# Assessment of Attention-based Deep Learning Architectures for Classifying EEG in ADHD and Typical Children

Mingzhu Han[1], Guoqin Jin[2], Wei Li[3]
Department of Industry-College Cooperation, Zhejiang Business College, Hangzhou 310053, People's Republic of China[1]
Organization and Personnel Department, Zhejiang Business College, Hangzhou 310053, People's Republic of China[2]
President's Office, Zhejiang Business College, Hangzhou 310053, People's Republic of China[3]

***Abstract*—Although limited research has explored the integration of electroencephalography (EEG) and deep learning approaches for attention deficit hyperactivity disorder (ADHD) detection, using deep learning models for actual data, including EEGs, remains a difficult endeavour. The purpose of this work was to evaluate how different attention processes affected the performance of well-established deep-learning models for the identification of ADHD. Two specific architectures, namely long short-term memory (LSTM)+ attention (Att) and convolutional neural network (CNN)s+Att, were compared. The CNN+Att model consists of a dropout, an LSTM layer, a dense layer, and a CNN layer merged with the convolutional block attention module (CBAM) structure. On top of the first LSTM layer, an extra LSTM layer, including T LSTM cells, was added for the LSTM+Att model. The information from this stacked LSTM structure was then passed to a dense layer, which, in turn, was connected to the classification layer, which comprised two neurons. Experimental results showed that the best classification result was achieved using the LSTM+Att model with 98.91% accuracy, 99.87% accuracy, 97.79% specificity and 98.87% F1-score. After that, the LSTM, CNN+Att, and CNN models succeeded in classifying ADHD and Normal EEG signals with 98.45%, 97.74% and 97.16% accuracy, respectively. The information in the data was successfully utilized by investigating the application of attention mechanisms and the precise position of the attention layer inside the deep learning model. This fascinating finding creates opportunities for more study on large-scale EEG datasets and more reliable information extraction from massive data sets, ultimately allowing links to be made between brain activity and specific behaviours or task execution.**

***Keywords—ADHD; EEG; deep learning; attention mechanisms; CNN; LSTM***

## I. Introduction

The well-being of children's minds is incredibly important, and it's essential to address their mental health needs promptly (1). Several factors, like genetics, environment, and experiences, can influence how children's mental health develops [1, 2]. Young individuals commonly face psychological challenges such as anxiety, attention-deficit/hyperactivity disorder (ADHD), and depression [3]. ADHD is a mental condition characterized by hyperactivity, inattention, and impulsive behaviours. Studies show that around five per cent of children have ADHD, with a higher prevalence among boys [4, 5]. The symptoms of ADHD can vary, with some individuals showing more hyperactivity and impulsivity while others experience difficulties with attentiveness [6]. Generally, ADHD symptoms emerge during preschool years, but significant struggles can occur during a child's school years. One of the main difficulties for children with ADHD is controlling and regulating their behaviours, often resulting in inappropriate responses to their surroundings [7]. Managing and regulating their behaviours poses a significant challenge for them. This struggle may manifest as difficulty staying seated, constant fidgeting, or excessive physical activity, making it hard for them to concentrate in a classroom setting. Additionally, they may encounter problems sustaining attention as they easily get distracted by external stimuli or their own thoughts. These difficulties can negatively impact their ability to focus on tasks, leading to difficulties with organizing work and completing assignments [8].

Detecting ADHD in a timely manner is crucial for preventing potential complications and ensuring the well-being of children's social interactions. Traditionally, ADHD diagnosis has relied on diagnostic assessments based on criteria outlined in various editions of the International Classification of Diseases (ICD) or the Diagnostic and Statistical Manual of Mental Disorders (DSM) [9]. However, this method heavily relies on parents and teachers understanding psychologists' or psychiatrists' questions and providing accurate responses. To address these challenges, researchers have been exploring and implementing objective techniques for ADHD diagnosis, such as electroencephalography (EEG) [10]. These approaches analyze neurophysiological irregularities and provide valuable insights into identifying ADHD [11]. Neurophysiological examinations like EEG offer a deeper understanding of brain structure and functioning [12, 13], enabling healthcare professionals to gather significant information [14, 15]. Studies show that individuals with ADHD often exhibit distinct brain wave activity patterns, including increased theta waves and decreased beta waves. These specific patterns indicate difficulties related to attention management and impulse control. By leveraging these neurophysiological findings, healthcare providers can better comprehend and diagnose ADHD, leading to more targeted and effective treatments and interventions for those affected by this condition [16, 17].

Extensive research has been conducted on various aspects of EEG signals in individuals with ADHD, including power spectrum density, event-related potentials, multivariate and univariate EEGs, complexity analysis, and alpha asymmetry [18, 19]. While machine learning (ML) algorithms like logistic regression, LDA, SVM, KNN, principle component analysis, and various neural network models have commonly been used to classify EEG patterns in ADHD [20], deep learning models in this field have received relatively less attention and require further investigation. Some studies have focused on applying convolutional neural networks (CNNs) to detect ADHD using functional and structural MRI [21, 22]. However, limited research has explored the integration of EEG and deep learning approaches for ADHD detection. Traditional ML methods typically employ shallow architectures with limited capacity for nonlinear feature transformation [23]. For example, SVMs utilize a shallow linear pattern separation model that requires a larger number of computational elements and struggles to model complex concepts and multi-level abstractions. Moreover, due to their single-layer construction, traditional ML methods lack effectiveness in identifying anomalous points in the deep hidden layers.

The extraction of preexisting designed characteristics and intensive preprocessing were key components of previous ML methods [24]. Nonetheless, a number of deep learning models have been effectively launched in the last ten years [25]. Consequently, the challenge has shifted from developing relevant engineered features to the need for large-scale data collection, which is crucial for effectively training optimal deep learning models. Finding the most important information has become a critical task due to the growing number of data. One of the newest and most important deep learning principles is attention, which makes it possible to understand which portions of the data are pertinent to the output and to seamlessly integrate outside information into a deep learning model [26]. This approach seeks to facilitate the adoption of parallel computing while improving a deep learning network's explainability and interpretability [27]. Hence, over the past few years, several diverse attention techniques have been implemented in EEG-based recognition [28-30]. Therefore, in this research, the potential of employing different attention strategies is investigated. Indeed, this study focused on the application of attention in various deep learning models for the EEG classification of ADHD and typical children. For this purpose, commonly utilized deep learning models for EEG recognition, namely CNN and LSTM, were re-implemented. Each of these models was augmented with attention mechanisms, and the influence of attention on the resulting classification accuracy was assessed. In Section II, a detailed explanation of the methodology is presented. Section III provides the experimental results and findings. The findings of the study are discussed in the Section IV and finally Section V concludes the paper.

## II. Methods

### A. Dataset

A freely available dataset from the "First EEG Data Analysis Competition with Clinical Applications" was employed for the study [31]. This dataset comprises EEG recordings collected from 61 children aged between 7 and 12 years. In the ADHD group, there were 31 children, consisting of 22 boys and 9 girls, with an average age of 9.64±1.73. Conversely, the control group consisted of 30 children, including 25 boys and 5 girls, with an average age of 9.85±1.77. None of the subjects in the control group exhibited any psychiatric conditions. In order to maintain consistency, specific criteria were established to exclude children with ADHD and those who were healthy. These criteria encompassed a history of significant neurological disorders or cortical damage (e.g., epilepsy), major physical illnesses, learning or speech disabilities, other psychiatric issues, and the use of barbiturates and benzodiazepines.

During the EEG recording, the 10-20 standard was followed, and a total of 19 channels were utilized. The specific channels employed were F7, Cz, Fz, T3, Pz, Fp1, C3, T5, C4, F8, T4, Fp2, F3, P4, F4, P3, T6, O1, and O2. Reference channels A1 and A2 were placed on the ears. The signals were digitized at a sampling rate of 256 Hz and captured within the frequency range of 0.1 to 60 Hz. To eliminate unwanted noise and interference, a FIR band-pass filter with cut-off frequencies of 0.4 and 60 Hz was applied, along with a notch filter set at 50 Hz to cancel out any electrical interference from the city. Throughout the EEG recording, the child was presented with various images of animal figures or cartoon characters displayed on a nearby monitor. These images were shown both at the top and bottom of the screen (see Fig. 1). The child's task was to count the characters at the top, then count the pictures at the bottom, and finally add the two numbers together to announce the total. The accuracy of the sum was not a crucial factor in this protocol; the primary objective was to keep the child consistently engaged in a cognitive state throughout the EEG recording process.

### B. Feature Extraction

In this study, the focus was on analyzing EEG data, which consisted of both ADHD and typical frames or segments. A recent study found that nonlinear and frequency features are better markers of EEG patterns for diagnosing ADHD [32]. Therefore, this study focused on nonlinear and frequency features as input to deep classification models. 15 well-established characteristics were evaluated in the frequency and temporal domains for each unique EEG channel. Specifically, in the time domain via different nonlinear analysis approaches, the following features were extracted: Higuchi fractal dimension, Hurst exponent, correlation dimension, Lempel-Ziv complexity, sample entropy, permutation entropy, Katz fractal dimension, Lyapunov exponent, detrended fluctuation analysis, and Petrosian fractal dimension, as mentioned in prior studies [32-34]. Moving on to the frequency domain, the spectral power within clinically relevant frequency bands was calculated. These bands include delta band ranging from 0.5 Hz to 4 Hz, theta band ranging from 4 Hz to 8 Hz, alpha band ranging from 8 Hz to 13 Hz, beta band ranging from 13 Hz to 30 Hz, and gamma band ranging from 30 Hz to 45 Hz. To collectively refer to the set of features extracted from each channel in both time and frequency domains, it is termed the vector:

$$S_c(t) = [F_1(t), F_2(t), \dots, F_n(t)] \quad (1)$$

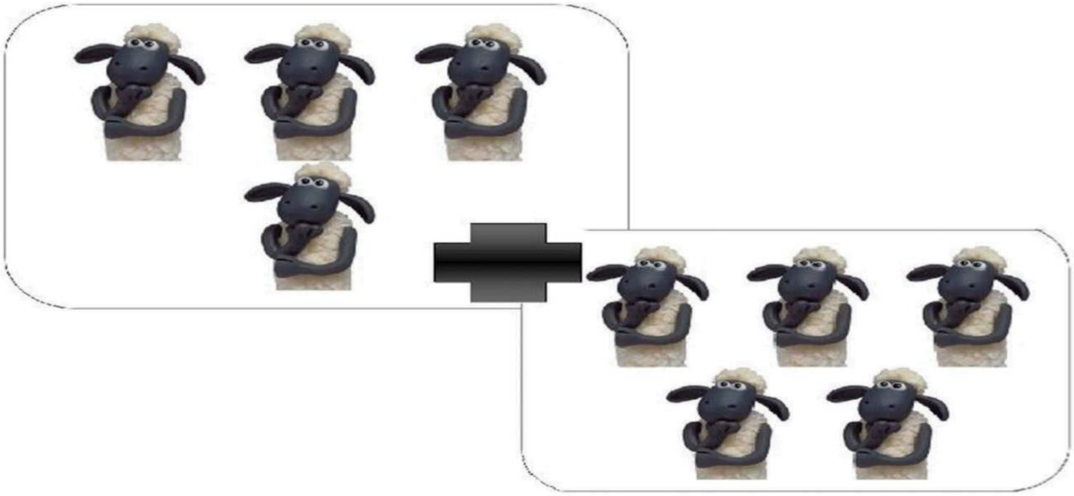

Fig. 1. An instance of images depicted to subjects during signal capturing.

where, n = 15 and t = 1, 2, …, E, where E denotes the data segment count. Furthermore, for every time segment t, the Spearman's correlation coefficient among all EEG electrodes was calculated, resulting in a distinct correlation matrix m for every given time segment t:

$$m(t) = \begin{bmatrix} m_{11}(t) & \cdots & m_{1C}(t) \\ \vdots & \ddots & \vdots \\ m_{C1}(t) & \cdots & m_{CC}(t) \end{bmatrix} \quad (2)$$

For each time segment, the deep learning networks used in this study receive inputs consisting of the correlation matrix m and the feature vector $S_c$ for all EEG electrodes, where $c$ runs from 1 to C. Here, $m_{ij}(t)$ denotes the correlation coefficient in the segment t between channels i and j. To prevent any confusion, unless stated otherwise, the reliance on the time segment, denoted as t, will be disregarded.

## C. Attention Models

Within this study, two deep learning models that benefit from attention mechanisms exhibit similar structures, with the variation occurring in the initial layer. To efficiently handle time-related information in the input data, the LSTM with attention model includes an LSTM unit in the first layer. In contrast, the CNN with attention model processes the input using a one-dimensional convolution operation. The LSTM layer, the dense layer, and the classification layer are the next three layers that both models have in common. In each model, the attention mechanism is designed to meet the specific processing needs of the corresponding initial layers. With the exception of the LSTM model with attention, which places the attention mechanism after the second LSTM layer, the attention mechanism is typically positioned between the initial layer and the LSTM layer. In all of the models, cross-entropy was implemented as the loss function for optimizing the parameters, determined as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(Y_{i,j}\log\left(P_{i,j}\right)\right) \quad (3)$$

Here, $Y_{i,j}$ denotes the desired class label for the segment i, $P_{i,j}$ denotes the estimated outcome for that class, N denotes the total sample count, and M denotes the count of classes. During this research, sole focus was placed on two classes, and one-hot encoding was utilized for the output. Every model had the softmax function in its final layer. The settings were updated using a mini-batch gradient descent method. This method updated the model's parameters using a batch of B samples, where B is the batch size that was determined empirically.

*1) CNN with attention:* The model utilized in this research is known as CNN with Attention (CNN+Att). This network was inspired by a previous work [35] introducing the Convolutional Block Attention Module (CBAM), an attention process particularly adjusted for convolutional architectures. The spatial attention and channel attention sub-modules, which functioned in tandem, made up the two different attention processes that made up the CBAM module. The channel attention focused on identifying relevant information in the input, whereas the spatial attention determined the meaningful placement of that information. The relevance was established by the attention coefficients matrix, which was represented by the symbols $A_s$ for spatial attention (arising from the convolution operations) and $A_a$ for channel attention (derived using a shared MLP). These processes were applied to this model in a sequential fashion, starting with the channel module and moving on to the spatial module. Fig. 2 illustrates the overall architecture of CNN+Att, which includes a CNN layer integrated with the CBAM structure, a dense layer, an LSTM layer, and a dropout. This structure conducts one-dimensional convolutional operations on every input vector, indicating time segment (t) from 1 to E. Two improvements were applied to the input feature matrix: one included multiplication with the channel attention sub-module ($A_a$) and the other with the spatial attention sub-module ($A_s$). After integration, the CNN layers' outputs were fed into the LSTM.

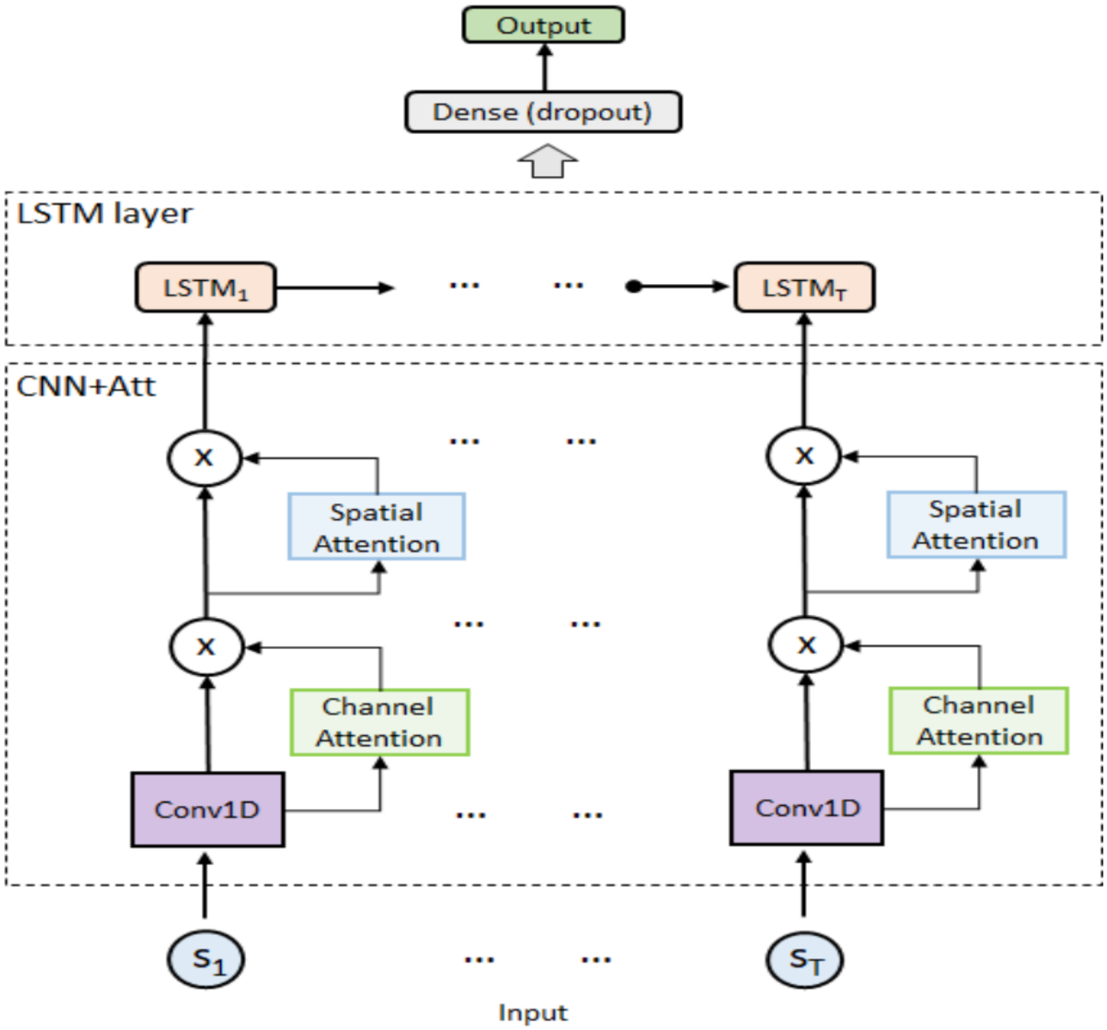


Fig. 2. The structure of the CNN+Att model.

*2) LSTM with attention:* For the second attention-based framework, inspiration was drawn from a structure introduced in the previous work [36], and the LSTM with attention (LSTM+Att) was implemented. In the implementation, a two-layer LSTM structure was opted for instead of the original three-layer version to maintain consistency with the other model examined in the current work. An additional LSTM layer with T LSTM cells was added on top of the initial LSTM layer. The information from this stacked LSTM structure was then passed to a dense layer, which, in turn, was connected to the classification layer, which comprised two neurons. During training, the described loss function was utilized. To create the input vector for every EEG segment (t), Spearman's correlation coefficients from m(t) were concatenated with the extracted feature vector, $S_c(t)$. The integrated vector representing one segment is denoted as:

$$s(t) = [s_1||s_2||..||s_C] \quad (4)$$

Every $s_i$ is explained through Eq. (3). The attention layer was positioned above the second LSTM layer, as seen in Fig. 3. The attention layer designates suitable weights, represented by $\alpha_i$, to every i$^{th}$ cell's output ($h_i$) in the LSTM layer. Each vector $h_i$ was multiplied by the weight $\alpha_i$ that corresponded to it. E vectors were concatenated to create a single vector, which was then transmitted to a dense layer without any dropout. The last layer, which made use of the softmax activation function, carried out the EEG categorization. In this model, each cell of the LSTM layer constructed its own delineation of the input segment. The attention process in this model specifically relied on segments/time steps that contained more distinguishing information, assigning higher coefficients $\alpha_i$ to these time steps. To calculate the attention coefficients, a transformation function was applied, $u_i = \tanh(W_s h_i)$, where i belongs to the set 1, 2, …, E and $W_s$ represented the weight matrix. Subsequently, softmax($u_i$) was utilized to determine the attention weights $\alpha_i$ after normalizing the attention coefficients. Furthermore, as mentioned in Eq. (3), the model was trained using the cross-entropy loss function, which is different from the original work.

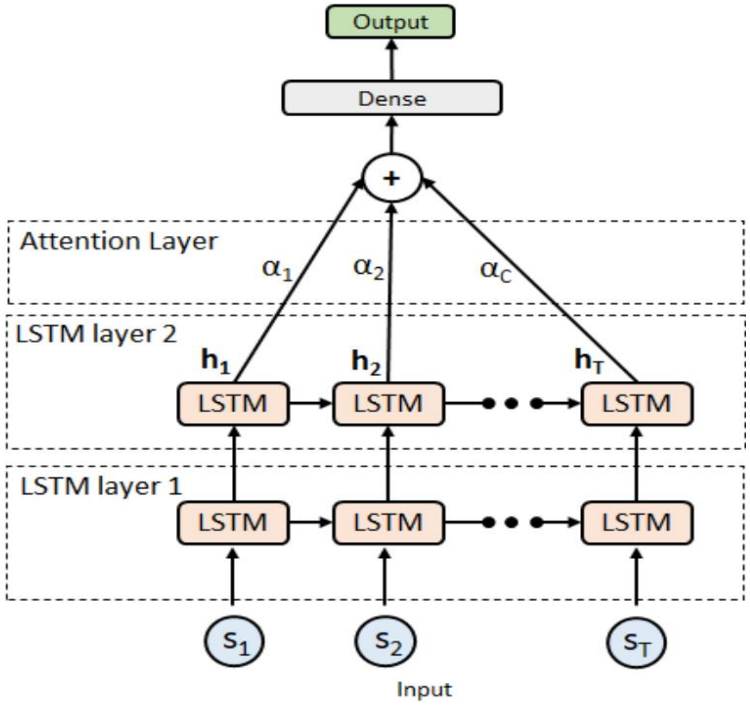


Fig. 3. The structure of the LSTM+Att model.

### D. Baseline Deep Learning Networks

In order to compare the models discussed earlier with attention and also to investigate the effect of attention mechanisms included in deep structures on the classification performance of models, two additional deep learning models without attention mechanisms were considered in this work: a CNN and an LSTM. These models had identical structures to their respective attention-enhanced counterparts, with the exception of the removal of the attention layer. Whole networks were executed in Python through the Tensorflow 2 approach. To optimize the performance of the models, the hyper-parameter values were carefully selected to obtain the highest F1-score averaged over all data. For parameter optimization, the Stochastic Adam optimizer was employed. The optimal parameters for CNN+Att and LSTM+Att networks can be found in Tables I and II.

TABLE I. SELECTED HYPER-PARAMETER VALUES FOR CNN+ATT CLASSIFICATION NETWORK

| Hyper-parameter | Range | CNN | CNN+Att |
|---|---|---|---|
| Convolution kernel | 3, 5, 7, 9, 11 | 3 | 3 |
| Convolution filters | 8, 16, 32, 64 | 64 | 64 |
| LSTM hidden layers | 8, 128, 256 | 8 | 256 |
| Dropout level | [0.1, 0.5] | 0.5 | 0.4 |
| Learning rate | [0.0001, 0.001] | 0.0002 | 0.0002 |
| CBAM reduction ratio | 4, 8, 16 | - | 16 |
| CBAM spatial kernel | 5, 7, 9, 11 | - | 7 |

TABLE II. SELECTED HYPER-PARAMETER VALUES FOR THE LSTM+ATT CLASSIFICATION NETWORK

| Hyper-parameter | Range | LSTM | LSTM+Att |
|---|---|---|---|
| LSTM hidden layers | 8, 128, 256 | 128 | 128 |
| LSTM L2 reg | [0.001, 0.05] | 0.002 | 0.001 |
| Input dropout level | [0.1, 0.5] | 0.4 | 0.4 |
| LSTM layer 1 dropout | [0.1, 0.5] | 0.4 | 0.2 |
| LSTM layer 2 dropout | [0.1, 0.5] | 0.3 | 0.2 |
| Learning rate | [0.0001, 0.001] | 0.0001 | 0.0001 |

### E. Evaluation of Models

A 10-fold cross-validation approach was used to increase the power of the estimate of error and guarantee the validity of the results. All models were trained with a batch size of 16 for 50 epochs inside each fold. The models were evaluated using recognized classification measures, such as F1-score, sensitivity, specificity, and accuracy. Sensitivity and specificity are especially important when assessing how well a classifier works to detect uncommon but important samples. TP (True Positive) indicates the positively categorized samples that were correctly identified, with N being the total sample count for classification; TN (True Negative) indicates the accurately classified negative samples; FP (False Positive) represents the incorrectly classified positive samples, and FN (False Negative) representing the incorrectly classified negative samples, the accuracy, sensitivity, specificity, and F1-score values were determined.

## III. Results

In the current work, the performance of two attention-based models was evaluated in comparison to baseline models for a two-group classification problem for ADHD diagnosis. Fig. 4 shows the scatterplots of nonlinear features extracted from the FP2 channel of ADHD and normal subjects.

Table III presents a summary of all results for each classification model in terms of F1-score, sensitivity, specificity, and accuracy. As shown, the best classification result was achieved using the LSTM+Att model with 98.91% accuracy, 99.87% accuracy, 97.79% specificity and 98.87% F1-score. After that, the LSTM, CNN+Att, and CNN models succeeded in classifying ADHD and Normal EEG signals with 98.45%, 97.74% and 97.16% accuracy, respectively.

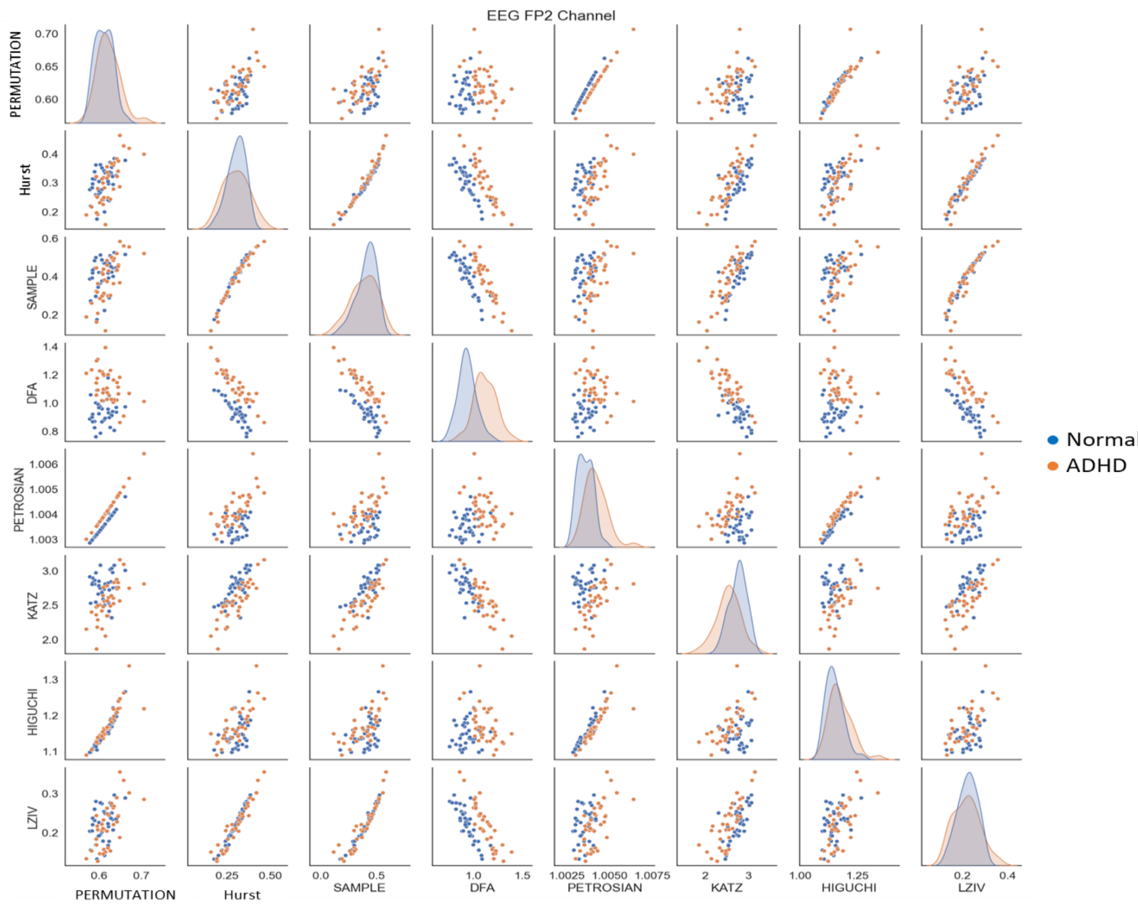


Fig. 4. Scatterplots of nonlinear features extracted from the FP2 channel of ADHD and normal subjects.

TABLE III. Mean and Standard Deviation of Classification Results for All Models for ADHD Detection

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|
| CNN | 97.16 ± 0.91 | 95.33 ± 1.03 | 98.89 ± 1.31 | 97.05 ± 0.70 |
| CNN+Att | 97.74 ± 1.07 | 96.88 ± 0.85 | 98.57 ± 1.25 | 97.61 ± 0.99 |
| LSTM | 98.45 ± 1.05 | 98.10 ± 1.02 | 98.82 ± 1.04 | 98.40 ± 0.95 |
| LSTM+Att | 98.91 ± 0.64 | 99.87 ± 0.22 | 97.79 ± 1.03 | 98.87 ± 0.72 |

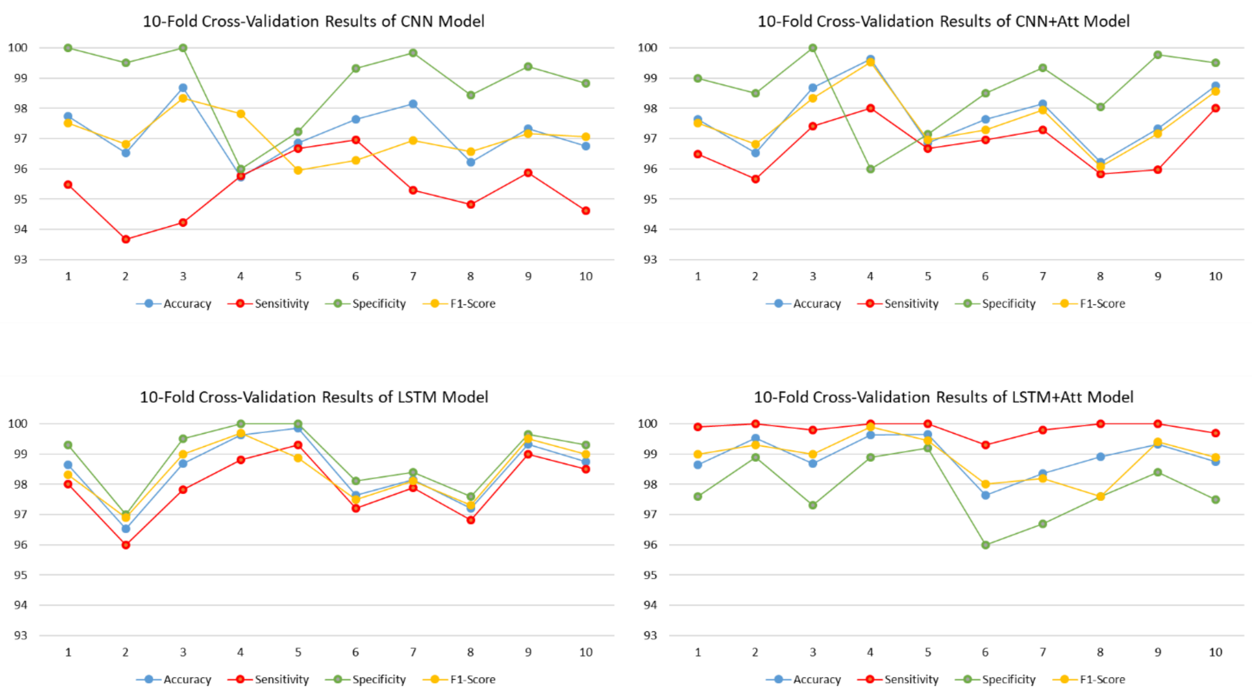


Fig. 5. Obtained classification results of a 10-fold cross-validation algorithm for all classification models for ADHD detection.

Fig. 5 shows the Obtained classification results of a 10-fold cross-validation algorithm for all classification models for ADHD detection. As can be seen, the CNN and CNN+Att models had more variance than the LSTM and LSTM+Att models.

## IV. Discussion

Using deep learning models for actual data, including EEGs, remains a difficult endeavour. These datasets encapsulate intricate scenarios where various factors, such as technological instruments, recording interference, and both emotional and physical states, intertwine. Consequently, the classification performance can be heavily influenced by disparities between subjects and within individuals themselves. Reflecting on this observation, two attention-enhanced deep learning models were executed and juxtaposed (alongside their respective counterparts lacking attention) across an EEG dataset for ADHD detection. This approach aimed to explore how attention can augment deep learning models in identifying ADHD EEG patterns. The accuracy and F1 scores for all models were remarkably close, surpassing the 97% threshold. Compared to previous deep learning models without an attention mechanism, this study improved the accuracy of ADHD diagnosis. Chen et al. reported an accuracy of 94.67% in diagnosing ADHD using a novel connectivity matrix and a CNN model [37]. Vahid et al achieved 83% accuracy in diagnosing ADHD using EEGNet deep model [38]. Using a four-layer CNN model, Dubreuil-Vall et al. achieved an accuracy of 88% in diagnosing ADHD [39]. Cisotto et al. also showed that attention-based deep learning models can improve the classification performance of EEG datasets [40].

It is important to remember that attention processes were purposefully kept simple in order to evaluate their influence on each suggested model. Each model was composed of an LSTM layer, a dense layer for output generation, and a single attention layer that stored a model-specific attention mechanism. This simple yet efficient design made it easier to compare various attention-enhanced versions. Every attention mechanism was created to make use of input properties in a unique way. The LSTM+Att model employed attention in the temporal dimension to filter out irrelevant information. On the other hand, the CNNs+Att model utilized the CBAM module to apply attention to each EEG channel individually. Interestingly, models primarily focused on spatial features demonstrated performance improvements when attention was introduced, such as with CNNs+Att outperforming CNN. Jiang et al. improved the performance of their CNN model in the EEG-based emotion recognition task by incorporating the temporal-channel attention mechanism into the designed deep model [41]. Altuwaijri and Muhammad improved the performance of their CNN model by adding CBAM structure to multi-branch EEGNet through attention mechanism and fusion methods for EEG-based motor imagery classification [42]. Notably, the proposed attention-enhanced models demonstrated versatility in leveraging different EEG descriptions that consider time, frequency, and spatial information (sensor locations) interchangeably or in conjunction. These considerations offer valuable insights for devising suitable experimental protocols and data processing pipelines based on the specific behaviours or task performances under study. For instance, in cognitive tasks where individuals are expected to respond promptly to external stimuli, architectures like LSTM+Att can effectively filter time-dependent features. Zhou et al. showed that the attention-based LSTM performs better than the LSTM structure without the attention mechanism in detecting abnormal behavior [43]. It is important to emphasize that despite their simple designs, the attention mechanisms enabled the models to achieve high levels of accuracy in a range of real-world scenarios with minimal preprocessing. This statement has been shown in previous studies for EEG-based sleep stage classification [44, 45], clinical events prediction in the intensive care unit [46], and diagnosis of various diseases [47, 48]. Preprocessing is usually directed by domain expertise or knowledge, and depending on the analyst performing the data analysis, it may provide non-reproducible findings. As a result, minimizing the need for preprocessing offers a big benefit over traditional ML or other deep learning techniques. However, it's worth mentioning that this work still requires further investigation on larger datasets impacted by artefacts, where preprocessing is often crucial. Nonetheless, it paves the way for future research aiming to minimize preprocessing in extensive EEG datasets empirically.

Similar researches have investigated the application of different deep learning models in EEG for epilepsy diagnosis [49, 50], psychiatric disorder diagnosis [20, 51], motion imagery classification [52, 53] and mental workload classification [54]. In Table IV, a comparison is made between the proposed approach and other leading ML methods for diagnosing ADHD using automated EEG data on the same dataset. The results revealed that this approach outperformed previous studies, showcasing a higher accuracy value. Specifically, it surpassed conventional ML techniques employed on unipolar EEG signals. Furthermore, when compared to other deep learning methods applied to the same EEG signals, the approach presented here produced satisfactory outcomes. This study introduces a newly developed deep learning model that utilizes EEG data for ADHD diagnosis.

TABLE IV. COMPARING THE PERFORMANCE OF THE PROPOSED APPROACH WITH SOME STATE-OF-THE-ART RESEARCH IN ADHD DIAGNOSIS THROUGH EEG ANALYSIS ON THE SAME DATASET

| References | Dataset | Approach | Accuracy |
|---|---|---|---|
| [55] | Same as this study | Nonlinear features, MLP neural network | 96.70% |
| [56] | Same as this study | Nonlinear features, MLP neural network | 93.65% |
| [31] | Same as this study | EEG image generation based on spectral features, Deep CNN model | 98.48% |
| The proposed technique | 31 ADHD and 30 Normal children | Nonlinear and spectral features and LSTM+Att model | 98.91% |

The insufficient clinical implications of this paper and similar studies constitute a significant drawback. In general, there is a need for further evidence regarding the effectiveness of employing EEG-based ML techniques in diagnosing ADHD. For instance, it remains unexplored how these methods perform when applied to individuals who have undergone treatment for ADHD in the past. Furthermore, in order to utilize these approaches effectively, it is crucial to obtain a broader range of EEG datasets specific to ADHD. This is particularly significant for deep learning techniques as they necessitate extensive datasets to achieve optimal results. Furthermore, the segmentation of EEG signals on a second-to-second basis for data augmentation, which was employed in this study and previous similar studies, may not possess clinical justification. In addition, the proposed models were only tested on a cross-sectional dataset, and it is necessary to examine their validity through longitudinal studies. Nevertheless, the proposed approach can serve as a CAD tool for clinical purposes.

## V. CONCLUSION

The purpose of this work was to evaluate how different attention processes affected the performance of well-established deep-learning models for the identification of ADHD. Two specific architectures, namely LSTM+Att and CNNs+Att, were compared. These models were employed for the classification of EEG patterns, including ADHD and Normal patterns. Notably, despite the simplicity of the suggested attention-enhanced models, the results showed state-of-the-art performance across all categorization models. The information in the data was successfully utilized by investigating the application of attention mechanisms and the precise position of the attention layer inside the deep learning model. This fascinating finding creates opportunities for more study on large-scale EEG datasets and more reliable information extraction from massive data sets, ultimately allowing links to be made between brain activity and specific behaviours or task execution. Hence, attention is a viable method for evaluating the accuracy and applicability of EEG data in the identification of ADHD. Additionally, attention mechanisms facilitate parallel computation, thereby accelerating the analysis of significant electrophysiological datasets such as EEG. These promising results could encourage stakeholders to offer a CAD system for diagnosing ADHD through the suggested method. For future research, collecting more diverse EEG samples, exploring alternative ML and deep learning techniques, incorporating psychophysiological attributes and other neurophysiological recordings with EEG, and developing ML methods for automatically scaling the severity of ADHD is recommended.

## REFERENCES

[1] M. R. Mohammadi et al., "Prevalence of autism and its comorbidities and the relationship with maternal psychopathology: a national population-based study," Arch Iran Med, vol. 22, no. 10, 2019.

[2] S. Talepasand et al., "Psychiatric disorders in children and adolescents: prevalence and sociodemographic correlates in Semnan Province in Iran," Asian journal of psychiatry, vol. 40, pp. 9-14, 2019.

[3] M. R. Mohammadi, R. Badrfam, A. Khaleghi, Z. Hooshyari, N. Ahmadi, and A. Zandifar, "Prevalence, comorbidity and predictor of separation anxiety disorder in children and adolescents," Psychiatric Quarterly, vol. 91, pp. 1415-1429, 2020.

[4] M.-R. Mohammadi et al., "Prevalence of ADHD and its comorbidities in a population-based sample," Journal of Attention Disorders, vol. 25, no. 8, pp. 1058-1067, 2021.

[5] M. R. Mohammadi et al., "Prevalence and correlates of psychiatric disorders in a national survey of Iranian children and adolescents," Iranian journal of psychiatry, vol. 14, no. 1, p. 1, 2019.

[6] M. R. Mohammadi, N. Malmir, A. Khaleghi, and M. Aminiorani, "Comparison of sensorimotor rhythm (SMR) and beta training on

selective attention and symptoms in children with attention deficit/hyperactivity disorder (ADHD): A trend report," Iranian journal of psychiatry, vol. 10, no. 3, p. 165, 2015.

[7] A. Khaleghi, H. Zarafshan, and M. R. Mohammadi, "Visual and auditory steady-state responses in attention-deficit/hyperactivity disorder," European archives of psychiatry and clinical neuroscience, vol. 269, pp. 645-655, 2019.

[8] M. Daneshparvar et al., "The role of lead exposure on attention-deficit/hyperactivity disorder in children: a systematic review," Iranian journal of psychiatry, vol. 11, no. 1, p. 1, 2016.

[9] M. R. Mohammadi, A. Khaleghi, K. Shahi, and H. Zarafshan, "Attention Deficit Hyperactivity Disorder: Behavioral or Neuro-developmental Disorder? Testing the HiTOP Framework Using Machine Learning Methods," Journal of Iranian Medical Council, vol. 6, no. 4, pp. 652-657, 2023.

[10] M. Adamou, T. Fullen, and S. L. Jones, "EEG for diagnosis of adult ADHD: a systematic review with narrative analysis," Frontiers in Psychiatry, vol. 11, p. 871, 2020.

[11] A. Khaleghi, M. R. Mohammadi, G. P. Jahromi, and H. Zarafshan, "New ways to manage pandemics: Using technologies in the era of covid-19: A narrative review," Iranian journal of psychiatry, vol. 15, no. 3, p. 236, 2020.

[12] A. Khaleghi, M. R. Mohammadi, M. Moeini, H. Zarafshan, and M. Fadaei Fooladi, "Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder," Clinical EEG and neuroscience, vol. 50, no. 5, pp. 311-318, 2019.

[13] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Possible Neuropathological Mechanisms Underlying the Increased Complexity of Brain Electrical Activity in Schizophrenia: A Computational Study," Iranian Journal of Psychiatry, pp. 1-7, 2023.

[14] A. Khaleghi, A. Sheikhani, M. R. Mohammadi, and A. M. Nasrabadi, "Evaluation of cerebral cortex function in clients with bipolar mood disorder I (BMD I) compared with BMD II using QEEG analysis," Iranian Journal of Psychiatry, vol. 10, no. 2, p. 93, 2015.

[15] A. Khaleghi et al., "EEG classification of adolescents with type I and type II of bipolar disorder," Australasian physical & engineering sciences in medicine, vol. 38, pp. 551-559, 2015.

[16] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Computational neuroscience approach to psychiatry: A review on theory-driven approaches," Clinical Psychopharmacology and Neuroscience, vol. 20, no. 1, p. 26, 2022.

[17] H. Zarafshan, A. Khaleghi, M. R. Mohammadi, M. Moeini, and N. Malmir, "Electroencephalogram complexity analysis in children with attention-deficit/hyperactivity disorder during a visual cognitive task," Journal of clinical and experimental neuropsychology, vol. 38, no. 3, pp. 361-369, 2016.

[18] S. Kaur, S. Singh, P. Arun, D. Kaur, and M. Bajaj, "Event-related potential analysis of ADHD and control adults during a sustained attention task," Clinical EEG and neuroscience, vol. 50, no. 6, pp. 389-403, 2019.

[19] A. Lenartowicz et al., "Alpha modulation during working memory encoding predicts neurocognitive impairment in ADHD," Journal of Child Psychology and Psychiatry, vol. 60, no. 8, pp. 917-926, 2019.

[20] A. Afzali, A. Khaleghi, B. Hatef, R. Akbari Movahed, and G. Pirzad Jahromi, "Automated major depressive disorder diagnosis using a dual-input deep learning model and image generation from EEG signals," Waves in Random and Complex Media, pp. 1-16, 2023.

[21] L. Zhu and W. Chang, "Application of deep convolutional neural networks in attention-deficit/hyperactivity disorder classification: Data augmentation and convolutional neural network transfer learning," Journal of Medical Imaging and Health Informatics, vol. 9, no. 8, pp. 1717-1724, 2019.

[22] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," Ieee Access, vol. 5, pp. 23626-23636, 2017.

[23] H. Chen, Y. Song, and X. Li, "Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD," Journal of neural engineering, vol. 16, no. 6, p. 066046, 2019.

[24] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of EEG signals based on pattern recognition approach," Frontiers in computational neuroscience, vol. 11, p. 103, 2017.

[25] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," Journal of neural engineering, vol. 16, no. 5, p. 051001, 2019.

[26] K. Cho, A. Courville, and Y. Bengio, "Describing multimedia content using attention-based encoder-decoder networks," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 1875-1886, 2015.

[27] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," IEEE access, vol. 6, pp. 52138-52160, 2018.

[28] T. Zhu, W. Luo, and F. Yu, "Convolution-and attention-based neural network for automated sleep stage classification," International Journal of Environmental Research and Public Health, vol. 17, no. 11, p. 4152, 2020.

[29] I. Zoppis et al., "An Attention-based Architecture for EEG Classification," in BIOSIGNALS, 2020, pp. 214-219.

[30] Y. Yuan and K. Jia, "FusionAtt: deep fusional attention networks for multi-channel biomedical signals," Sensors, vol. 19, no. 11, p. 2429, 2019.

[31] M. Moghaddari, M. Z. Lighvan, and S. Danishvar, "Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG," Computer Methods and Programs in Biomedicine, vol. 197, p. 105738, 2020.

[32] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," Research on Biomedical Engineering, vol. 36, pp. 1-11, 2020.

[33] S. Pahuja and K. Veer, "Recent approaches on classification and feature extraction of EEG signal: A review," Robotica, vol. 40, no. 1, pp. 77-101, 2022.

[34] W. Xiao, G. Manyi, and A. Khaleghi, "Deficits in auditory and visual steady-state responses in adolescents with bipolar disorder," Journal of Psychiatric Research, vol. 151, pp. 368-376, 2022.

[35] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.

[36] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," IEEE Sensors Journal, vol. 20, no. 6, pp. 3113-3122, 2019.

[37] H. Chen, Y. Song, and X. Li, "A deep learning framework for identifying children with ADHD using an EEG-based brain network," Neurocomputing, vol. 356, pp. 83-96, 2019.

[38] A. Vahid, A. Bluschke, V. Roessner, S. Stober, and C. Beste, "Deep learning based on event-related EEG differentiates children with ADHD from healthy controls," Journal of clinical medicine, vol. 8, no. 7, p. 1055, 2019.

[39] L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon, "Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG," Frontiers in neuroscience, vol. 14, p. 251, 2020.

[40] G. Cisotto, A. Zanga, J. Chlebus, I. Zoppis, S. Manzoni, and U. Markowska-Kaczmar, "Comparison of attention-based deep learning models for eeg classification," arXiv preprint arXiv:2012.01074, 2020.

[41] L. JIANG, P. Siriaraya, D. Choi, F. Zeng, and N. Kuwahara, "Electroencephalogram Signals Emotion Recognition Based on CNN-RNN Framework with Channel-Temporal Attention Mechanism for Older Adults," Frontiers in Aging Neuroscience, 2022.

[42] G. A. Altuwaijri and G. Muhammad, "Electroencephalogram-Based Motor Imagery Signals Classification Using a Multi-Branch Convolutional Neural Network Model with Attention Blocks," Bioengineering, vol. 9, no. 7, p. 323, 2022.

[43] K. Zhou, B. Hui, J. Wang, C. Wang, and T. Wu, "A study on attention-based LSTM for abnormal behavior recognition with variable pooling," Image and Vision Computing, vol. 108, p. 104120, 2021.

[44] E. Eldele et al., "An attention-based deep learning approach for sleep stage classification with single-channel EEG," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 809-818, 2021.

[45] C. Yang, B. Li, Y. Li, Y. He, and Y. Zhang, "LWSleepNet: A lightweight attention-based deep learning model for sleep staging with singlechannel EEG," Digital Health, vol. 9, p. 20552076231188206, 2023.

[46] D. A. Kaji et al., "An attention based deep learning model of clinical events in the intensive care unit," PloS one, vol. 14, no. 2, p. e0211057, 2019.

[47] E. Sibilano et al., "An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG," Journal of Neural Engineering, vol. 20, no. 1, p. 016048, 2023.

[48] A. Affes, A. Mdhaffar, C. Triki, M. Jmaiel, and B. Freisleben, "Personalized attention-based EEG channel selection for epileptic seizure prediction," Expert Systems with Applications, vol. 206, p. 117733, 2022.

[49] R. Jana and I. Mukherjee, "Deep learning based efficient epileptic seizure prediction with EEG channel optimization," Biomedical Signal Processing and Control, vol. 68, p. 102767, 2021.

[50] I. Ahmad et al., "EEG-based epileptic seizure detection via machine/deep learning approaches: A Systematic Review," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[51] W. A. Campos-Ugaz, J. P. P. Garay, O. Rivera-Lozada, M. A. A. Diaz, D. Fuster-Guillén, and A. A. T. Arana, "An Overview of Bipolar Disorder Diagnosis Using Machine Learning Approaches: Clinical Opportunities and Challenges," Iranian Journal of Psychiatry, vol. 18, no. 2, pp. 237-247, 2023.

[52] H. Zhu, D. Forenzo, and B. He, "On the deep learning models for EEG-based brain-computer interface using motor imagery," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2283-2291, 2022.

[53] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," Neural Computing and Applications, vol. 35, no. 20, pp. 14681-14722, 2023.

[54] S. Shao, G. Han, T. Wang, C. Lin, C. Song, and C. Yao, "EEG-Based Mental Workload Classification Method Based on Hybrid Deep Learning Model Under IoT," IEEE Journal of Biomedical and Health Informatics, 2023.

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

[56] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," Biomedical Engineering Letters, vol. 6, pp. 66-73, 2016.