

EEG-Based ADHD Diagnosis in Children Using Channel Attention and Lightweight Deep Learning

Zeynep Garip¹, Ekin Ekinci^{2*}

^{1,2}Computer Engineering Department, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Türkiye

Abstract – In this study, EEGNet-CA, a lightweight deep learning model with an integrated attention mechanism, is proposed to diagnose Attention Deficit and Hyperactivity Disorder (ADHD) in children via EEG data. The proposed model is able to learn the contribution of each EEG channel in the classification and focus on important signals by integrating the channel attention (CA) mechanism into the classical EEGNet architecture. In this context, 19-channel EEG recordings from 121 children aged 7–12 years, obtained from the IEEE dataset, were used. The raw EEG signals were scaled with z-score normalization to improve the stability of the learning process and segmented using the 50 % overlap sliding window method. The EEGNet-CA model, developed using the obtained EEG segments, was utilised for classification purposes. Experimental results show that the proposed EEGNet-CA model achieved a 99.79 % F1-score and exhibited statistically significantly superior performance compared to standard EEGNet, 1D CNN, 2D CNN, and artificial neural network-based methods. In addition, high accuracy even with only frontal region channels supports the neurophysiological importance of this region in ADHD diagnosis. The results show that the CA mechanism is an effective method to increase the classification success in EEG-based diagnostic systems.

Keywords – ADHD, channel attention, classification, deep learning.

I. INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is defined as a neurodevelopmental disorder that begins in childhood and often continues into adulthood [1]. It affects approximately 5–12 % of children worldwide [2]. The etiology of ADHD is assumed to be a multifactorial disorder involving genetic, environmental, neuroanatomical, and neurochemical factors [3].

The current gold standard for diagnosing ADHD is clinical assessment based on comprehensive information collected from multiple sources [4]. However, no single test, biomarker, or study is currently sufficient to definitively confirm or reject a diagnosis of ADHD [5]. Although clinical expertise remains central to diagnosis and treatment, limited access to specialists and inter-clinic variability in assessment may hinder accurate and early diagnosis [5]. Given these challenges, there is an increasing need for supportive tools that can increase the objectivity and efficiency of diagnostic processes. Delayed

diagnosis can lead to significant long-term consequences such as conduct disorder, substance use disorders, and personality disorders, all of which can impose significant economic and social burdens on a national level [6]. In recent years, artificial intelligence (AI) has emerged as a promising tool to address such challenges in healthcare by improving diagnostic accuracy and reducing reliance on specialist access [7]–[9]. Given its high prevalence and diagnostic complexity, ADHD has become an important focus of AI. AI, especially machine learning and deep learning (DL), has been widely used in both diagnostic and differential diagnosis tasks [10]. Researchers have integrated AI with various data sources such as neuroimaging, electroencephalography (EEG), cognitive questionnaires, biochemical tests, electronic health records, genetic analyses, computer games, and virtual reality environments [11]. Among these, EEG and neuroimaging modalities are the most widely used [12].

In this context, EEG-based approaches are particularly noteworthy, as EEG has a 35-year history in ADHD research and an ongoing clinical history [13]. It also provides valuable information about brain function when combined with machine learning and deep learning techniques [14]. EEG studies in ADHD have been extensively used to explore features such as nonlinear dynamics, power spectral densities, and entropy measurements of EEG signals [12]. EEG captures the natural electrical activity of the brain as recorded at the scalp and is classified into five main frequency bands: delta (1–4 Hz), theta (4.5–8 Hz), alpha (8.5–14 Hz), beta (14.5–25 Hz), and gamma (>25 Hz) [12]. These frequency bands have been shown to demonstrate specific changes in individuals with ADHD compared to neurotypical controls. These changes have been consistently reported in the literature as increased theta activity, decreased alpha power, and atypical patterns in beta and gamma waves. In addition, features derived from nonlinear dynamics and entropy measurements provide greater discriminatory power [15], [16]. It is thought that the evaluation of these EEG-based features using machine learning and deep learning models will facilitate and potentially improve the diagnostic process.

Numerous machine learning algorithms have been employed to analyse these EEG-derived features for ADHD classification.

*Corresponding author's e-mail: ekinekinici@subu.edu.tr
Article received 2026-04-15; accepted 2026-07-02

These include Linear Discriminant Analysis (LDA), Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and various architectures of Artificial Neural Networks (ANNs) [16]. Reported accuracy rates in EEG-based ADHD classification tasks using algorithms such as SVM, KNN, and neural networks (NN) have generally been found to range between 79 % and 97 %, demonstrating the feasibility of automatic diagnostic support systems [4], [17]–[19].

II. MATERIAL AND METHODS

A. Dataset

In this study, the ADHD EEG dataset is obtained from a publicly available data repository from IEEE [20]. A total of

121 right-handed, school-aged children between 7 and 12 years old participated in the EEG data collection process. Among them, 61 were diagnosed with ADHD by an experienced child and adolescent psychiatrist based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria, while the remaining 60 were healthy controls with no history of neuropsychological disorders. The children diagnosed with ADHD had been undergoing treatment with Ritalin, a commonly prescribed medication for ADHD, for a period of up to six months. The dataset included 48 males and 13 females in the ADHD group, and 50 males and 10 females in the control group.

TABLE I

COMPARISON OF RECENT EEG-BASED ADHD DIAGNOSIS STUDIES AND THEIR REPORTED CLASSIFICATION PERFORMANCE

Ref.	Proposed Algorithms	Description	Perf. Criteria (Acc %)
[21]	CNN	They used a lightweight CNN model to support accurate detection of ADHD.	97.08
[22]	Siamese based CNN	They utilized CNN model based on the Siamese architecture in the analysis of EEG-based brain maps.	99.17
[23]	MEMD-GA-ANN	In this study, the effectiveness of three multivariate decomposition methods, namely MEMD, MEWT and MVMD, was investigated in order to diagnose ADHD using EEG signals. Genetic Algorithm (GA) and NCA methods were used for feature selection; classification was performed with SVM-RBF and ANN models.	99.16
[24]	CNN+ LR	They trained a CNN model with raw EEG data for ADHD diagnosis, then the features extracted from CNN were evaluated with classical classifiers such as SVM, LR and RF.	95.83
[25]	ConvLSTM+ attention framework	They performed feature extraction with dynamic connection tensors from EEG data and discriminated between ADHD and control groups with a model including ConvLSTM and attention mechanism.	99.34
[26]	Gaussian SVM.	They automatically identified children with ADHD from their characteristic brain waves using Gaussian SVM.	94.00
[27]	GRU	In this study, a GRU-based classifier working with Weighted Distribution Entropy - based features was developed to distinguish ADHD and typically developing children.	96.70
[28]	CNN 2D	In this study, EEG signals were separated into sub-bands by robust local mode decomposition and variational mode decomposition methods and these data were given as input to the CNN 2D model developed for ADHD diagnosis.	87.00
[29]	ML	In this study, an improved Topological Data Analysis method for multi-channel EEG data in ADHD diagnosis is presented. After EEG signals are processed with phase space reconstruction and k-PDTM, topological features are extracted; then the analysis is completed with Gaussian-based multivariate density estimation and persistence imaging methods.	78.27
[30]	Resnet	They designed a ResNet model to address the problems of late and unclear diagnosis in diagnosing ADHD from EEG signals.	98.60
[31]	DL	In this study, Multitaper and MVMD methods were used for feature extraction from EEG data; then, the most effective features for the deep learning model were selected by the NCA method.	95.54
[32]	LSTM+GRU	They developed an ensemble method based on the average summation of LSTM and GRU models to classify children with ADHD with an EEG-based visual attention task.	95.33
[15]	AM-FDCNN	In this study, a deep learning model called AM-FDCNN, which combines Channel, Spatial and Position Attention Modules, is proposed to increase the accuracy in ADHD detection with EEG signals.	97.60
Proposed Study	EEGNet + Attention	In this study, EEGNet, a lightweight deep learning model with integrated CA mechanism, was developed to distinguish the difference between ADHD and control groups using EEG signals.	99.98

EEG data were recorded using a 19-channel placement system according to the international 10–20 system, as shown in Fig. 1. The recordings were obtained from the electrode locations Fp1, Fp2, Fz, Cz, C3, T3, C4, T4, F3, F4, F7, F8, P3, P4, T5, T6, O1 and O2 with a sampling frequency of 128 Hz, as shown in Fig. 2. The reference electrodes were the A1 and A2 points placed on the earlobes. This arrangement provides accurate and balanced monitoring of brain activity from different cortical regions.

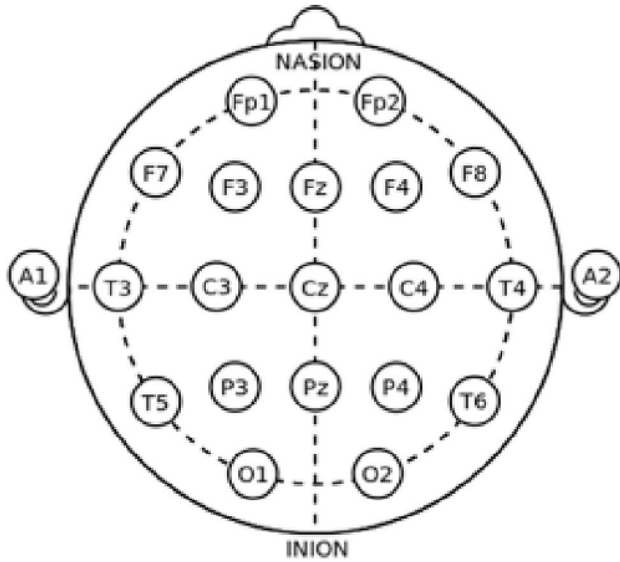


Fig. 1. International 10–20 EEG electrode placement showing the 19 recording channels used for ADHD diagnosis [9].

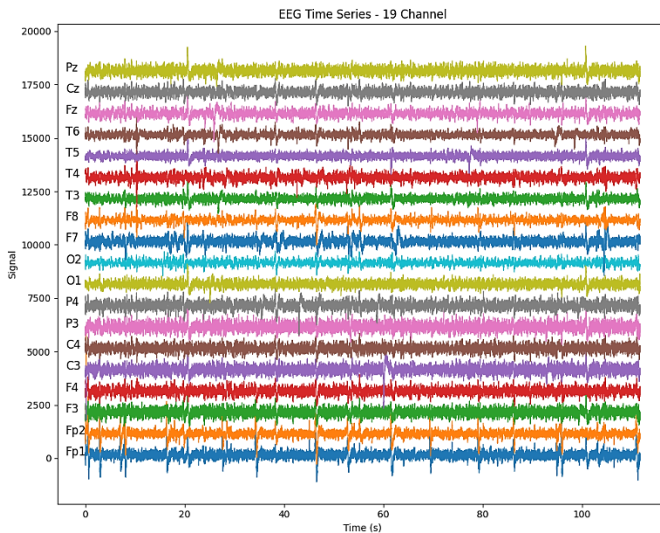


Fig. 2. Representative multichannel EEG recording from a child diagnosed with ADHD.

As visual attention deficits are frequently observed in children with ADHD, the EEG recording protocol was designed to incorporate a task centred on visual attention. Throughout the recording session, children were shown a sequence of images, each displaying between 5 and 16 characters, and were

instructed to count the number of characters in each image. To maintain continuous visual engagement, a new image was presented immediately after the child responded. Consequently, the total duration of EEG recording varied according to each child's task performance.

B. Preprocessing and Segmentation

Raw EEG signals should be preprocessed before being given as input to classification models directly due to their high variance and large inter-individual variability. In this study, in order to ensure stable learning during the training process of the model, z-score normalization was first applied in the preprocessing step. With this method, each sample was rescaled so that its mean was zero and its standard deviation was one. Thus, inter-individual measurement differences were minimised, signals were made more homogeneous and the learning process of the model was stabilised.

The z-score normalized value $z_{i,j}$ for a sample $x_{i,j}$ is calculated using the formula (1) below.

$$z_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i} \quad (1)$$

In (1), $x_{i,j}$ is j^{th} time sample of i^{th} EEG channel. While μ_i is the mean of i^{th} EEG channel, σ_i is the standard deviation of the same channel. $z_{i,j}$ is the normalized value. As a result of this process, the mean of each channel is rescaled to zero and the standard deviation to 1. Thus, measurement differences from different individuals or records are suppressed, the signals are made statistically homogeneous, and thus the learning process of the model becomes more stable.

Another important stage of the data preprocessing process, data segmentation, involves dividing EEG signals into windows of certain length. In this study, EEG signals were divided into segments of 256 samples using the fixed-length sliding windows method. Since the sampling frequency of the records was 128 Hz, each segment represented approximately 2 seconds of EEG signal.

During the segmentation process, 50 % overlap was applied between the windows; in other words, each new segment was shifted to include the last 128 samples of the previous segment. This segmentation approach offers two main advantages: First, thanks to overlapping segmentation, a much larger number of training samples were obtained from limited-length EEG records, allowing the model to encounter more diverse patterns, thus increasing its generalization ability. Second, short-term signal could be captured within the segments, allowing the model to learn the changes in the EEG signal over time more effectively.

The segments are arranged in such a way that each one is the number of channels $\times$ window length. Since the EEG data used in this study contains 19 channels, the size of each segment is (19, 256). In order to make it compatible with deep learning models, an additional channel size was added to the end of the segments, and the input tensors were converted to the size of (19, 256, 1). The labelling process was performed by considering the class label corresponding to the sample at the

beginning of each segment. This method also makes it possible to monitor classes that change over time. In Fig. 3, the first 10 segments, each representing approximately 2 seconds of EEG signal (256 samples @ 128 Hz sampling frequency) and

containing data from 19 independent EEG channels, are visually presented. The colour intensity used in the visualisation reflects the normalized signal amplitude distribution across the time series of each channel.

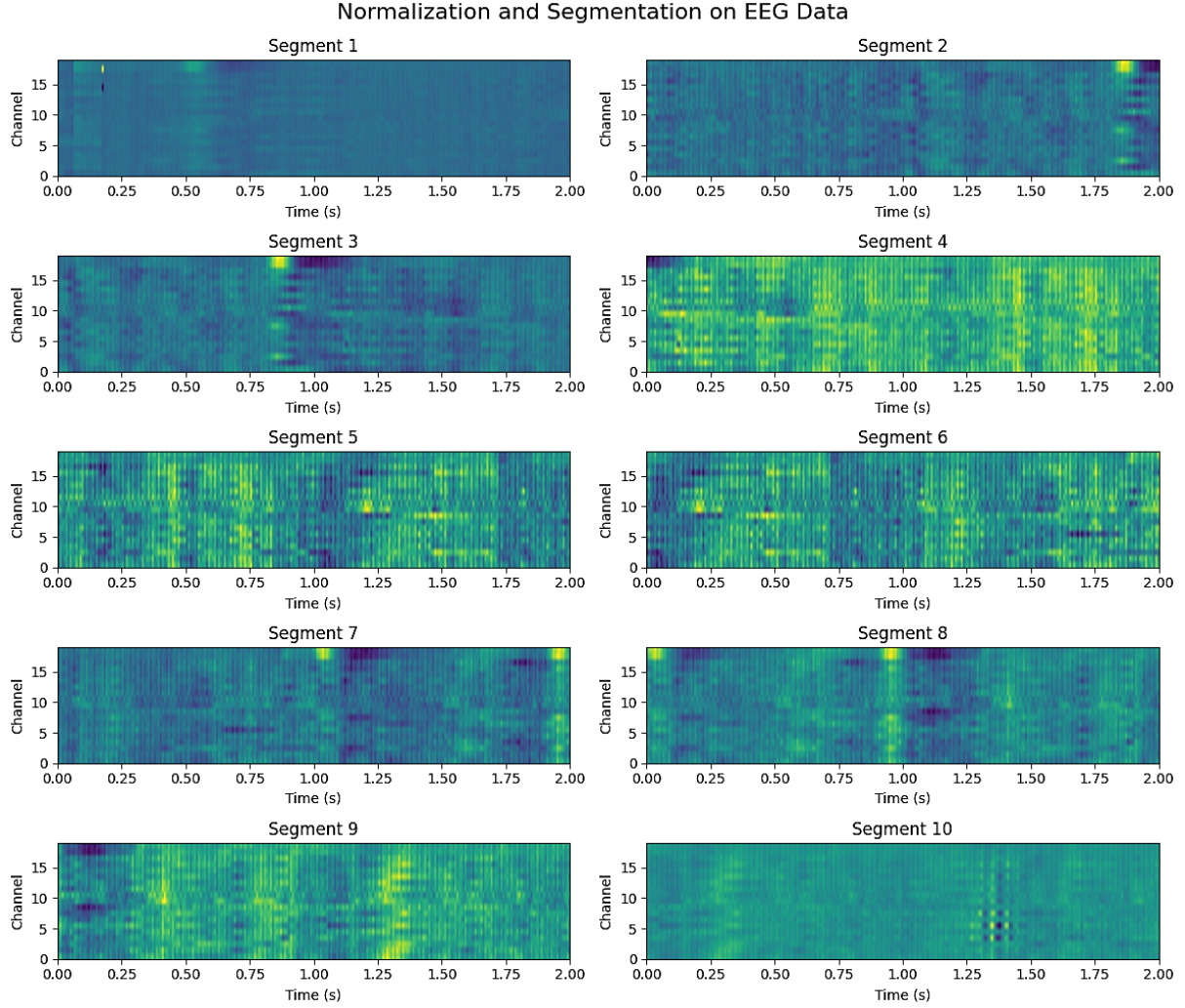


Fig. 3. Visualisation of EEG signals after z-score normalization and 50 % overlapping sliding-window segmentation into 256-sample segments.

III. PROPOSED EEGNET MODEL WITH AN INTEGRATED CHANNEL ATTENTION MECHANISM

In this study, we propose a new model for EEG signal classification called EEGNet-CA by integrating the CA mechanism into the lightweight CNN architecture EEGNet as shown in Fig. 4.

A. EEGNet

EEGNet is a compact CNN architecture specifically designed for the classification of EEG signals. Despite the low number of parameters, this proposed architecture exhibits acceptable decoding performance on EEG data, making it a suitable base model for EEG-based brain-computer interface applications [33].

The EEGNet architectural structure consists of consecutive convolutional layers that enable the extraction of both temporal and spatial features. The convolutional blocks of EEGNet

consist of standard Conv2D, Depthwise Conv2D and Separable Conv2D operations, respectively, and thanks to this structure, high classification performance can be achieved with a low number of parameters. While the first two layers capture time- and space-specific information, respectively, the third layer uses a depthwise separable convolution structure to improve parameter efficiency [34].

B. Channel Attention Mechanism

The CA mechanism can improve the feature extraction capacity of the network model with a small computation and parameter increase by providing a more effective representation of certain features. This mechanism aims to highlight meaningful details and neurophysiologically important information in the input feature map by learning the contribution of each channel.

The mechanism first obtains the global representation of each channel by applying the Global Average Pooling operation on the input feature map. This representation vector is used to calculate CA scores by passing it through two consecutive fully connected (Dense) layers. The first layer acts as a projection by reducing the number of channels by a certain ratio and is activated with the ReLU activation function. The second layer produces attention scores for each channel by expanding this low-dimensional representation back to the original channel size.

The obtained attention scores are normalized with the sigmoid activation function and rescaled by multiplying the input feature maps on a channel basis. This operation aims to increase the classification performance by allowing the model to focus on more information-rich and meaningful channels.

C. Proposed EEGNet-CA Model

The proposed model effectively extracts both spatial and temporal features from EEG data, while also having a structure that can learn the relative importance of each channel in terms of classification. The assignment of equal weights to all input channels of classical convolutional neural network layers can cause significant information loss, especially in data such as EEG, where each channel reflects different neurophysiological information. In order to overcome this problem and direct the

attention of the model to channels that are more critical for classification, the CA mechanism has been integrated into the architecture. In this way, the model aims to increase the classification performance by prioritizing channels with more information density.

In this study, a lightweight deep learning architecture based on EEGNet and strengthened with CA mechanism is proposed. The model takes EEG segments of size $(19 \times 256 \times 1)$ as input and extracts temporal and spatial features with Conv2D and DepthwiseConv2D layers, respectively. Then, an attention module similar to Squeeze-and-Excitation is integrated that learns channel-level importance weights. This structure processes the channel features summarised by global average pooling with fully connected layers, rescales the original feature maps and highlights the channels with high contribution to the classification. Temporal deepening is performed with the low-parameter SeparableConv2D layer, and the generalizability of the model is increased with the following regularization layers: batch normalization, ELU, average pooling, and dropout. The features obtained in the last layer are flattened and transmitted to the softmax activation Dense layer for two-class (ADHD - Control) classification. This structure, which contains approximately 2174 parameters, offers an effective solution in terms of both computational efficiency and feature representation in EEG-based classification tasks.

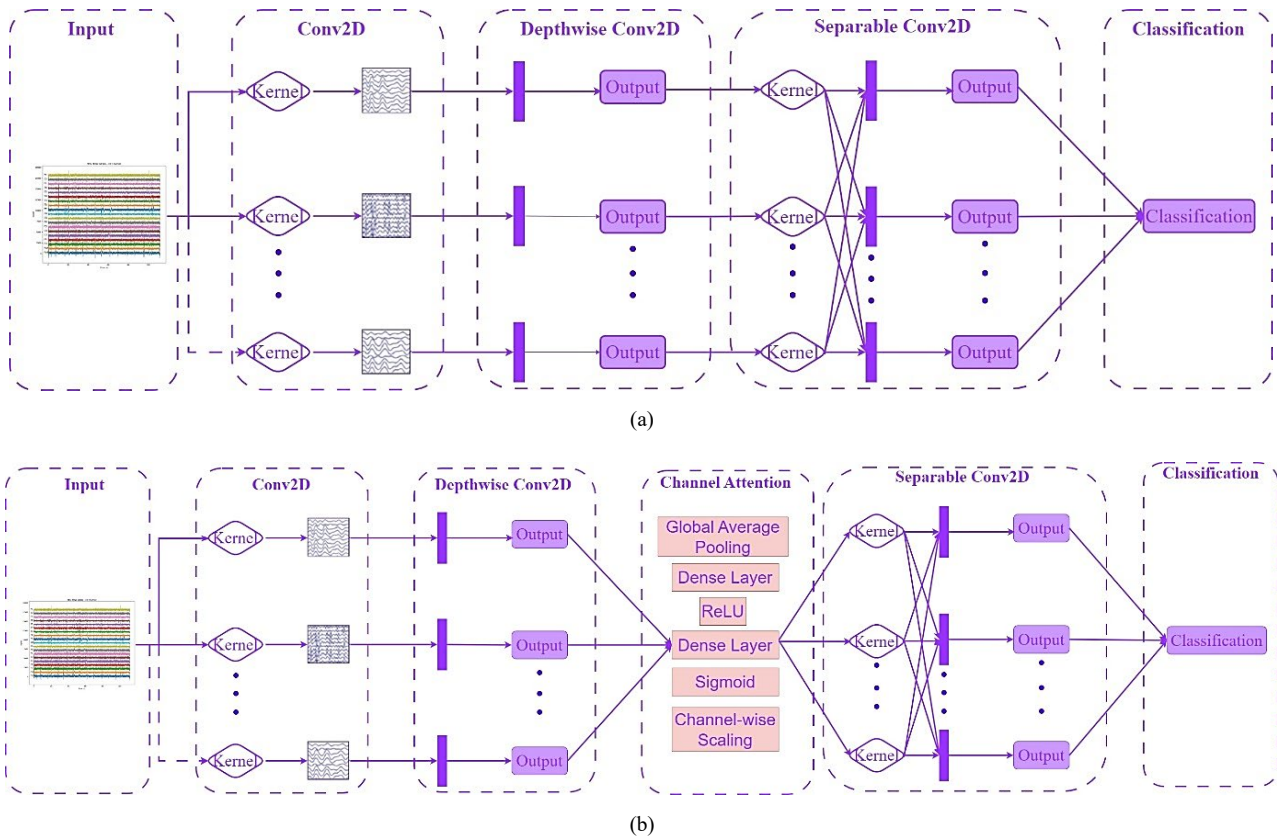


Fig. 4. Comparison between the classical EEGNet architecture and the proposed EEGNet-CA model with the integrated channel attention mechanism: (a) Classical EEGNet model, (b) Proposed EEGNet-CA mode.

IV. EXPERIMENTS

In this study, a deep learning-based classification model based on EEGNet and integrated with the CA mechanism was developed to distinguish between ADHD and the control group via EEG signals.

A. Experimental Setup

The experiments were conducted using the Google Colab Python development environment on a server running Windows operating system. This server was equipped with a Tesla K80 GPU with 12 GB GDDR5 VRAM, a dual-core Intel Xeon processor with a clock speed of 2.20 GHz, and 13 GB of RAM.

In this study, a parametrically lightweight and highly explainable deep neural network architecture integrated with a CA mechanism is proposed to perform two-class classification with multi-channel data obtained from EEG signals. The input of the model consists of three-dimensional tensors of size (19, 256, 1) (19, 256, 1) consisting of 19 EEG channels, each containing 256-time samples. In the first stage, Conv2D layer with (1×64) dimensional kernels is applied to extract temporal features, and learning is provided with 512 parameters using 8 filters. Following this, Batch Normalization is applied, and then DepthwiseConv2D layer is activated to model spatial relationships. This layer performs separate learning for each channel and contains a total of 304 parameters with MaxNorm regularization. Then, Batch Normalization and the ELU activation function are used once again. AveragePooling2D (1×4) is applied to reduce the length dimension, and Dropout is applied at a rate of 50 % to prevent over-learning.

The second stage of the model includes the CA mechanism. This mechanism starts by taking the global averages of the feature maps in the channel dimension. The obtained vector is passed through two fully connected dense layers containing 2 and 16 neurons, respectively, with a compression ratio of 8, and attention weights are calculated. This structure, which contains 82 parameters in total, learns the relationships between channels and emphasises the prominent information. The obtained attention weights are multiplied with the original feature maps at the channel level to obtain a refined representation.

In the section following the attention mechanism, the SeparableConv2D (1×16) layer is used to increase the temporal resolution and obtain more abstract representations, then Batch Normalization, ELU activation, AveragePooling2D (1×8) and Dropout layers are applied again, respectively. Finally, the extracted feature maps are flattened with the Flatten layer and transferred to the two-neuron, softmax activation Dense layer. This layer generates class probabilities and has a total of 258 parameters. The number of epochs is selected as 10. For classification, the dataset has been split into a training set and a test set with a ratio of 80 % and 20 %, respectively. While training set includes 96 subjects and test set includes 25 subjects. As EEG signals exhibit high correlation between segments belonging to the same subject, it is known that segment-level reported performance metrics may be optimistic and insufficient for direct clinical generalization. Therefore, to completely eliminate the risk of potential subject leakage, the

training and test sets were separated by subject, and the segmentation process was performed after this separation. This allowed the model to evaluate EEG data from subjects it had never seen before during the testing phase, preserving the model's true generalization ability.

As a result, the proposed model has only 1670 learnable parameters and has an efficient structure that can effectively process both temporal and spatial information, while also highlighting meaningful channels thanks to the attention mechanism. This architecture provides computational advantages with a low number of parameters, while also having the potential to increase classification performance.

B. Baseline Methods

We evaluate EEGNet-CA against the following models. The ANN classifier is a multilayer perceptron consisting of input nodes, hidden layer nodes, and output nodes [35]. 1D-CNN is a structure that includes convolution layers, pooling layers, dropout layers, and activation functions to process 1D data [36]. This structure allows for the automatic learning of distinctive temporal features from the raw signal and reduces the risk of overlearning. A 2-D CNN is a type of deep, feed-forward neural network composed of multiple convolutional layers, pooling layers, and fully connected layers [37]. Thanks to their multi-layered hierarchical structure, 2D-CNNs can model both the spatial relationships between EEG channels and the temporal dynamics of the signal. EEGNet is a compact CNN structure developed for classifying EEG signals [33]. By utilising depthwise and separable convolutional layers, it significantly reduces the number of parameters while maintaining the ability to learn temporal and spatial EEG features together. Thanks to these features, EEGNet offers high performance and generalization capabilities in EEG studies with limited datasets.

C. Evaluation Metrics

To evaluate the performance of the model, metrics such as precision, recall, accuracy, and F1 score are used. In this context, true positive (TP) represents the number of EEG signals correctly classified as ADHD. False positive (FP) represents the number of signals predicted as ADHD but not ADHD. False negative (FN) represents the number of EEG signals incorrectly classified as not ADHD despite being ADHD. True negative (TN) is the number of EEG signals correctly identified as non-ADHD. The formulas of these metrics calculated based on TP, FP, FN, and TN are presented in (2)–(5).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TF + TN} \quad (4)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

D. Evaluation of Results

In this study, comprehensive experimental and statistical evaluations were conducted to assess the effectiveness and efficiency of the proposed EEGNet-CA model. The performance of the model was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. In addition, the proposed approach was systematically compared with baseline methods such as conventional EEGNet, 1D CNN, 2D CNN, and ANN models to ensure a fair and meaningful performance assessment. To

further validate the reliability of the obtained results, non-parametric statistical tests were employed. The Friedman test was applied to identify overall performance differences among the models, followed by pairwise Wilcoxon signed-rank tests to determine statistically significant improvements achieved by the proposed EEGNet-CA architecture. These analyses provide strong evidence that the integration of the CA mechanism significantly enhances classification performance while maintaining a lightweight and computationally efficient model structure.

TABLE II

QUANTITATIVE COMPARISON OF THE PROPOSED EEGNET-CA MODEL WITH BASELINE DEEP LEARNING MODELS USING CLASSIFICATION PERFORMANCE METRICS AND STATISTICAL TESTS

Models	Precision	Recall	Accuracy	F1-score	Wilcoxon (p-value)	Friedman Sort
EEGNet-CA	0.9995	0.9978	0.9998	0.9979	0.0020	1
EEGNet	0.9626	0.9544	0.9548	0.9578		2
1D CNN	0.6818	0.7956	0.6793	0.7042		5
2 D CNN	0.6517	0.7619	0.6422	0.7439		4
ANN	0.6543	0.7581	0.6422	0.7036		3

Based on Table II, the proposed EEGNet-CA model achieved the highest performance, attaining an F1-score of 99.79 %. This indicates that incorporating an attention mechanism substantially enhances the model's ability to distinguish between classes. The conventional EEGNet model also performed well, reaching an F1-score of 95.78 %. In contrast, the 1D CNN, 2D CNN, and ANN models yielded lower F1-scores of 70.42 %, 74.39 %, and 70.36 %, respectively. These results suggest that simpler architectures are insufficient to fully capture the complex patterns inherent in EEG signals. The statistical analyses show that the proposed EEGNet-CA model provides a significant superiority in classification performance. This is because the CA mechanism allows the model to learn the relative importance of each EEG channel. Since each channel in EEG signals corresponds to a different region of the brain, the equal weighting learning strategy in classical CNNs can lead to information loss. Thanks to the attention mechanism integrated in this study, the model was able to create more discriminative representations by giving more weight to information-rich channels. This mechanism played a critical role especially in EEG signals where the distinction between classes is ambiguous.

As a result of the Friedman test, it was determined that there were statistically significant differences between the models ($\chi^2 = 37.84, p < 0.001$). In order to determine the source of these differences, pairwise Wilcoxon signed-rank tests revealed that the EEGNet-CA model performed statistically significantly better than the ANN, 2D CNN, 1D CNN and standard EEGNet models where p value is 0.002 for each comparison. These findings show that the attention mechanism strengthens the discrimination of the model and significantly increases the overall classification performance.

The confusion matrix given in Fig. 5 shows that the proposed EEGNet-based deep learning approach can distinguish EEG segments from ADHD and control groups with high accuracy.

In the segment-level evaluation, high sensitivity and specificity values were obtained for the ADHD class, and the false negative and false positive rates were observed to be quite low.

		Predicted	
		ADHD	Control
True	ADHD	2464	25
	Control	15	1472

Fig. 5. Confusion matrix demonstrating the classification performance of the proposed EEGNet-CA model on ADHD and control EEG segments.

TABLE III

CLASSIFICATION PERFORMANCE OF THE PROPOSED EEGNET-CA MODEL USING DIFFERENT EEG BRAIN REGIONS

Proposed Model	Regions	Channels	F1-Score
EEGNet-CA	Full	Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz	99.79
	Frontal	Fp1, Fp2, F3, F4, F7, F8, Fz	94.96
	Temporal	T7, T8	72.86
	Parietal	P3, P4, P7, P8, Pz	84.00

According to Table III, the proposed model exhibited high accuracy with an F1-score of 99.79 % when all EEG channels were used. When only frontal channels were used within the scope of regional analyses, the performance of the model remained quite high with an F1-score of 94.96 %. This finding shows that the frontal region, which is associated with attention and executive functions, is the main determinant in ADHD

classification. It has been frequently reported in the literature that ADHD is associated with structural and functional abnormalities in these regions [38]. This finding suggests that the proposed model produces neurophysiologically consistent outcomes, not only in terms of performance. On the other hand, the classification accuracy with channels belonging to the temporal region decreased significantly, and an F1-score of 72.86 % was obtained. This situation shows that the temporal lobe makes a limited contribution to ADHD discrimination. The temporal lobe is mostly associated with auditory processing and memory [39]; therefore, the contribution of this region may be limited in an EEG protocol based on a visual attention task. The F1-score of 84.00 % obtained with parietal region channels reveals that this region carries moderate information in classification. In general, although the use of all channels together provides the highest accuracy, the fact that high accuracy can be maintained with a limited number of channels belonging to the frontal region reveals that this region is neurophysiologically critical in ADHD diagnosis. These results offer important implications in terms of both system designs based on channel reduction and clinical applications.

V. CONCLUSION

In this study, we propose EEGNet-CA, a new model that integrates the EEGNet architecture with the CA mechanism to diagnose ADHD in children using EEG signals. The model efficiently extracts both temporal and spatial information, while learning the relative importance of each EEG channel in classification and directing attention to more meaningful features.

Experimental evaluations demonstrate that EEGNet-CA achieves a high classification performance, reaching an F1-score of 99.79 % with only 1670 learnable parameters, highlighting its efficiency in terms of both accuracy and computational complexity. Comparative analyses against conventional EEGNet, 1D CNN, 2D CNN, and ANN models reveal that the proposed approach provides statistically significant improvements in classification performance, as confirmed by Wilcoxon signed-rank tests ($p < 0.05$). Furthermore, regional channel analysis indicates that the model maintains a high level of performance when trained using frontal EEG channels alone, achieving an F1-score of 94.96 %, which underscores the neurophysiological relevance of frontal brain regions in ADHD diagnosis.

Overall, the findings of this study suggest that the integration of CA into compact EEG-based deep learning architectures constitutes an effective strategy for enhancing diagnostic performance while preserving model interpretability and computational efficiency. The proposed EEGNet-CA framework therefore represents a promising and clinically applicable approach for EEG-based ADHD diagnostic support systems.

REFERENCES

- [1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., 2013.
<https://doi.org/10.1176/appi.books.9780890425596>
- [2] D. Landínez-Martínez *et al.*, "Neuropsychological and academic performance in Colombian children with attention-deficit hyperactivity disorder: A comparative study with a control group," *Children*, vol. 12, no. 5, Apr. 2025, Art. no. 561. <https://doi.org/10.3390/children12050561>
- [3] G. Polanczyk *et al.*, "The worldwide prevalence of ADHD: a systematic review and metaregression analysis," *American Journal of Psychiatry*, vol. 164, no. 6, pp. 942–948, June 2007.
<https://doi.org/10.1176/ajp.2007.164.6.942>
- [4] T. Chen *et al.*, "Diagnosing attention-deficit hyperactivity disorder (ADHD) using artificial intelligence: a clinical study in the UK," *Frontiers in Psychiatry*, vol. 14, June 2023, Art. no. 1164433.
<https://doi.org/10.3389/fpsy.2023.1164433>
- [5] M. Mulraney *et al.*, "Systematic review and meta-analysis: screening tools for attention-deficit/hyperactivity disorder in children and adolescents," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 61, no. 8, pp. 982–996, Aug. 2021.
<https://doi.org/10.1016/j.jaac.2021.11.031>
- [6] J. Schein *et al.*, "Economic burden of attention-deficit/hyperactivity disorder among children and adolescents in the United States: a societal perspective," *Journal of Medical Economics*, vol. 25, no. 1, pp. 193–205, Feb. 2022. <https://doi.org/10.1080/13696998.2022.2032097>
- [7] F. Atban, E. Ekinci, and Z. Garip, "Traditional machine learning algorithms for breast cancer image classification with optimized deep features," *Biomedical Signal Processing and Control*, vol. 81, Mar. 2023, Art. no. 104534. <https://doi.org/10.1016/j.bspc.2022.104534>
- [8] Ş. Ay, E. Ekinci, and Z. Garip, "A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases," *Journal of Supercomputing*, vol. 79, pp. 11797–11826, Mar. 2023. <https://doi.org/10.1007/s11227-023-05132-3>
- [9] Z. Garip, E. Ekinci, K. Serbest, and S. Eken, "Chaotic marine predator optimization algorithm for feature selection in schizophrenia classification using EEG signals," *Cluster Computing*, vol. 27, pp. 11277–11297, May 2024. <https://doi.org/10.1007/s10586-024-04511-6>
- [10] Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," *Biomedical Engineering Online*, vol. 24, Feb. 2025, Art. no. 23. <https://doi.org/10.1186/s12938-025-01349-w>
- [11] E. Sarigedik *et al.*, "Diagnosis of attention deficit hyperactivity disorder with machine learning methods: A systemic review," *Turkish Journal of Child and Adolescent Mental Health*, vol. 31, no. 3, pp. 179–185, 2024. <https://doi.org/10.4274/tjcamh.galenos.2022.02486>
- [12] N. Ahire, R. N. Awale, and A. Wagh, "Electroencephalogram (EEG) based prediction of attention deficit hyperactivity disorder (ADHD) using machine learning," *Applied Neuropsychology: Adult*, vol. 32, no. 4, pp. 966–977, Aug. 2025.
<https://doi.org/10.1080/23279095.2023.2247702>
- [13] J. F. Lubar, "Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders," *Biofeedback and Self-Regulation*, vol. 16, no. 3, pp. 201–225, 1991.
<https://doi.org/10.1007/BF01000016>
- [14] E. Gkintoni, A. Vantarakis, and P. Gourzis, "Neuroimaging Insights into the public health burden of neuropsychiatric disorders: A systematic review of electroencephalography-based cognitive biomarkers," *Medicina*, vol. 61, no. 6, May 2025, Art. no. 1003.
<https://doi.org/10.3390/medicina61061003>
- [15] N. K. Ahire, "Attention-driven deep learning framework for EEG analysis in ADHD detection," *Applied Neuropsychology: Child*, pp. 1–11, May 2025. <https://doi.org/10.1080/21622965.2025.2512919>
- [16] 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, Dec. 2020, Art. no. 105738.
<https://doi.org/10.1016/j.cmpb.2020.105738>

- [17] A. Parashar, N. Kalra, J. Singh, and R. K. Goyal, "Machine learning based framework for classification of children with ADHD and healthy controls," *Intelligent Automation & Soft Computing*, vol. 28, no. 3, pp. 669–682, Apr. 2021. <https://doi.org/10.32604/iasc.2021.017478>
- [18] H. T. Tor *et al.*, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Computer Methods and Programs in Biomedicine*, vol. 200, Mar. 2021, Art. no. 105941. <https://doi.org/10.1016/j.cmpb.2021.105941>
- [19] Y. Hu *et al.*, "Functional connectivity anomalies in medication-naïve children with ADHD: Diagnostic potential, symptoms interpretation, and a mediation model," *Clinical Neurophysiology*, vol. 174, pp. 212–219, June 2025. <https://doi.org/10.1016/j.clinph.2025.04.011>
- [20] A. M. Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi, "EEG data for ADHD/Control children," IEEE DataPort, 2020. [Online]. Available: <https://iee-dataport.org/open-access/eeeg-data-adhd-control-children>
- [21] U. Hassan and A. Singhal, "Convolutional neural network framework for EEG-based ADHD diagnosis in children," *Health Information Science and Systems*, vol. 12, no. 1, Aug. 2024, Art. no. 44. <https://doi.org/10.1007/s13755-024-00305-7>
- [22] B. Latifi, A. Amini, and A. M. Nasrabadi, "Siamese based deep neural network for ADHD detection using EEG signal," *Computers in Biology and Medicine*, vol. 182, Nov. 2024, Art. no. 109092. <https://doi.org/10.1016/j.compbiomed.2024.109092>
- [23] Y. Sharma and B. K. Singh, "Attention deficit hyperactivity disorder detection in children using multivariate empirical EEG decomposition approaches: A comprehensive analytical study," *Expert Systems with Applications*, vol. 213, Mar. 2023, Art. no. 119219. <https://doi.org/10.1016/j.eswa.2022.119219>
- [24] B. TaghiBeyglou *et al.*, "Detection of ADHD cases using CNN and classical classifiers of raw EEG," *Computer Methods and Programs in Biomedicine Update*, vol. 2, 2022, Art. no. 100080. <https://doi.org/10.1016/j.cmpbup.2022.100080>
- [25] M. Bakhtyari and S. Mirzaei, "ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework," *Biomedical Signal Processing and Control*, vol. 76, July 2022, Art. no. 103708. <https://doi.org/10.1016/j.bspc.2022.103708>
- [26] A. Alim and M. H. Imtiaz, "Automatic identification of children with ADHD from EEG brain waves," *Signals*, vol. 4, no. 1, pp. 193–205, Feb. 2023. <https://doi.org/10.3390/signals4010010>
- [27] M. Z. Ullah, N. Ta, and D. Yu, "A weighted dispersion entropy based EEG analysis for ADHD diagnosis," *Biomedical Signal Processing and Control*, vol. 108, Oct. 2025, Art. no. 107993. <https://doi.org/10.1016/j.bspc.2025.107993>
- [28] M. Y. Esas and F. Latifoğlu, "Detection of ADHD from EEG signals using new hybrid decomposition and deep learning techniques," *Journal of Neural Engineering*, vol. 20, no. 3, June 2023, Art. no. 036028. <https://doi.org/10.1088/1741-2552/acc902>
- [29] T. Cai *et al.*, "Topological feature search method for multichannel EEG: Application in ADHD classification," *Biomedical Signal Processing and Control*, vol. 100, Feb. 2025, Art. no. 107153. <https://doi.org/10.1016/j.bspc.2024.107153>
- [30] H. Jahani and A. A. Safaei, "Efficient deep learning approach for diagnosis of attention-deficit/hyperactivity disorder in children based on EEG signals," *Cognitive Computation*, vol. 16, no. 5, pp. 2315–2330, May 2024. <https://doi.org/10.1007/s12559-024-10302-3>
- [31] Ö. Kasim, "Identification of attention deficit hyperactivity disorder with deep learning model," *Physical and Engineering Sciences in Medicine*, vol. 46, pp. 1081–1090, May 2023. <https://doi.org/10.1007/s13246-023-01275-y>
- [32] S. Aggarwal, N. Chugh, and A. Balyan, "Identification of ADHD disorder in children using EEG based on visual attention task by ensemble deep learning," in *Proc. Int. Conf. Data Sci. Appl.: ICDSA 2022*, vol. 2, pp. 243–259, Singapore: Springer Nature Singapore, Feb. 2023. https://doi.org/10.1007/978-981-19-6634-7_18
- [33] V. J. Lawhern *et al.*, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, 2018, Art. no. 056013. <https://doi.org/10.1088/1741-2552/aace8c>
- [34] M. Zuo, X. Y. Chen, and L. Sui, "A novel STA-EEGNet combined with channel selection for classification of EEG evoked in 2D and 3D virtual reality," *Medical Engineering & Physics*, 2025, Art. no. 104363. <https://doi.org/10.1016/j.medengphy.2025.104363>
- [35] P. P. Mini, T. Thomas, and R. Gopikakumari, "EEG based direct speech BCI system using a fusion of SMRT and MFCC/LPCC features with ANN classifier," *Biomedical Signal Processing and Control*, vol. 68, July 2021, Art. no. 102625. <https://doi.org/10.1016/j.bspc.2021.102625>
- [36] E. U. H. Qazi, A. Almorjan, and T. Zia, "A one-dimensional convolutional neural network (1D-CNN) based deep learning system for network intrusion detection," *Applied Sciences*, vol. 12, no. 16, Aug. 2022, Art. no. 7986. <https://doi.org/10.3390/app12167986>
- [37] J. Gan, W. Wang, and K. Lu, "A new perspective: Recognizing online handwritten Chinese characters via 1-dimensional CNN," *Information Sciences*, vol. 478, pp. 375–390, Apr. 2019. <https://doi.org/10.1016/j.ins.2018.11.035>
- [38] M. Wang, J. Yu, H. D. Kim, and A. B. Cruz, "Neural correlates of executive function and attention in children with ADHD: An ALE meta-analysis of task-based functional connectivity studies," *Psychiatry Research*, vol. 345, Mar. 2025, Art. no. 116338. <https://doi.org/10.1016/j.psychres.2024.116338>
- [39] M. Dhoisne *et al.*, "SEEG guided hippocampus-sparing resection in mesial temporal lobe epilepsy," *Neurophysiologie Clinique*, vol. 55, no. 3, June 2025, Art. no. 103073. <https://doi.org/10.1016/j.neucli.2025.103073>

Zeynep Garip received the PhD degree from the Department of Mechatronics Engineering, Sakarya University, Türkiye, in 2018. She is currently an Associate Professor with Sakarya University of Applied Sciences, Türkiye. Her current research interests include artificial intelligence, robotics, and image processing.

E-mail: zbatik@subu.edu.tr

ORCID iD: <https://orcid.org/0000-0002-0420-8541>

Ekin Ekinci received the PhD degree from the Department of Computer Engineering, Kocaeli University, Türkiye, in 2019. She is currently an Associate Professor with Sakarya University of Applied Sciences, Türkiye. Her current research interests include artificial intelligence, machine learning, and natural language processing.

E-mail: ekinekinici@subu.edu.tr

ORCID iD: <https://orcid.org/0000-0003-0658-592X>