PF are applied. In order to select the appropriate features, once PCA (to reduce the complexity of the calculations and

select the best features), and once again No-PCA (to check the performance of PCA) are performed.

The output of this section entered the three classifiers of CNN, KNN and MLP and the recognition rates in all three categories were examined. The results revealed that the most effective parietal electrode with a recognition rate of 66%. Therefore, it could be proven that parietal lobe neurons play an important role in the etiology of this disorder. Also, the best method was to use the PCA feature and CNN classification applied to the subjects in the database and the recognition rate was equal to 91%.

To justify the proposed model and the weakness of the competing models, it could be concluded that in the KNN classification, since the algorithm was very sensitive to the amount of value and was suitable for multivariate environments with small space, in this experiment it did not provide a high rate diagnosis. Also in the MLP network and due to the low rate of this model, we can point to problems such as failure to learn or retain information. This happens when the network parameters do not converge to a certain value after a long time or sometimes reach the state of data retention due to over-training. One of the biggest advantages of CNN and its good performance is that it does not change much against small input errors and applies the weight sharing principles, which drastically reduce the number of free parameters. Therefore, they increase generalizability.

In the conclusion part and according to the obtained results, it could be proved that CNN is suitable for implementing large and complicated issues, and the structure presented in the proposed model reduces training time, the number of trainable parameters, and increases the classification accuracy. Also, according to high accuracy of the algorithm, it could be used to automatically diagnosis of ADHD patients on EEG signals.

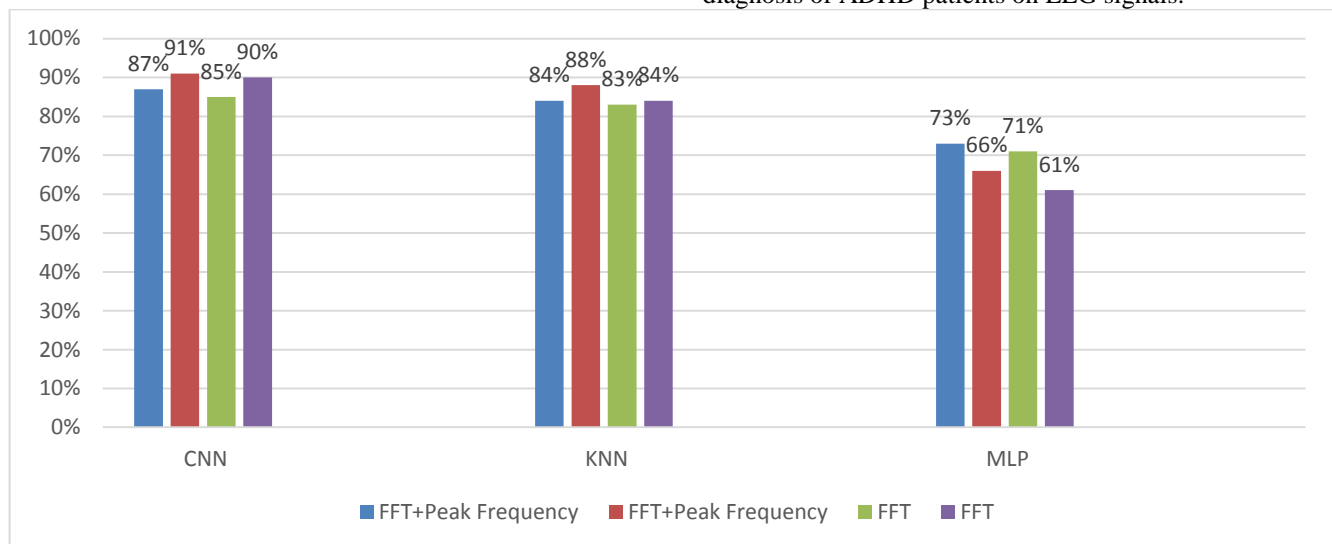


Fig. 5 Review of Fast Fourier Transform and peak frequency performance on the classification rates

References

- [1] A. Shirazi, J. Alaghandrad, "Treatment of Attention Deficit Hyperactivity Disorder (ADHD) With a Cognitive Behavioral Approach", *Journal of Advances in Cognitive Science*. Vol. 2, 1379, pp. 29- 34.
- [2] C. H. Chen, J. D. Lee, M. C. Lin, "Classification of Underwater Signals Using Neural Networks", *Journal of Science and Engineering*. Vol. 3, 2000, pp. 31-48.
- [3] B. P. Howell, S. Wood, S. Koksai, "Passive Sonar Recognition and Analysis Using Hybrid Neural Networks", *IEEE Proceedings Oceans*. Vol. 4, 2003, pp. 1917- 1924.
- [4] H. H. Jasper, P. Solomon, C. Bradley, "Electroencephalographic Analyses of Behavior Problem Children", *Am. J. Psychiatry*. Vol. 95, 1938, pp. 641–658.
- [5] A. Lenartowicz, S. K. Loo, "Use of EEG to Diagnose ADHD", *Curr. Psychiatry Rep*. Vol. 16, 2014.
- [6] M. Fabiani, G. Gratton, K. D. Federmeier, "Event-Related Brain Potentials: Methods, Theory, and Applications, in *Handbook of Psychophysiology*", 3rd Edn (New York, NY: Cambridge University Press), 2007, pp. 85–119.
- [7] B. Kim, J. Roh, S. Dong, S. Lee, "Hierarchical Committee of Deep Convolutional Neural Networks for Robust Facial Expression Recognition", *Journal on Multimodal User Interfaces*. 2016, pp. 1–17.
- [8] G. F. Woodman, "A Brief Introduction to the Use of Event-Related Potentials in Studies of Perception and Attention", *Psychophysics*, 2010, Vol. 72, pp. 2031–2046.
- [9] E. Kroupi, S. Frisch, A. Castellano, M. David, I. S. Montplaisir, J. Gagnon, "Deep Networks Using Auto-Encoders for PD prodromal analysis", In *Proceedings of the HBP Student Conference on Transdisciplinary Research Linking Neuroscience, Brain Medicine and Computer Science*, Vienna, 2017.
- [10] T. H. Grandy, M. Werkle-Bergner, C. Chicherio, F. Schmiedek, M. Lövdén, U. Lindenberger. "Peak Individual Alpha Frequency Qualifies as a Stable Neurophysiological Trait Marker in Healthy Younger and Older Adults", *Psychophysiology*, 2013, pp. 570–582.
- [11] <https://school.brainhackmtl.org/project/adhdsubtype-project>.
- [12] S. Haykin, "Neural Networks, a Comprehensive Foundation", Prentice Hall, Vol. 4n No. 2, 1999, pp. 191-192.
- [13] H. Abdi, L. J. Williams, "Principal Component Analysis", *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 2, 2010, pp. 433-459.
- [14] K. Boyd, K. H. Eng, "Area Under the Precision-recall Curve: Point Estimates and Confidence Intervals", Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 451–466.
- [15] J. Davis, M. Goadrich, "The Relationship between Precision-Recall and ROC Curves", *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 233–240.
- [16] F. Li, X. Li, F. Wang, D. Zhang, Y. Li, F. He, "A Novel P300 Classification of Algorithm based on a Principal Component Analysis-Convolutional Neural Network", *Applied Science*, 2019, pp. 2-15.
- [17] M. Robertson, S. Furlong, B. Voytek, C. Boettiger, M. Sheridan, "EEG Power Spectral Slope Differs by ADHD Status and Stimulant Medication Exposure in Early Childhood", *Journal of Neurophysiology*, 2019, pp. 2427-2437.
- [18] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, "EEG Classification of ADHD and Normal Children using Non-Linear Features and Neural Network", *The Korean Society of Medical & Biological Engineering and Springer*, Vol. 6, 2016, pp. 66-73.