



Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes

Amirmasoud Ahmadi ^{a,*¹}, Mehrdad Kashefi ^{a,1}, Hassan Shahrokhi ^b, Mohammad Ali Nazari ^{c,d,*}

^a Biomedical Engineering Department, School of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

^b Research Center of Psychiatry and Behavioral Sciences, Tabriz University of Medical Sciences, Tabriz, Iran

^c Department of Neuroscience, Iran University of Medical Sciences, Tehran, Iran

^d Division of Cognitive Neuroscience, University of Tabriz, Iran

ARTICLE INFO

Keywords:

Attention deficit hyperactivity disorder (ADHD)
Computer-aided diagnosis (CAD)
EEG
Deep learning
Convolutional neural network (CNN)

ABSTRACT

Background: Attention deficit hyperactivity disorder (ADHD) is a ubiquitous neurodevelopmental disorder affecting many children. Therefore, automated diagnosis of ADHD can be of tremendous value. Unfortunately, unlike many other applications, the use of deep learning algorithms for automatic detection of ADHD is still limited.

Method: In this paper, we proposed a novel computer aided diagnosis system based on deep learning approach to classify the EEG signal of Healthy children (Control) from ADHD children with two subtypes of Combined ADHD (ADHD-C) and Inattentive ADHD (ADHD-I). Inspired by the classical approaches, we proposed a deep convolutional neural network that is capable of extracting both spatial and frequency band features from the raw electroencephalograph (EEG) signal and then performing the classification.

Result: We achieved the highest classification accuracy with the combination of β_1 , β_2 , and γ bands. Accuracy Recall, Precision, and Kappa values were %99.46, %99.45, %99.48, and 0.99, respectively. After investigating the spatial channels, we observed that electrodes in the Posterior side had the most contribution.

Conclusions: To the best of our knowledge, all previous multiclass studies were based on fMRI and MRI imaging. Therefore, the presented research is novel in terms of using a deep neural network architecture and EEG signal for multiclass classification of ADHD and healthy children with high accuracy.

1. Introduction

Attention deficit hyperactivity disorder (ADHD) is one of the most pervasive disorders affecting approximately 5% of children and 2–4% of adults [1–3]. ADHD can negatively impact academic performance and social functions of the patient because it makes it difficult for them to focus their attention and control their behavior [4,5]. Right now, the clinical diagnosis of ADHD is challenging and usually with high misdiagnose rate [6]. That is why developing, and automatic diagnosing method with high accuracy is crucial. Diagnostic and Statistical Manual of Mental Disorders (DSM-5) classifies three types of ADHD based on the main symptoms: I) predominantly inattentive presentation, II) predominantly hyperactive-impulsive presentation, and III) Combined presentation [7]. Although many studies, like [8–10], carefully investigated these symptoms, few of them [11,12] tried to discriminated

features of different subtypes using neuroimaging methods.

There are different clinical procedures to address patients suffering from different sub-types of ADHD. Compared to two class classification, three class diagnosis is more complex, and therefore, more prone to diagnosis error, and consequently, incorrect diagnosis of ADHD subtypes can significantly decrease the effectiveness of the treatment process. Hence, having a high accuracy diagnosis system can meaningfully improve the treatment process. Social and academic difficulties are well-known clinical symptoms for both ADHD-C and ADHD-I types. However, these symptoms manifest variably in each of the sub-types. For instance, comorbid external features (like conduct disorders and oppositional defiant), response inhibition, and impairment in social functioning are more associated with ADHD-C. On the other hand, ADHD-I is more associated with internalizing comorbid disorders (like depression, self-esteem difficulties, and anxiety), shyness and passive social behavior

* Corresponding authors.

E-mail addresses: amirmasoud.ahmadi@alumni.iust.ac.ir (A. Ahmadi), nazaripsycho@yahoo.com (M.A. Nazari).

¹ These authors contributed equally to this work.

[10,13]. Additionally, in spite of the long-lasting nature of inattentive symptoms, impulsivity and hyperactivity will no longer be valid clinical symptoms for diagnosis in late adolescent and adult ADHD [14]. Therefore, ADHD-C seems to be more diagnostically unstable during development compared with ADHD-I which suggests that different brain areas are affected in ADHD-I and ADHD-C. Consequently, clinical diagnosis of sub-types of ADHD remains a challenge [10].

Started from 80 years ago [15], electroencephalograph (EEG) has been one of the most widely used neuroimaging techniques due to its accessibility and low cost [16–22]. Multiple signal processing methods were used to discriminate electrophysiological alterations in children with ADHD, including but not limited to power analysis [17,23,24], complexity analysis [25–27], and synchronization [25,28]. Using these handcrafted feature engineering, many machine learning algorithms like logistic regression [29,30], linear discriminant analysis [31], and support vector machines (SVM) [32–34] were developed as a complementary tool for ADHD diagnosis.

Yet convolutional neural networks (CNN) have an advantage in comparison to the models mentioned above, which is their ability to learn features automatically using a large dataset [35–37]. CNN had enormous success in fields like computer vision and natural language processing thanks to the vast annotated datasets [38–40]. Over the years, multiple architectures were introduced, such as AlexNet [41] and GoogLeNet [42], which some of them outperformed human-level accuracy. This astonishing performance leads to a rising expectation that CNNs should also be able to solve the ADHD diagnosis problem. In the study of Kuang et al. [43], ADHD and normal control groups were classified using a deep belief network. Resting-state fMRI data was used to classify different subtypes of ADHD from healthy subjects. The results were validated with an inter-subject cross-validation method with 72% accuracy. Zou et al. [44] presented a model that combined fMRI and sMRI data; they used a three-dimensional CNN to extract the local spatial patterns of sMRI data and another CNN to combine the information from sMRI feature extraction network and fMRI data. With inter-subject cross-validation method, the accuracy of classifying different subtypes of ADHD from healthy was reported 69%. In another study, Atif et al. [45] designed deep structure, consisting of three sub-networks, capable of classifying raw fMRI time-series data. With the same validation method, the accuracy of classifying ADHD from healthy was reported 68%.

In this study, the classification process was performed using a resting-state EEG signal, i.e., the brain activities recorded while no sensory stimulation or task performance is present. One should always consider that interpretation of task-related fluctuation of brain functions is somewhat tricky, without having sufficient knowledge of individual functional differences among patients who have ADHD during rest state. Furthermore, in task-related studies, merely the activates occurring at a certain time window are investigated, and all other brain activities, most of the time, are ignored and treated as background noise [46–48]. Additionally, it should be considered that the brain is a system with intrinsic activities; this means that the brain's intrinsic activity is usually altered, and not caused by the external stimuli [48–50]. As an example, Gruber et al. [51] showed that visual stimuli could be predicted by pre-stimulus EEG activities; further, Mazaheri et al. [52] argued that pre-stimulus EEG activities could predict motor responses.

Our study uses machine-learning algorithms to provide a differential

diagnosis of the different subtypes of ADHD. Currently, the diagnostic process of psychiatric diseases fully depends on experienced psychiatrists, which is not always without mistakes. Therefore, there is a great tendency to automate the diagnostic process using machine-learning algorithms. Deep Convolutional neural networks (CNNs) both offer high performance and show meaningful differences between individuals suffering from ADHD in terms of spatial-frequency abnormalities.

In many previous studies [53–58], the architecture design of CNNs for EEG classification is inspired by classical EEG feature extraction and classification methods. As can be seen in Fig. 1, the typical EEG feature extraction and decoding pipeline consist of a band-pass filtering, spatial filtering, and a classifier. The band-pass block filters the input signal into a series of predetermined bands. In the spatial filter method, like a common spatial filter (CSP), finds the optimum combination of channels. The variance of the signal is calculated, and finally, these values are used as features for classification.

One hopes that these observed individual abnormalities could improve targeted treatment of children afflicted with ADHD and, in general, improve our knowledge about the nature of the disease itself. We organized this paper in the following manner: Section 2 explains the collection and preprocessing of the EEG data and described the proposed method in detail. Sections 3, 4 and 5 will contain results, discussion, and conclusion, respectively.

2. Materials and methods

2.1. Subjects

Forty right-handed children (age between 6 and 11 years old) took part in this study. ADHD-diagnosed children were all selected from Hamrah Child and Adolescent Multidisciplinary Neuropsychiatric Center, Tabriz, Iran. None of the children had any neuro-feedback, transcranial direct current stimulation, or any other neuro-modulation treatments; similarly, none had been treated with methylphenidate. ADHD children were identified as two combined ADHD (ADHC-C) and inattentive ADHD (ADHD-I) using DSM-5 criteria. For all the participants, both their parents and teachers were asked to fill the Swanson, Nolan, and Pelham IV questionnaire [59]; Additionally, parents filled the child behavior checklist [60]. The final diagnosis of the children was independently performed by a child psychologist and a child psychiatrist who both were blinded to the previous findings. The subject was assigned to a group in case both the psychologist and the psychiatrist agreed on the diagnosis. The subject was excluded from the study in case any confounding neuro-psychiatric disorders were diagnosed [61]. Finally, we had 13 ADHD-C [9 boys, 4 girls; age: 8.5 ± 0.7 y (mean \pm standard error)], 12 ADHD-I [7 boys, 5 girls; age: 8.75 ± 0.65 y] and 14 controls [8 boys, 6 girls; mean age: 8.92 ± 1.38 y].

2.2. EEG recording and preprocessing

We used Mitsar® amplifier with 21 channels and WinEEG® software for recording. EEG was sampled at 250 Hz with an online 0.1–40 Hz band-pass filter. EEG was recorded using 19 channels and Electrocap® according to the 10–20 international system. Linked earlobes were selected as the reference, and the ground was placed on AFZ. While recording, the electrodes' impedance was maintained below 10 k Ω .



Fig. 1. A typical feature extraction pipeline in classical EEG classification methods.

Finally, EEG signals were recorded for 5 min in an open-eye condition. One of the most challenging parts of analyzing and interpretation of EEG data is the sensitivity of the process to the noise and artifacts. Ocular signals, for instance, must be discarded by denoising methods Independent Component Analysis (ICA) to mitigate the negative effect of eye blink. Firstly, the EEG signal was filtered (zero-phase Butterworth filter, 3th order) between the most important frequency bands (1–40 Hz). In the next step, a 50 Hz phase-shift free second-order IIR notch filter was used to remove power line noise. Then, the filtered EEG data was referenced to the common average and decomposed into independent components using ICA method. To remove unwanted artifacts, eye blinks and muscle artifacts were identified by brain-related independent components (ICs) and manually removed based on their spectra, scalp maps, and time courses. The continuous-time series was cut to non-overlapping 10-second epochs. An epoch was rejected if the EEG signal of any channel exceeded a threshold of ± 250 μ V or had an abnormal probability or kurtosis (more than five standard deviations beyond the mean). Finally, after preprocessing, the total number of epoch 905 remains, which is divided into three classes, with the number of epochs in the normal class, ADHD-C, ADHD-I, 362, 246, and 297, respectively.

2.3. Classification

In recent years, CNNs have been successfully applied in image processing [62,63] and natural language processing applications [64,64]. Thanks to their convolution-based structure, they can handle high-dimensional input data structures, like images, without creating a large number of weights in each layer; this feature and high accuracy of these networks make them an apt candidate also for classification of EEG data in different paradigms.

In this study, a CNN structure was designed based on previous similar structures like EEGNet [65]. In addition to spatial and temporal filters, we used Dropout [66] and Batch normalization [67] layers to make the network less prone to overfitting. The model was trained to minimize the categorical cross-entropy using Adam optimizer [68]. The model was validated using five times five-fold cross-validation; in each fold, 20% of

train data was selected for validation and finding optimum values for hyper-parameters. We run 200 train iterations, and the model performing best on validation was saved for the test. Models were trained with NVIDIA GTX 1060 GPU with CUDA 9 in Keras API [69]. The structure of the network can be seen in Fig. 2.

The detail of each layer is described in the next section:

1- Input layer: The input of the network for each trial is a tensor with the shape of (1, 2500, 19) containing 2500 time samples of 19 EEG channels.

2- Band-Filter (Conv2D): This layer performs six convolution filters with size (125, 1) on the first dimension of the input tensor. The output will be a tensor with shape (6, 2500, 19).

Initialization:

The weights of this layer are initialized with band-pass FIR filters (Kaiser Window method) in 6 well-known EEG bands, i.e., delta 1~4 Hz, theta 4~8 Hz, alpha 8~13 Hz, beta1 13~20 Hz, beta2 20~30 Hz, and gamma 30~40 Hz. The order of the filter was the same as the length of the convolution layer. To make sure that the output of each filter is always a band-pass filter, the weights in this layer are not updated during optimization.

Investigating the contribution of each frequency band:

The contribution of each band can be investigated by only applying one of the frequency bands at the time. For instance, by applying one (125, 1) filter, initialized with band-pass FIR in beta 13~20 Hz frequency band, the network is merely using this single band in the classification process. We used the same method to investigate the contribution of each frequency band in classification.

3- Spatial-Filtering (Depthwise Convolution): In this layer, multiple (1, 19) kernels are applied to each frequency band of input tensor resulting in an output tensor with the shape of (6*depth_multiplier, 2500, 1). The hyper-parameter "depth_multiplier" controls the number of filters that are applied for each frequency band. The results were tested for a different number of depth_multiplier, and the final value was set to 11; therefore, with "valid" padding, the output of this layer will be a tensor with the shape of (66, 2500, 1).

4- Batch Normalization and Activation: The output of the last layer was passed through a Batch Normalization layer and an "ELU"

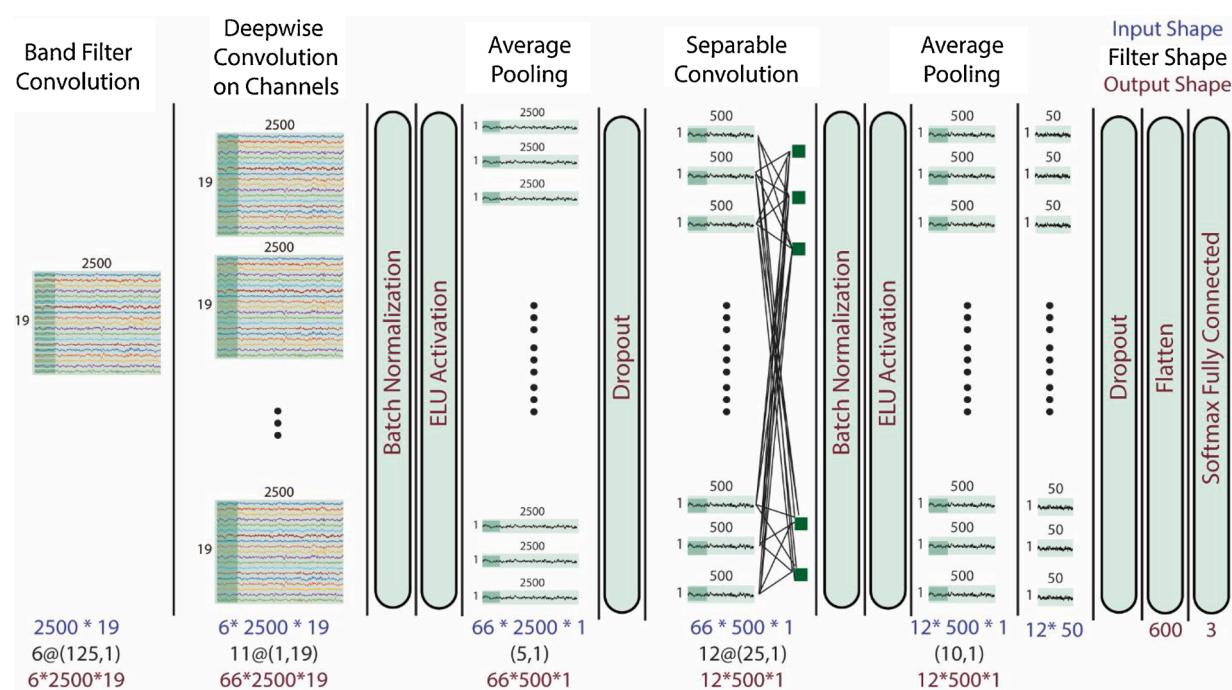


Fig. 2. Structure of the convolutional neural network. The layers are separated with vertical lines. In the bottom, the input shape, filter shape, and output shape are shown in blue, black, and red text for each layer.

activation.

5- Average pooling: To reduce the size of the data tensor, an average pooling with the size of (5, 1) was applied, resulting in an output tensor with the size of (66, 500, 1).

6- Dropout Layer

7- Temporal-Filter (Separable Convolution): To have a richer feature map containing information from different time samples of data, multiple 2D Separable Convolution filters with size (25, 1) are applied. The number of Separable Convolution filters is a hyperparameter of the network. With setting the number of Separable filters to 12 and using padding, the output will be a tensor with the size of (12, 500, 1). Separable convolution consists of two steps. The first is a Depthwise convolution with size (25, 1) and then followed by 12 (1,1) pointwise convolution [65]. This approach will reduce the number of trainable parameters and separates the learning process of Depthwise feature maps.

8- Batch Normalization and Activation: The output of the last layer was passed through a Batch Normalization layer and an “ELU” activation.

9- Average pooling: To reduce the size of the data tensor, an average pooling with the size of (10, 1) was applied, resulting in an output tensor with the size of (12, 50, 1).

10 - Dropout Layer

11- Flatten: This layer transforms the tensor to a vector with the length 600.

12- Dense Layer: A fully connected layer with 3 neurons and “Softmax” activation.

Detailed information about different layers of the neural network can be seen in Table 1.

Multiple metrics were used to evaluate the model. First, accuracy is used to shows how well the model can generally discriminate against the healthy cases from the patient cases. Moreover, two other metrics, recall, and precision, are used to evaluate the model’s determination for each of the classes. In addition, Cohen’s kappa static measure [70] is used to compare the model’s accuracy with the expected accuracy (Chance level) (more details in [70,71]).

3. Result

The primary purpose of this study was to discriminate the EEG signal of healthy children from that of children with Inattentive ADHD and Combined ADHD. To achieve this goal, we recorded the EEG signal of 40 children in the rest state with their eyes open. After preprocessing, the continuous signal was segmented. Each segment, along with its proper label, was classified using the proposed deep neural network. The first layer of the networks decomposed the input signal to commonly known EEG bands. An important hyper-parameter of the network is the number of spatial filters in the second layer of the network. This parameter

indicates how many spatial filters are created per band. We examined the performance of different architectures with 1–15 (with steps of 2) spatial filters per band. As can be seen in Fig. 3, the performance of the network improved with increasing the number of spatial filters in all the frequency bands. However, the rate of improvement decreased for more than nine spatial filters per band. The performance of the network is maximized when β_1 , β_2 , and γ bands were used. In addition, in these three bands, the difference in classification accuracy is negligible for 11, 13, and 15 spatial filters. With β_1 and β_2 bands, a minute loss of performance is observed when we used 15 spatial filters.

Increasing the number of spatial filters will increase the number of trainable parameters and, consequently, the computational load. Therefore, selecting the optimum number of spatial filters is essential. The number of spatial filters was considered optimum when increasing it did not significantly improve the performance of the network. The statistical comparison between classification results for all frequency bands and the number of spatial filters is shown in Fig. 4. We used Friedman statistical test with Bonferroni Correlation; significant values are indicated with the red asterisks. The architecture with 11 spatial filters was selected as optimum since the performance of the classifier with 13 or 15 spatial filters were not significantly different.

The performance of the proposed method was validated by four criteria in each frequency band. The results are shown in Fig. 5. As seen in Fig. 5, all four performance criteria are significantly (Friedman’s test with Bonferroni Correction, P-Value<0.05) higher in high-frequency bands (β_1 , β_2 , γ) compared to lower ones (δ , θ , α). The Accuracy of the classifier was 94.3, 97.2, and 96.4 for β_1 , β_2 , and γ band, respectively; Kappa values for these bands are higher than 0.9, which is considered in the category of “Almost Perfect Agreement,” according to [70].

We used confusion matric to investigate the performance of the classifier in Normal, ADHD-I, and ADHD-C classes. In Fig. 6, the confusion matrix for each frequency band is presented. When we merely used the β_1 band, the Normal class was detected the best, and the lowest classification performance was related to ADHD-I class with %91.04 accuracy; the missed ADHD-I samples were mostly confused with normal samples. In the β_2 band, the performance of all three classes is close together, and again, the ADHD-I class was missed the most. In the case of merely using the γ band, the normal and ADHD-I classes had the highest performance with %97.63 and %95.08 accuracy, respectively.

EEG signal shows the difference in electrical potential between electrodes and a reference. Since these electrodes are placed in different areas over the scalp, the electrical activities can indicate the activities of different brain areas. In the proposed neural network, 11 spatial filters are created for each frequency band; these filters assign a weight for each channel in the given frequency band. The absolute value of each channel’s weight indicates the importance of the given channel in the classification process. To calculate the contribution of each channel in

Table 1

Detailed information about different layers of the neural network. For each layer, the number of filters, size of each filter, number of trainable parameters, input shape, output shape, and activation type are reported. For some layers, additional information is reported in the Detail column.

Layer Name	# filters	Size	# Parameters	Input shape	Output shape	Activation	Detail
Input			0	1,2500,19	1,2500,19		
Band Filter Conv2D	6	(125, 1)	750	1,2500,19	6,2500,19		Not Trainable
Deep wise on Channels	11	(1,19)	209	(6, 2500, 19)	(66, 2500, 1)	Linear	No Bias, Channels first, depth multiplier = 11
Batch Normalization 1			44	(66, 2500, 1)	(66, 2500, 1)		On Feature Map
Activation Layer			0	(66, 2500, 1)	(66, 2500, 1)	Elu	
Average Pooling		(5, 1)	0	(66, 2500, 1)	(66, 500, 1)		Channel First
Dropout			0	(66, 500, 1)	(66, 500, 1)		dropout rate = 0.25
Separable Convolution	12	(25, 1)	407	(66, 500, 1)	(12, 500, 1)	Elu	No Bias, Channels first, padding = same,
Batch Normalization 2			48	(12, 500, 1)	(12, 500, 1)		On Feature Map
Activation Layer			0	(12, 500, 1)	(12, 500, 1)	Elu	
Average Pooling		(10, 1)	0	(12, 500, 1)	(12, 50, 1)		Channel First
Dropout			0	(12, 50, 1)	(12, 50, 1)		dropout rate = 0.25
Flatten			0	(12, 50, 1)	600		
Dense	1	3	1803	600	3	Softmax	

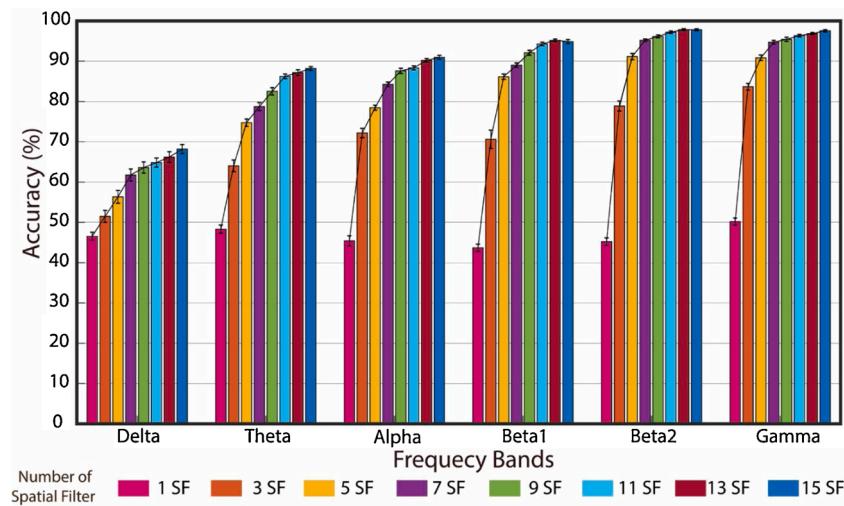


Fig. 3. Performance of the 3-class classification for each frequency band and the various number of spatial filters.

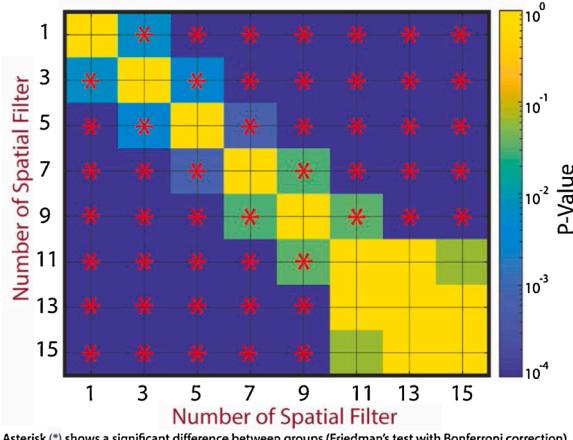


Fig. 4. Statistical comparison between the performance and the different number of spatial filters. Horizontal and vertical axes show the number of spatial filters, and statistical significance between two numbers of spatial filters is shown with asterisks.

each fold, we calculated the absolute value of all the spatial weights, and to normalize the contribution value, we divided the absolute values of weights by their maximum value. Then, we calculated the mean of the contribution value overall 11 spatial filters and all the folds. These values are normalized between 0 and 1. The closer this value was close to 1, the more it had a contribution in discriminating different classes. In Fig. 7, topographic maps show the contribution of EEG channels in β_1 , β_2 , and γ bands. In all three β_1 , β_2 and γ bands, O1 and F_z showed more contribution. Furthermore, in β_2 and γ bands, the contribution of the posterior side was significantly more than in other areas (Anterior and Central side).

Additionally, to have a closer investigation on the contribution of different brain regions, the contribution of three main brain sides in β_1 , β_2 , and γ bands are shown in Fig. 8. As can be seen, for all three investigated bands, the contribution of the posterior side is significantly more than the other sides, and this effect is more significant in β_2 and γ compared to β_1 . Furthermore, for β_2 and γ bands, the contribution of the central side was more than the anterior side; In contrast, for β_1 , the anterior side contributed more than the central side.

In previous sections, we showed that the performance of the proposed method is higher when β_1 , β_2 , and γ bands are used for classification. Here, we want to investigate whether using a combination of these bands will improve the performance of the classifier. In Fig. 9, we measured the performance of the proposed method in four cases (only β_1 , only β_2 , only γ , and the combination of three bands). As it can be

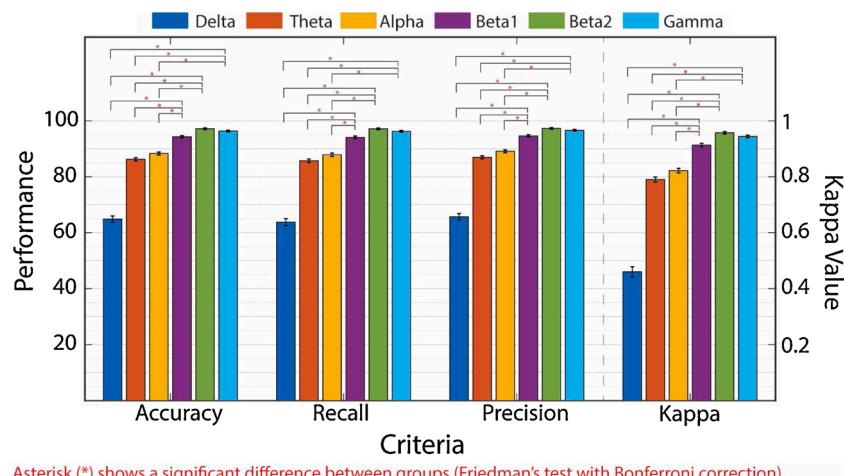


Fig. 5. Multiclass performance of the classifier for four performance criteria and different frequency bands.

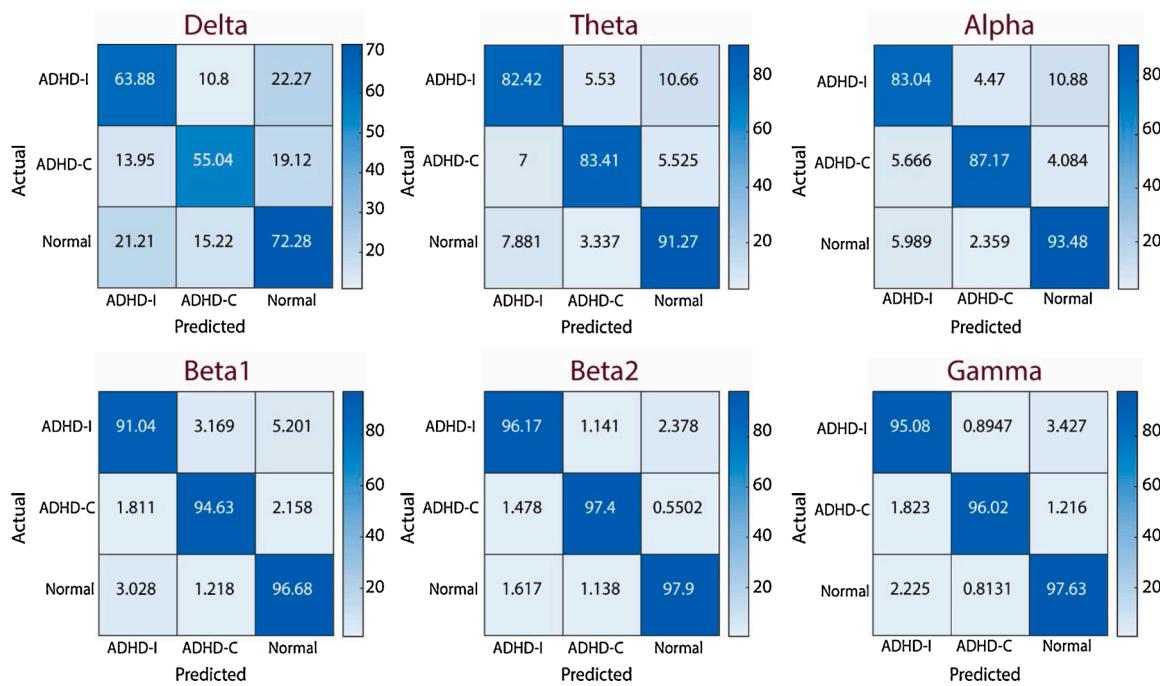


Fig. 6. The confusion matrix for 3-class classification. The horizontal and vertical axes indicate the predicted and the true class, respectively.

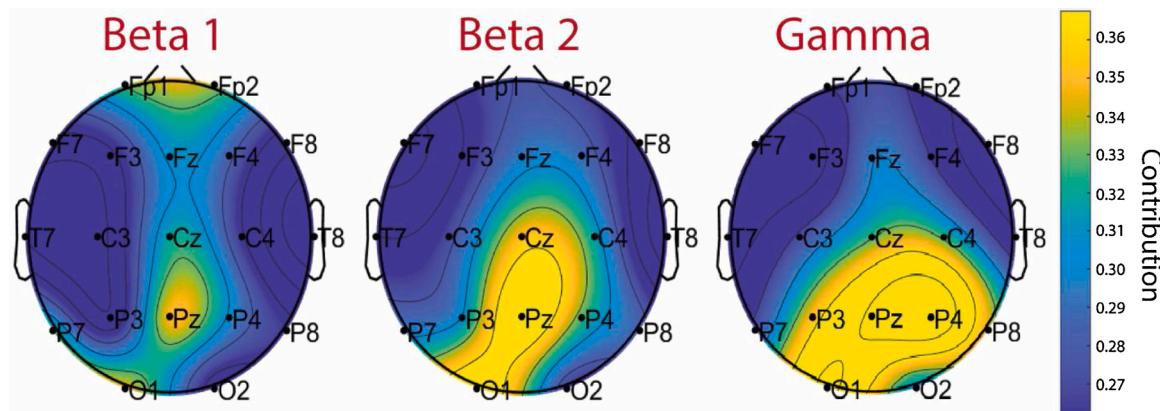


Fig. 7. Topographical map for the contribution of EEG channels in classifying healthy and ADHD children for β_1 , β_2 , and γ bands. The warmer colors indicate a higher contribution.

seen, the performance of the proposed method was significantly higher when three bands were combined; Accuracy, recall, precision, and Kappa were %0.9946, %99.45, %99.48 and %99 respectively; hence, the performance of the proposed method was significantly higher when we combined β_1 , β_2 and γ bands.

4. Discussion

In this paper, we proposed a novel deep learning approach to classify the EEG signal of Healthy children (Control) from ADHD children with two subtypes of Combined ADHD (ADHD-C) and Inattentive ADHD (ADHD-I). Inspired by the classical approaches for classification of EEG signal, we proposed a deep neural network that is capable of extracting the relevant features from the raw EEG signal and then performing the classification. The first layer of the network performs band-pass filtering on the raw signal, and the second layer performs channel selection by applying a spatial filter on EEG channels for each band passed signal of the first layer. This systematic feature extraction in the network architecture can make the classification process more explainable. The

classification method proposed in this work is based on the EEG signal, and to the best of our knowledge, all previous multiclass (ADHD-C, ADHD-I, and Control) studies were based on fMRI and MRI imaging [72–76]. Therefore, the presented research is novel in terms of using a deep neural network architecture and EEG signal for multiclass classification of ADHD and healthy children.

In recent years, many studies have been focusing on Computer-Aided Diagnosing (CAD) systems for automatic detection of neurological abnormalities. Many imaging techniques like fMRI, MRI, MEG, EEG, and optical imaging approaches were used for developing such CAD systems. All the techniques mentioned above have their advantages and disadvantages. For instance, fMRI and optical imaging methods rely on hemodynamic response variations, which is rather slow compared to neural responses. Furthermore, fMRI, MRI and MEG techniques are expensive and not easily accessible. Thus, in this study, we used the EEG signal because of its high temporal resolution, lower hardware, and maintenance costs, and, more importantly, because it is more easily available for clinical use [18].

The performance of the proposed method was higher when β_1 , β_2 ,

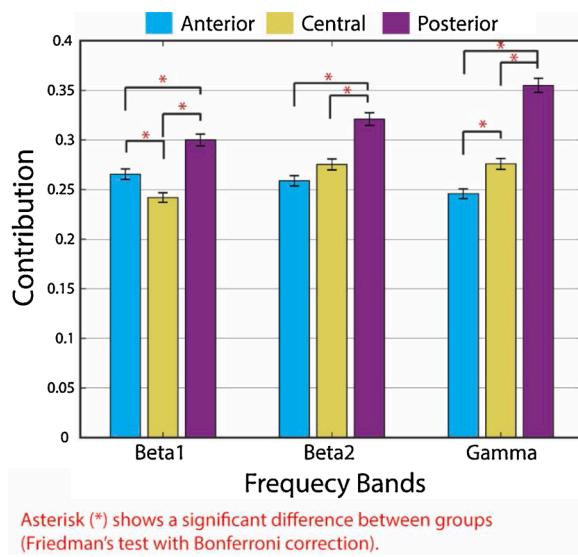


Fig. 8. Mean contribution of EEG electrodes in posterior, central, and anterior sides for in β_1 , β_2 , and γ bands.

and γ bands were used. (Fig. 5) Also, we achieved maximum performance when all the three higher frequency bands ($\beta_1 + \beta_2 + \gamma$) were used. This combined-bands performance was significantly higher than all the cases that only one frequency band was used. We used four criteria to evaluate the performance. In the case of using all the frequency bands, the Accuracy Recall, Precision, and Kappa values were % 99.46, %99.45, %99.48, and 0.99, respectively. After investigating the spatial channels, we observed that electrodes in the Posterior side had the most contribution. Both the observed frequency band and spatial contribution results are in line with previous studies on ADHD patients. Over the years, many studies showed a relation between high-frequency EEG activities and attentional processes in both healthy individuals and those with clinical conditions [77]. For instance, Gola et al. [78] found that difficulties in activating and sustaining attentional processes are correlated with decreased β band activities during a visual attention

task. As another instance, MacLean et al. [56] argued that during an attentional blink task, high resting β band power and increased α band power were associated with high and low-performance accuracy, respectively. In the case of individuals with clinical conditions, Matsuura et al. [79]; Clarke et al. [23]; and Roh et al. [80] showed that ADHD patients had lower β and γ band power compared to a healthy control group with the matched age. Uhlhaas et al. [81] associated γ band activities (EEG filtered >30 Hz) with fast and synchronous spiking action potentials of inhibitory parvalbumin-positive interneurons, which is capable of forming both local and long-range cortical circuits. Many studies, including Basar et al. [82], Herrmann et al. [83], and Herrmann et al. [84] and Buzsaki and Draguhn [85] stated that high-order cognitive processes like working memory, attention, language, and executive functions would increase the γ band activities. Furthermore, many studies correlated abnormalities in stimulus-driven EEG γ band activity of individuals suffering from neurological and psychiatric disorders [86]. For instance, Kwon et al. [87] studied the relationship between γ band activation and schizophrenia, Rampp and Stefan [88] showed γ band abnormalities in epilepsy penitents, and Yordanova et al. [89] investigated the γ band abnormalities in ADHD patients.

Despite considerable success in Image Classification and Natural Language Processing, deep learning methods face a significant challenge when used for clinical applications. In clinical applications, the interpretability of the tool is vital. Unfortunately, in the case of deep learning methods, the designers usually see the method as a black box and therefore, when the classifier makes a decision, neither the designer nor the clinician will understand which part of the input (EEG signal) had the most contribution to the decision-making process. To address the problem mentioned above, we proposed a method that can show the frequency bands and the brain areas with the most contribution in the decision-making process in addition to having a high performance in terms of classification accuracy.

To the best of our knowledge, in all the previous studies, which applied a deep learning approach to fMRI data, the results were below satisfactory. Qureshi et al. [70] classified subtypes of ADHD from rest-state sMRI data using a hierarchical machine learning method; three classes (two types of ADHD and healthy) were classified with 69% accuracy (k-fold cross-validation). In [35], sMRI and fMRI were used to

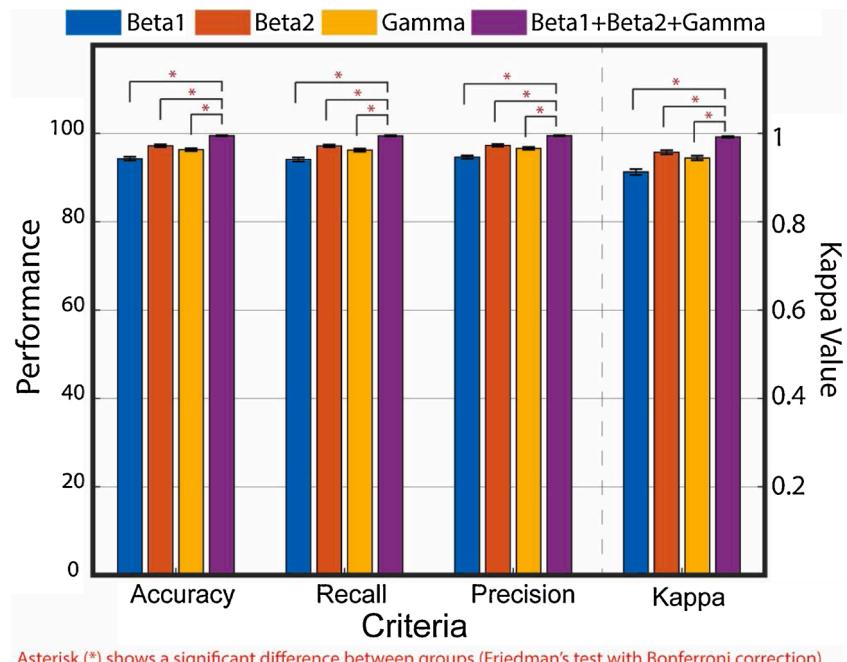


Fig. 9. Performance of the proposed method for β_1 , β_2 , γ and $(\beta_1, \beta_2, \gamma)$ combined. Accuracy, Recall, Precision, and Kappa were used as the measure of performance.

classify ADHD patients into three classes, i.e., ADHD inattentive, ADHD combined, and normally developing. Features were extracted from rest-state fMRI, and additionally, sMRI data were used to extract parcellation-based features. Then, the most discriminative features of fMRI data were selected using an ANOVA analysis over each modality and all the voxels. Furthermore, multi-modal and multi-measure features of sMRI data, including surface area, cortical thickness, image intensity, volume, and cortical thickness standard deviation were added to feature space. Using an ELM, the classification accuracy for the binary setting and the multi-class settings were 76.2% and 92.86% respectively. ANOVA-based fMRI features showed the most contribution to the classification process. In another study [86], two classes of ADHD and normal children were discriminated using their rest-state EEG signal. They used a convolutional neural network (CNN) structure combined with a visualization technique named Gradient-weighted Class Activation Mapping (Grad-CAM) as the classifier. They achieved 90% classification accuracy under k-fold cross-validation. Ahmadlou and Adeli [87], in a rest-state EEG study, classified two classes of ADHD and Healthy subjects using an RBF neural network and wavelet-synchronization feature extraction method. The EEG signal was decomposed to a set of sub-band and then the synchronization features between different channels were calculated. The final classification accuracy was 96% with the k-fold cross-validation method. In recent years, Deep Learning methods were applied for ADHD detection. For instance, Chen et al. [88] proposed a Convolutional Neural Network (CNN), which utilized EEG, based brain network for diagnosing ADHD subjects. They compared the proposed method with a Multi-Layer Perceptron (MLP) and a Support Vector Machine. The proposed method outperformed others, achieving 94.67% classification accuracy.

We compared the models used in previous works with the proposed method. As can be seen, the proposed method outperformed both EEG-based and fMRI-based methods. In order to have a fair comparison, in the Table 2, we only presented studies that used resting-state brain data for classification and k-fold cross-validation for validating the method. Table 2 has five columns. In Methods column, we summarized the feature extraction and method used in previous studies. In the Data column, we indicated whether the EEG or fMRI data were used in classification. In the scenario column, two classification scenarios were noted in the table; multiclass and binary respectively refer to three-class classification (ADHD-I, ADHD-C, and Control) and two-class classification (ADHD vs. Control).

The performance of the proposed method was investigated for various epoch lengths (2, 4, 6 ... 16 s were tested) and based on the results, we considered 10 s trial length to be optimal. In future works automatic methods such as CWGAN [93] can be used to find the optimal epoch length and improving the overall system performance.

5. Conclusion

In this study, we designed a deep neural network architecture to classify two subsets of ADHD children (ADHD-I, ADHD-C) from healthy children (Control Group) using their rest EEG signal. Diagnosis of patients suffering from different ADHD subtypes from healthy children is a new problem. Most of the proposed methods are based on fMRI brain imaging, which is harder to obtain and more expensive compared to the EEG signal. Inspired by the classical methods, the proposed structure firstly decomposes the signal into the well-known frequency bands. It then constructs a set of optimal spatial filters for each frequency band. The network showed a high performance merely using three higher frequency bands (β_1 , β_2 , γ); accuracy, recall, precision, and Kappa values were %99.46, %99.45, %99.48, and %99, which indicates the capability of the system in discrimination of three classes (ADHD-I, ADHD-C, Normal). Furthermore, investigation of optimized spatial filters showed a higher contribution of posterior side electrodes for β_1 , β_2 , γ bands. In future works, the training process of the network can be improved by increasing EEG data samples from more subjects. With

Table 2

Comparing the performance of the proposed method with previous studies.

Author	Methods	Data	Scenario	Accuracy
Sidhu et al. [72]	Components of FFT and kernel PCA as features	fMRI	Multiclass	68.6%
Qureshi et al. [73]	Cortical thickness features and hierarchical extreme learning machine classifier	fMRI	Multiclass	60.8%
Qureshi et al. [74]	Structural features and extreme learning machine classifier	fMRI	Multiclass	76.2%
Chen et al. [90]	Convolutional neural network and a visualization technique named gradient weighted class activation mapping.	EEG	Binary	90.29%
Ahmadolou [91]	Wavelet-Synchronization and RBF neural network	EEG	Binary	95.6%
Chen [92]	Combining and EEG-Based Brain network with Convolutional neural network	EEG	Binary	94.67%
Proposed Method	Frequency band and spatial filtering feature extraction and classification using a Deep Convolutional Neural Network	EEG	Multiclass	99.46%

more data, the classification system can be evaluated with more exacting validation methods. Multiclass classification of ADHD patients using the EEG signal is a challenging subject, and we hope that this work will be inspiring for future studies.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the University of Tabriz research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

All the children's parents provided written informed consent forms after receiving detailed information about the study.

CRediT authorship contribution statement

Amirmasoud Ahmadi: Conceptualization, Methodology, Validation, Software, Resources, Writing - review & editing, Visualization, Data curation, Formal analysis. **Mehrdad Kashefi:** Conceptualization, Methodology, Validation, Software, Resources, Writing - review & editing, Visualization, Data curation, Formal analysis. **Hassan Shahrokhi:** Writing - original draft, Resources, Investigation, Formal analysis, Data curation. **Mohammad Ali Nazari:** Writing - review & editing, Writing - original draft, Resources, Investigation, Formal analysis, Data curation, Project administration, Supervision.

Declaration of Competing Interest

None Declared

References

- [1] S.J.J. Kooij, S. Bejerot, A. Blackwell, H. Caci, M. Casas-Brugué, P.J. Carpentier, D. Edvinsson, J. Fayyad, K. Foeken, M. Fitzgerald, V. Gaillac, Y. Ginsberg, C. Henry, J. Krause, M.B. Lensing, I. Manor, H. Niederhofer, C. Nunes-Filipe, M. D. Ohlmeier, P. Oswald, S. Pallanti, A. Pehlivaniidis, J.A. Ramos-Quiroga, M. Rastam, D. Ryffel-Rawak, S. Stes, P. Asherson, European consensus statement on diagnosis and treatment of adult ADHD: the European Network Adult ADHD, *BMC Psychiatry* 10 (2010) 67, <https://doi.org/10.1186/1471-244X-10-67>.
- [2] R.C. Kessler, L. Adler, R. Berkley, J. Biederman, C.K. Conners, O. Demler, S. V. Faraone, L.L. Greenhill, M.J. Howes, K. Scenik, T. Spencer, T.B. Ustun, E. E. Walters, A.M. Zaslavsky, The prevalence and correlates of adult ADHD in the

- United States: results from the national comorbidity survey replication, *Am. J. Psychiatry* 163 (2006) 716–723, <https://doi.org/10.1176/ajp.2006.163.4.716>.
- [3] M. Jouzizadeh, R. Khanbabae, A.H. Ghaderi, A spatial profile difference in electrical distribution of resting-state EEG in ADHD children using sLORETA, *Int. J. Neurosci.* (2020) 1–14, <https://doi.org/10.1080/00207454.2019.1709843>.
 - [4] L.J. Seidman, E.M. Valera, N. Makris, Structural brain imaging of attention-deficit/hyperactivity disorder, *Biol. Psychiatry* 57 (2005) 1263–1272, <https://doi.org/10.1016/j.biopsych.2004.11.019>.
 - [5] D.S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, E. Bullmore, Adaptive reconfiguration of fractal small-world human brain functional networks, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 19518–19523, <https://doi.org/10.1073/pnas.0606005103>.
 - [6] E.G. Willcutt, The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review, *Neurotherapeutics* 9 (2012) 490–499.
 - [7] V. Del Barrio, Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Pub, 2016, <https://doi.org/10.1016/B978-0-12-809324-5.05530-9>.
 - [8] F.X. Castellanos, R. Tannock, Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes, *Nat. Rev. Neurosci.* 3 (2002) 617–628, <https://doi.org/10.1038/nrn896>.
 - [9] M.V. Solanto, The predominantly inattentive subtype of Attention-Deficit/Hyperactivity disorder, *CNS Spectr.* 5 (2000) 45–51, <https://doi.org/10.1017/S1092852900007069>.
 - [10] E.G. Willcutt, J.T. Nigg, B.F. Pennington, M.V. Solanto, L.A. Rohde, R. Tannock, S. K. Lee, C.L. Carlson, K. McBurnett, B.B. Lahey, Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes, *J. Abnorm. Psychol.* 121 (2012) 991–1010, <https://doi.org/10.1037/a0027347>.
 - [11] D.A. Pineda, M.A. Restrepo, R.J. Sarmiento, J.E. Gutierrez, S.A. Vargas, Y. T. Quiroz, G.W. Hynd, Statistical analyses of structural magnetic resonance imaging of the head of the caudate nucleus in Colombian children with attention-deficit hyperactivity disorder, *J. Child Neurol.* 17 (2002) 97–105, <https://doi.org/10.1177/088307380201700202>.
 - [12] S.R. Miller, C.J. Miller, J.S. Bloom, G.W. Hynd, J.G. Craggs, Right hemisphere brain morphology, attention-deficit hyperactivity disorder (ADHD) subtype, and social comprehension, *J. Child Neurol.* 21 (2006) 139–144, <https://doi.org/10.1177/08830738060210021901>.
 - [13] D. Baeyens, H. Roeyers, J. Vande Walle, Subtypes of Attention-Deficit/Hyperactivity Disorder (ADHD): Distinct or related disorders across measurement levels? *Child Psychiatry Hum. Dev.* 36 (2006) 403–417, <https://doi.org/10.1007/s10578-006-0011-z>.
 - [14] S.V. Faraone, P. Asherson, T. Banaschewski, J. Biederman, J.K. Buitelaar, J. A. Ramos-Quiroga, B. Franke, Attention-deficit/hyperactivity disorder, *Nat. Rev. Dis. Primers* 1 (2015), 15020.
 - [15] H.H. Jasper, P. Solomon, C. Bradley, Electroencephalographic analyses of behavior problem children, *Am. J. Psychiatry* 95 (1938) 641–658, <https://doi.org/10.1176/ajp.95.3.641>.
 - [16] M. Van Lieshout, M. Luman, J. Buitelaar, N.N.J. Rommelse, J. Oosterlaan, Does neurocognitive functioning predict future or persistence of ADHD? A systematic review, *Clin. Psychol. Rev.* 33 (2013) 539–560, <https://doi.org/10.1016/j.cpr.2013.02.003>.
 - [17] S. Markovska-Simoska, N. Pop-Jordanova, Quantitative EEG in children and adults with attention deficit hyperactivity disorder, *Clin. EEG Neurosci.* 48 (2017) 20–32, <https://doi.org/10.1177/1550059416643824>.
 - [18] A. Ahmadi, S. Davoudi, M.R. Daliri, Computer Aided Diagnosis System for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention, *Comput. Methods Programs Biomed.* 169 (2019) 9–18, <https://doi.org/10.1016/j.cmpb.2018.11.006>.
 - [19] A. Ahmadi, M. Behroozi, V. Shalchyan, M.R. Daliri, Classification of epileptic EEG signals by wavelet based CFC, 2018 Electric Electronics, Computer Science, Biomedical Engineers' Meeting, EBBT 2018, IEEE (2018) 1–4, <https://doi.org/10.1109/EBBT.2018.8391471>.
 - [20] A. Ahmadi, S. Davoudi, M. Behroozi, M.R. Daliri, Decoding covert visual attention based on phase transfer entropy, *Physiol. Behav.* 222 (2020) 112932, <https://doi.org/10.1016/j.physbeh.2020.112932>.
 - [21] A. Ahmadi, V. Shalchyan, M.R. Daliri, A new method for epileptic seizure classification in EEG using adapted wavelet packets, in: 2017 Electric Electronics, Computer Science, Biomedical Engineers' Meeting, EBBT 2017, IEEE, 2017, pp. 1–4, <https://doi.org/10.1109/EBBT.2017.7956756>.
 - [22] S. Davoudi, A. Ahmadi, M.R. Daliri, Frequency-amplitude coupling: a new approach for decoding of attended features in covert visual attention task, *Neural Comput. Appl.* (2020) 1–16, <https://doi.org/10.1007/s00521-020-05222-w>.
 - [23] A.R. Clarke, R.J. Barry, R. McCarthy, M. Selikowitz, Electroencephalogram differences in two subtypes of Attention-Deficit/Hyperactivity Disorder, *Psychophysiology* 38 (2001) 212–221, <https://doi.org/10.1111/1469-8986.3820212>.
 - [24] G. Ogrim, J. Kropotov, K. Hestad, The quantitative EEG theta/beta ratio in attention deficit/hyperactivity disorder and normal controls: sensitivity, specificity, and behavioral correlates, *Psychiatry Res.* 198 (2012) 482–488, <https://doi.org/10.1016/j.psychres.2011.12.041>.
 - [25] J.J. González, G. Alba, S. Mañas, A. González, E. Pereda, Assessment of ADHD Through Electroencephalographic Measures of Functional Connectivity, ADHD - New Directions in Diagnosis and Treatment, 2015, pp. 35–54, <https://doi.org/10.5772/60559>.
 - [26] H. Sohn, I. Kim, W. Lee, B.S. Peterson, H. Hong, J.H. Chae, S. Hong, J. Jeong, Linear and non-linear EEG analysis of adolescents with attention-deficit/hyperactivity disorder during a cognitive task, *Clin. Neurophysiol.* 121 (2010) 1863–1870, <https://doi.org/10.1016/j.clinph.2010.04.007>.
 - [27] L. Chenxi, Y. Chen, Y. Li, J. Wang, T. Liu, Complexity analysis of brain activity in attention-deficit/hyperactivity disorder: a multiscale entropy analysis, *Brain Res. Bull.* 124 (2016) 12–20, <https://doi.org/10.1016/j.brainresbull.2016.03.007>.
 - [28] K. Geladé, M. Bink, T.W.P. Janssen, R. van Mourik, A. Maras, J. Oosterlaan, An RCT into the effects of neurofeedback on neurocognitive functioning compared to stimulant medication and physical activity in children with ADHD, *Eur. Child Adolesc. Psychiatry* 26 (2017) 457–468, <https://doi.org/10.1007/s00787-016-0902-x>.
 - [29] A. Fernández, J. Quintero, R. Hornero, P. Zuluaga, M. Navas, C. Gómez, J. Escudero, N. García-Campos, J. Biederman, T. Ortiz, Complexity analysis of spontaneous brain activity in Attention-Deficit/Hyperactivity disorder: diagnostic implications, *Biol. Psychiatry* 65 (2009) 571–577, <https://doi.org/10.1016/j.biopsych.2008.10.046>.
 - [30] I. Buyck, J.R. Wiersma, Resting electroencephalogram in attention deficit hyperactivity disorder: developmental course and diagnostic value, *Psychiatry Res.* 216 (2014) 391–397, <https://doi.org/10.1016/j.psychres.2013.12.055>.
 - [31] M. Duda, R. Ma, N. Haber, D.P. Wall, Use of machine learning for behavioral distinction of autism and ADHD, *Transl. Psychiatry* 6 (2016), <https://doi.org/10.1038/tp.2015.221> e732–e732.
 - [32] A. Mueller, G. Candrian, J.D. Kropotov, V.A. Ponomarev, G.M. Baschera, Classification of ADHD patients on the basis of independent ERP components using a machine learning system, *Nonlinear Biomedical Physics*, Springer, 2010, <https://doi.org/10.1186/1753-4631-4-p>.
 - [33] J. Anuradha, V. Tisha, K.V. Ramachandran, B.K. Arulalan, Tripathy, diagnosis of ADHD using SVM algorithm, *COMPUTE 2010 - The 3rd Annual ACM Bangalore Conference* (2010) 1–4, <https://doi.org/10.1145/1754288.1754317>.
 - [34] A. Teney, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, G. Candrian, Machine learning approach for classification of ADHD adults, *Int. J. Psychophysiol.* 93 (2014) 162–166, <https://doi.org/10.1016/j.ijpsycho.2013.01.008>.
 - [35] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, *IEEE Trans. Med. Imaging* 35 (2016) 1285–1298, <https://doi.org/10.1109/TMI.2016.2528162>.
 - [36] M. Mahmud, M.S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (2018) 2063–2079, <https://doi.org/10.1109/TNNLS.2018.2790388>.
 - [37] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, K.R. Müller, Evaluating the visualization of what a deep neural network has learned, *IEEE Trans. Neural Netw. Learn. Syst.* 28 (2017) 2660–2673, <https://doi.org/10.1109/TNNLS.2016.2599820>.
 - [38] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, Li Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255, <https://doi.org/10.1109/cvpr.2009.5206848>.
 - [39] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, Understanding neural networks through deep visualization, *ArXiv Preprint ArXiv:1506.06579* (2015). <http://arxiv.org/abs/1506.06579>.
 - [40] R.G. Snyder, Vibrational spectra of crystalline n-paraffins. II. Intermolecular effects, *J. Mol. Spectrosc.* 7 (1961) 116–144, [https://doi.org/10.1016/0022-2852\(61\)90347-2](https://doi.org/10.1016/0022-2852(61)90347-2).
 - [41] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Communications of the ACM* (2017) 84–90, <https://doi.org/10.1145/3065386>.
 - [42] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. (2015).
 - [43] D. Kuang, X. Guo, X. An, Y. Zhao, L. He, Discrimination of ADHD based on fMRI data with deep belief network, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, 2014, pp. 225–232, https://doi.org/10.1007/978-3-319-09330-7_27.
 - [44] L. Zou, J. Zheng, C. Miao, M.J. McKeown, Z.J. Wang, 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI, *IEEE Access* 5 (2017) 23626–23636, <https://doi.org/10.1109/ACCESS.2017.2762703>.
 - [45] A. Riaz, M. Asad, S.M.M.R. Al Arif, E. Alonso, D. Dima, P. Corr, G. Slabaugh, Deep fMRI: AN end-to-end deep network for classification of fMRI data, in: *Proceedings - International Symposium on Biomedical Imaging*, IEEE, 2018, pp. 1419–1422, <https://doi.org/10.1109/ISBI.2018.8363838>.
 - [46] S. Makeig, A. Delorme, M. Westerfield, T.P. Jung, J. Townsend, E. Courchesne, T. J. Sejnowski, Electrocorticographic brain dynamics following manually responded visual targets, *PLoS Biol.* 2 (2004), <https://doi.org/10.1371/journal.pbio.0020017>.
 - [47] M.D. Fox, A.Z. Snyder, J.M. Zacks, M.E. Raichle, Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses, *Nat. Neurosci.* 9 (2006) 23–25, <https://doi.org/10.1038/nn1616>.
 - [48] M.D. Fox, M.E. Raichle, Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging, *Nat. Rev. Neurosci.* 8 (2007) 700–711, <https://doi.org/10.1038/nrn2201>.
 - [49] B.A. Olshausen, D.J. Field, How close are we to understanding V1? *Neural Comput.* 17 (2005) 1665–1699, <https://doi.org/10.1162/0899766054026639>.
 - [50] M.E. Raichle, A.Z. Snyder, A default mode of brain function: a brief history of an evolving idea, *NeuroImage* 37 (2007) 1083–1090, <https://doi.org/10.1016/j.neuroimage.2007.02.041>.

- [51] W.R. Gruber, W. Klimesch, P. Sauseng, M. Doppelmayr, Alpha phase synchronization predicts P1 and N1 latency and amplitude size, *Cereb. Cortex* 15 (2005) 371–377, <https://doi.org/10.1093/cercor/bhh139>.
- [52] A. Mazaheri, I.L.C. Nieuwenhuis, H. Van Dijk, O. Jensen, Prestimulus alpha and mu activity predicts failure to inhibit motor responses, *Hum. Brain Mapp.* 30 (2009) 1791–1800, <https://doi.org/10.1002/hbm.20763>.
- [53] R. San-Segundo, M. Gil-Martín, L.F. D'Haro-Enríquez, J.M. Pardo, Classification of epileptic EEG recordings using signal transforms and convolutional neural networks, *Comput. Biol. Med.* 109 (2019) 148–158, <https://doi.org/10.1016/j.combiomed.2019.04.031>.
- [54] S.L. Oh, J. Vicnesh, E.J. Ciaccio, R. Yuvaraj, U.R. Acharya, Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals, *Appl. Sci.* 9 (2019) 2870, <https://doi.org/10.3390/app9142870>.
- [55] H. Khajehpour, F. Mohagheghian, H. Ektiari, B. Makkabadi, A.H. Jafari, E. Eqlimi, M.H. Harirchian, Computer-aided classifying and characterizing of methamphetamine use disorder using resting-state EEG, *Cogn. Neurodyn.* 13 (2019) 519–530, <https://doi.org/10.1007/s11571-019-09550-z>.
- [56] B. Pitchford, K.M. Arnell, Resting EEG in alpha and beta bands predicts individual differences in attentional breadth, *Conscious. Cogn.* 75 (2019) 218–229, <https://doi.org/10.1016/j.concog.2019.102803>.
- [57] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, W. Kong, EEG classification of driver mental states by deep learning, *Cogn. Neurodyn.* 12 (2018) 597–606, <https://doi.org/10.1007/s11571-018-9496-y>.
- [58] Y.C. Liu, C. Liang, Design exploration predicts designer creativity: a deep learning approach, *Cogn. Neurodyn.* 14 (2020) 291–300, <https://doi.org/10.1007/s11571-020-09569-7>.
- [59] A. Delavar, *The Construction and Norm-Finding of a Rating Scale for Diagnosing Attention Deficit Hyperactivity Disorder in Children*, Psychiatry: Interpersonal and Biological Processes, 2008, pp. 9–15.
- [60] M. Tehrani-Doost, Z. Shahrivar, B. Pakbaz, A. Rezaie, F. Ahmadi, Normative data and psychometric properties of the child behavior checklist and teacher rating form in an Iranian community sample, *Iran. J. Pediatr.* 21 (2011) 331–342.
- [61] A.H. Ghaderi, M.A. Nazari, H. Shahrokh, A.H. Darooneh, Functional brain connectivity differences between different ADHD presentations: impaired functional segregation in ADHD-combined presentation but not in ADHD-inattentive presentation, *Basic Clin. Neurosci.* 8 (2017) 267–278, <https://doi.org/10.18869/nirp.bcn.8.4.267>.
- [62] O. Yıldırım, M. Talo, B. Ay, U.B. Baloglu, G. Aydin, U.R. Acharya, Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals, *Comput. Biol. Med.* 113 (2019) 103387, <https://doi.org/10.1016/j.combiomed.2019.103387>.
- [63] A.A. Abbasi, L. Hussain, I.A. Awan, I. Abbasi, A. Majid, M.S.A. Nadeem, Q. A. Chaudhary, Detecting prostate cancer using deep learning convolution neural network with transfer learning approach, *Cogn. Neurodyn.* 14 (2020) 523–533, <https://doi.org/10.1007/s11571-020-09587-5>.
- [64] W. Wang, J. Gang, Application of convolutional neural network in natural language processing, in: Proceedings of 2018 International Conference on Information Systems and Computer Aided Education, ICISCAE 2018, IEEE, 2019, pp. 64–70, <https://doi.org/10.1109/ICISCAE.2018.8666928>.
- [65] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces, *J. Neural Eng.* 15 (2018) 56013, <https://doi.org/10.1088/1741-2552/aace8c>.
- [66] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* 15 (2014) 1929–1958.
- [67] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, 32nd International Conference on Machine Learning, ICML 2015, 1 (2015) 448–456.
- [68] D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings (2015).
- [69] F. Chollet, et al., Keras, 2015.
- [70] M.L. McHugh, Interrater reliability: the kappa statistic, *Biochem. Med.* 22 (2012) 276–282, <https://doi.org/10.11613/bm.2012.031>.
- [71] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manag.* 45 (2009) 427–437, <https://doi.org/10.1016/j.ipm.2009.03.002>.
- [72] G. Sidhu, N. Asgarian, R. Greiner, M.R.G. Brown, Kernel principal component analysis for dimensionality reduction in fMRI-based diagnosis of ADHD, *Front. Syst. Neurosci.* 6 (2012) 1–17, <https://doi.org/10.3389/fnsys.2012.00074>.
- [73] M.N.I. Qureshi, J. Oh, B. Min, H.J. Jo, B. Lee, Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI, *Front. Hum. Neurosci.* 11 (2017) 157, <https://doi.org/10.3389/fnhum.2017.00157>.
- [74] M.N.I. Qureshi, B. Min, H.J. Jo, B. Lee, Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study, *PLoS One* 11 (2016), <https://doi.org/10.1371/journal.pone.0160697>.
- [75] S. Dey, A. Ravishankar Rao, M. Shah, Exploiting the brain's network structure in identifying ADHD subjects, *Front. Syst. Neurosci.* 6 (2012) 75, <https://doi.org/10.3389/fnsys.2012.00075>.
- [76] S. Sartipi, H. Kalbkhani, P. Ghasemzadeh, M.G. Shayesteh, Stockwell transform of time-series of fMRI data for diagnoses of attention deficit hyperactive disorder, *Appl. Soft Comput.* J. 86 (2020) 105905, <https://doi.org/10.1016/j.asoc.2019.105905>.
- [77] R.J. Barry, A.R. Clarke, S.J. Johnstone, A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography, *Clin. Neurophysiol.* 114 (2003) 171–183, [https://doi.org/10.1016/S1388-2457\(02\)00362-0](https://doi.org/10.1016/S1388-2457(02)00362-0).
- [78] M. Gola, M. Magnuski, I. Szumska, A. Wróbel, EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects, *Int. J. Psychophysiol.* 89 (2013) 334–341, <https://doi.org/10.1016/j.ijpsycho.2013.05.007>.
- [79] M. Matsuura, Y. Okubo, T. Kojima, R. Takahashi, Y.-F. Wang, Y.-C. Shen, C.K. Lee, A cross-national prevalence study of children with emotional and behavioural problems—a WHO collaborative study in the Western Pacific Region, *J. Child Psychol. Psychiatry* 34 (1993) 307–315, <https://doi.org/10.1111/j.1469-7610.1993.tb00994.x>.
- [80] S.C. Roh, E.J. Park, Y.C. Park, S.K. Yoon, J.G. Kang, D.W. Kim, S.H. Lee, Quantitative electroencephalography reflects inattention, visual error responses, and reaction times in male patients with attention deficit hyperactivity disorder, *Clin. Psychopharmacol. Neurosci.* 13 (2015) 180–187, <https://doi.org/10.9758/cpn.2015.13.2.180>.
- [81] P.J. Uhlihaas, G. Pipa, B. Lima, L. Melloni, S. Neuenschwander, D. Nikolić, W. Singer, Neural synchrony in cortical networks: history, concept and current status, *Front. Integr. Neurosci.* 3 (2009) 17, <https://doi.org/10.3389/neuro.07.017.2009>.
- [82] E. Başar, C. Başar-Eroğlu, S. Karakaş, M. Schürmann, Brain oscillations in perception and memory, *Int. J. Psychophysiol.* 35 (2000) 95–124, [https://doi.org/10.1016/S0167-8760\(99\)00047-1](https://doi.org/10.1016/S0167-8760(99)00047-1).
- [83] C.S. Herrmann, M.H.J. Munk, A.K. Engel, Cognitive functions of gamma-band activity: memory match and utilization, *Trends Cogn. Sci.* 8 (2004) 347–355, <https://doi.org/10.1016/j.tics.2004.06.006>.
- [84] C.S. Herrmann, I. Fründ, D. Lenz, Human gamma-band activity: a review on cognitive and behavioral correlates and network models, *Neurosci. Biobehav. Rev.* 34 (2010) 981–992, <https://doi.org/10.1016/j.neubiorev.2009.09.001>.
- [85] A. Draguhn, G. Buzsáki, Neuronal oscillations in cortical networks, *Science* 304 (2004) 1926–1930.
- [86] P.J. Uhlihaas, W. Singer, Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology, *Neuron* 52 (2006) 155–168, <https://doi.org/10.1016/j.neuron.2006.09.020>.
- [87] J.S. Kwon, B.F. O'Donnell, G.V. Wallenstein, R.W. Greene, Y. Hirayasu, P. G. Nestor, M.E. Hasselmo, G.F. Potts, M.E. Shenton, R.W. McCarley, Gamma frequency-range abnormalities to auditory stimulation in schizophrenia, *Arch. Gen. Psychiatry* 56 (1999) 1001–1005, <https://doi.org/10.1001/archpsyc.56.11.1001>.
- [88] S. Rampp, H. Stefan, Fast activity as a surrogate marker of epileptic network function? *Clin. Neurophysiol.* 117 (2006) 2111–2117, <https://doi.org/10.1016/j.clinph.2006.02.023>.
- [89] J. Yordanova, T. Banaschewski, V. Kolev, W. Woerner, A. Rothenberger, Abnormal early stages of task stimulus processing in children with attention-deficit hyperactivity disorder - evidence from event-related gamma oscillations, *Clin. Neurophysiol.* 112 (2001) 1096–1108, [https://doi.org/10.1016/S1388-2457\(01\)00524-7](https://doi.org/10.1016/S1388-2457(01)00524-7).
- [90] H. Chen, Y. Song, X. Li, Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD, *J. Neural Eng.* 16 (2019) 66046, <https://doi.org/10.1088/1741-2552/ab3a0a>.
- [91] M. Ahmadiou, H. Adeli, Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD, *Clin. EEG Neurosci.* 41 (2010) 1–10, <https://doi.org/10.1177/155005941004100103>.
- [92] H. Chen, Y. Song, X. Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network, *Neurocomputing* 356 (2019) 83–96, <https://doi.org/10.1016/j.neucom.2019.04.058>.
- [93] Y. Luo, EEG data augmentation for emotion recognition using a conditional wasserstein GAN, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, IEEE, 2018, pp. 2535–2538, <https://doi.org/10.1109/EMBC.2018.8512865>.