Biomedical Signal Processing and Control 76 (2022) 103708

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework

Mohammadreza Bakhtyari, Sayeh Mirzaei *

School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
ADHD
EEG
Dynamic Connectivity Tensor
ConvLSTM
Attention mechanism

ABSTRACT

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental behavioral disorder. It is common in children, can be carried over into adulthood, and is associated with inattention, hyperactivity, and impulsive behavior. Physicians typically use the patient's description and questionnaires to diagnose this disorder. Due to its subjective nature, this procedure can lead to false diagnoses, which may cause irreparable distress to the patient's life. Since mental disorders are dependent on the brain function, researchers use biological signals such as electroencephalography (EEG) to help diagnose ADHD.

In this study, we propose a new feature extraction scheme based on evaluating dynamic connectivity tensors among EEG channels for constructing the input formulation of the classification model. The tensors contain correlations among the EEG channels over different time frames. This method allows preserving both temporal and spatial structures of the EEG data while reducing the input dimensions of the model. We then employ a neural network model consisting of a convolutional long short-term memory (ConvLSTM) and an attention mechanism to classify ADHD patients and the control group. This model can encode the spatiotemporal representation of EEG recordings and identify dependencies between temporal segments. Convolution is responsible for encoding and finding spatial dependencies between electrodes. LSTM explores the relationships between different time blocks, and finally, attention focuses on the most relevant parts of the sequence. We evaluate the proposed approach by performing experiments on the EEG dataset, including 400 instances with 30 s length collected from 46 children with ADHD and 45 children in the control group. We achieved an average accuracy of 99.34% on this dataset, and our best model has an accuracy of 99.75%, both are the highest among the work done in this field.

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a common behavioral neurodevelopmental disorder in children and adolescents and may have lifelong effects [1]. Investigations show that about 5% of children and 2.5% of adults suffer from this disorder [2]. ADHD is associated with inattention, hyperactivity, and impulsive behaviors or a combination of these symptoms and may accompany a psychological problem such as depression or bipolar disorder. These symptoms may increase the risk of criminal behavior, divorce, unemployment, learning disabilities, low social achievement, and suicide in adults. Early diagnosis of ADHD can prevent it from adversely affecting the patients' life. Diagnosis of ADHD is usually made based on the information given by the person, school teachers, parents, and questionnaires. The doctor's experience impacts this subjective diagnosis. Regarding these issues, diagnosing ADHD disorder is a challenging task, and the probability of misdiagnosis is high [3]. Given that ADHD is directly related to brain function, many studies have focused on bio-feedbacks to diagnose ADHD objectively and help the procedure.

There are generally two categories of methods for recording bio-feedbacks: invasive and non-invasive. Most studies utilize non-invasive bio-feedbacks like EEG, fNIRS, fMRI, EOG, and MEG to diagnose ADHD. For example, [4] uses the fMRI signals in the ADHD-200 dataset and achieves an accuracy of 71.3%. In another study, [5] investigated the relationship between the amplitude of the hemodynamic response and ADHD patients using fNIRS data. EEG signal recorders are inexpensive and portable, with no medical risks. In addition, the non-invasiveness and high temporal resolution of these signals make them preferable to other bio-feedbacks in the field of neuroscience [6].

EEG is a method of recording the electrical activity of the human

* Corresponding author.
E-mail addresses: mo.bakhtyari@ut.ac.ir (M. Bakhtyari), s.mirzaei@ut.ac.ir (S. Mirzaei).

https://doi.org/10.1016/j.bspc.2022.103708
Received 24 December 2021; Received in revised form 12 March 2022; Accepted 9 April 2022
Available online 18 April 2022
1746-8094/© 2022 Elsevier Ltd. All rights reserved.

brain using electrodes placed on the human scalp. Hans Berger used EEG signals on humans in 1924 for the first time [7]. These signals span a wide range of frequency spectrums. In general, there are five frequency bands for analysis of EEG signals: Delta (<4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (greater than30 Hz). Each of these bands has its unique biological characteristic. They have various applications, such as diagnosing sleep problems, cancer, tumors, and other brain-related disorders [8]. JF Lubar used EEG bio-feedback for the first time to diagnose ADHD. In his study, he observed that event-related potentials (ERPs) in ADHD patients are different from the control group. In his 15-year study, he found that patients with ADHD had increased theta activity and decreased beta power [9].

To analyze the EEG data, we first render the preprocessing step. The purpose of preprocessing is to remove internal and external artifacts from the EEG data [10]. In most studies, preprocessing is carried out manually, which requires expert knowledge. Automated methods such as Independent Component Analysis (ICA), and machine learning based approach, are among other techniques to remove these artifacts [11]. The next step is to prepare the input for the classifier model. There are different input formulations for this task. Common Spatial Patterns (SCP) [12], Discrete Wavelet Transform (DWT) [13], Lyapunov Exponent (LE) [14] and Fractal Dimension (FD) [15] are among the most common feature extraction methods. Some studies use features such as Spectrograms [16], Fast Fourier Transform maps [17], and 3D Grids [18].

It is hard to detect high-level patterns in the complex data extracted from the human brain. For years, machine learning has been developing methods for recognizing patterns with the objective to predict and classify brain data. Machine learning algorithms like Support Vector Machine (SVM) [19], K-Nearest Neighbor (KNN) [20,21], and Feedforward neural network [22] are applied to diagnose ADHD patients utilizing EEG data. As a consequence of the increasing computing power of today's processors and the high performance of neural network models, these models have been used frequently in recent years in various fields, and basic methods such as SVM and KNN are less favorable.

Neural networks are a type of machine learning algorithm that attempts to mimic the process of the human brain. Neural networks were first introduced in 1943 [23]. The computational load of the neural network distributes to cells called neurons. These cells are in a network structure consisting of three layers: input, hidden, and output, and they send the results of their computations on the data through their connections with other cells to carry out the learning process and obtain the coefficients needed to predict the class of new samples [24]. Many studies have used Multi-Layer Perceptron (MLP) to classify ADHD patients and control groups. For example, in 2016, Mohammedi et al. employed an MLP with non-linear features of EEG samples and achieved an accuracy of 93.65% [22].

Deep neural networks (DNNs) are a type of neural network known to have multiple layers in their network structure. In recent years, deep learning has been an area of interest for clinical diagnosis. Due to the end-to-end architecture and the power of deep neural networks to extract representation from the data, they do not need a separate feature extraction phase. Convolutional neural networks (CNNs) are deep neural networks with convolutional layers [24]. Due to the success of this model in the field of machine vision, researchers are interested in this type of neural network in clinical applications as well. For example, [25] first processes the EEG data and then uses the wavelet transform to acquire the spectrogram. Finally, these spectrograms are input to a CNN with four layers of filtering and pooling. The model accuracy in this paper is 88%, and it is superior to similar tasks since it does not require manual channel selection. [26] extracts a 3-channel RGB color image from different bands of EEG signal and uses this transformed input to a three-layer CNN with Max-pooling and dropout. Their model achieved an accuracy of 98.48.

This article introduces a new method for feature extraction based on computing dynamic connectivity tensors. Our input formulation uses a framing approach to segment the data into time windows to preserve the temporal dynamics of the EEG signal. The dot product between these segments across different EEG channels then extracts the spatial correlations among the electrodes. In addition to preserving the original structure of the data, these spatiotemporal features significantly reduce the input dimensions of the model. We propose a new framework consisting of two main parts for classification: convolutional long short-term memory (ConvLSTM) and attention mechanism. ConvLSTM encodes spatial features with convolutions and temporal dependencies with LSTM. We exploit the attention mechanism to give higher weight to the more crucial time frames of the EEG signal, which contain more discriminatory information for the ADHD detection task. We show that our proposed method leads to superior performance for the ADHD classification within the experiments.

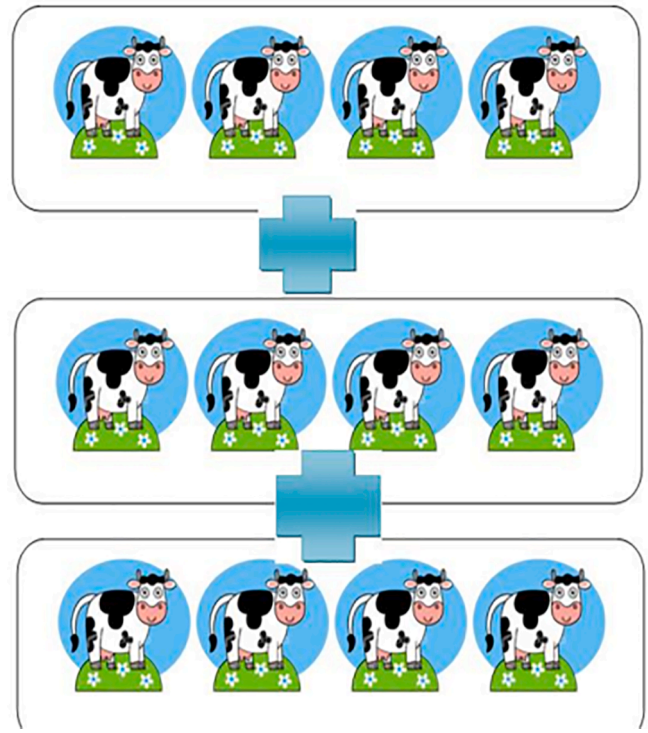

Fig. 1. An example of pictures displayed to children.

The rest of this paper is organized as follows: Section 2 introduces the ADHD dataset used in the experiment. We explain the feature extraction method in section 3. Section 4 is dedicated to the classification framework. The experiment setting and the performance results are presented in section 5. Section 6 provides a discussion about the different parameters of the model. Ultimately, conclusive remarks are presented in section 7.

2. Materials

The National Brain Mapping Organization (NBML) provided the dataset of this study [27]. The dataset includes EEG recordings of 46 children with ADHD and 45 healthy controls (boys and girls, ages 7–12). Children in the control group have no history of mental illness, disorders, or epilepsy, and there are no high-risk behaviors in the recordings of these children. The diagnosis of children with ADHD was under the DSM-IV protocol. They have taken Ritalin for up to 6 months.

Given that one of the problems for ADHD patients is performing activities related to visual stimuli, NBML provided a set of images for the children, including some animal figures or cartoon characters (randomly selected between 5 and 16). Some of these animal figures were at the top of the images, and some were at the bottom. Fig. 1 represents an

Table 1
Dataset Details.

	Number of subjects	Number of intervals	Duration	Sample rate
ADHD	46	233	30 s	512 Hz
Healthy controls	45	167	30 s	512 Hz

example of these images. They asked children to count the number of animal figures in the picture above and then count the number of animal figures at the bottom. It does not matter whether the children's answer is right or wrong. The length of the recorded EEG signals during this cognitive visual task depends on the time each child needs to respond.

Two neurologists extracted an average number of 38 ± 5 epochs from the data. As a result, the output of this procedure was 30 s instances of data. The data of 46 children with ADHD and 45 children in the control group were divided into 30-second intervals to train the model which is publicly available [28]. The data include 400 instances, 233 intervals for ADHD patients, and 167 samples for healthy controls (see Table 1). The sampling frequency of EEG recordings is 512 Hz. The children performed a cognitive task while recording the whole 30 s intervals. Electrodes were of the normal type (the electrodes were not active), and they registered all the electrodes considering the device reference.

This dataset is acquired based on the standard 10–20 with 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2).

3. Methodology

3.1. Problem definition

EEG signals are multivariate time series. In our case, this time series consists of $C = 19$ channels (number of recording sensors), and each recording contains $S = 30$ s of brain data with $F = 512$ Hz frequency. The number of samples in each EEG channel is $P = S \times F = 15360$. We use $\mathbf{x}_t \in \mathbb{R}^C$ to denote the recording signal of all channels at time t. We use $\mathbf{X}$ to indicate the raw data recorded by the EEG device in 30 s with the sampling frequency of 512 Hz. One dimension of this matrix is related to EEG channels, and the other dimension is associated with the time series recorded by these channels.

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P]^{\mathrm{T}} \tag{1}$$

There are 400 labelled instances of these multivariate time series in our dataset, 233 instances with ADHD and 167 instances in the control group. Our goal is to use this data to train a model which classifies new examples of ADHD and healthy group.

3.2. Data preprocessing

Due to the high temporal resolution of EEG signals, there is a possibility of occurring artifacts during the recording process [29]. There are internal and external artifacts in EEG data. Experimental errors and environmental noise can create external artifacts, and filtering can eliminate their effects on the data [30]. Data in the frequency range below 4 Hz may be contaminated by sweating or shaking the neck and jaw muscles. These artifacts have a significant impact on other preprocessing steps. So, we remove them utilizing a Finite Impulse Response (FIR) high-pass filter with automatic order. Besides, gamma-band frequencies are sensitive and easily affected by artifacts. By plotting the power spectral density of the dataset, we observed that the gamma-band is affected by the artifacts in the range of more than 60–70 Hz, and we also find that there is line noise around the 45–55 Hz frequency. Fig. 2 exhibits the power spectral density of an instance of the dataset. To eliminate these artifacts, we filter the dataset with a low-pass filter at a frequency of 40 Hz with the same configuration in the high-pass filter.

Internal artifacts, also known as physiological artifacts, can reduce the quality of EEG signal recording. In many studies of EEG signals, it is common to apply manual preprocessing to remove internal artifacts. The need for expert knowledge in manual preprocessing, limits EEG applications. For this reason, there are many methods for automatic preprocessing of data and different pipelines to standardize the process. We tried different automatic pipelines such as PREP [31], but they did not lead to better results. As Aforementioned, the dataset recording protocol was under a visual stimuli activity. Therefore, it is necessary to remove the eye movement and blinking artifacts. Considering that ICA is the most common way to suppress ocular artifacts [32], we utilize it to find components related to them. After ICA separates sources of EEG recordings, an automatic method is required to detect artifacts from brain sources. There are different statistical methods for flagging components automatically, but the most reliable approach is ICLabel [11], which we employ. We eliminate eye artifacts components with a probability of more than 0.9. Applying this preprocessing step leads to better results in terms of evaluation metrics.

In this study, we analyze EEG data with connectivity analysis. In this case, one of the potential sources of confusion that need attention is volume conduction [33]. If there was no volume conduction, it wasn't viable to collect electrical signals from the brain using non-invasive methods like EEG. In particular, the potential source can project electrical waves onto all of the electrodes of EEG, which permits us to have topographical map analysis and interpretation. But from another point of view, volume conduction can be an artifact source. In connectivity analysis, we examine the neural connectivity between one population of neurons and another. Ideally, we would like to record the activity of each of these populations with distinct electrodes. Regardless, there is a

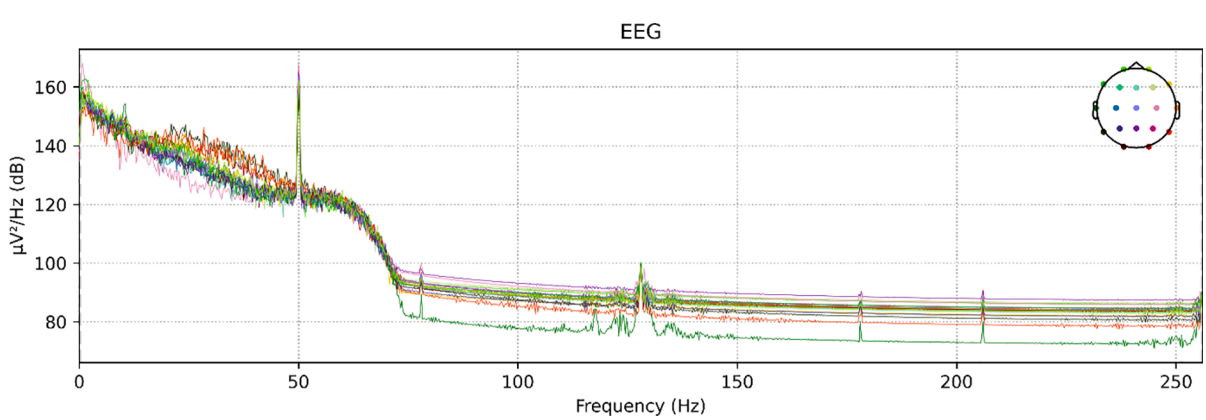

Fig. 2. Power spectral density of an instance of EEG recordings before filtering.

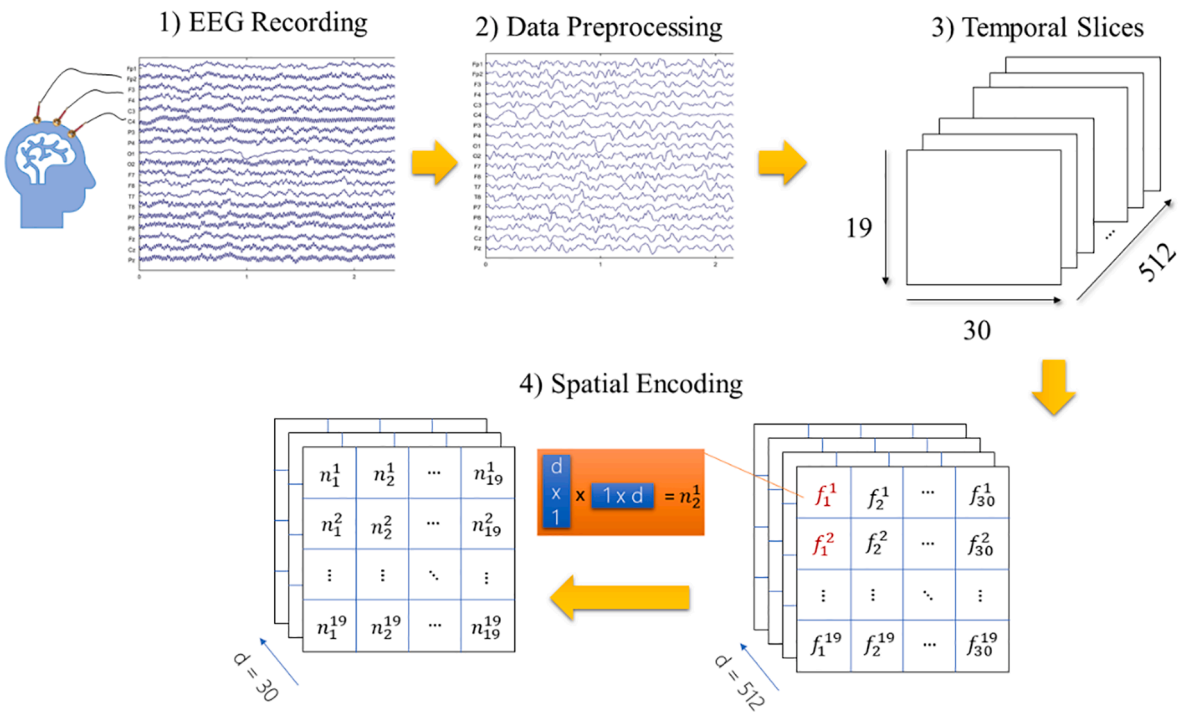

Fig. 3. The proposed feature extraction scheme.

possibility that both electrodes record a single source of potential electrics. In this circumstance, both electrodes are measuring the same thing. To avoid these issues, we need to reduce the effect of volume conduction artifacts. In this study, we employ current source density (CSD) for this purpose [34]. This method takes the spatial Laplacian of the EEG signal and reduces the volume conduction. Using CSD to mitigate volume conduction artifacts results in a more robust and general model, which enhances the average accuracy of the model.

3.3. Dynamic connectivity tensors

There are 400 instances stored in a matrix with (19 × 15360) size in the dataset consisting of 291,840 points. When we multiply this number by 400, we get 116,736,000 samples. With this large amount of training data, our model must learn millions of parameters, so using them as model input is not computationally feasible and may cause overfitting. Therefore, instead of the End-to-End approach, we consider a two-step approach with a feature extraction scheme before the classifier.

In recent years, connectivity measures developed an interest in literature and showed significant success in finding distinctive features for neuro markers in the EEG recordings of ADHD [35,36]. There are several ways to measure the connectivity and relationships between signals. The two main categories of these methods are time-domain analysis and frequency domain analysis [37]. Although frequency domain analysis is conventional, it is best to consider the full potential of the signals in the time domain before transforming them to the frequency domain. Hence, we suggest a novel feature extraction method based on the connectivity in the time domain. In addition, the frequency domain analysis follows the O(*nlogn*) time complexity of short-time Fourier Transform. While calculating the proposed temporal correlation has a time complexity of O(*n*) and imposes less computational load on the framework. Here *n* is the number of signals' samples.

Our method consists of two steps. The first step is to split each channel of an instance into 30 frames with a size of 1 s (512 samples) and 0% overlap. At this point, data structure changes from a 2D matrix of size (19 × 15360) to a 3D tensor of size (19 × 30 × 512). The advantage of this temporal slicing is that it retains continuously interconnected features over time. We analyze the effect of various parameters such as frame length and overlap on the performance in the discussion section. In the second step, we construct the spatial features by applying dot product. To do this, we compute the Pearson correlation between the temporal segments of different channels. Hence, we obtain a dynamic

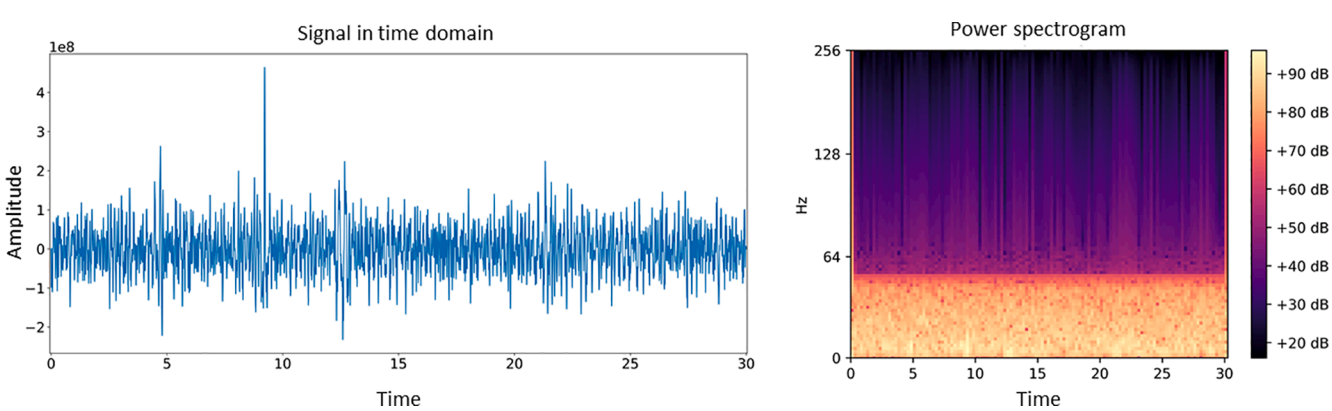

Fig. 4. Time domain and Power spectrogram of an instance of the data.

connectivity tensor (DCT) with size (30 × 19 × 19). This tensor is a compact representation of the spatial and temporal structure of the EEG instance and can be regarded as a 30 s video with (19 × 19) pixel images.

Performing the aforementioned steps reduces the size of the input features significantly compared to the raw data. The proposed feature extraction scheme has some key advantages. First, it decreases the number of learning parameters of the model to a great extent, thus reducing the computational burden and time consumption of the algorithm. This feature size reduction can avoid overfitting issues as well. Besides, it preserves the spatiotemporal structure of the EEG data in the acquired DCT. Fig. 3 provides an overview of the proposed feature extraction technique.

In addition to the method we presented, we apply two other input formulations to provide comparative material for the model. To compare the DCT results with frequency-domain techniques, here we use Welch's coherence [38] method to measure connectivity between different signal channels. To make the results comparable, using the same parameters of DCT, we convert the raw data into time segments and then calculate the coherence between different channels in each temporal portion. For another input formulation, we compute spectrograms of raw data. It is a popular approach for representing EEG data input in machine learning models but has received little attention in the ADHD detection field. We use the Mel spectrogram to examine the effectiveness of our proposed feature extraction method. The Mel spectrogram is obtained by calculating discrete Fourier transform over short time frames of data [40]. Then, it is mapped onto the Mel scale. We apply an FFT window with 1024 samples, and these windows have a hop length of 512. Fig. 4 displays the power spectrogram of an instance of the EEG data with its time domain signal.

4. Models

This section introduces a deep neural network architecture containing convolution, long short-term memory (LSTM), and an attention mechanism. In the proposed model, the convolution layer encodes the spatial features among channels. The LSTM layer encodes time dependencies, and an attention layer learns the importance of the dependencies among the temporal segments with another neural network. Finally, data goes through a dense layer for classification. The following subsections provide a detailed description of each of these layers.

4.1. ConvLSTM

Recurrent neural networks (RNNs) are a type of network designed to learn temporal dynamic behaviours [41]. Gradient vanishing is the main issue of vanilla RNNs. The idea of LSTMs was proposed in 1997 to mitigate recurrent neural network problems [42]. Although LSTMs can learn the temporal features of time series and encode long-term dependencies, they are not able to represent spatial properties of time series such as EEG signals with topological information. This weakness of LSTMs leads us to find a better architecture that can handle spatiotemporal data.

ConvLSTM was proposed by Xingjian Shi et al. [43]. This neural network exploits convolutional layers instead of fully connected layers in LSTM. We leverage the ConvLSTM model for learning the spatiotemporal input structure. Inspired by this network, we will build a different ConvLSTM that does not use cells output for the feedback loop. The formula used in this network for learning is as follows. In this formula, * denotes convolution operation and ∘ indicates Hadamard product:

$$\mathbf{I}_t = \boldsymbol{hardsigmoid}(\mathbf{K}_{xi}*\mathbf{X}_t + \mathbf{b}_i + \mathbf{K}_{hi}*\mathbf{H}_{t-1}) \tag{2}$$

$$\mathbf{F}_t = \boldsymbol{hardsigmoid}\left(\mathbf{K}_{xf}*\mathbf{X}_t + \mathbf{b}_f + \mathbf{K}_{hf}*\mathbf{H}_{t-1}\right)$$

$$\mathbf{C}_t = \mathbf{F}_t \circ \mathbf{C}_{t-1} + \mathbf{I}_t \circ \boldsymbol{tanh}(\mathbf{K}_{xc}*\mathbf{X}_t + \mathbf{b}_c + \mathbf{K}_{hc}*\mathbf{H}_{t-1})$$

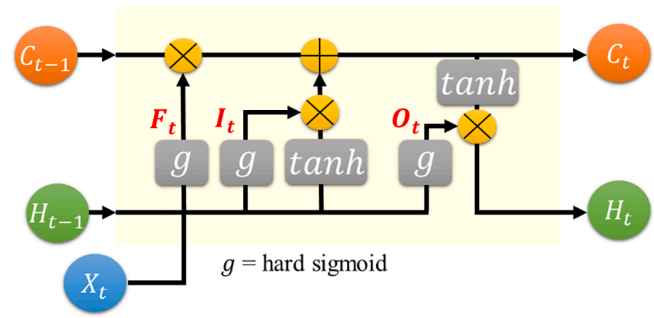

Fig. 5. Internal structure of a ConvLSTM cell.

$$\mathbf{O}_t = \boldsymbol{hardsigmoid}(\mathbf{K}_{xo}*\mathbf{X}_t + \mathbf{b}_o + \mathbf{K}_{ho}*\mathbf{H}_{t-1})$$

$$\mathbf{H}_t = \mathbf{O}_t \circ \boldsymbol{tanh}(\mathbf{C}_t)$$

$\mathbf{I}_t$, $\mathbf{F}_t$ and $\mathbf{O}_t$ are designed to control cell information memoizing. If $\mathbf{I}_t$ is activated, it adds the information that comes in as input, to the cell information. Enabling forget gate $\mathbf{F}_t$, lead to forgetting information entered through the previous cell $\mathbf{C}_{t-1}$, and gate $\mathbf{O}_t$ will control whether $\mathbf{C}_t$ transmits to state $\mathbf{H}_t$ or not. An important property of this implementation is that all inputs $\mathbf{X}_1, \cdots, \mathbf{X}_S$, cell outputs $\mathbf{C}_1, \cdots, \mathbf{C}_S$, hidden states $\mathbf{H}_1, \cdots, \mathbf{H}_S$, and the gates $\mathbf{I}_t, \mathbf{F}_t, \mathbf{O}_t$ are 3D tensors with the last two dimensions related to the spatial feature maps. $\mathbf{K}_{xi}, \mathbf{K}_{xf}, \mathbf{K}_{xc}, \mathbf{K}_{xo}$ are kernel weights of the model, and $\mathbf{K}_{hi}, \mathbf{K}_{hf}, \mathbf{K}_{ho}, \mathbf{K}_{hc}$ with subscript h corresponds to recurrent kernel weights. We initialized recurrent weights with orthogonal initializer, and the kernel initializer is glorot uniform. $\mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_c, \mathbf{b}_o$ are bias vectors initialized with zeros. Fig. 5 shows the internal structure of a ConvLSTM. As far as we know, this is the first time that this architecture has been applied to EEG data for ADHD detection. We benefit from the ConvLSTM framework to learn joint spatial and temporal representations.

4.2. Attention mechanism

Reference [44] introduced the initial idea of attention for translation machines. Attention models focus on long-term dependencies. This idea simulates the relationships between different points in time, and the model can focus on crucial portions of time. In this study, we implement the method presented in [45]. To do this, for each of the hidden states $\mathbf{H}_1, \cdots, \mathbf{H}_S$ generated by the ConvLSTM model, we derive a context vector $\mathbf{c}$ as follows.

$$\mathbf{c} = \sum_{t=1}^{S} \alpha_t \mathbf{H}_t \tag{3}$$

$\mathbf{c}$ is a weighted sum of $\mathbf{H}_1, \cdots, \mathbf{H}_S$. α_t is the weight calculated for t^{th} state $\mathbf{H}_t$, and S is the number of time steps in the time series. To obtain the coefficient α_t corresponding to each $\mathbf{H}_t$, we do the following:

$$e_t = tanh(\mathbf{H}_t) \tag{4}$$

$$\alpha_t = \frac{\boldsymbol{exp}(e_t)}{\sum_{k=1}^{S} \boldsymbol{exp}(e_k)}$$

In this formula, e_t is a scalar that specifies the importance of $\mathbf{H}_t$. e_t is acquired as the output of a single layer feedforward neural network with *tanh* activation function taking $\mathbf{H}_t$ as the network input. Eventually, we give these values to a softmax function to obtain α_t. The weight and bias parameters of the network are updated such as other layers' parameters within the learning process. Hence, we add the Attention mechanism to our classification model by computing the context vectors. Fig. 6 outlines an overview of our proposed classification model, and the parameters of the model are listed in Table 2.

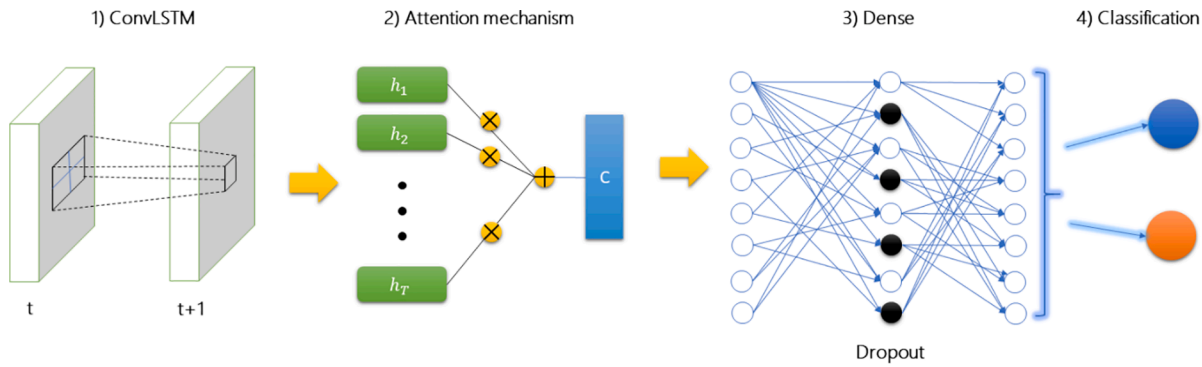

Fig. 6. Proposed ConvLSTM with Attention mechanism.

Table 2
Details of the model parameters and implementation.

Layer	Output shape	Number of parameters
Input Layer	(None, 30, 19, 19, 1)	0
ConvLSTM2D	(None, 30, 19, 19, 60)	132,000
Reshape	(None, 30, 21660)	0
Attention	(None, 21660)	21,690
Dense	(None, 128)	2,772,608
Dropout	(None, 128)	0
Dense	(None, 2)	258
Trainable params	2,926,556	

5. Experiments and results

In this section, we evaluate the proposed ADHD classification framework by performing experiments on the ADHD dataset described in Section 2. We compute four evaluation metrics:

- $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$
- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F_1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$

In these formulas, TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively.

One of the best ways to evaluate the model is to use the K-fold cross-validation method. Towards this, we divide data into K folds, one of these folds is test data, and the other $K-1$ folds are training data. This procedure is performed individually for each fold to make the results more realistic and generalized. In this way, we evaluate the model on all data folds. We divide the data into 5 folds to make the results comparable to the previous studies. We consider one of these folds containing 80 instances as test data and the other 320 instances as training data. In deep learning models, the initialization of some parameters is random. Regarding this, we run the model 10 times to acquire more robust results.

To investigate the impact of different model parameters, we first specify the parameters of the default model. We set the batch size to 16 and epochs to 100 for fitting the data to the model. In the first layer, the number of filters for ConvLSTM is 60 with a kernel size of (3×3) and padding = same. After the data passes through the ConvLSTM layer, the attention layer attempts to extract the score for each hidden state of the ConvLSTM. After these layers, we apply a dense layer with 128 neurons. We utilize the dropout technique to avoid overfitting and create different random architecture. The keep probability of the dropout is 0.5. Finally, we calculate the cost of our model with cross-entropy as a loss function.

First, we demonstrate the superiority of the results when using CSD.

Table 3
Evaluation metrics for 10 iterations without CSD.

	Average accuracy	STD accuracy	Average loss	STD loss
Iteration 1	0.9800	0.0061	0.1006	0.0243
Iteration 2	0.9825	0.0127	0.1082	0.0563
Iteration 3	0.9800	0.0100	0.0984	0.0364
Iteration 4	0.9825	0.0187	0.1089	0.0530
Iteration 5	0.9900	0.0049	0.0643	0.0214
Iteration 6	0.9900	0.0049	0.0749	0.0151
Iteration 7	0.9800	0.0061	0.0816	0.0286
Iteration 8	0.9900	0.0049	0.0620	0.0226
Iteration 9	0.9774	0.0145	0.0920	0.0208
Iteration 10	**0.9975**	**0.0049**	**0.0432**	**0.0097**
Average	0.9850	0.0088	0.0834	0.0288
STD	0.0061	0.0046	0.0209	0.0145

Table 4
Evaluation metrics for 10 iterations with CSD.

	Accuracy	STD accuracy	Loss	STD loss
Iteration 1	0.9900	0.0049	0.0836	0.0288
Iteration 2	**0.9975**	**0.0049**	**0.0528**	**0.0411**
Iteration 3	0.9875	0.0079	0.0860	0.0602
Iteration 4	**0.9975**	**0.0049**	**0.0651**	**0.0430**
Iteration 5	0.9875	0.0136	0.0872	0.0659
Iteration 6	**0.9975**	**0.0049**	**0.0728**	**0.0580**
Iteration 7	0.9925	0.0099	0.0837	0.0462
Iteration 8	0.9950	0.0061	0.0735	0.0510
Iteration 9	0.9950	0.0099	0.0707	0.0535
Iteration 10	0.9950	0.0061	0.0643	0.0397
Average	0.9935	0.0073	0.0740	0.0487
STD	0.0037	0.0028	0.0106	0.0105

Table 3 reports the resulted evaluation measures for the dataset without applying CSD. Each row shows the average and standard deviation of accuracy and loss over 5-folds per iteration. We find that our model achieves an average accuracy of 98.50% over ten executions of the algorithm, and the best model in iteration 10 has an average accuracy of 99.75%. Table 4 is the same experiment with CSD preprocessing to avoid volume conduction effects. By comparing Tables 3 and 4, we can see that when using CSD instead of the raw EEG signals, the average accuracy of the model in 10 iterations increases from 98.50% to 99.35%, but the highest accuracy remains 99.75%. This result demonstrates that the model provides better predictive performance and generalization power. Table 5 lists the details of the results we achieved for the metrics of each fold in the best model, and Fig. 7 illustrates the average loss and accuracy of 5-fold cross-validation for the best results.

5.1. Model Comparison

We provide a comparison with other classification frameworks to study the validity and effectiveness of the proposed model. The first

Table 5
Evaluation metrics for the best model.

	Class	Support	Precision	Recall	F1-score	Accuracy	Loss
Fold 1	Healthy	30	1.0	1.0	1.0	1.0	0.0486
	ADHD	50	1.0	1.0	1.0		
	Weighted Avg	80	1.0	1.0	1.0		
Fold 2	Healthy	37	1.0	1.0	1.0	1.0	0.0441
	ADHD	43	1.0	1.0	1.0		
	Weighted Avg	80	1.0	1.0	1.0		
Fold 3	Healthy	33	1.0	1.0	1.0	1.0	0.0436
	ADHD	47	1.0	1.0	1.0		
	Weighted Avg	80	1.0	1.0	1.0		
Fold 4	Healthy	36	1.0	1.0	1.0	1.0	0.0253
	ADHD	44	1.0	1.0	1.0		
	Weighted Avg	80	1.0	1.0	1.0		
Fold 5	Healthy	31	1.0	0.9677	0.9836	0.9875	0.0546
	ADHD	49	0.9800	1.0	0.9898		
	Weighted Avg	80	0.9877	0.9874	0.9873		
Average			0.9975	0.9974	0.9974	**0.9975**	**0.0432**
STD						**0.0049**	**0.0097**

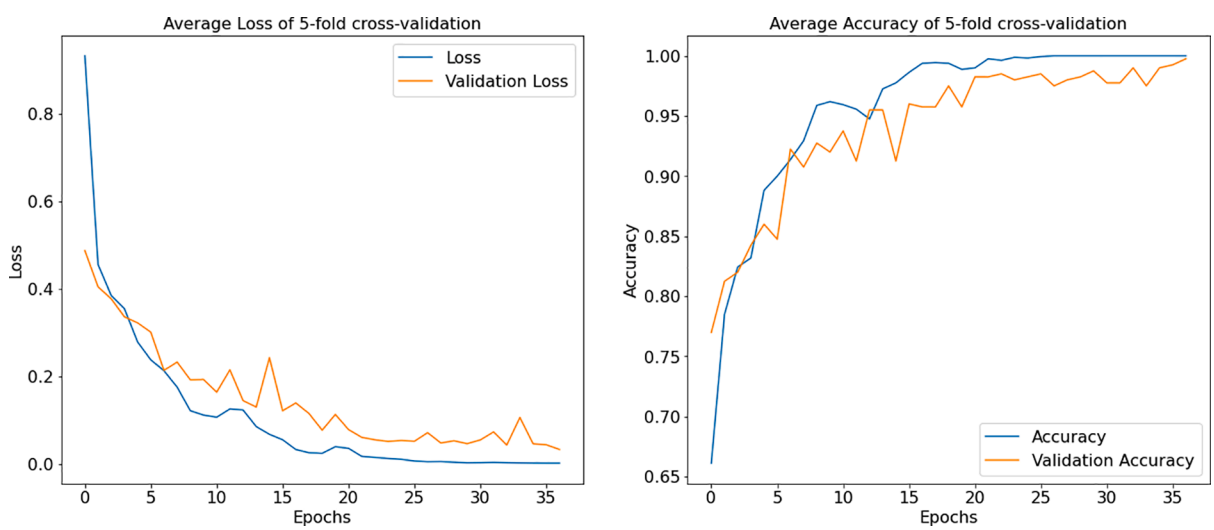

Fig. 7. Average loss (left) and average accuracy (right) of 5-fold cross-validation for the best iteration.

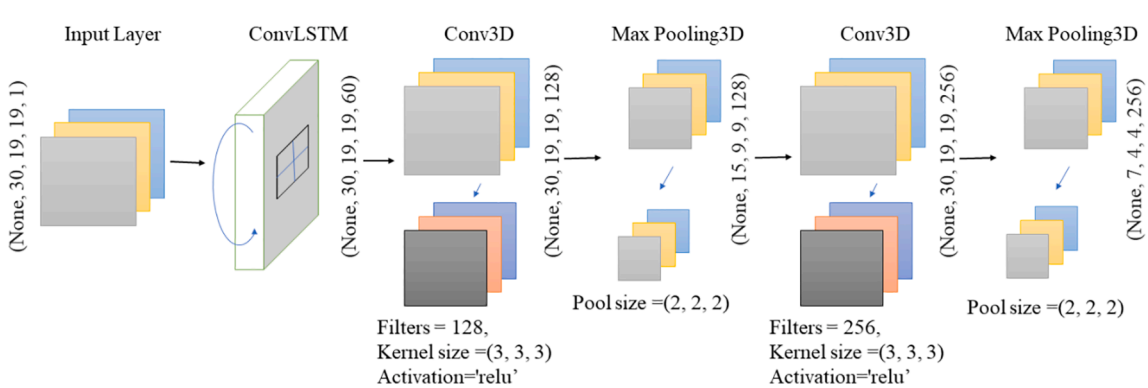

Fig. 8. ConvLSTM with two Conv3D and MaxPooling3D.

Table 6
Comparison of ConvLSTM with Attention, ConvLSTM and ConvLSTM with Conv3D models.

	Average on 5 folds				
	Accuracy	Loss	Precision	Recall	F1-score
ConvLSTM with Attention	**0.9975**	**0.0432**	**0.9975**	**0.9974**	**0.9974**
ConvLSTM	0.9024	0.4469	0.9145	0.9024	0.8994
ConvLSTM with Conv3D	0.9450	0.2533	0.9459	0.9449	0.9445

Table 7
Comparison between coherence and the proposed input formulation.

	Average on 5 folds				
	Accuracy	Loss	Precision	Recall	F1-score
DCT	**0.9975**	**0.0432**	**0.9975**	**0.9974**	**0.9974**
Coherence	0.9775	0.1577	0.9777	0.9773	0.9774

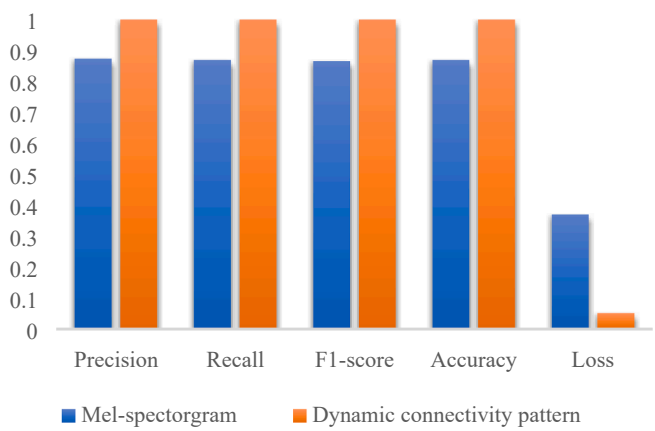

Fig. 9. .Comparison of Mel-spectrogram and Dynamic connectivity pattern applied as input features of ConvLSTM with Attention model.

framework is similar to the ConvLSTM with attention mechanism, and the only difference is that we remove the attention layer to examine how it would affect the results. We replace the attention mechanism with two Conv3D and a Max Pooling layer for creating another model. Fig. 8 depicts the details of this model implementation.

Table 6 shows the results of comparing three models: ConvLSTM with Attention, ConvLSTM and ConvLSTM with Conv3D. We find that the ConvLSTM with Attention model outperforms other models in all metrics. According to the results, it is clear that the attention module plays an essential role in the proposed model. The ConvLSTM without an attention module is about 9% less accurate, and the ConvLSTM module with a 3D convolution drops 5% accuracy. The performance improvement caused by incorporating the attention mechanism can be associated with the fact that exploiting attention enables us to give higher weight to more relevant temporal segments in EEG data recorded from ADHD patients. Therefore, it is advisable to focus more on some of the time portions of data to provide more discriminatory features for the ADHD classification task.

5.2. Input formulation comparison

To inspect the effectiveness of the proposed feature extraction technique, we compare the performance of the ConvLSTM with Attention model with the three different input formulations introduced in Section 3. First, we compare the time-domain versus coherence feature

Table 8
Comparison with other articles on the same dataset.

Study	Year	Method	Classifier	Accuracy
Allahverdy et al. [14]	2011	Lyapunov exponent Katz fractal dimension Sevcik fractal dimension Higuchi fractal dimension Symbolic dynamic	MLP	0.86
Allahverdy et al. [15]	2016	Lyapunov exponent Katz fractal dimension Higuchi fractal dimension Sevcik fractal dimension	MLP	0.967
Mohammadi et al. [22]	2016	Approximate entropy Largest Lyapunov Exponent Katz fractal dimension Higuchi fractal dimension Petrosian Fractal Dimension	MLP	0.9365
Moghaddari et al. [26]	2020	Frequency band separation Making RGB images	CNN	0.9848
Our proposed approach	**2021**	**Dynamic connectivity tensor**	**ConvLSTM with Attention**	**0.9975**

extraction methods. Thus, we investigate the coherence and DCT results. Table 7 shows the performance of these methods within the experiment. For the coherence method, the size of the FFT window is 32. 1The result is a 4-dimensional tensor with (30 × 19 × 19 × 16) dimensions, and these dimensions are equal to (30 × 19 × 19 × 1) for the proposed spatiotemporal input formulation.

As expected, the frequency-domain connectivity input for the proposed classification model has less accuracy than temporal correlation inputs. The reason for the superiority of the results in time-domain input is that the proposed model is specifically designed and has a better ability to exploit temporal information.

We want to compare the time-domain connectivity measure with another popular feature extraction technique in the literature, Mel-spectrogram. Fig. 9 represents five performance metrics for the Mel-spectrogram and the dynamic connectivity tensors. The input size of the Mel spectrogram feature is equal to (31 × 19 × 32 × 1). In this experiment, the FFT windows contain 1024 samples with 50% overlap. It is observable that the proposed input formulation is superior to the features of the Mel spectrogram in terms of all measures. The results reveal the advantage of the proposed dynamic connectivity tensor features for ADHD classification. The DCT contains joint temporal and spatial information of the EEG data. Therefore, it improves evaluation metrics compared to Mel-spectrogram features. In addition, the number of learning parameters of Mel Spectrum input is more than 4,800,000, which decreases the learning procedure speed, and complicates the model.

For further validation, we need to utilize statistical tests to show that the better performance of the proposed DCT over the Mel- spectrogram is not due to random factors of the model. The paired student *t*-test is a common way to ensure that the superior results of the DCT input are reliable. However, for K-fold cross-validation, the paired *t*-test is unreliable because it assumes that the data sampled independently. To solve this problem, we use [46] as it deals with this issue by taking the dependency into account. Here no difference between the model performance corresponds to the null hypothesis. If the P-value is less than the considered significant level, we reject the null hypothesis, and the alternative hypothesis holds. In this case, the performance of the models is different. We repeat the training process ten times for both feature extraction methods. The P-value for accuracy is 2.36e-15. This value for precision, recall, and f1-score is 1.56e-14, 2.36e-15, and 2.46e-15, respectively. These P-values reject the null hypothesis, and the

Table 9
Accuracy and loss of different frequency bands.

	Average on 5 folds	
	Accuracy	Loss
Theta (4–7 Hz)	0.9374	0.1998
Alpha (8–15 Hz)	0.9724	0.1346
Beta (16–30 Hz)	0.9825	0.0828
Gamma (31–40 Hz)	0.9725	0.1444
4 – 40 Hz	**0.9975**	**0.0432**

superior performance of the DCT method is approved.

5.3. Related works

In this part, we compare the performance of the proposed method with other studies on the same dataset. Table 8 shows the accuracy values reported by the various studies. Reference [14] divides the dataset into train and test sets. Similarly, [15] and [22] have used this method to obtain results. However, this is not a reliable approach, and the results are prone to overfitting. Reference [26], the latest study on the dataset, employs 5-fold cross-validation, and for this reason, we also utilized this method to have comparable results. It is observed that our proposed feature extraction and classification framework results in better classification accuracy.

6. Discussion

This section analyses the impact of principal model parameters on the performance. We set the default hyperparameters corresponding to the best model we aforesaid. First, let's examine the effect of considering different frequency bands. We do not consider the delta frequency band due to the high probability of being noisy with sweating artifacts [47]. We run the experiment on the other four frequency bands, i.e., theta, alpha, beta, and gamma. As we mentioned in section 3.2, there are artifacts in the dataset at frequencies above 40 Hz. Therefore, it is better to consider the low-gamma band. Table 9 displays the results of building the DCT with different frequency bands. It is manifested that considering the individual frequency bands does not improve the results. Extracting the dynamic connectivity tensors from the raw data containing 4–40 Hz frequency bands achieved the best performance.

The next step is to analyse the impact of the window size and the amount of overlap in temporal segmentation on the performance. As the window size decreases or the overlap increases, the number of training parameters increases. In addition, the attention mechanism will be more complex. Table 10 represents the results of this analysis. Choosing a window size of 256 leads to more computational and time cost of the model and does not improve the accuracy. We checked the window length 1024 with different overlap parameters as well, and as shown in Table 10, the best performance is obtained for the window length of 512 (1 s) without overlap. Batch size is another parameter that potentially can have a significant impact on the results. We tested three values of 8, 16, and 32 to find the optimal batch size. Table 11 reports the metrics achieved for different batch sizes. We can deduce that setting the batch size to 16, results in better performance. We examine three different values to determine the effect of filter size and the number of dense layer neurons. According to Table 11, selecting a filter size of 60 and setting the fully connected layer size to 128 leads to the best results.

7. Conclusion

This study sought to provide an objective method to support the diagnostic process for ADHD. As mentioned earlier in section 3.3, the two-step method is more applicable and attracts more attention than the end-to-end approach. There are several methods for the feature extraction step. Using Fourier transform signals from the time domain to the frequency domain is of great interest, and domain conversion seems to be a common practice in the literature. However, conversion to the frequency domain puts an additional computational burden on the system and may reduce the efficiency of the methods. We believe that the potential of signals in the time domain has not yet been fully investigated, and we seek to fill this gap with our contributions. To achieve this, we first introduce a novel feature extraction technique that derives the dynamic connectivity tensors containing a joint spatial and temporal structure of the potential electrics. This approach preserves the essential features of the signals and significantly reduces the input dimension.

We then employ a compatible model with the DCT input based on ConvLSTM and attention mechanism to classify the data. This model exploits the extracted spatiotemporal features to achieve superior accuracy for discriminating between ADHD patients and the healthy group. We concluded within the experiments that the ConvLSTM model effectively learns discriminant representations for ADHD classification tasks. Although the ConvLSTM model leads to satisfactory results utilizing the proposed dynamic connectivity tensors as input features, the participants in the experiments might render more informative and relevant brain activity at specific time frames. Consequently, we decided to pay more attention to these time frames. We conclude that by including an attention mechanism in the model, we can improve classification performance by giving higher weight to the more crucial temporal segments. The combination of this model with the input based on correlation in the time domain shows satisfactory results in the experiments.

Table 11
Accuracy and loss for different batch size, dense layer size and filter size parameters.

Average on 5 folds			
		Accuracy	Loss
Batch size	8	0.9850	0.0802
	16	**0.9975**	**0.0432**
	32	0.9850	0.0915
Dense layer size	64	0.9750	0.0998
	128	**0.9975**	**0.0432**
	256	0.9825	0.0927
Filter size	30	0.9750	0.1002
	60	**0.9975**	**0.0432**
	120	0.9925	0.0591

Table 10
Accuracy and loss for different window size and overlap parameters.

Average on 5 folds							
		Accuracy			Loss		
Window size		256	512	1024	256	512	1024
Overlap	0	0.9799	**0.9975**	0.9875	0.1242	**0.0432**	0.0875
	25%	0.9850	0.9850	0.9800	0.0859	0.0911	0.0726
	50%	0.8700	0.8875	0.9174	0.3801	0.3485	0.3134

CRediT authorship contribution statement

Mohammadreza Bakhtyari: Investigation, Software, Writing – original draft. **Sayeh Mirzaei:** Conceptualization, Methodology, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R.C. Kessler, L. Adler, M. Ames, O. Demler, S. Faraone, E. Hiripi, M.J. Howes, R. Jin, K. Secnik, T. Spencer, T.B. Ustun, E.E. Walters, The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population, Psychol. Med. 35 (2) (2005) 245–256.

[2] D. Demontis, R.K. Walters, J. Martin, M. Mattheisen, T.D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, C. Churchhouse, A. Dumont, N. Eriksson, M. Gandal, J.I. Goldstein, K. L. Grasby, J. Grove, O.O. Gudmundsson, C.S. Hansen, M.E. Hauberg, M. V. Hollegaard, D.P. Howrigan, H. Huang, J.B. Maller, A.R. Martin, N.G. Martin, J. Moran, J. Pallesen, D.S. Palmer, C.B. Pedersen, M.G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, E.B. Robinson, F.K. Satterstrom, H. Stefansson, C. Stevens, P. Turley, G.B. Walters, H. Won, M.J. Wright, O.A. Andreassen, P. Asherson, C. L. Burton, D.I. Boomsma, B. Cormand, S. Dalsgaard, B. Franke, J. Gelernter, D. Geschwind, H. Hakonarson, J. Haavik, H.R. Kranzler, J. Kuntsi, K. Langley, K.-P. Lesch, C. Middeldorp, A. Reif, L.A. Rohde, P. Roussos, R. Schachar, P. Sklar, E.J. S. Sonuga-Barke, P.F. Sullivan, A. Thapar, J.Y. Tung, I.D. Waldman, S.E. Medland, K. Stefansson, M. Nordentoft, D.M. Hougaard, T. Werge, O. Mors, P.B. Mortensen, M.J. Daly, S.V. Faraone, A.D. Børglum, B.M. Neale, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder, Nat. Genet. 51 (1) (2019) 63–75.

[3] F. Edition, Diagnostic and statistical manual of mental disorders, Am. Psychiatric. Assoc. 21 (2013), https://doi.org/10.1176/appi.books.9780890425596.

[4] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, N. Xiong, Spatio-temporal deep learning method for adhd fmri classification, Inf. Sci. 499 (2019) 1–11.

[5] Y. Monden, I. Dan, M. Nagashima, H. Dan, M. Uga, T. Ikeda, D. Tsuzuki, Y. Kyutoku, Y. Gunji, D. Hirano, T. Taniguchi, H. Shimoizumi, E. Watanabe, T. Yamagata, Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS, NeuroImage: Clinical 9 (2015) 1–12.

[6] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural Eng. 16 (3) (2019), 031001, https://doi.org/10.1088/1741-2552/ab0ab5.

[7] D.L. Schomer, F.L.D. Silva, Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields, Lippincott Williams & Wilkins (2012), https://doi.org/10.1093/med/9780190228484.001.0001.

[8] Hu, Li, and Zhiguo Zhang, eds. EEG Signal Processing and Feature Extraction. Springer Singapore, 2019. https://doi.org/10.1007/978-981-13-9113-2.

[9] J.F. Lubar, Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders, Biofeedback Self-regulation 16 (3) (1991) 201–225, https://doi.org/10.1007/bf01000016.

[10] X. Jiang, G.-B. Bian, Z. Tian, Removal of artifacts from EEG signals: a review, Sensors 19 (5) (2019) 987.

[11] L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, ICLabel: an automated electroencephalographic independent component classifier, dataset, and website, NeuroImage 198 (2019) 181–197, https://doi.org/10.1016/j.neuroimage.2019.05.026.

[12] Hamadicharef, Brahim, Haihong Zhang, Cuntai Guan, Chuanchu Wang, Kok Soon Phua, Keng Peng Tee, Kai Keng Ang. Learning EEG-based spectral-spatial patterns for attention level measurement. In 2009 IEEE International Symposium on Circuits and Systems, pp. 1465-1468. IEEE, 2009. https://doi.org/10.1109/ISCAS.2009.5118043.

[13] R.C. Joy, S.T. George, A.A. Rajan, M.S.P. Subathra, Detection of Attention Deficit Hyperactivity Disorder from EEG Signal using Discrete Wavelet Transform, in: In 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 2019, pp. 1–5, https://doi.org/10.1109/ICCUBEA47591.2019.9128977.

[14] A. Allahverdy, A.M. Nasrabadi, M.R. Mohammadi, Detecting ADHD children using symbolic dynamic of nonlinear features of EEG, in: In 2011 19th Iranian Conference on Electrical Engineering, 2011, pp. 1–4.

[15] A. Allahverdy, A.K. Moghadam, M.R. Mohammadi, A.M. Nasrabadi, Detecting adhd children using the attention continuity as nonlinear feature of eeg, Front. Biomed. Technol. 3 (1–2) (2016) 28–33.

[16] B. Mandhouj, M.A. Cherni, M. Sayadi, An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis, Analog Integr. Circ. Sig. Process 108 (1) (2021) 101–110.

[17] K. Polat, S. Güneş, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, Appl. Math. Comput. 187 (2) (2007) 1017–1026, https://doi.org/10.1016/j.amc.2006.09.022.

[18] X. Wei, L. Zhou, Z. Chen, L. Zhang, Y. Zhou, Automatic seizure detection using three-dimensional CNN based on multi-channel EEG, BMC Med. Inf. Decision Making 18 (5) (2018) 71–80.

[19] J. Anuradha, Varun Ramachandran, K.V. Arulalan, B.K. Tripathy, Diagnosis of ADHD using SVM algorithm, in: In Proceedings of the Third Annual ACM Bangalore Conference, 2010, pp. 1–4, https://doi.org/10.1145/1754288.1754317.

[20] S. Kaur, S. Singh, P. Arun, D. Kaur, M. Bajaj, Phase space reconstruction of EEG signals for classification of ADHD and control adults, Clin. EEG Neurosci. 51 (2) (2020) 102–113.

[21] H.T. Tor, C.P. Ooi, N.SJ. Lim-Ashworth, J.K.E. Wei, V. Jahmunah, S.L. Oh, U. R. Acharya, D.S.S. Fung, Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals, Comput. Methods Programs Biomed. 200 (2021) 105941.

[22] M.R. Mohammadi, A. Khaleghi, A.M. Nasrabadi, S. Rafieivand, M. Begol, H. Zarafshan, EEG classification of ADHD and normal children using non-linear features and neural network, Biomed. Eng. Lett. 6 (2) (2016) 66–73.

[23] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (4) (1943) 115–133, https://doi.org/10.1007/BF02478259.

[24] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep Learning, Vol. 1, no. 2, MIT press, Cambridge, 2016.

[25] L. Dubreuil-Vall, G. Ruffini, J.A. Camprodon, Deep learning convolutional neural networks discriminate adult adhd from healthy individuals on the basis of event-related spectral eeg, Front. Neurosci. 14 (2020) 251, https://doi.org/10.3389/fnins.2020.00251.

[26] M. Moghaddari, M.Z. Lighvan, S. Danishvar, Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG, Comput. Methods Programs Biomed. 197 (2020) 105738.

[27] Ali Motie Nasrabadi, Armin Allahverdy, Mehdi Samavati, Mohammad Reza Mohammadi. "EEG data for ADHD/Control children." https://dx.doi.org/10.21227/rzfh-zn36.

[28] "EEG Competition Data." National Brain Mapping Laboratory of Iran https://nbml.ir/FA/scientific-tournament/First-Iranian-EEG-competition.

[29] T. Radüntz, J. Scouten, O. Hochmuth, B. Meffert, Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features, J. Neural Eng. 14 (4) (2017), 046004, https://doi.org/10.1088/1741-2552/aa69d1.

[30] K.T. Sweeney, T.E. Ward, S.F. McLoone, Artifact removal in physiological signals—Practices and possibilities, IEEE Trans. Inf Technol. Biomed. 16 (3) (2012) 488–500, https://doi.org/10.1109/TITB.2012.2188536.

[31] N. Bigdely-Shamlo, T. Mullen, C. Kothe, S.u. Kyung-Min, K.A. Robbins, The PREP pipeline: standardized preprocessing for large-scale EEG analysis, Front. Neuroinformatics 9 (2015) 16, https://doi.org/10.3389/fninf.2015.00016.

[32] O. Dimigen, Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments, NeuroImage 207 (2020), 116117, https://doi.org/10.1016/j.neuroimage.2019.116117.

[33] S. Haufe, V.V. Nikulin, K.-R. Müller, G. Nolte, A critical assessment of connectivity measures for EEG data: a simulation study, Neuroimage 64 (2013) 120–133, https://doi.org/10.1016/j.neuroimage.2012.09.036.

[34] J. Kayser, C.E. Tenke, On the benefits of using surface Laplacian (current source density) methodology in electrophysiology, Int. J. Psychophysiol. 97 (3) (2015) 171.

[35] H. Kiiski, L.M. Rueda-Delgado, M. Bennett, R. Knight, L. Rai, D. Roddy, K. Grogan, J. Bramham, C. Kelly, R. Whelan, Functional EEG connectivity is a neuromarker for adult attention deficit hyperactivity disorder symptoms, Clin. Neurophysiol. 131 (1) (2020) 330–342, https://doi.org/10.1016/j.clinph.2019.08.010.

[36] S. Furlong, J.R. Cohen, J. Hopfinger, J. Snyder, M.M. Robertson, M.A. Sheridan, Resting-state EEG connectivity in young children with ADHD, J. Clin. Child Adolesc. Psychol. 50 (6) (2021) 746–762, https://doi.org/10.1080/15374416.2020.1796680.

[37] B. He, L. Astolfi, P.A. Valdes-Sosa, D. Marinazzo, S.O. Palva, C.-G. Benar, C. M. Michel, T. Koenig, Electrophysiological brain connectivity: theory and implementation, IEEE Trans. Biomed. Eng. 66 (7) (2019) 2115–2137.

[38] P. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoustics 15 (2) (1967) 70–73, https://doi.org/10.1109/TAU.1967.1161901.

[40] H.J. Nussbaumer, The fast Fourier transform, in: Fast Fourier Transform and Convolution Algorithms, Springer, Berlin, Heidelberg, 1981, pp. 80–111, https://doi.org/10.1007/978-3-662-00551-4_4.

[41] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, https://doi.org/10.1038/323533a0.

[42] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[43] Shi, Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting, arXiv preprint arXiv:1506.04214 (2015). https://arxiv.org/abs/1506.04214.

[44] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014). https://arxiv.org/abs/1409.0473.

[45] Colin Raffel, Daniel PW Ellis. "Feed-forward networks with attention can solve some long-term memory problems." arXiv preprint arXiv:1512.08756 (2015). https://arxiv.org/abs/1512.08756.
[46] C. Nadeau, Y. Bengio, Inference for the generalization error, Adv. Neural Inf. Process. Syst. 12 (1999), https://doi.org/10.1023/A:1024068626366.
[47] P. Reis, F. Hebenstreit, F. Gabsteiger, V. von Tscharner, M. Lochmann, Methodological aspects of EEG and body dynamics measurements during motion, Front. Hum. Neurosci. 8 (2014) 156, https://doi.org/10.3389/fnhum.2014.00156.