

Article

Electroencephalogram-Based ConvMixer Architecture for Recognizing Attention Deficit Hyperactivity Disorder in Children

Min Feng ^{1,2,*}  and Juncai Xu ^{3,*} ¹ Nanjing Rehabilitation Medical Center, The Affiliated Brain Hospital, Nanjing Medical University, Nanjing 210029, China² School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210024, China³ School of Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

* Correspondence: minfeng@njmu.edu.cn (M.F.); juncai.xu@gmail.com (J.X.)

Abstract: Attention deficit hyperactivity disorder (ADHD) is a neuro-developmental disorder that affects approximately 5–10% of school-aged children worldwide. Early diagnosis and intervention are essential to improve the quality of life of patients and their families. In this study, we propose ConvMixer-ECA, a novel deep learning architecture that combines ConvMixer with efficient channel attention (ECA) blocks for the accurate diagnosis of ADHD using electroencephalogram (EEG) signals. The model was trained and evaluated using EEG recordings from 60 healthy children and 61 children with ADHD. A series of experiments were conducted to evaluate the performance of the ConvMixer-ECA. The results showed that the ConvMixer-ECA performed well in ADHD recognition with 94.52% accuracy. The incorporation of attentional mechanisms, in particular ECA, improved the performance of ConvMixer; it outperformed other attention-based variants. In addition, ConvMixer-ECA outperformed state-of-the-art deep learning models including EEGNet, CNN, RNN, LSTM, and GRU. t-SNE visualization of the output of this model layer validated the effectiveness of ConvMixer-ECA in capturing the underlying patterns and features that separate ADHD from typically developing individuals through hierarchical feature learning. These outcomes demonstrate the potential of ConvMixer-ECA as a valuable tool to assist clinicians in the early diagnosis and intervention of ADHD in children.

Keywords: attention deficit hyperactivity disorder (ADHD); ConvMixer; efficient channel attention (ECA); electroencephalogram (EEG); deep learning; early diagnosis



Citation: Feng, M.; Xu, J. Electroencephalogram-Based ConvMixer Architecture for Recognizing Attention Deficit Hyperactivity Disorder in Children. *Brain Sci.* **2024**, *14*, 469. <https://doi.org/10.3390/brainsci14050469>

Academic Editors: Evanthia Bernitsas, Jesús Pastor and Lorena Vega-Zelaya

Received: 9 April 2024

Revised: 29 April 2024

Accepted: 6 May 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by persistent inattention, hyperactivity, and impulsivity that adversely affects daily functioning and development [1,2]. ADHD affects approximately 5–10% of school-aged children worldwide [3,4], and early diagnosis and intervention are essential to improve the quality of life of patients and their families [5,6]. However, traditional diagnostic methods such as clinical interviews, behavioral observations, and rating scales can be subjective and time-consuming [7–9], highlighting the need for objective and valid diagnostic tools.

Recent advances in machine learning, particularly deep learning, have shown promise in utilizing electroencephalography (EEG) signals to diagnose ADHD [10,11]. EEG is a non-invasive and inexpensive method of measuring electrical activity in the brain, and individuals with ADHD exhibit distinctive patterns compared to typically developing (TD) individuals [12]. Various deep learning architectures including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks have been applied to EEG-based ADHD diagnosis with promising results [13–17].

However, these methods often face the challenge of effectively capturing the spatial and channel dependence of EEG signals, which is crucial for accurate diagnosis [18]. Some studies have explored the use of EEG analysis in children with attention-deficit/hyperactivity disorder (ADHD), utilizing various techniques such as spatial normalization, smoothing, and multivariate empirical decomposition [19–21]. These studies highlight the lack of biomarkers for ADHD diagnosis and the complexity and challenges associated with capturing the spatial and channel dependence of EEG signals in this disorder. Recent surveys indicate the potential of deep learning techniques in uncovering complex patterns and biomarkers associated with mental disorders from non-invasive brain signals such as EEG [22]. The growing interest in machine learning approaches for precision psychiatry aims to personalize diagnosis and treatment based on individual characteristics and brain activity patterns [23]. Moreover, findings on the use of resting-state EEG functional connectivity to identify symptom dimensions in autism suggest potential parallels to the use of EEG-based deep learning models for characterizing ADHD subtypes or symptom dimensions [24]. These recent advances in EEG and deep learning for mental disorders contextualize the current study within the research landscape and emphasize the importance of developing novel deep learning architectures that can effectively capture the spatial and temporal dependencies in EEG signals for accurate ADHD diagnosis.

ConvMixer, a recently proposed architecture that combines deep and point-by-point convolution [25], has shown great potential in various computer vision tasks due to its ability to efficiently capture spatial and channel information [26–28]. Despite its success in computer vision, ConvMixer has not been applied to EEG-based ADHD diagnosis. In addition, attention mechanisms such as efficient channel attention (ECA) have gained significant momentum in deep learning due to their ability to focus on relevant features and suppress irrelevant ones [29,30]. Combining ECA with ConvMixer has the potential to improve the model's ability to capture discriminative patterns in EEG data for ADHD diagnosis.

To address these research gaps, this study proposes ConvMixer-ECA, a novel deep learning architecture that combines ConvMixer with ECA blocks for the accurate diagnosis of ADHD using EEG signals. The incorporation of the ECA blocks allows the model to adaptively recalibrate the channel features and improves its ability to capture discriminative patterns in EEG data. The main contributions of the work are as follows:

- Proposing ConvMixer-ECA, a novel deep learning architecture that combines ConvMixer with ECA blocks for the accurate EEG-based diagnosis of ADHD.
- Conduct extensive experiments to evaluate the performance of ConvMixer-ECA and demonstrate its superior accuracy compared to state-of-the-art deep learning models.
- Investigate the impact of different attentional mechanisms, in particular ECA, on the performance of ConvMixer and highlight the effectiveness of ECA in improving categorization performance.
- Insights into the feature learning process of ConvMixer-ECA are provided through t-distributed stochastic neighbor embedding (t-SNE) visualization, validating its ability to capture discriminative patterns in EEG data for ADHD diagnosis.

The rest of the paper is organized as follows. Section 2 describes the methodology including the ConvMixer-ECA architecture and its components, with a focus on the integration of ECA blocks. Section 3 describes the experiments and results including the training process, performance evaluation, impact of the attention mechanism, and comparative analysis with other models. Section 4 discusses the results, implications, and limitations of the study. Finally, Section 5 summarizes the paper and outlines future research directions.

2. Principles and Methodology

2.1. Participants and Data Preprocessing

This study employed an openly accessible EEG dataset obtained from IEEE DataPort. The sample included 121 children aged 7–12 years, comprising 61 children diagnosed with ADHD and 60 children serving as TD controls [31]. The cohort included both male and

female participants. The ADHD group consisted of 48 boys and 13 girls with a mean age of 9.62 ± 1.75 years, while the TD control group included 50 boys and 10 girls with a mean age of 9.85 ± 1.77 years. The ADHD diagnoses were confirmed by a qualified psychiatrist in accordance with the criteria outlined in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., DSM-IV) [32]. The participants in the ADHD group had been undergoing treatment with Ritalin for a period not exceeding 6 months prior to their involvement in the study. On the other hand, the control group consisted of individuals without psychiatric disorders, epilepsy, or any reported involvement in high-risk behaviors. This group was drawn from two sources: 50 boys were selected from a primary school, while 10 girls were chosen from an all-girls' primary school. Following assessment by a child and adolescent psychiatrist, it was confirmed that none of the children in the control group displayed any psychiatric issues.

EEG recordings were performed with the International 10–20 system, consisting of 19 channels: Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, and O2 (Figure 1) [33]. Recordings were sampled at a frequency of 128 Hz and reference electrodes A1 and A2 were placed on the participants' earlobes. The EEG recording protocol consisted of tasks designed to elicit visual attention, thus assessing the main deficits in children with ADHD. During the task, participants were presented with images of cartoon characters and asked to count them (Figure 2). The number of characters presented ranged from 5 to 16, and the image size was calibrated to ensure legibility. Images were presented sequentially, and each subsequent image was displayed immediately after the participant responded, with the duration of the EEG recording session depending on the participant's performance and response rate.

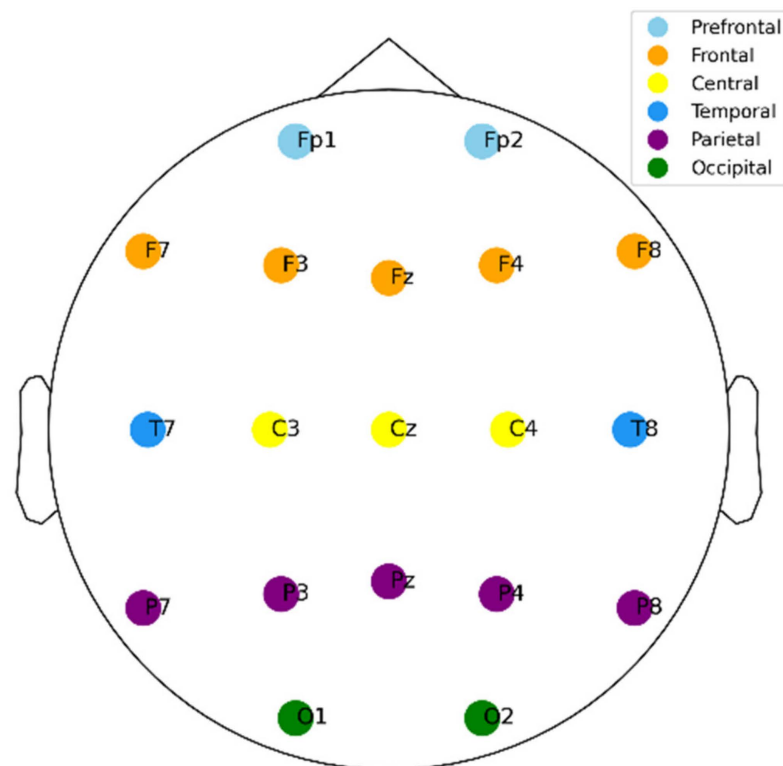


Figure 1. The 19-channel EEG recording based on the International 10–20 system.

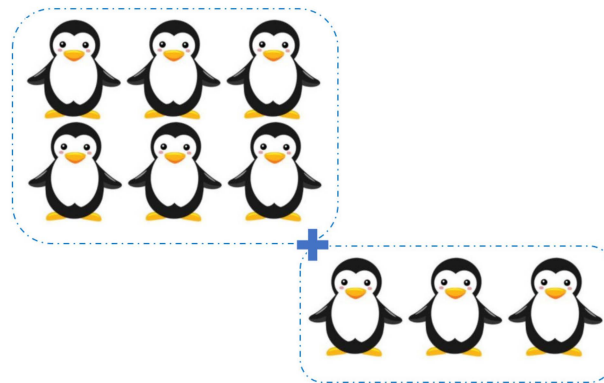


Figure 2. An example of the pictures displayed to the children.

Preprocessing is critical in EEG signal processing to ensure data quality and reliability for subsequent analysis by removing artifacts, reducing noise, and separating relevant frequency bands. This study focused on two key preprocessing techniques: the Butterworth bandpass filter and the segmentation process. EEG signals were subjected to preprocessing using a sixth-order Butterworth bandpass filter with a frequency range of 4–45 Hz. This was conducted in order to remove low-frequency drift and high-frequency noise while preserving the relevant EEG frequency bands. The filter coefficients were determined based on the filter order and cutoff frequencies. The lower and upper cutoff frequencies were set at 4 Hz and 45 Hz, respectively. Following filtration, the continuous EEG signals were divided into 4-s epochs, with each epoch representing a specific temporal window of the EEG recording.

The Butterworth bandpass filter is a widely used technique in signal processing to extract a specific range of frequencies from a signal. In EEG analysis, it removes unwanted frequencies and artifacts that may interfere with the desired brain activity. The transfer function of a sixth-order Butterworth bandpass filter is given by:

$$H(s) = \frac{\omega_{c1} \cdot \omega_{c2}}{s^6 + a_1 \cdot s^5 + a_2 \cdot s^4 + a_3 \cdot s^3 + a_4 \cdot s^2 + a_5 \cdot s + a_6} \quad (1)$$

where ω_{c1} and ω_{c2} are the lower and upper cutoff frequencies, and a_1, a_2, a_3, a_4, a_5 , and a_6 are the filter coefficients determined by the filter order and cutoff frequencies. In this study, a Butterworth bandpass filter was designed with a frequency range of 4–45 Hz. This range was selected to remove low-frequency drift and high-frequency noise while preserving the relevant EEG frequency bands. The rationale for this approach was to capture the primary activity associated with the P300 component and other cognitive processes of interest in the study of ADHD, while minimizing the influence of artifacts [34,35].

The segmentation process divides a continuous EEG signal into manageable segments or epochs; this process is essential for further analysis and feature extraction. The segmentation process can be mathematically represented as:

$$x_i(n) = x(n + (i - 1) \cdot N), \text{ for } n = 1, 2, \dots, N \quad (2)$$

where $x_i(n)$ represents the i -th segment of the EEG signal, $x(n)$ is the original EEG signal, and N is the number of samples in each segment, determined by the sampling frequency and segment duration.

In this study, the dataset was divided into a training set and a testing set, each of which contained EEG epochs from two distinct groups of subjects. Notably, the training and testing epochs were extracted from different subjects to ensure that there was no overlap between the sets. Each EEG signal was segmented into 4-s epochs. The preprocessing steps cleaned, filtered, and prepared the EEG dataset for further analysis. The Butterworth bandpass filter effectively removed unwanted frequencies and artifacts, and the segmenta-

tion process divided the continuous EEG signal into manageable epochs, ready for feature extraction and classification tasks.

2.2. ConvMixer Architecture

The ConvMixer architecture is a groundbreaking approach to visual modeling that seamlessly integrates the strengths of CNNs and transformers. This innovative architecture employs a series of strategic design choices to effectively process and learn from visual data. The key strategies adopted by ConvMixer include patch embedding, a ConvMixer layer, a mixing mechanism, layer stacking, and global pooling [25].

Patch embedding: The input image is divided into fixed-size patches, which are then mapped to a high-dimensional space. This process is mathematically represented as:

$$N = \frac{HW}{P^2} \quad (3)$$

where H , W , and P denote the height, width, and patch size of the image, respectively. Each patch is flattened and linearly projected using an embedding matrix W_e to obtain the embedded patch matrix E :

$$E = \text{Reshape}(X) \cdot W_e \quad (4)$$

where $X \in \mathbb{R}^{H \times W \times C}$ represents the input image, $\text{Reshape}(\cdot)$ represents the operation of dividing the image into patches and flattening them, and $W_e \in \mathbb{R}^{(P^2 C) \times D}$ represents the embedding matrix.

ConvMixer layer: The core building block of ConvMixer is a ConvMixer layer, which consists of depthwise separable convolutions and pointwise convolutions. Depthwise convolutions capture local spatial relationships within each channel of the input tensor X :

$$X_d = \text{DepthwiseConv}(X, W_d) \quad (5)$$

where X represents the input tensor, W_d represents the depthwise convolution kernel, and X_d represents the output of the depthwise convolution. Pointwise convolutions, applied subsequently, mix information across channels:

$$X_p = \text{PointwiseConv}(X_d, W_p) \quad (6)$$

where W_p represents the pointwise convolution kernel, and X_p represents the output of the pointwise convolution.

Mixing mechanism: ConvMixer employs a mixing mechanism that applies depthwise separable convolutions and pointwise convolutions independently to each patch:

$$p_i^d = \text{DepthwiseConv}(p_i, W_d) \quad (7)$$

$$p_i^p = \text{PointwiseConv}(p_i^d, W_p) \quad (8)$$

where p_i represents the i -th patch, p_i^d denotes the output of the depthwise convolution, and p_i^p represents the output of the pointwise convolution. The mixed output is represented as:

$$P = [p_1^p; p_2^p; \dots; p_N^p] \quad (9)$$

where $P \in \mathbb{R}^{N \times D}$ represents the matrix of mixed patches.

Layer stacking: To learn hierarchical feature representations, ConvMixer stacks multiple ConvMixer layers together:

$$X^{(l+1)} = \text{ConvMixerLayer}(X^{(l)}) \quad (10)$$

where $X^{(l)}$ represents the output of the l -th ConvMixer layer, and $X^{(l+1)}$ represents the output of the $(l + 1)$ -th ConvMixer layer.

Global pooling and classification: After processing through multiple ConvMixer layers, the model performs global average pooling (GAP) on the features of all patches to obtain a pooled feature vector:

$$f = \frac{1}{N} \sum_{i=1}^N p_i^p \quad (11)$$

where p_i^p represents the pointwise convolution output of the i -th patch, and $f \in \mathbb{R}^D$ represents the pooled feature vector. Finally, the pooled features are fed into a classifier for prediction.

2.3. ECA Mechanism

The ECA mechanism is a groundbreaking approach in deep learning that refines channel-wise attention within CNNs while maintaining a balance between model complexity and performance [29]. ECA employs several key strategies.

GAP: GAP is applied to the input feature map $X \in \mathbb{R}^{C \times H \times W}$ to obtain a vector $z \in \mathbb{R}^C$ that encapsulates channel-wise information:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_{c,i,j} \quad (12)$$

Adaptive kernel size determination: ECA determines the kernel size for the one-dimensional convolution based on the number of channels C :

$$k = \psi\left(\frac{C}{\gamma}\right) + 1 \quad (13)$$

where $\psi(\cdot)$ is the rounding operation and γ is a hyperparameter.

Convolution: A convolution with kernel size k is applied to z to generate attention weights $\hat{z} \in \mathbb{R}^C$:

$$\hat{z}_c = \sum_{i=1}^k w_i \cdot z_{\text{mod}(c+i-\lfloor \frac{k}{2} \rfloor - 1, C) + 1} \quad (14)$$

where w_i are learnable convolution weights.

Sigmoid activation: A sigmoid activation function is applied to \hat{z} to obtain the final attention weights $a \in \mathbb{R}^C$:

$$a_c = \frac{1}{1 + e^{-\hat{z}_c}} \quad (15)$$

Channel-wise attention scaling: The input feature map X is scaled with the attention weights a to obtain a refined feature map $\tilde{X} \in \mathbb{R}^{C \times H \times W}$:

$$\tilde{X}_{c,i,j} = a_c \cdot X_{c,i,j} \quad (16)$$

ECA enhances the representational capacity of CNNs while maintaining efficiency. The mathematical formulations provide a clear understanding of the operations involved, facilitating integration into various CNN architectures for improved feature learning and task-specific adaptability.

2.4. Implementation of ConvMixer-ECA

ConvMixer-ECA combines the ConvMixer model with the ECA mechanism to enhance the model's ability to capture and refine channel-wise attention. The ConvMixer-ECA architecture consists of patch embedding, a stack of ConvMixer blocks with integrated ECA blocks, GAP, layer normalization, and a fully connected layer for classification.

Patch embedding: The input tensor $X \in \mathbb{R}^{B \times C \times L}$ is divided into patches and projected into a higher-dimensional space using a convolution operation:

$$X_p = \text{Conv}(X, W_p, p) \quad (17)$$

where $X_p \in \mathbb{R}^{B \times C_p \times L_p}$ is the patch embedding, $W_p \in \mathbb{R}^{C_p \times C \times p}$ is the patch embedding kernel with size p , and $\text{Conv}(\cdot)$ denotes the convolution operation.

ConvMixer block with ECA: Each ConvMixer block consists of a depthwise convolution, Gaussian error limit unit (GELU) activation, pointwise convolution, and an ECA block for channel-wise attention refinement:

$$X' = X + \text{ECABlock}(\text{Conv}(\text{GELU}(\text{DepthwiseConv}(X, W_d)), W_p)) \quad (18)$$

where $X' \in \mathbb{R}^{B \times C_p \times L_p}$ is the output of the ConvMixer block, $W_d \in \mathbb{R}^{C_p \times 1 \times k}$ is the depthwise convolution kernel with size k , $W_p \in \mathbb{R}^{C_p \times C_p \times 1}$ is the pointwise convolution kernel, and $\text{ECABlock}(\cdot)$ represents the ECA block.

The ECA block applies GAP, convolution, and sigmoid activation to generate channel-wise attention weights:

$$z = \frac{1}{L} \sum_{i=1}^L X_i \quad (19)$$

$$\hat{z} = \text{Conv}(z, W, k) \quad (20)$$

$$a = \sigma(\hat{z}) \quad (21)$$

where $z \in \mathbb{R}^{B \times C \times 1}$ is the channel-wise descriptor, $\hat{z} \in \mathbb{R}^{B \times C \times 1}$ is the output of the convolution, $a \in \mathbb{R}^{B \times C \times 1}$ represents the channel-wise attention weights, and $\sigma(\cdot)$ is the sigmoid activation function.

GAP and classification: The output of the ConvMixer blocks is subjected to GAP to obtain a global representation:

$$y = \frac{1}{L_p} \sum_{i=1}^{L_p} X'_i \quad (22)$$

where y is the global representation.

The global representation is then normalized using layer normalization and passed through a fully connected layer for classification:

$$\hat{y} = \text{LayerNorm}(y) \quad (23)$$

$$\hat{y} = \text{FC}(\hat{y}, W_{fc}) \quad (24)$$

where $\hat{y} \in \mathbb{R}^{B \times N}$ is the output of the fully connected layer, $W_{fc} \in \mathbb{R}^{C_p \times N}$ is the fully connected weight matrix, and N is the number of classes.

ConvMixer-ECA uses deep convolution, point-by-point convolution, and ECA blocks to capture local and global dependencies while efficiently focusing on correlated channels. Mathematical formulas clearly illustrate the operations and transformations used to process and optimize the input data to produce accurate predictions.

As shown in Figure 3, ConvMixer-ECA extends the ConvMixer architecture by combining ECA blocks. The model consists of a patch embedding layer, followed by a series of mixer blocks, and finally a classification layer. While ConvMixer-ECA is typically implemented at depth 8, Figure 3 shows the model at depth 1 for clarity.

The patch embedding layer is realized by a one-dimensional (1D) convolutional layer that transforms the input tensor into patches and projects them into a higher dimensional space. The heart of the model lies in the mixer blocks, each of which consists of three key components: a deep convolutional layer for spatial mixing, a point-by-point convolutional layer for channel mixing, and an ECA block for channel feature recalibration. The ECA block uses adaptive mean pooling and a 1D convolutional layer to capture cross-channel interactions and generate attention weights that are multiplied by the input features to emphasize the most relevant channels.

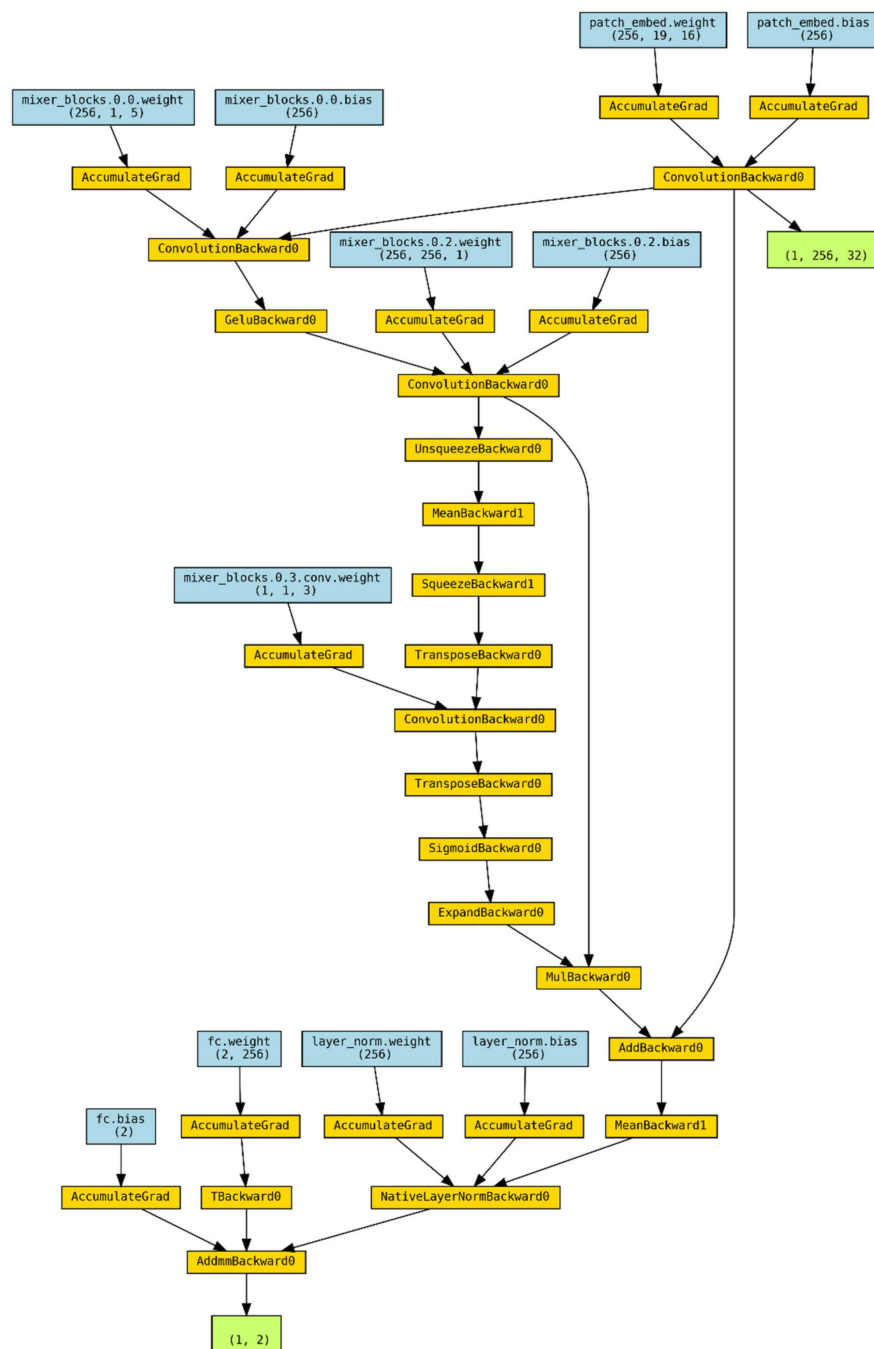


Figure 3. The ConvMixer-ECA architecture.

After the mixing block, the feature map is subjected to global average pooling (GAP) and layer normalization to generate a compact and normalized representation. Finally, a fully connected layer generates output logits for classification.

ConvMixer-ECA is initialized with several key hyperparameters including the number of input channels, number of categories, patch size, hidden dimension, depth, and kernel size. During the forward pass, the inputs are processed sequentially through the patch embedding layer, the mixer block, the GAP, the layer normalization, and the fully connected layer.

In essence, ConvMixer-ECA seamlessly combines the strengths of the ConvMixer architecture with the capabilities of the ECA block. This synergistic combination enables the model to efficiently process input data through deep and point-by-point convolution while

enhancing its ability to capture and exploit channel dependencies. The final architecture strikes a balance between computational efficiency and representational power, making it well-suited for a wide range of computer vision tasks.

3. Experiments and Results

3.1. Training ConvMixer-ECA

To ensure that there was no overlap between the training and test sets, data from different subjects were assigned to each set. The training set came from 80% of the subjects and consisted of 3352 samples, while the test set came from the remaining 20% of the subjects and consisted of 821 samples. A total of 70% of the training data was used to train ConvMixer-ECA, and the remaining 30% of the data was used to validate the training process. Although the proportions of the training and validation protocols were the same, different data segments were selected for each training cycle due to the random interruption of the data.

Binary cross entropy (BCE) was used as the loss function for training, and the Adam optimizer was used to determine the weights and filters for the network, minimizing the loss function using a back-propagation technique with a learning rate of 0.0001. The total number of cycles was set to 30, and the number of iterations per cycle depended on the batch size of 64, as described above.

To evaluate the training performance on the training and validation sets, the loss and accuracy of the validation set were calculated at the end of each cycle and in the final iteration. All implementations were performed on an *NVIDIA V100 GPU* (NVIDIA Corporation, Santa Clara, CA, USA) with *32 GB of DDR4 RAM* (Samsung Group, Seoul, Republic of Korea) running *Ubuntu 20.04.1 LTS* (Canonical Ltd., London, UK) and *Python 3.11.6* (Python Software Foundation, Wilmington, NC, USA).

The training results show that the model exhibited good performance and convergence over 30 cycles (Figure 4). The steady decrease in training loss from about 0.7 to 0.1, along with a corresponding decrease in validation loss, indicates that the model learns effectively from the training data and generalizes well to unseen data. The continued improvement in training and validation accuracy to 0.98 and 0.95, respectively, further highlights the model's success in learning to correctly categorize samples. Overall, the results show a well-trained model that performed well on both the training and validation sets, demonstrating its effectiveness in learning from the given data and generalizing to unseen samples.

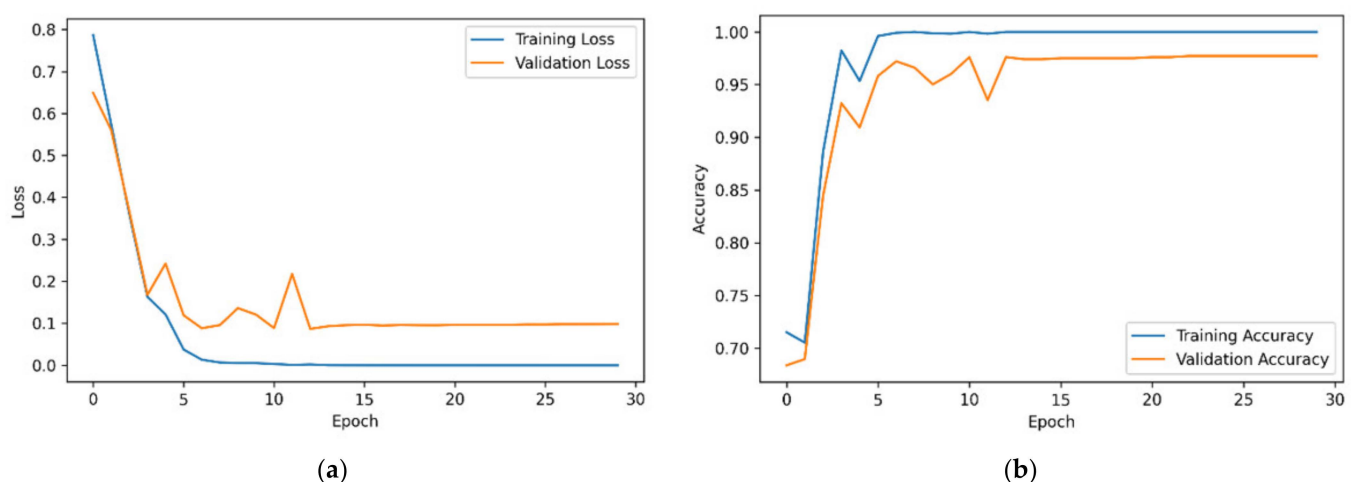


Figure 4. Training and validation of the ConvMixer-ECA model: (a) loss and (b) accuracy across the epochs.

3.2. Results and Analysis

The performance of ConvMixer-ECA was evaluated with several metrics including the accuracy, the F1 score, recall, and precision [36]. These metrics are defined by the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (25)$$

$$Precision = \frac{TP}{TP + FP} \quad (26)$$

$$Recall = \frac{TP}{TP + FN} \quad (27)$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (28)$$

where TP is the number of instances correctly predicted as positive, TN is the number of instances correctly predicted as negative, FP is the number of instances incorrectly predicted as positive, and FN is the number of instances incorrectly predicted as negative.

The performance of the trained ConvMixer-ECA model was evaluated on a test set consisting of 821 samples from 20% of the subjects not used for training. The model achieved 94.52% accuracy, 94.69% precision, 96.40% recall, and 95.54% F1 score. These metrics indicate strong performance, with the model correctly identifying the majority of positive instances while maintaining a low false positive rate.

The confusion matrix shown in Figure 5 provides a detailed assessment of the performance of the trained ConvMixer-ECA model in classifying ADHD and TD individuals. The model demonstrates strong classification ability with a true positive rate of 96.4% for the ADHD cases and a true negative rate of 91.59% for the TD cases. The low false negative rate of 3.6% indicates the high sensitivity of the model in detecting ADHD cases. The ConvMixer-ECA architecture demonstrated its effectiveness in capturing the distinguishing features of ADHD as evidenced by the high accuracy in correctly identifying ADHD cases. The model's ability to discriminate between ADHD and non-ADHD cases is further supported by its high true negative rate.

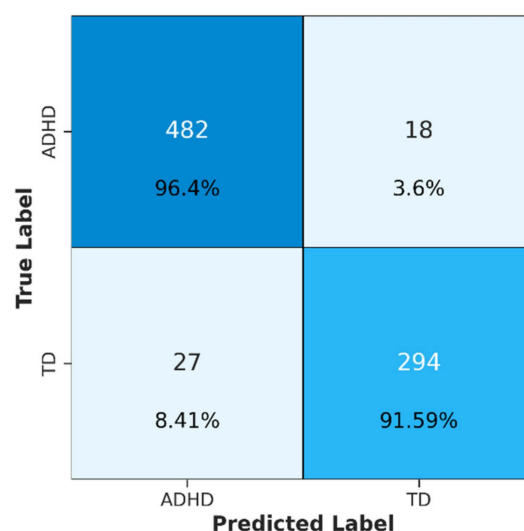


Figure 5. The confusion matrix for ADHD and TD classification.

To enhance the robustness of the performance evaluation, we implemented a 5-fold cross-validation procedure in testing the ConvMixer-ECA model. The dataset was randomly partitioned into five equally sized folds, ensuring a balanced representation of both ADHD and TD subjects. In each iteration, four folds were used for training, and the remaining fold was used for testing, with the process repeated five times. The accuracy results were obtained from the 5-fold cross-validation (Table 1), yielding an average accuracy of

0.9290. The consistent accuracy values across the different folds demonstrated the stability and robustness of the model's performance, validating the effectiveness of the ConvMixer-ECA architecture in accurately classifying ADHD and TD subjects based on their EEG data. By employing this rigorous methodological approach, we enhanced the reliability of the reported results and strengthened the evaluation of the model's performance.

Table 1. Accuracy from the 5-fold cross-validation of the ConvMixer-ECA.

Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	0.9245	0.9038	0.9567	0.9245	0.9354

To further analyze the functionality of ConvMixer-ECA, t-SNE visualizations of the model's first and last layer outputs were generated, revealing the separation of ADHD within the learned feature space. Significant improvements were seen in the ADHD and TD samples (Figure 6). The first layer showed initial discriminative ability, although there was some overlap between categories. In contrast, the last layer showed significant separation, highlighting the hierarchical feature learning process of the model. The clear separation observed in the last layer validates the ability of ConvMixer-ECA in ADHD diagnosis, successfully capturing the underlying patterns that distinguish ADHD individuals from TD individuals. In summary, the t-SNE output demonstrates the model's ability to learn increasingly discriminative features, and the final layer confirms its effectiveness in distinguishing between the two categories.

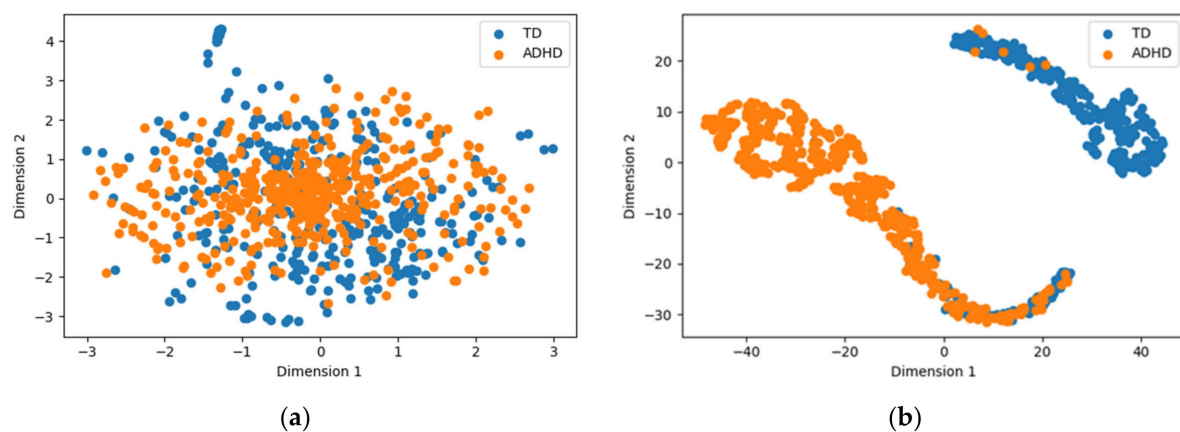


Figure 6. Visualizing the t-SNE-reduced outputs of the proposed model's (a) first layer and (b) final layer.

The activation maps of all layers of the ConvMixer-ECA model, as visualized in Figure 7, demonstrate the progressive feature extraction process. The initial layer (Layer 0) captures local patterns and textures after the patch embedding. As ConvMixer blocks progress from Layer 1 to Layer 7, the model learns and refines features through the use of depthwise convolutions to capture the spatial relationships within channels, pointwise convolutions to mix information across channels, and ECA blocks to adaptively recalibrate channel-wise attention. The activation maps become increasingly abstract and capture higher-level features in the deeper layers, as evidenced by the transition from fine-grained, dense activations in the earlier layers to coarse-grained, sparse activations in the deeper layers. This suggests that the deeper layers are selective to specific higher-level features and capture more complex and semantically meaningful patterns. The final ConvMixer block (Layer 7) outputs a rich set of discriminative features that are then aggregated and used for classification. This hierarchical feature learning process, where earlier layers learn low-level features and deeper layers progressively learn more abstract representations, enables the ConvMixer-ECA model to effectively identify the distinguishing characteristics of ADHD

and TD samples, leading to accurate classification. These findings indicate that the model is capable of effectively capturing the underlying patterns and features that distinguish individuals with ADHD from those without, thereby achieving accurate categorization.

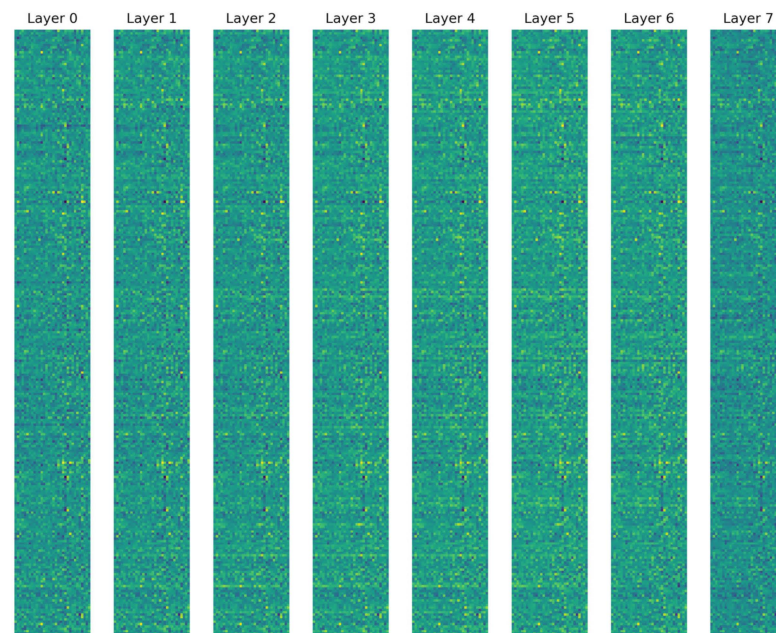


Figure 7. Activation maps from each ConvMixer block layer.

3.3. The Impact of Attention Mechanisms on ConvMixer Performance

A comprehensive analysis was performed to investigate the impact of different attention mechanisms on the performance of ConvMixer. The attention mechanisms evaluated include the convolutional block attention (CBA) module, squeeze and excite attention (SEA), non-localized attention (NLA), and ECA. Each ConvMixer variant with a specific attention mechanism is trained and evaluated on the same dataset, with the original ConvMixer model serving as the benchmark for comparison.

Table 2 shows the evaluation metrics (i.e., accuracy, precision, recall, and F1 score) for each ConvMixer variant. The baseline ConvMixer model achieved an accuracy of 0.9160, providing a reference point for assessing the impact of attentional mechanisms. The ConvMixer-CBA with CBA improved slightly over the baseline, achieving an accuracy of 0.9196. The ConvMixer-SEA further improved the performance to an accuracy of 0.9318. In addition, ConvMixer-NLA showed excellent results with an accuracy of 0.9415. However, ConvMixer-ECA achieved the highest performance among the variants with an accuracy of 0.9452.

Table 2. Evaluation results of ConvMixer variants with different attention mechanisms.

Model	Accuracy	Precision	Recall	F1 Score
ConvMixer	0.9160	0.9016	0.9544	0.9273
ConvMixer-CBA	0.9196	0.9022	0.9610	0.9307
ConvMixer-SEA	0.9318	0.9299	0.9501	0.9399
ConvMixer-NLA	0.9415	0.9460	0.9501	0.9481
ConvMixer-ECA	0.9452	0.9469	0.9640	0.9554

The experimental results provide compelling evidence that the combination of attentional mechanisms consistently improved ConvMixer's performance across all evaluation metrics. In particular, ConvMixer-ECA has emerged as the most effective method, highlighting the potential of ECA to significantly improve the model's classification or prediction capabilities. These results underscore the importance of exploring and exploiting atten-

tional mechanisms to improve the performance of deep learning models. The integration of attentional mechanisms, particularly ECA, shows promise for improving the effectiveness of ConvMixer.

3.4. Comparative Evaluation of Recognition Models

ConvMixer-ECA was compared with five commonly used deep learning models (EEGNet, CNN, RNN, LSTM, and gated recurrent unit (GRU)) to evaluate its effectiveness in EEG signal classification. EEGNet, a compact CNN architecture specifically designed for EEG-based brain–computer interfaces [37], has shown promising results in a variety of EEG classification tasks.

As shown in Figure 8a, ConvMixer-ECA achieved the highest accuracy of 94.52%, which was 7.19 percentage points higher than the second best model, EEGNet. CNN, RNN, and LSTM achieved 85.38%, 50.91%, and 50.18% accuracy, respectively, while GRU had the lowest accuracy of 55.90%.

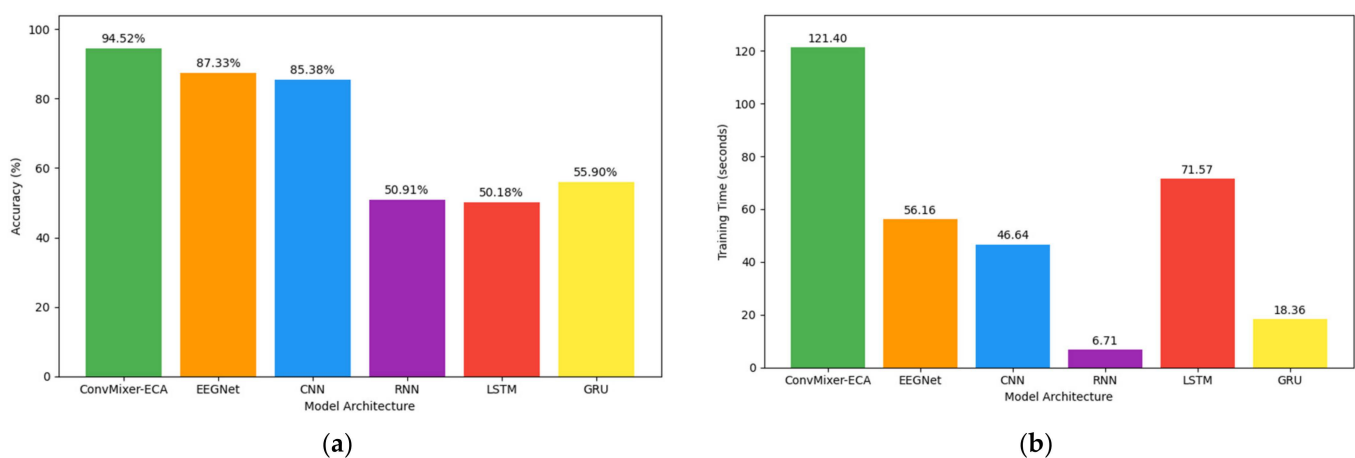


Figure 8. Comparison of the model architectures: (a) accuracy and (b) training time.

The excellent performance of ConvMixer-ECA can be attributed to its combination of deep convolution, pooling, and ECA mechanisms, which improves its discriminative power and ability to learn salient features from EEG data. In contrast, recurrent models (RNN, LSTM, and GRU) performed poorly, possibly due to the difficulty in capturing long-term dependencies in EEG signal sequences. This comparative analysis establishes ConvMixer-ECA as a promising architecture for accurate EEG-based recognition, outperforming specialized EEG classification models such as EEGNet and several standard deep learning architectures.

A comparative analysis of computational time for different deep learning models was conducted on a platform comprising an Intel Core i5-9500T (Intel Corporation, Santa Clara, CA, USA) processor with 16 GB of DDR4 RAM running Ubuntu 20.04.1 LTS and Python 3.11.6. The training condition was set to 10 epochs.

As illustrated in Figure 8b, the outcomes indicate that ConvMixer-ECA and LSTM were more computationally demanding, whereas GRU and RNN were the most efficient in terms of training time on the specified platform and configuration. Although ConvMixer-ECA necessitated more computational resources in comparison to the other models, the hardware configuration utilized in this analysis was sufficient to accommodate its training requirements, rendering it a viable option for use in a PC environment.

4. Discussion

The experimental results of this study demonstrate the effectiveness of the proposed ConvMixer-ECA architecture in detecting ADHD using EEG signals. The model achieved a remarkable accuracy of 94.52%, outperforming several state-of-the-art deep learning models including EEGNet, CNN, RNN, LSTM, and GRU. The high precision (94.69%),

recall (96.40%), and F1 score (95.54%) further emphasize the model's ability to correctly identify ADHD cases while maintaining a low false-positive rate ability.

The t-SNE visualization of the model layer output provides valuable insight into the feature learning process, demonstrating the model's ability to learn increasingly discriminative features. This hierarchical feature learning process demonstrates the effectiveness of ConvMixer-ECA in capturing the underlying patterns and features that distinguish ADHD from TD.

The effects of different attentional mechanisms on the performance of the ConvMixer model were extensively investigated. The combination of attentional mechanisms consistently improved the model's performance, and ConvMixer-ECA using ECA achieved the highest accuracy among the variants. This finding highlights the importance of exploring and exploiting attentional mechanisms to improve the classification capabilities of deep learning models in EEG-based recognition tasks.

The proposed ConvMixer-ECA method, summarized in Table 3, showed excellent performance in identifying ADHD using EEG signals compared to the existing literature. The excellent performance of ConvMixer-ECA can be attributed to its ability to efficiently capture spatial and channel dependencies in EEG signals through a combination of deep convolution, pooling, and ECA mechanisms.

Table 3. Comparison of the proposed ConvMixer-ECA model with findings from the literature.

Author	Year	Dataset	Method	Accuracy (%)
Tenev et al. [38]	2014	50 healthy, 67 ADHD	SVM and voting	82.3
Khoshnoud et al. [39]	2015	10 healthy, 12 ADHD	LLE, ApEn, PNN	87.5
Mohammadi et al. [40]	2016	30 healthy, 31 ADHD	MLP neural network	93.65
Chen et al. [18]	2019	57 healthy, 50 ADHD	Deep CNN	90.29
Dubreuil-Vall et al. [15]	2020	20 healthy, 20 ADHD	Spectrogram and CNN	88.0
Tosun [16]	2021	16 subject	Data augmentation, PSD, SE, and LSTM	92.15
Saini et al. [17]	2022	80 healthy, 77 ADHD	PCA, KNN	86.0
Our proposed approach	2024	60 healthy, 61 ADHD	ConvMixer with ECA	94.52

The results of this study have important implications for the early diagnosis and intervention of ADHD in children. The high accuracy and reliability of ConvMixer-ECA demonstrates its potential as a valuable tool to assist clinicians in the diagnostic process, facilitating timely intervention and individualized treatment plans. The ConvMixer-ECA model can capture EEG abnormalities associated with ADHD such as increased theta activity, reduced beta activity, and altered fronto-central connectivity [41,42]. These abnormalities are robust markers of ADHD and are linked to impaired inhibitory control and hypoarousal in the prefrontal cortex [43]. Functional imaging studies have also highlighted abnormal functioning in the prefrontal cortex and striatum in individuals with ADHD [44]. Moreover, there is evidence of altered fronto-subcortical intrinsic functional connectivity in ADHD. The model's ability to extract complex patterns and relationships from EEG data may potentially reveal the distinct EEG signatures of ADHD, providing valuable insights for diagnostic and therapeutic purposes.

The present study has some limitations that need to be addressed in future studies. The false positive rate of 8.41% suggests that there is room for improvement in reducing the misclassification of TD individuals as ADHD individuals. A comparison of the performance of our EEG-based ConvMixer-ECA model with that of non-EEG-based ADHD diagnostic tools such as clinical interviews, behavioral assessments, and neuropsychological tests could provide a more comprehensive understanding of its potential utility and complementary role in the diagnostic process. However, the ROC curve analysis, with an area under the curve (AUC) of 0.98, demonstrated the high overall discriminative ability of our ConvMixer-ECA model in distinguishing between ADHD and TD children (Figure 9). This strong performance highlights the potential of our approach to assist clinicians in the accurate diagnosis of ADHD, while future work should focus on fur-

ther improving the model's specificity to minimize false positives. Subsequent research endeavors should also explore the specificity and applicability of the model to various EEG classification tasks such as distinguishing between ADHD subtypes or predicting treatment response. Investigating the generalizability of the ConvMixer-ECA model by validating its performance on larger and more diverse ADHD datasets including subjects from different age groups, ethnicities, and clinical subtypes would further establish its robustness. Furthermore, the integration of multimodal data sources such as functional magnetic resonance imaging (fMRI) or genetic information with EEG signals could facilitate the development of more comprehensive and accurate models for the diagnosis and subtyping of ADHD. Moreover, future research should focus on further validating and refining the model for broader clinical utility, facilitating its integration into clinical practice and decision-making processes.

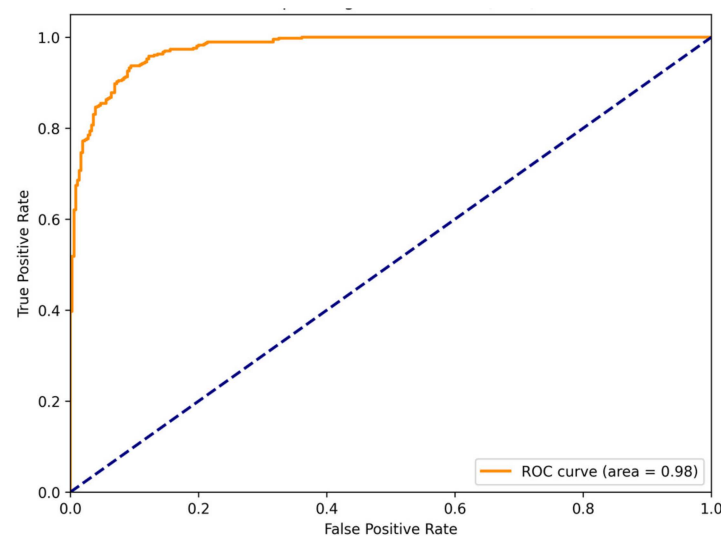


Figure 9. Receiver operating characteristic (ROC) curve of the ConvMixer-ECA model.

5. Conclusions

This study proposes ConvMixer-ECA, an innovative deep learning framework for identifying ADHD from EEG signals. The architecture integrates the ConvMixer model, which uses deep convolution and point-by-point convolution for optimized feature extraction, combined with ECA blocks to recalibrate features in a channelized manner. The model was trained and evaluated on EEG data from 60 healthy children and 61 children with ADHD. Experiments and analyses led to the following main conclusions:

- (1) ConvMixer-ECA performed well in detecting ADHD with an accuracy of 94.52%. This highlights its effectiveness in recognizing discriminative features and accurately classifying ADHD individuals from TD individuals.
- (2) The integration of attentional mechanisms, especially ECA, significantly improved the performance of the ConvMixer model. It outperformed other attention-based variants, highlighting the importance of incorporating attentional mechanisms in EEG-based recognition tasks.
- (3) ConvMixer-ECA outperformed existing state-of-the-art deep learning models including EEGNet, CNN, RNN, LSTM, and GRU, establishing ConvMixer-ECA as an accurate EEG-based ADHD detection method.
- (4) The t-SNE visualization of the output of the ConvMixer-ECA layer confirmed the model's ability to learn to distinguish between the intrinsic patterns and features of individuals with ADHD and those with TD through hierarchical feature learning.

These results highlight the potential of ConvMixer-ECA as a valuable tool to assist clinicians in the early diagnosis and intervention of ADHD in children.

Author Contributions: Conceptualization, J.X. and M.F.; Methodology, J.X. and M.F.; Software, J.X. and M.F.; Validation, M.F. and J.X.; Formal analysis, M.F.; Investigation, M.F.; Resources, J.X.; Data curation, M.F.; Writing—original draft preparation, M.F.; Writing—review and editing, J.X.; Visualization, M.F.; Supervision, J.X.; Project administration, J.X.; Funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a National Key Research Key Special Sub-Project on the Development Plan for Major Chronic Non-Communicable Disease Prevention and Control (2016YFCI306205), the Project of the Nanjing Rehabilitation Medical Center in Jiangsu Province (879), and the Anhui Province Natural Science Foundation of China (2208085US02).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We thank the numerous contributors to the IEEE DataPort for their efforts in the collection, organization, and sharing of their datasets. The data that support the findings of this study are openly available in the IEEE DataPort at <https://ieee-dataport.org/open-access/eeg-data-adhd-control-children> (accessed on 9 April 2024).

Acknowledgments: We are grateful to the three anonymous reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kollins, S.H.; DeLoss, D.J.; Cañadas, E.; Lutz, J.G.; Findling, R.L.; Keefe, R.S.; Epstein, J.N.; Cutler, A.J.; Faraone, S.V. A Novel Digital Intervention for Actively Reducing Severity of Paediatric Adhd (Stars-Adhd): A Randomised Controlled Trial. *Lancet Digit. Health* **2020**, *2*, e168–e178. [\[CrossRef\]](#)
2. Thapar, A.; Livingston, L.A.; Eyre, O.; Riglin, L. Practitioner Review: Attention-deficit Hyperactivity Disorder and Autism Spectrum Disorder—The Importance of Depression. *J. Child Psychol. Psychiatry* **2022**, *64*, 4–15. [\[CrossRef\]](#)
3. Fuermaier, A.B.M.; Hüpen, P.; de Vries, S.M.; Müller, M.; Kok, F.M.; Koerts, J.; Heutink, J.; Tucha, L.; Gerlach, M.; Tucha, O. Perception in Attention Deficit Hyperactivity Disorder. *Atten. Deficit Hyperact. Disord.* **2017**, *10*, 21–47. [\[CrossRef\]](#)
4. Chen, B.; Zhan, Y.; Kessi, M.; Chen, S.; Xiong, J.; Deng, X.; Yang, L.; Peng, J.; Yin, F.; He, F. Urine Organic Acids as Metabolic Indicators for Global Developmental Delay/Intellectual Disability in Chinese Children. *Front. Mol. Biosci.* **2021**, *8*, 792319. [\[CrossRef\]](#)
5. Xu, G.; Strathearn, L.; Liu, B.; Yang, B.; Bao, W. Twenty-Year Trends in Diagnosed Attention-Deficit/Hyperactivity Disorder among Us Children and Adolescents, 1997–2016. *JAMA Netw. Open* **2018**, *1*, e181471. [\[CrossRef\]](#)
6. Feng, M.; Xu, J.; Zhai, M.; Wu, Q.; Chu, K.; Xie, L.; Luo, R.; Li, H.; Xu, Q.; Xu, X.; et al. Behavior Management Training for Parents of Children with Preschool ADHD Based on Parent-Child Interactions: A Multicenter Randomized Controlled, Follow-Up Study. *Behav. Neurol.* **2023**, *2023*, 3735634. [\[CrossRef\]](#)
7. Hall, C.L.; Valentine, A.; Groom, M.J.; Walker, G.; Sayal, K.; Daley, D.; Hollis, C. The Clinical Utility of the Continuous Performance Test and Objective Measures of Activity for Diagnosing and Monitoring Adhd in Children: A Systematic Review. *Eur. Child Adolesc. Psychiatry* **2015**, *25*, 677–699. [\[CrossRef\]](#)
8. Lenartowicz, A.; Loo, S.K. Use of Eeg to Diagnose Adhd. *Curr. Psychiatry Rep.* **2014**, *16*, 498. [\[CrossRef\]](#)
9. Feng, M.; Xu, J. Detection of ASD Children through Deep-Learning Application of fMRI. *Children* **2023**, *10*, 1654. [\[CrossRef\]](#)
10. Chen, I.; Chang, C.; Chang, Y.; Lin, D.; Lin, C.; Ko, L. Neural Dynamics for Facilitating Adhd Diagnosis in Preschoolers: Central and Parietal Delta Synchronization in the Kiddie Continuous Performance Test. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 1524–1533. [\[CrossRef\]](#)
11. Joy, R.C.; George, S.T.; Rajan, A.A.; Subathra, M.S.P.; Sairamya, N.J.; Prasanna, J.; Mohammed, M.A.; Al-Waisy, A.S.; Jaber, M.M.; Al-Andoli, M.N. Detection and Classification of Adhd from Eeg Signals Using Tunable Q-Factor Wavelet Transform. *J. Sens.* **2022**, *2022*, 3590973. [\[CrossRef\]](#)
12. Loo, S.K.; Makeig, S. Clinical Utility of Eeg in Attention-Deficit/Hyperactivity Disorder: A Research Update. *Neurotherapeutics* **2012**, *9*, 569–587. [\[CrossRef\]](#)
13. Vahid, A.; Bluschke, A.; Roessner, V.; Stober, S.; Beste, C. Deep Learning Based on Event-Related Eeg Differentiates Children with Adhd from Healthy Controls. *J. Clin. Med.* **2019**, *8*, 1055. [\[CrossRef\]](#)
14. Chang, Y.; Stevenson, C.; Chen, I.; Lin, D.; Ko, L. Neurological State Changes Indicative of Adhd in Children Learned via Eeg-Based Lstm Networks. *J. Neural Eng.* **2022**, *19*, 016021. [\[CrossRef\]](#)
15. Dubreuil-Vall, L.; Ruffini, G.; Camprodon, J.A. Deep Learning Convolutional Neural Networks Discriminate Adult ADHD from Healthy Individuals on the Basis of Event-Related Spectral EEG. *Front. Neurosci.* **2020**, *14*, 515034. [\[CrossRef\]](#)
16. Tosun, M. Effects of Spectral Features of EEG Signals Recorded with Different Channels and Recording Statuses on ADHD Classification with Deep Learning. *Phys. Eng. Sci. Med.* **2021**, *44*, 693–702. [\[CrossRef\]](#)

17. Saini, S.; Rani, R.; Kalra, N. Prediction of Attention Deficit Hyperactivity Disorder (ADHD) Using Machine Learning Techniques Based on Classification of EEG Signal. In Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March 2022; Volume 1, pp. 782–786.
18. Chen, H.; Song, Y.; Li, X. Use of Deep Learning to Detect Personalized Spatial-Frequency Abnormalities in EEGs of Children with ADHD. *J. Neural Eng.* **2019**, *16*, 066046. [\[CrossRef\]](#)
19. Ahmadi, M.; Kazemi, K.; Kuc, K.; Cybulska-Kłosowicz, A.; Helfroush, M.S.; Aarabi, A. Resting State Dynamic Functional Connectivity in Children with Attention Deficit/Hyperactivity Disorder. *J. Neural Eng.* **2021**, *18*, 0460d1. [\[CrossRef\]](#)
20. Sharma, Y.; Singh, B.K. Attention Deficit Hyperactivity Disorder Detection in Children Using Multivariate Empirical Eeg Decomposition Approaches: A Comprehensive Analytical Study. *Expert Syst. Appl.* **2023**, *213*, 119219. [\[CrossRef\]](#)
21. Zanus, C.; Miladinović, A.; De Dea, F.; Skabar, A.; Stecca, M.; Ajčević, M.; Accardo, A.; Carrozzi, M. Sleep Spindle-Related Eeg Connectivity in Children with Attention-Deficit/Hyperactivity Disorder: An Exploratory Study. *Entropy* **2023**, *25*, 1244. [\[CrossRef\]](#)
22. Kim, J.; Lee, J.; Park, E.; Han, J. A Deep Learning Model for Detecting Mental Illness from User Content on Social Media. *Sci. Rep.* **2020**, *10*, 11846. [\[CrossRef\]](#)
23. Chen, Z.S.; Galatzer-Levy, I.R.; Bigio, B.; Nasca, C.; Zhang, Y. Modern Views of Machine Learning for Precision Psychiatry. *Patterns* **2022**, *3*. [\[CrossRef\]](#)
24. Tong, X.; Xie, H.; Fonzo, G.A.; Zhao, K.; Satterthwaite, T.D.; Carlisle, N.B.; Zhang, Y. Symptom Dimensions of Resting-State Electroencephalographic Functional Connectivity in Autism. *Nat. Ment. Health* **2024**, *2*, 287–298. [\[CrossRef\]](#)
25. Trockman, A.; Kolter, J.Z. Patches Are All You Need? *arXiv* **2022**, arXiv:2201.09792. [\[CrossRef\]](#)
26. Wang, J.; Yin, P.; Wang, Y.; Yang, W. CMAT: Integrating Convolution Mixer and Self-Attention for Visual Tracking. *IEEE Trans. Multimed.* **2023**, *26*, 326–338. [\[CrossRef\]](#)
27. Sun, L.; Pan, J.; Tang, J. Shufflemixer: An Efficient Convnet for Image Super-Resolution. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 17314–17326.
28. Huang, X.; Azzam, R.; Javed, S.; Gan, D.; Seneviratne, L.; Abdelqader, A.; Zweiri, Y. Cm-Unet: Convmixer Unet for Segmentation of Unknown Objects in Cluttered Scenes. *IEEE Access* **2022**, *10*, 123622–123633. [\[CrossRef\]](#)
29. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Eca-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [\[CrossRef\]](#)
30. Karthik, R.; Vaichole, T.S.; Kulkarni, S.K.; Yadav, O.; Khan, F. Eff2Net: An Efficient Channel Attention-Based Convolutional Neural Network for Skin Disease Classification. *Biomed. Signal Process. Control* **2022**, *73*, 103406. [\[CrossRef\]](#)
31. Nasrabadi, A.M.; Allahverdy, A.; Samavati, M.; Mohammadi, M.R. EEG Data for ADHD/Control Children. *IEEE Dataport* **2020**, *10*.
32. Feng, M.; Zhai, M.; Xu, J.; Wang, M.; Wang, C.; Yu, L.; Huang, Y.; Ke, X. Kindergarten-Based Screening of ADHD in Preschool Children. *J. Nanjing Med. Univ.* **2022**, *42*, 569–572. [\[CrossRef\]](#)
33. Congedo, M.; Lubar, J.F.; Joffe, D. Low-Resolution Electromagnetic Tomography Neurofeedback. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2004**, *12*, 387–397. [\[CrossRef\]](#)
34. Bakhtyari, M.; Mirzaei, S. ADHD Detection Using Dynamic Connectivity Patterns of EEG Data and ConvLSTM with Attention Framework. *Biomed. Signal Process. Control* **2022**, *76*, 103708. [\[CrossRef\]](#)
35. Mafi, M.; Radfar, S. High Dimensional Convolutional Neural Network for EEG Connectivity-Based Diagnosis of ADHD. *J. Biomed. Phys. Eng.* **2022**, *12*, 645. [\[CrossRef\]](#)
36. Xu, J.; Yu, X. Detection of Concrete Structural Defects Using Impact Echo Based on Deep Networks. *J. Test. Eval.* **2021**, *49*, 20190801. [\[CrossRef\]](#)
37. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain–Computer Interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [\[CrossRef\]](#)
38. Tenev, A.; Markovska-Simoska, S.; Kocarev, L.; Pop-Jordanov, J.; Müller, A.; Candrian, G. Machine Learning Approach for Classification of ADHD Adults. *Int. J. Psychophysiol.* **2014**, *93*, 162–166. [\[CrossRef\]](#)
39. Khoshnoud, S.; Shamsi, M.; Nazari, M.A. Non-Linear EEG Analysis in Children with Attention-Deficit/Hyperactivity Disorder during the Rest Condition. In Proceedings of the 2015 22nd Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 22–27 November 2015; pp. 87–92.
40. Mohammadi, M.R.; Khaleghi, A.; Nasrabadi, A.M.; Rafieivand, S.; Begol, M.; Zarafshan, H. EEG Classification of ADHD and Normal Children Using Non-Linear Features and Neural Network. *Biomed. Eng. Lett.* **2016**, *6*, 66–73. [\[CrossRef\]](#)
41. Lee, I.; Lee, J.; Lim, M.H.; Kim, K.M. Comparison of Quantitative Electroencephalography between Tic Disorder and Attention-Deficit/Hyperactivity Disorder in Children. *Clin. Psychopharmacol. Neurosci.* **2021**, *19*, 739–750. [\[CrossRef\]](#)
42. Cabaleiro, P.; Cueli, M.; Cañamero, L.M.; Castro, P.G. A Case Study in Attention-Deficit/Hyperactivity Disorder: An Innovative Neurofeedback-Based Approach. *Int. J. Environ. Res. Public Health* **2021**, *19*, 191. [\[CrossRef\]](#)

43. Scarpelli, S.; Gorgoni, M.; D'Atri, A.; Reda, F.; Gennaro, L.D. Advances in Understanding the Relationship between Sleep and Attention Deficit-Hyperactivity Disorder (Adhd). *J. Clin. Med.* **2019**, *8*, 1737. [[CrossRef](#)]
44. Rosch, K.S.; Mostofsky, S.H.; Nebel, M.B. Adhd-Related Sex Differences in Fronto-Subcortical Intrinsic Functional Connectivity and Associations with Delay Discounting. *J. Neurodev. Disord.* **2018**, *10*, 34. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.