Preprints.org

Article

Not peer-reviewed version

Temporal Attention and Convolutional Tokenization for Interpretable ADHD Classification from EEG

Julián David Pastrana-Cortés *, Alejandra Gomez-Rivera, Andres Marino Álvarez-Meza, Julian Gil-Gonzalez, David Cárdenas-Peña

Posted Date: 27 May 2026

doi: 10.20944/preprints202605.1894.v1

Keywords: ADHD; EEG; neurodevelopmental disorders; deep learning; convolutional model; transformer encoder; attention pooling

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

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.

Article

Temporal Attention and Convolutional Tokenization for Interpretable ADHD Classification from EEG

Julián David Pastrana-Cortés [1,*], Alejandra Gomez-Rivera [2], Andrés Marino Álvarez-Meza [2], Julian Gil-Gonzalez [1] and David Cárdenas-Peña [1]

[1] Automatics Research Group, Universidad Tecnológica de Pereira (UTP), Pereira 660003, Colombia
[2] Signal Processing and Recognition Group, Universidad Nacional de Colombia, 170003 Manizales, Colombia
* Correspondence: j.pastrana@utp.edu.co

Abstract

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental condition commonly assessed through clinical interviews, behavioral observation, and rating scales. Although electroencephalography (EEG) has emerged as a promising complementary tool for ADHD assessment, robust subject-independent classification remains challenging due to inter-subject variability, limited datasets, and the need for interpretable computational models. This work introduces EEG-TACT, a compact end-to-end deep learning architecture for identifying ADHD subjects from EEG epochs. The proposed model integrates an EEGNet-inspired convolutional embedding, a Transformer encoder operator, and an attention-based pooling mechanism to jointly capture local spatiotemporal EEG patterns, contextual temporal dependencies, and task-relevant latent representations. EEG-TACT was evaluated on a publicly available EEG dataset using strict, subject-independent stratified group partitions, ensuring that data from the same subject were never shared across the training, validation, and test subsets. Model interpretability is examined through learned temporal filter responses, class-conditioned self-attention maps, and latent-space projections, while an ablation study quantifies the contribution of each architectural component. Performance was assessed at the fold, subject, and epoch levels, together with statistical significance comparisons against representative state-of-the-art architectures. EEG-TACT outperformed the contrasted models, achieving subject-level accuracy of 87.5%, recall of 96.0%, and precision of 82.8%, while requiring only a few thousand trainable parameters. By exhaustively repeating the initialization, the proposed model demonstrated improved labeling reliability and achieved the best average ranking among the evaluated architectures. The reported results therefore prove that EEG-TACT provides a compact, robust, and interpretable model for EEG-based ADHD identification under subject-independent evaluation settings.

Keywords: ADHD; EEG; neurodevelopmental disorders; deep learning; convolutional model; transformer encoder; attention pooling

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD), one of the most prevalent neurodevelopmental conditions, commonly emerges during childhood with signs such as difficulties in academic achievement, social functioning, and emotional regulation [1]. Epidemiological evidence estimates that ADHD affects approximately 5.9% of youth worldwide, with recent meta-analytic estimates reporting prevalence rates of 7.6% in children aged 3-12 years and 5.6% in adolescents aged 12-18 years [2]. Clinically, ADHD is characterized by three main symptom domains: inattention, hyperactivity, and impulsivity, although many children exhibit a combined presentation involving features from more than one domain [3]. Beyond its cognitive and behavioral manifestations, ADHD is frequently accompanied by comorbid conditions such as anxiety, depression, and learning disorders. Therefore, early diagnosis and timely intervention are essential to reduce the long-term negative effects of the disorder on educational performance, social relationships, and overall quality of life [4].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

Despite the availability of standardized diagnostic criteria, ADHD assessment remains largely based on clinical interviews, behavioral observation, and parent-, teacher-, caregiver-, or clinician-completed behavioral rating scales [5]. Commonly used instruments, such as the Conners Rating Scales, the Vanderbilt ADHD Diagnostic Rating Scales, and the Swanson, Nolan, and Pelham Rating Scale (SNAP-IV), attempt to quantify inattentive and hyperactive-impulsive behaviors across different environments [6]. Despite their value for structured symptom screening and for collecting multi-informant evidence, the instruments are biased by informant perception, contextual variability, and differences between home and school observations [7]. In addition, behavioral manifestations of ADHD may overlap with other neurodevelopmental, emotional, or learning-related conditions, making differential diagnosis challenging [8]. Consequently, there is a growing interest in complementary objective approaches capable of characterizing neurophysiological patterns associated with ADHD and supporting clinical decision-making without replacing comprehensive clinical evaluation [9].

From a neurobiological perspective, ADHD has been associated with dysfunction in the prefrontal cortex and subcortical structures, particularly the basal ganglia, which are essential for attention regulation, behavioral control, and executive functioning [10].

Given such association with atypical brain function and neurodevelopmental alterations, brain-based neurophysiological and neuroimaging techniques have gained increasing attention as potential objective complementary tools for ADHD assessment [11]. Although functional magnetic resonance imaging has been widely used to characterize functional alterations in ADHD, its use in pediatric populations may be limited by high cost, reduced accessibility, long acquisition times, and sensitivity to motion artifacts [12]. By contrast, electroencephalography (EEG) has emerged as a promising complementary tool for ADHD assessment because it provides a portable, affordable, non-invasive, and high-temporal-resolution means of measuring brain electrical activity [13]. These properties make EEG suitable for characterizing neurophysiological spectral, temporal, and connectivity patterns related to attention, executive functioning, inhibitory control, and behavioral regulation [14].

More recently, the advent of machine learning and deep learning has further advanced the use of EEG for ADHD detection by enabling the analysis of complex, nonlinear, and high-dimensional signal patterns that may not be fully captured through conventional approaches [15]. In line with this, recent studies have shown that EEG-based machine learning approaches can achieve promising performance in distinguishing children with ADHD from typically developing controls, supporting the use of EEG as an objective and data-driven aid for early detection [16].

In this direction, EEG-based studies have explored machine-learning strategies to support ADHD diagnosis from subject-level neurophysiological dynamics [17]. Among machine learning approaches, deep learning methods have shown particular potential for automatically learning discriminative representations from EEG signals, reducing the dependence on handcrafted feature extraction and enabling end-to-end optimization for EEG classification tasks [18,19]. Convolutional neural networks (CNNs), for example, are effective in capturing local temporal and spatial patterns from EEG data and have been widely used for EEG decoding and visualization [20]. In particular, the EEGNet architecture has emerged as a compact and efficient architecture specifically designed for EEG analysis, thanks to its depthwise and separable convolutions, which extract meaningful features while maintaining a relatively low number of trainable parameters [21]. This makes EEGNet especially attractive for ADHD-related EEG studies, where datasets are often limited and robust feature extraction is essential.

Although CNN-based models are effective at capturing local structures, their finite receptive fields may limit their ability to model long-range dependencies and global contextual information across EEG segments [22]. In this regard, attention mechanisms have gained relevance because they allow the model to focus on the most informative parts of the signal and to capture relationships across distant temporal representations [23]. Building on this idea, Transformer-inspired and hybrid convolutional-Transformer architectures have emerged as promising alternatives for EEG decoding, since they combine local feature extraction with self-attention-based contextual modeling [24]. This type of architecture is particularly relevant for EEG analysis, where discriminative information may be

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

distributed across multiple temporal segments and spatial patterns rather than confined to isolated local features.

This work aims to bridge the gap between compact EEG-specific feature extraction and global sequence modeling for ADHD detection by introducing a novel end-to-end deep learning architecture, termed EEG-TACT. Unlike conventional approaches that rely exclusively on handcrafted features or purely convolutional processing, EEG-TACT combines an EEGNet-inspired convolutional embedding operator with a Transformer encoder to jointly learn local spatiotemporal patterns and short-term temporal dependencies directly from raw EEG signals. The convolutional front-end extracts compact and physiologically meaningful representations across channels and time, while the Transformer module leverages self-attention to model contextual interactions among temporal segments. In addition, an attention-based pooling strategy is incorporated to adaptively emphasize the most informative ADHD EEG representations prior to classification, thereby improving both feature aggregation and interpretability. The proposed model was evaluated on a publicly available EEG dataset composed of school-aged children and assessed using five fixed subject-wise folds under a stratified group cross-validation protocol. The results and comparisons against six state-of-the-art deep learning models show that EEG-TACT provides an effective and compact framework for EEG-based ADHD classification, outperforming several baseline approaches by combining local feature extraction, global dependency modeling, and adaptive attention-based aggregation.

The main contributions of this work are: (i) a compact hybrid convolutional-Transformer architecture that jointly learns frequency-selective spatial filters, short-term attention for contextualizing temporal tokens, and a pooling mechanism that weights the most relevant tokens before classification; (ii) a rigorous evaluation framework combining stratified group cross-validation, varying-seed testing, and subject-independent training to mitigate optimistic performance estimates; and (iii) a multi-faceted analysis encompassing interpretability (parameter visualization, latent-space projections, and ablation experiments) and statistical analysis (paired statistical significance tests, and risk-coverage curves).

The remainder of this paper is organized as follows . Section 2 describes the EEG dataset, preprocessing pipeline, proposed architecture, training strategy, and subject-independent evaluation protocol. Section 3 reports and discusses the experimental results, including hyperparameter tuning, interpretability analysis, ablation study, comparative performance evaluation, and statistical significance tests. Finally, Section 4 presents the concluding remarks and future research directions.

2. Materials and Methods

This work introduces EEG-TACT, an end-to-end deep learning architecture designed to jointly model local spatiotemporal patterns and long-range temporal dependencies directly from raw EEG signals. The model combines a convolutional front-end inspired by EEGNet, which extracts compact and physiologically representations across channels and time, with a Transformer encoder that captures contextual interactions among temporal segments. Furthermore, an attention-based pooling strategy is employed to adaptively aggregate temporal information, emphasizing the most informative EEG segments for classification. The following subsections present the mathematical formulation of the proposed model in detail.

2.1. Dataset and Preprocessing

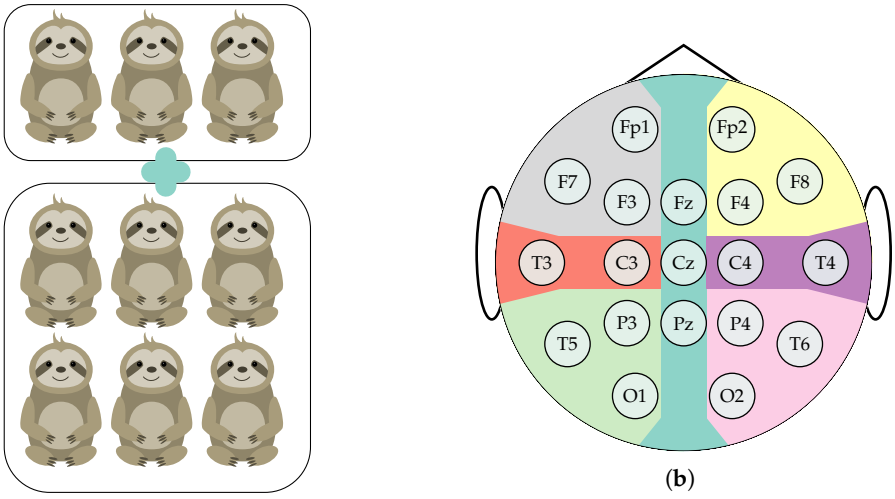
This study tests the proposed EEG-TACT using publicly available EEG data collected from children with ADHD and control subjects. The data were shared by Nasrabadi et al. via IEEE DataPort [25]. The database contains EEG recordings from 121 children aged 7 to 12 years. Among them, 61 children were diagnosed with ADHD, and 60 were typically developing controls. ADHD diagnoses were made by an experienced child psychiatrist using DSM-IV criteria [26]. At the time of recording, patients had been using methylphenidate (Ritalin) for up to six months. The control group had no history of psychiatric or neurological disorders, including epilepsy or high-risk behaviors [26,27].

EEG signals were recorded while participants performed a visually guided attention task designed to assess attentional deficits linked to ADHD [28]. During the task, children viewed cartoon images

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

containing 5 to 16 characters, as depicted in Figure 1(a), and were asked to silently count the number of figures in each scene. Images appeared continuously, each new image shown immediately following the participant's response, resulting in variable recording lengths based on individual response speeds [28]. EEG was recorded at a sampling rate of 128 Hz from 19 scalp electrodes placed according to the international 10-20 system (see Figure 1(b)).



(a) (b)

Figure 1. Acquisition setup for EEG-based ADHD assessment. **(a)** A representative example of the visual counting stimuli presented to participants during recording. **(b)** The 19-electrode scalp montage arranged according to the international 10–20 system, colour-coded by anatomical region: left frontal, right frontal, midline (Fz, Cz, Pz), left central-temporal, right central-temporal, left posterior, and right posterior.

Regarding preprocessing, the EEG signals were first re-referenced to the common average reference, aiming to reduce dependence on the physical recording reference and enhance the consistency of spatial EEG patterns across subjects [29]. This transformation subtracts the instantaneous mean scalp potential from each channel, thereby emphasizing channel-specific deviations from global activity and reducing reference-induced bias. Then, a 50 Hz notch filter attenuates narrow-band power-line contamination, and a $0.5 - 60$ Hz band-pass filter retains the physiologically relevant EEG spectrum while reducing slow drifts, movement-related low-frequency artifacts, and high-frequency muscular or instrumental noise [30]. It is worth noting that subject named v56p was removed from further analysis due to marked artifact contamination in the EEG recording. Recordings from remaining subjects were then segmented into 2-second, 50%-overlapping epochs, yielding a (C, T) shaped epochs, where $C = 19$ denotes the number of channels and $T = 256$ the number of temporal samples per epoch. Finally, the epochs from all retained subjects were concatenated into a global dataset of shape (N, C, T), where $N = 16{,}640$ represents the total number of epochs across all subjects, where each epoch inherits the label of its corresponding subject, while subject identifiers were preserved to enable training and evaluation strategy.

2.2. EEGNet-Based Feature Embedding

Let $\mathbf{X} \in \mathbb{R}^{T \times C}$ denote a multivariate time-series EEG epoch with C channels and T temporal samples, associated with a label $y \in \mathcal{Y}$ (ADHD: $y = 1$, Control: $y = 0$). The proposed model aims at classifying $\mathbf{X}$ through a parametric mapping $g_\theta(\mathbf{X})$, which estimates the posterior probability of belonging to the target class $P(y \mid \mathbf{X}) \in [0, 1]$. To this end, hierarchical embedding operator extracts informative representations from the input signal, capturing both temporal dynamics and inter-channel dependencies. First, temporal patterns are modeled by applying a set of learnable filters along the time dimension, producing feature maps that encode frequency-selective or time-localized characteristics. Subsequently, channel-wise interactions are incorporated through operations that

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

transform and aggregate information across the spatial (channel) dimension, enabling the model to capture cross-channel dependencies inherent to multivariate signals.

The intermediate representations are further refined using convolutional transformations that jointly model temporal and feature-wise interactions, typically through separable or factorized operations that improve parameter efficiency while preserving expressiveness. These stages include temporal downsampling and pooling operations, resulting in a compressed representation with reduced temporal resolution. Formally, the embedding process can be expressed as

$$\mathbf{H}_\varphi = \varphi(\mathbf{X}; \theta_\varphi) \in \mathbb{R}^{T' \times d}, \tag{1}$$

where $\varphi : \mathbb{R}^{T \times C} \to \mathbb{R}^{T' \times d}$ denotes the embedding operator parameterized by θ_φ, d the number of learned feature components, and T' the temporal instances after transformation and downsampling. The proposed representation encodes the input signal into a compact feature space that captures joint spatio-temporal structure, unraveling downstream classification.

2.3. Transformer-Based Contextual Encoding

Given the embedded representation $\mathbf{H}_\varphi \in \mathbb{R}^{T' \times d}$ resulting from the feature extraction stage, this framework aims to capture long-range temporal dependencies and global contextual interactions across the sequence. To this end, the embedded segment feeds a Transformer-based operator $\mathcal{T}(\mathbf{H}_\varphi; \theta_\tau)$, enabling each temporal position to incorporate information from all other positions in the sequence, resulting in a refined version through self-attention and channel-wise transformations.

The Multi-Head Self-Attention (MHSA) mechanism constitutes the core of the Transformer operator, which models pairwise interactions between temporal tokens [23]. Specifically, the input $\mathbf{H}_\varphi$ is linearly projected into sets of query $\mathbf{Q}_m \in \mathbb{R}^{T' \times d}$, key $\mathbf{K}_m \in \mathbb{R}^{T' \times d}$, and value $\mathbf{V}_m \in \mathbb{R}^{T' \times d}$ matrices:

$$\mathbf{Q}_m = \mathbf{H}_\varphi \mathbf{W}_Q^{(m)}, \quad \mathbf{K}_m = \mathbf{H}_\varphi \mathbf{W}_K^{(m)}, \quad \mathbf{V}_m = \mathbf{H}_\varphi \mathbf{W}_V^{(m)}, \tag{2}$$

with $\mathbf{W}_Q^{(m)}, \mathbf{W}_K^{(m)}, \mathbf{W}_V^{(m)} \in \mathbb{R}^{d \times d}$ as the learnable projection matrices for the m-th self-attention head ($m \in \{1, \ldots, M\}$). The attention output for the m-th head results from the weighted average of the values according to the

$$\mathbf{H}_m = \sigma\left(\frac{\mathbf{Q}_m \mathbf{K}_m^\top}{\sqrt{d}}\right)\mathbf{V}_m \in \mathbb{R}^{T' \times d}, \tag{3}$$

being $\sigma(\cdot)$ the softmax activation function that normalizes the relevance of each temporal position with respect to all others. This formulation enables the proposed model to dynamically weight contributions from different time steps, effectively capturing non-local dependencies in the sequence. Then, the concatenation of the head-wise outputs is linearly projected into the multihead self-attention output:

$$\bar{\mathbf{H}} = [\mathbf{H}_1 | \mathbf{H}_2 | \cdots | \mathbf{H}_M]\mathbf{W}_O \tag{4}$$

being $\bar{\mathbf{H}} \in \mathbb{R}^{T' \times d}$ a contextualized representation where each temporal embedding is expressed as a data-adaptive weighted combination of all embeddings in the sequence, with weights $\mathbf{W}_O \in \mathbb{R}^{(M \cdot d) \times d}$ determined by learned pairwise interactions. Lastly, a residual connection from the EEG embedding allows to preserve the original representation and improves the optimization of the first block $\tilde{\mathbf{H}} = \mathbf{H}_\varphi + \bar{\mathbf{H}}$.

Following the attention mechanism, a token-wise N-layered dense block, along with its respective residual connection, enhances the feature expressiveness of the transformer encoder:

$$\mathbf{H}_\tau = \mathcal{T}(\mathbf{H}_\varphi; \theta_\tau) = \tilde{\mathbf{H}} + f\big(f(\cdots f(\tilde{\mathbf{H}}\mathbf{W}_1 + \boldsymbol{b}_1)\mathbf{W}_2 + \boldsymbol{b}_2 \cdots)\mathbf{W}_{N'} + \boldsymbol{b}_{N'}\big), \tag{5}$$

where $\mathbf{W}_n \in \mathbb{R}^{F_{l-1} \times F_l}$, $\boldsymbol{b}_l \in \mathbb{R}^{F_l}$, and $F_l \in \mathbb{N}$ correspond to the trainable weighting matrix, bias vector, and number of neurons at the l-th dense layer, respectively; while $f(\cdot)$ stands for an element-wise

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

non-linear activation function. Further, $\theta_\tau = \{\mathbf{W}_Q^{(m)}, \mathbf{W}_K^{(m)}, \mathbf{W}_V^{(m)}, \mathbf{W}_O, \mathbf{W}_n, \boldsymbol{b}_n\}_{m,n=1}^{M,N'}$ aggregates the parameters of the full attention encoder to be trained. As a consequence, the attention-based mapping $\mathbf{H}_\tau \in \mathbb{R}^{T' \times d}$ yields a context-aware representation in which each temporal token encodes information aggregated from the entire sequence, effectively enriching the initial embedding $\mathbf{H}_\varphi$ with global temporal structure while preserving its dimensional organization.

2.4. Attention Pooling and Classification Head

Despite unraveling discriminative information for EEG classification, directly operating on the attention embedded trial $\mathbf{H}_\tau$ becomes computationally demanding due to the high dimensionality. Traditional strategies consist of aggregating the sequence into a compact, informative vector representation before classification. However, conventional pooling operators are fixed and non-trainable, limiting their ability to adaptively emphasize the most informative patterns. For that reason, we incorporate an attention pooling mechanism that introduces a learnable aggregation that performs an adaptive temporal aggregation by assigning learnable importance weights to the contextualized tokens and compressing the sequence into a single discriminative vector for classification as follows [31]:

$$\mathbf{z} = \sigma\left(\frac{\mathbf{w}\mathbf{H}_\tau^\top}{\sqrt{d}}\right)\mathbf{H}_\tau, \tag{6}$$

with $\boldsymbol{w}^\top \in \mathbb{R}^d$ as a learnable parameter vector and $\mathbf{z}^\top \in \mathbb{R}^d$ the resulting discriminative feature vector. Lastly, a classification layer maps z to the predicted probability:

$$\tilde{p}(y \mid \mathbf{X}) = \psi(z; \theta_\psi), \tag{7}$$

where $\psi : \mathbb{R}^{F_2} \to (0,1)^{|\mathcal{Y}|}$ denotes a feed-forward network parameterized by θ_ψ, followed by a softmax/sigmoid activation. Therefore, the proposed EEG-TACT defines a hierarchical composition of embedding (Equation (1)), transformer (Equation (5)), and classification (Equation (7)) operators, yielding a parametric mapping $g_{\boldsymbol{\theta}} = \psi \circ \mathcal{T} \circ \varphi$ from an input EEG trial $\mathbf{X}$ to the conditional probability $P(y \mid \mathbf{X})$, which is learned by optimizing the parameter set $\boldsymbol{\theta} = \theta_\varphi \cup \theta_\tau \cup \theta_\psi$.

2.5. Training Strategy and Evaluation Protocol

To ensure a fair and clinically meaningful evaluation, the model was trained and assessed under a strict, subject-independent evaluation protocol using a stratified group cross-validation scheme. A five-fold stratified group cross-validation scheme was employed, in which each subject was treated as an independent group. This guarantees that data from the same subject never appear in more than one subset among training, validation, and test sets, thereby preventing any form of information leakage, in accordance with best practices in cross-validation and model selection [32]. The stratification procedure preserves the class distribution (ADHD vs. Control) across all folds, while the grouping constraint enforces subject-level separation. At each outer split, subjects were separated into training/validation and test subsets. Subsequently, an inner split, also performed using a stratified group strategy, was applied to divide the training portion into training and validation subsets. This hierarchical splitting ensures that model selection and performance estimation remain unbiased with respect to unseen subjects.

During training, the model operates at the epoch level, where each EEG recording is segmented into multiple overlapping windows. Since the ground-truth label is defined at the subject level, a post-processing aggregation step is necessary: for each subject, the predicted probabilities for all their epochs are combined to yield a subject-level prediction. This aggregation enables subject-wise

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

classification from epoch-level outputs. For this work, a majority voting strategy is employed to obtain the final subject-level prediction

$$\tilde{y}_s = \underset{i \in \mathcal{I}_s}{\text{mode}} \left[\underset{y \in \mathcal{Y}}{\arg\max} \; \tilde{p}(y \mid \mathbf{X}_i) \right], \tag{8}$$

being $\mathbf{X}_i$ the i-th EEG epoch and $\mathcal{I}_s$ the index set of epochs belonging to subject s. The final subject-level prediction $\hat{y}_s$ is therefore defined as the most frequent predicted class among the subject's epochs. This aggregation step ensures that performance metrics are computed at the subject level rather than at the epoch level.

During validation, subject-level accuracy is used as the primary criterion for model selection. This choice is motivated by the nearly balanced class distribution of the dataset and by the fact that accuracy is the main metric reported in the comparative evaluation. Thus, the optimization objective was aligned with the final reporting protocol. During testing, performance metrics are computed by aggregating epoch-level predictions at the subject level. In particular, accuracy, recall, and precision are derived from a single final prediction per subject. This design on training, validation, and testing reflects a realistic clinical scenario in which each subject is assigned a single diagnostic label rather than multiple epoch-level decisions and avoids overestimating performance due to multiple correlated epochs from the same individual.

Finally, hyperparameter optimization is conducted in the inner validation loop using Bayesian search with the Optuna toolbox, with subject-level validation accuracy as the objective. In addition, early stopping and pruning strategies are employed during hyperparameter optimization to reduce overfitting and computational cost. According to Section 2, the hyperparameters to be optimized include the embedding block, number of filters, and their configuration parameters. The search also included the Transformer encoder hyperparameters, comprising the number of attention heads M, the embedding dimensionality d, and the dense block architecture in Equation (5).

3. Results and Discussion

3.1. Model Setup and Hyperparameter Tuning

The proposed EEG-TACT follows a sequential deep learning architecture implemented as an end-to-end model composed of the three main operators defined in Section 2, allowing convolutional feature-extraction capacity, contextual temporal modeling, regularization strength preserving strict subject-independent validation and testing partitions.

Table 1 summarizes the resulting optimal architecture for the proposed EEG-TACT model. The first block resembles an EEGNet-based feature embedding block, applying a temporal convolution to the input time series to extract frequency-selective patterns, followed by batch normalization. Next, a depthwise spatial convolution integrates information across channels and learns spatial filters. The extracted features are then refined through nonlinear activation, average pooling, and spatial dropout. A separable convolution further captures compact temporal representations while reducing the number of trainable parameters. After a second stage of normalization, activation, pooling, and dropout, the resulting feature maps are reshaped into a sequence of latent tokens. The second block corresponds to the Transformer encoder operator processing the embedded representation over temporal tokens. Two self-attention heads and a two-layer dense block exploit long-range temporal dependencies. In the third block, the attention pooling layer aggregates the sequence of encoded tokens using adaptive importance weights. The tuned one-layer Transformer encoder supports epoch classification.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

Table 1. Architecture of the proposed EEG-TACT model. The network comprises a convolutional feature extractor φ, a Transformer encoder $\mathcal{T}$, and a classification head ψ. Tensor dimensions and tuned hyperparameters are reported for each layer.

Layer	Variable	Output Shape	Hyperparameter	Setting
Feature Extractor φ				
Input epoch	$\mathbf{X}_n$	$C \times T$	C, T	19, 256
Reshape	$\mathbf{X}_{0,n}$	$C \times T \times 1$	–	–
Temporal Conv	$\mathbf{H}_1$	$C \times T \times F_1$	F_1, kernel	8, $(1, 64)$
Batch Norm	$\mathbf{H}'_1$	$C \times T \times F_1$	–	–
Depthwise Conv	$\mathbf{H}_2$	$1 \times T \times (F_1 D)$	D, kernel	3, $(C, 1)$
Batch Norm	$\mathbf{H}_2^a$	$1 \times T \times (F_1 D)$	Activation	ELU
Average Pooling	$\mathbf{H}_2^p$	$1 \times \frac{T}{p_1} \times (F_1 D)$	p_1	4
Spatial Dropout	$\mathbf{H}'_2$	$1 \times \frac{T}{p_1} \times (F_1 D)$	Dropout	0.1
Separable Conv	$\mathbf{H}_3$	$1 \times \frac{T}{p_1} \times F_2$	F_2, kernel	48, $(1, 16)$
Batch Norm	$\mathbf{H}_3^a$	$1 \times \frac{T}{p_1} \times F_2$	Activation	ELU
Average Pooling	$\mathbf{H}_3^p$	$1 \times \frac{T}{p_1 p_2} \times F_2$	p_2	8
Spatial Dropout	$\mathbf{H}'_3$	$1 \times \frac{T}{p_1 p_2} \times F_2$	Dropout	0.1
Transformer Encoder $\mathcal{T}$				
Reshape	$\mathbf{H}_\varphi$	$T' \times d$	T', d	8, 48
Transformer Encoder	$\mathbf{H}_\tau$	$T' \times d$	L, M, N', Dropout	1, 2, 2, 0.3
Classification Head ψ				
Attention Pooling	z	d	–	–
Dropout	$\tilde{z}$	d	Dropout	0.3
Dense classifier	$\tilde{p}$	1	Activation	Sigmoid

3.2. Model Interpretability and Ablation Analysis

To provide a comprehensive interpretation of the proposed EEG-TACT, this section analyzes the model from three complementary perspectives. First, the learned parameters are examined at two key layers: the first temporal convolutional filters, which describe how the front-end filters reweight the available EEG frequency content, and the transformer self-attention, which explains the model's distribution of relevance across latent temporal tokens. Second, the evolution of the learned data representation is qualitatively inspected through two-dimensional visualizations that compare the feature spaces at the outputs of the Feature Embedding $\varphi(\cdot)$ and the Transformer-Based Contextual Encoding $\mathcal{T}(\cdot)$ operators. Third, an ablation analysis quantifies the contribution of each architectural component by progressively evaluating the effect of temporal self-attention and attention-based pooling under the same subject-independent validation protocol.

Regarding the first temporal convolution, Figure 2 shows the magnitude responses of the learned filters. Taking into account the preprocessing stage, the learned convolutions are interpreted as supervised reweighting of the available spectral content within 4-40 Hz rather than an unconstrained discovery of the full EEG spectrum. As a first insight, the optimal layer uses temporal kernels shared across all 19 EEG channels, making the filters global temporal feature extractors applied uniformly across the 10-20 montage instead of targeting channel-specific or topographical effects, following the EEGNet-inspired principle of first learning temporal filters and then spatial channel representations [21]. At the filter level, narrow attenuations and abrupt variations in the frequency response emerge, indicating that the convolutional stage builds specialized spectral selectors before the subsequent spatial and attention-based layers. This interpretation is consistent with recent EEG decoding frameworks based on filter-bank, multiscale, and explainable temporal convolutional networks, which show that convolutional front-ends can capture task-relevant spectral, temporal, and spectral–spatial EEG representations [33,34]. Notice F1-F3 and F5 target low-to-mid frequency components, F4 and

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

F6 highlight beta-range components, and F7-F8 provide broader, complementary profiles across the passband. This filter setup, focusing on lower frequencies and extending into beta, shows that the model learns a data-driven filter bank that represents interactions among theta, alpha, and beta range activity, rather than relying on a single spectral marker. The resulting model thus offers a physiological explanation of the temporal layer as a set of complementary, trainable, frequency-selective transformations commonly linked to ADHD-related EEG changes [35].

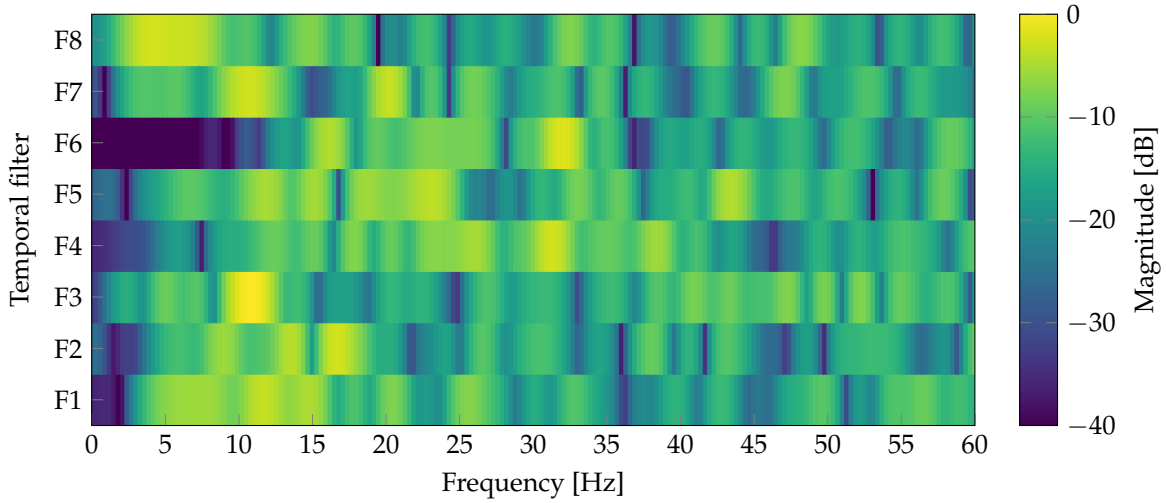


Figure 2. Frequency-domain magnitude responses of the eight temporal convolutional filters learned by EEG-TACT at the first embedding layer. Filters cluster into complementary low-to-mid frequency selectors (F1–F3, F5), beta-range emphasizers (F4, F6), and broadband profiles (F7–F8), forming a data-driven filter bank rather, indicating that the convolutional block captures task-relevant spectral diversity

To interpret the contextual encoding stage of EEG-TACT, average self-attention matrices ($\sigma\left(\frac{\mathbf{Q}_m\mathbf{K}_m^\top}{\sqrt{d}}\right)$ from Equation (3)) are extracted from the Transformer encoder on the held-out test set. For each attention head, attention scores are grouped by subject label (ADHD or Control) and averaged across test epochs to obtain class-specific attention maps. Since the Transformer operates over the latent temporal tokens produced by the convolutional embedding block, each attention matrix describes how strongly each encoded temporal segment attends to the others after the EEG signal has already been transformed into a compact spatiotemporal representation. Therefore, these maps describe class-conditioned patterns of interaction among learned temporal EEG representations, following the self-attention principle of modeling pairwise dependencies across sequence elements [23].

As shown in Figure 3, the two optimal attention heads (left and right) perform as complementary contextualizations within the transformer. Head 1 (left) shows a more localized structure, with attention concentrated over specific query–key interactions. This behavior is consistent with a discriminative mechanism that emphasizes particular temporal dependencies within the latent sequence. In turn, Head 2 (right) exhibits a more distributed organization, with attention spread across broader regions of the matrix. This pattern suggests a more global temporal encoding mechanism, in which each token integrates information from a wider portion of the embedded sequence. Regarding the discriminative behavior (top to bottom), the average attention maps reveal noticeable differences between classes. The ADHD group (top) exhibits a more structured, higher-contrast organization, with attention weights concentrated in specific query–key regions. This indicates that, during ADHD epochs, the Transformer tends to assign greater relevance to selected interactions among latent temporal tokens rather than distributing attention uniformly across the sequence. Such localized concentration suggests that discriminative information may be carried by specific temporal relationships within the embedded EEG representation, possibly reflecting more irregular or task-relevant temporal dynamics that the model learns to emphasize. In contrast, the Control group (bottom) displays a smoother, more diffuse attention profile, corresponding to a more homogeneous integration of information across temporal

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

tokens. Such a difference suggests that the contextual encoding required to classify Control epochs may rely on broader localized temporal relationships, whereas ADHD epochs induce a more selective attention pattern. Therefore, the self-attention analysis shows that the Transformer operator enables EEG-TACT to distinguish between ADHD and Control by reorganizing local EEG features into class-dependent temporal patterns, combining both localized and distributed attention mechanisms to capture distinctive temporal interactions within each group.Although attention maps should not be interpreted as direct causal or neurophysiological explanations, they provide useful model-level evidence about which latent temporal relationships are emphasized during classification [36,37].

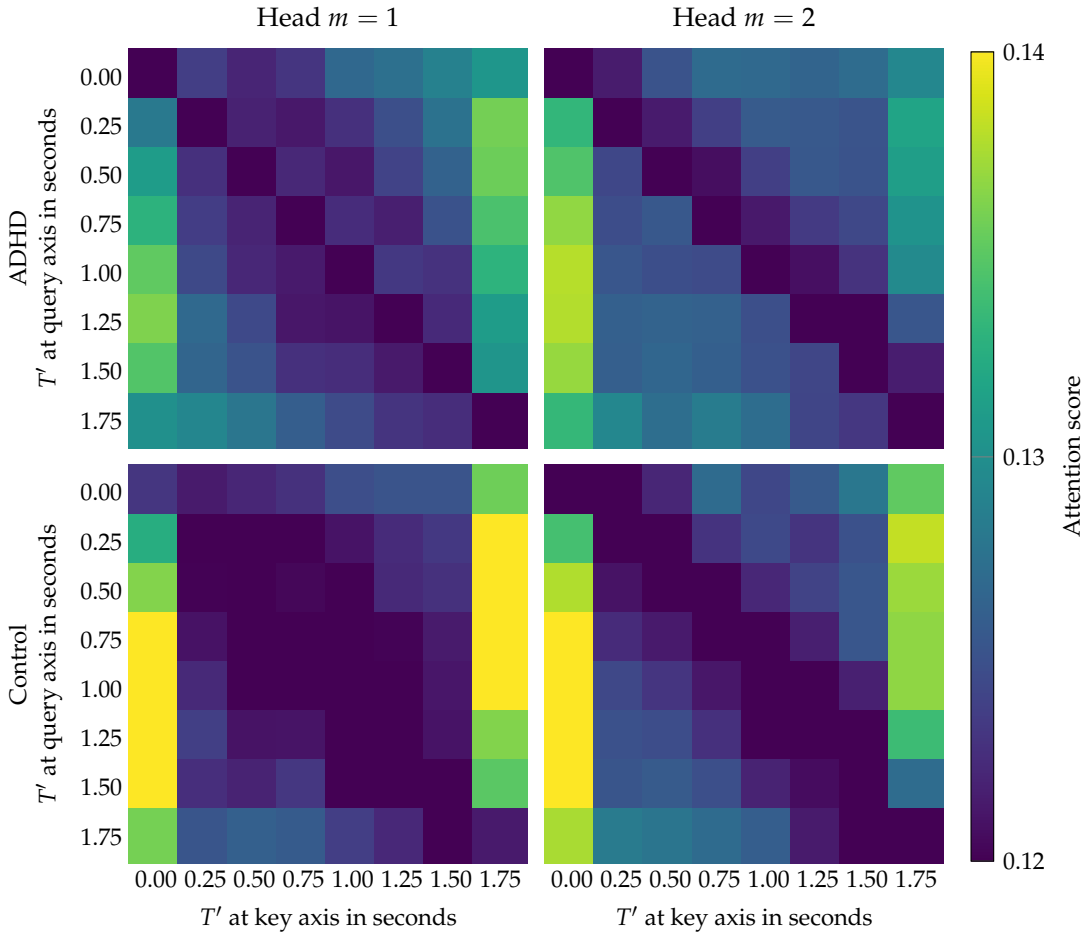


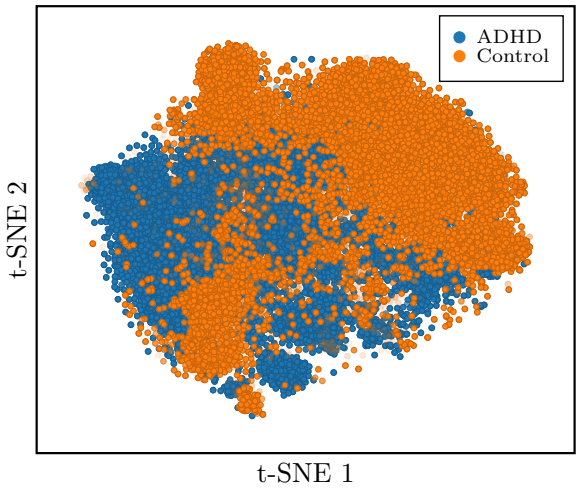
Figure 3. Average of self-attention matrices $\sigma\left(\frac{\mathbf{Q}_m\mathbf{K}_m^\top}{\sqrt{d}}\right)$ of the Transformer encoder on the test set for each class and attention head. Key and query axes represent latent temporal tokens T' within the epochs, expressed in seconds. Subjects with ADHD produce contrast patters, while controls show a more diffuse distribution among tokes. Heads complement more localized and distributed patters.

Beyond individual learned parameters, Figure 4 illustrates how the internal data representation evolves across the EEG-TACT pipeline. The t-SNE projection of the feature embedding $\mathbf{H}_\varphi$ (top) exhibits substantial overlap between ADHD and Control epochs [38]. This indicates that the convolutional embedding already extracts informative signal structure, but the resulting latent space remains only partially separable when considered before contextual temporal modeling. After the Transformer encoder, the t-SNE projection of $\mathbf{H}_\tau$ (bottom left) more clearly organizes the two classes, with ADHD and Control samples occupying more distinguishable regions of the embedded space. This change suggests that self-attention contributes not only by weighting temporal interactions, as observed in Figure 3, but also by reshaping the latent representation into a more discriminative geometry. Further, the subject-colored representation in bottom right provides an additional interpretation of the latent organization, noting that epochs from the same subject tend to form local neighborhoods or coherent

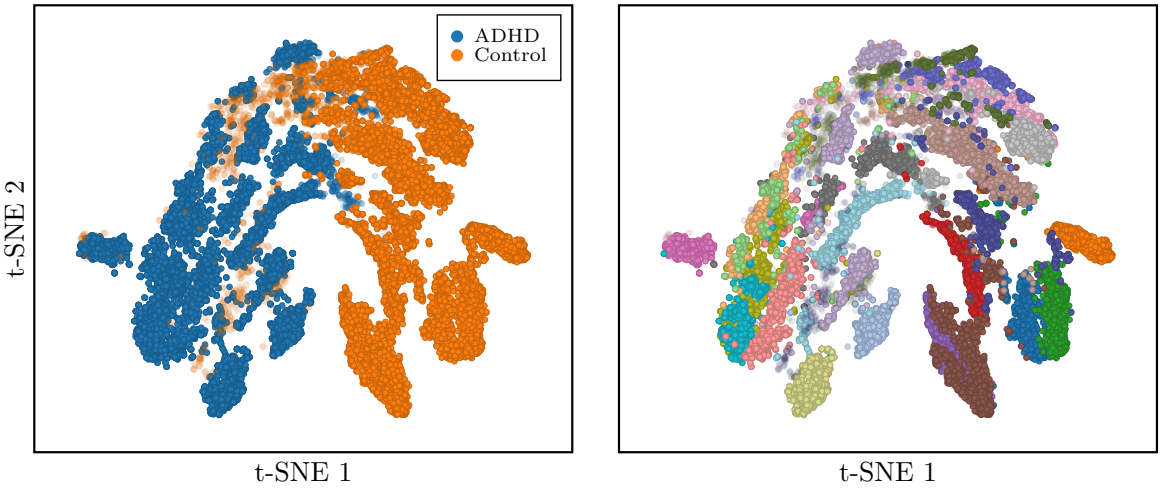
© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

trajectories in the projected space. Those neighborhoods are expected in EEG data because epochs from the same participant share subject-specific neurophysiological and recording characteristics. Such an insight supports the need for subject-grouped cross-validation to prevent subject-specific structure from causing information leakage between the training and test partitions [39]. As a result, the proposed EEG-TACT moves local convolutional features towards a more robust class organization while localizing subject-specific patterns.



(**a**) Feature embedding $\mathbf{H}_{\varphi}$.



(**b**) Transformer encoding $\mathbf{H}_{\mathcal{T}}$

(**c**) Transformer encoding $\mathbf{H}_{\mathcal{T}}$ by subject.

Figure 4. t-SNE visualization of the learned EEG representations for the temporal tokens at (a) the feature embedding $\mathbf{H}_{\varphi}$ (a) and (b,c) after the transformer encoding $\mathbf{H}_{T}$. (a) and (b) identify classes by color. (c) color codes tokens according to the subject.

The last interpretability perspective concerns the contribution of the main architectural components. To this end, an ablation study progressively evaluates the effects of temporal self-attention and attention-based pooling under a common validation protocol. This design ensures that the observed differences are attributable to architectural changes rather than to data-splitting variability, since all configurations are assessed using the same subject-grouped, five-fold stratified cross-validation protocol, a fixed random seed, and identical subject-wise partitions. The ablation design evaluates four configurations: an EEGNet-like baseline using flatten-based aggregation [21], an EEGNet variant equipped with attention pooling, a Transformer-enhanced variant using global average pooling [23], and the complete EEG-TACT architecture combining the Transformer encoder with attention-based aggregation. In this setting, the Transformer block evaluates the benefit of temporal self-attention

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

for contextual sequence modeling, whereas the pooling strategy evaluates how temporal features are summarized before classification.

Table 2. Progressive ablation analysis under subject-grouped, five-fold stratified cross-validation protocol using a fixed random seed. Accuracy, recall, and precision are reported as mean ± standard deviation (%) across folds.

	Architecture		Performance (%)		
Configuration	**Transformer**	**Pooling**	**Accuracy**	**Recall**	**Precision**
EEGNet-like Baseline	✗	Flatten	80.8 ± 8.1	91.0 ± 12.6	76.8 ± 9.6
Baseline + Attention Pooling	✗	Attention	81.7 ± 7.6	91.5 ± 14.4	78.3 ± 10.7
Baseline + Transformer	✓	Global Average	82.5 ± 4.6	90.8 ± 9.6	78.8 ± 7.3
EEG-TACT (Proposal)	✓	Attention	$\mathbf{87.5 \pm 5.1}$	$\mathbf{96.0 \pm 8.9}$	$\mathbf{82.8 \pm 4.0}$

Table 2 reports the mean fold-wise performance of each configuration. The results show a consistent improvement from the most straightforward configuration to the complete proposed model, supporting the complementary roles of temporal contextualization and adaptive aggregation. The EEGNet-like baseline provides the lowest overall performance, indicating that flatten-based aggregation has limited capacity to summarize temporally distributed EEG features in a subject-independent setting. Then, introducing attention pooling slightly improves classification, suggesting that learnable aggregation can help emphasize informative temporal segments even when the preceding representation is generated only by convolutional operations. Adding the Transformer block while using global average pooling further improves performance and reduces fold-to-fold variability, indicating that self-attention provides a more stable and discriminative temporal representation by modeling dependencies across embedded EEG segments. Nevertheless, the complete EEG-TACT outperforms all previous ablated configurations by providing a task-adaptive attention pooling to weight the most relevant temporal representations before classification. These results support the architectural premise that the synergistic contributions of convolutional feature extraction, temporal self-attention, and attention-based pooling improve the identification of ADHD from EEG signals.

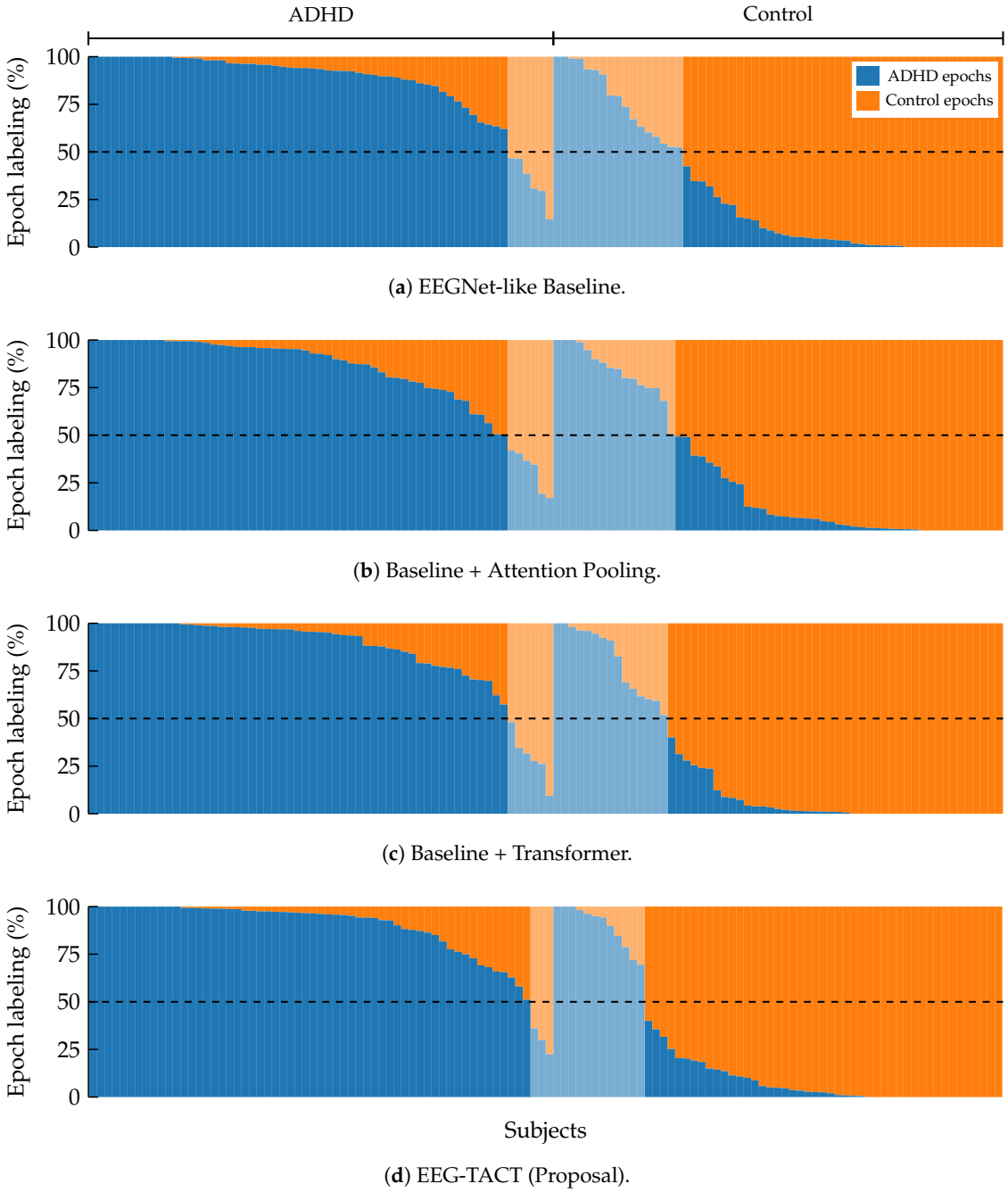
Complementarily, Figure 5 presents the subject-level epoch classification obtained by the four configurations evaluated in the ablation study. In each panel, vertical bars correspond to subjects and summarize the percentage of epochs assigned to the ADHD (blue) and Control (orange) classes. The dashed horizontal line indicates the 50% majority-vote threshold used to convert epoch-level predictions into a final subject-level decision. Subjects are ordered to emphasize classification confidence: true ADHD subjects are placed on the left and sorted from the highest to the lowest percentage of ADHD votes, whereas true Control subjects are placed on the right and sorted from the highest to the lowest percentage of Control votes. Therefore, correctly classified subjects appear as bars dominated by the expected class color on each side of the plot, while subjects near the center represent the most ambiguous cases. The pale central band highlights misclassified subjects; consequently, its width provides a direct visual indication of the subject-level classification error for each configuration.

As in Table 2, a positive quality progression is observed when introducing the architectural components. The EEGNet-like baseline exhibits the least favorable epoch-labeling profile, with a wider pale central band and several subjects near the 50% decision threshold, indicating that flatten-based aggregation produces less reliable subject-level evidence. Despite improving the baseline by reducing epoch mislabeling, adding the attention pooling still presents a persistent misclassification band, indicating that adaptive pooling alone is insufficient when the temporal representation is generated solely by convolutional operations. In turn, incorporating the Transformer block with global average pooling further improves the epoch-labeling structure, as the correct-class percentages decrease more gradually toward the center and fewer subjects fall into the mislabeled region, supporting the role of self-attention in producing more coherent contextual representations across epochs. Lastly, the complete EEG-TACT yields the most favorable pattern: the pale band is narrower, many correctly classified subjects retain high true-class vote percentages, and the remaining errors are less dominated

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

by mislabeled epochs. Such a labeling quality progression reinforces the ablation findings by showing that the combination of Transformer-based temporal contextualization and attention-based pooling leads to more reliable majority-vote behavior than either component alone, and indicating that EEG-TACT improves both the number and the quality of subject-level decisions.



(**a**) EEGNet-like Baseline.

(**b**) Baseline + Attention Pooling.

(**c**) Baseline + Transformer.

(**d**) EEG-TACT (Proposal).

Figure 5. Subject-wise epoch-labeling profiles for the ablation study under the SGKF-CV protocol. Each panel corresponds to one compared configuration. For each subject, stacked bars show the percentage of epochs classified as ADHD and Control. True ADHD subjects are displayed on the left and sorted from highest to lowest ADHD-vote percentage. True Control subjects are displayed on the right and sorted from highest to lowest Control-vote percentage toward the center. The pale central band indicates misclassified subjects.

3.3. Performance and Significance Analysis

For comparative performance evaluation, the proposed EEG-TACT is benchmarked against representative EEG deep-learning architectures previously used for EEG decoding and ADHD-related classification tasks. EEGNet is included as a compact EEG-specific convolutional network that com-

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

bines temporal convolutions, depthwise spatial filtering, and separable convolutions to efficiently learn frequency-specific spatial representations with a reduced number of trainable parameters [21]. The architecture of the ShallowConvNet, inspired by the filter-banked common spatial patterns, emphasizes band-power modulations via temporal and spatial convolutions, followed by squaring, average pooling, and logarithmic activation [20]. CNN-LSTM combines convolutional layers for local spatiotemporal feature extraction with recurrent LSTM units to model longer temporal dependencies in ERP-related ADHD responses [40]. In addition, the residual convolutions of the Multi-Stream model exploit skip connections to stabilize deeper temporal feature learning and mitigate degradation or vanishing gradient effects [41]. The integrated IM-CBGT combines convolutional layers for local feature extraction, bidirectional LSTM units for modeling forward–backward temporal dependencies, and GRU-Transformer blocks with multi-head self-attention to capture global sequence relationships before final classification [42]. Finally, T-GARNet constitutes a recent ADHD-oriented architecture that integrates Transformer-based temporal attention, multi-scale Gaussian-kernel functional connectivity, and Rényi-information regularization to capture long-range temporal dependencies, spatial interactions among channels, and interpretable connectivity patterns [43].

All state-of-the-art models are trained and validated under the same experimental conditions as EEG-TACT, including identical preprocessing, subject-wise data partitioning, training criteria, and subject-level performance estimation, thereby ensuring that the reported differences reflect architectural and representational factors rather than protocol-dependent biases. Specifically, each model is validated using a varying-seed, subject-independent protocol. First, five fixed subject-wise folds are generated using stratified group cross-validation. Then, the training procedure is repeated ten times with different random seeds, yielding 50 test evaluations. This design provides a more robust and reproducible estimate of model performance than a single-run evaluation, since it accounts not only for variability across unseen subject partitions but also for stochastic effects arising from weight initialization, mini-batch ordering, dropout, and optimizer dynamics.

Table 3 summarizes the comparative performance of the models across two complementary statistical settings: the varying-seed evaluation over 50 test outcomes and a subject-level matched comparison using the best seed. The varying-seed setting reports the mean and standard deviation across the five fixed stratified group folds and the ten independent random seeds, while the best-seed setting averages the five folds from the best initialized model. Each setting is supported with a statistical test. For the varying-seed analysis, the Wilcoxon signed-rank test is used to evaluate statistical differences between EEG-TACT and each baseline over the 50 paired scores, without assuming a parametric form for the distributions of the reported metrics [44]. For the latter, the McNemar test validates the significance of the odds ratio for paired subject-level decisions, that is, whether the number of subjects correctly classified by EEG-TACT but misclassified by the baseline differs significantly from the number of subjects correctly classified by the baseline but misclassified by EEG-TACT [44]. Filled cells indicate statistically significant differences ($p < 5\%$) relative to EEG-TACT, determined using the Wilcoxon or McNemar test.

In the first setting, EEG-TACT achieves the most robust overall performance, only reached by the CNN-LSTM accuracy. Nonetheless, the smaller standard deviation of EEG-TACT indicates that the proposal is less sensitive to stochastic training factors such as weight initialization, mini-batch ordering, dropout, and optimizer dynamics. This stability is further reflected in its best average rank across accuracy scores, suggesting that the proposed model performs consistently well across repetitions rather than depending on a favorable seed. In terms of class-wise performance, EEG-TACT achieves the highest precision while maintaining high recall, yielding a more balanced discrimination between ADHD and control. By contrast, T-GARNet and CNN-LSTM obtain very high recall but noticeably lower precision, suggesting a stronger tendency to classify subjects as ADHD and therefore to increase false positives. The Wilcoxon tests confirm the significance of the improvement against EEGNet, ShallowConvNet, Multi-Stream, IM-CBGT, and T-GARNet. It should be noted that the significant difference in the Recall score between EEG-TACT and T-GARNet favors the latter.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

Table 3. Performance statistics for contrasted models under ten varying seeds (top) and at the best seed (bottom). Metrics are reported as mean ± standard deviation. Bold-face indicates the best model in the row. For varying seed, filled cell indicates significative difference ($p < 5\%$) between EEG-TACT and the baseline using the Wilcoxon signed-rank tests on paired scores. In the odds ratio row, filled cell indicates significative difference ($p < 5\%$) according to the McNemar test.

Model	EEGNet	ShallowConvNet	CNN-LSTM	Multi-Stream	IM-CBGT	T-GARNet	EEG-TACT
Year	2017	2017	2022	2023	2024	2025	2026
Parameters	1,746	36,522	9,052	574,082	1,195,266	9,071	7,322
Varying seed (average over 50 folds)							
Accuracy (%)	80.7 ± 1.3	78.7 ± 3.2	$\mathbf{84.2 \pm 7.4}$	69.2 ± 2.9	70.6 ± 9.6	74.8 ± 2.9	$\mathbf{84.2 \pm 1.8}$
Recall (%)	88.8 ± 2.0	84.9 ± 12.2	93.8 ± 8.9	79.3 ± 3.6	80.0 ± 12.0	$\mathbf{96.2 \pm 1.3}$	92.8 ± 2.6
Precision (%)	77.7 ± 1.2	78.9 ± 5.3	79.9 ± 7.7	67.9 ± 3.3	68.9 ± 9.6	68.3 ± 2.1	$\mathbf{80.0 \pm 1.6}$
Average Rank	3.53	3.63	2.44	5.85	5.42	4.70	**2.43**
Best seed (average over 5 folds)							
Accuracy (%)	79.2 ± 9.3	80.8 ± 3.7	80.0 ± 10.8	73.3 ± 5.6	69.2 ± 9.6	70.0 ± 14.3	$\mathbf{87.5 \pm 5.1}$
Recall (%)	88.4 ± 16.5	91.1 ± 9.0	93.3 ± 14.9	79.1 ± 13.4	77.4 ± 11.0	$\mathbf{96.9 \pm 4.4}$	96.0 ± 8.9
Precision (%)	76.2 ± 8.5	76.7 ± 8.2	76.0 ± 11.3	74.1 ± 12.8	68.4 ± 10.4	64.8 ± 12.2	$\mathbf{82.8 \pm 4.0}$
Odds ratio	11/1	11/3	11/2	24/7	28/6	23/2	–

In the best-seed setting, EEG-TACT achieves the highest subject-level accuracy and precision, while maintaining one of the highest recall values, indicating that under the most favorable initialization, EEG-TACT not only detects ADHD subjects effectively but also better controls false positive ADHD predictions than the competing models. Such a dissimilarity is particularly relevant because several baselines, especially T-GARNet and CNN-LSTM, reach comparable or even slightly higher recall but at the cost of substantially lower precision, reflecting a less balanced diagnostic behavior. Further, the odds ratios support the subject-level superiority of EEG-TACT. For example, the 11/2 odds ratio against CNN-LSTM indicates that EEG-TACT correctly classified eleven subjects missed by CNN-LSTM, whereas CNN-LSTM corrected only two subject missed by EEG-TACT. The McNemar test reveals several statistically significant differences, indicating that the improvement is not only reflected in aggregate metrics but also translates into concrete gains in subject-level decisions. As a general finding, EEG-TACT achieves a favorable accuracy–reliability–complexity trade-off, demonstrating that performance advantages arise from an effective EEG-specific architectural bias rather than increased model size, and supporting its suitability for robust subject-independent ADHD detection.

Due to the majority voting scheme for subject labeling, a risk-coverage analysis is finally conducted to assess the reliability of the model predictions at the epoch level using the best-performing seed for each architecture [45]. Figure 6 illustrates the coverage-risk curves for each model, where a well-calibrated classifier is expected to produce low selective risk at low coverage levels and a gradual increase as coverage approaches 100%. The risk-coverage curves reveal marked differences in the confidence-ranking behavior of the evaluated models. Among the baselines, Multi-Stream and IM-CBGT exhibit substantially higher selective risk across nearly the entire coverage range, which is associated with poor confidence calibration and a larger fraction of high-confidence errors. The non-monotonic growing of T-GARNet, with a sharp increase in risk at low coverage and a partial decrease at intermediate coverage, implies that the model may assign excessive confidence to a small number of incorrect predictions. Remaining models exhibit a typical evolution: their risk remains relatively low at low coverage values but increases more rapidly as coverage increases, suggesting weaker separation between reliable and unreliable epochs. Finally, EEG-TACT yields the lowest selective risk, with the smoothest risk increment over coverage. Therefore, the proposed model provides confidence estimates with a reliable ordering of easy and difficult trials, demonstrating that epoch-level predictions are not only more accurate but also better ordered by confidence, yielding a more favorable selective classification profile under uncertainty.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

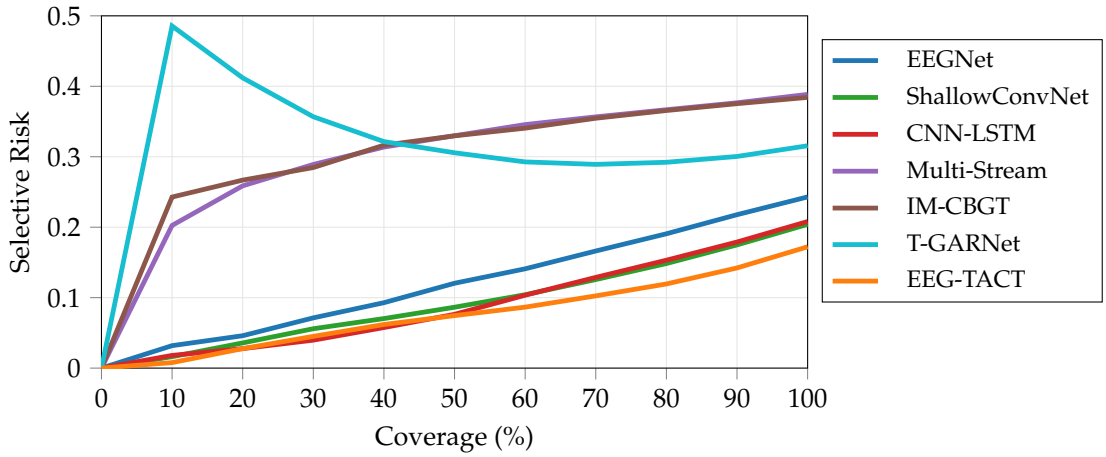


Figure 6. Selective risk versus coverage for the evaluated models using the best-performing seed. Coverage levels are evaluated every 10%. Selective risk is computed as the error rate over the retained epochs. EEG-TACT achieves the lowest selective risk and the smoothest increase across the coverage range, indicating improved confidence calibration and a more reliable ranking of easy and difficult trials under uncertainty.

4. Concluding Remarks and Future Work

This work introduces EEG-TACT, a compact end-to-end architecture for subject-level identification of ADHD from EEG epochs by integrating convolutional feature extraction, transformer-based temporal contextualization, and attention-based aggregation into a single differentiable framework. This design aims to learn frequency-selective temporal filters, spatial channel representations, contextual dependencies among latent temporal tokens, and adaptive temporal summaries for classification. The improvement on EEG decoding under clinically meaningful subject-level evaluation validates the competitive, statistically robust performance of EEG-TACT while maintaining a reduced number of trainable parameters, indicating that its performance gains arise from an appropriate architectural bias rather than increased model complexity alone.

The main findings support the relationship of convolutional feature extraction, temporal self-attention, and adaptive pooling, which improved both representation learning and subject-level classification performance. The learned temporal filters work as data-driven spectral reweighting mechanisms within the preprocessed EEG band, capturing complementary low-, mid-, and beta-range patterns. In addition, the self-attention maps highlight class-dependent temporal interactions among latent tokens, while the latent-space projections reveal local EEG feature reorganization from the Transformer-based contextual encoding into a more discriminative representation. The ablation analysis further confirmed that neither temporal self-attention nor attention pooling alone fully explains the observed improvement. Instead, the complete EEG-TACT configuration provided the most favorable subject-level labeling profiles.

From a methodological and practical perspective, this study reinforces the importance of evaluating EEG-based ADHD classifiers under strict subject-independent protocols. By using stratified group partitions, nested validation, and statistical testing, the proposed framework reduces the risk of information leakage and provides a more realistic estimate of model behavior on unseen participants. These aspects are particularly relevant for translational EEG research, where a model should not only achieve high accuracy but also produce stable, interpretable, and confidence-consistent predictions. Therefore, EEG-TACT has the potential to serve as a computational decision-support approach to complement comprehensive clinical assessment.

Despite the promising and reliable results, the following identified limitations offer future research directions. First, the model was trained at the epoch level, requiring post-hoc majority voting for subject labeling and constraining the time-horizon analysis. From a methodological perspective, new experiments may include efficient attention mechanisms, temporal conformer-like modules, hierarchical transformers, and spatial-temporal attention schemes to integrate whole-recording classification and

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

improve long-range dependency modeling while preserving computational efficiency [46]. Second, the model was evaluated on a single public binary ADHD/control EEG dataset, which limits conclusions about generalization across recording conditions (e.g., multi-site datasets, electrode montages, and recording devices) and clinical populations (e.g., ADHD subtypes, comorbidities, medication effects, and longitudinal clinical variability). Future work should not only externally validate on varying protocol settings, but also prioritize extending the EEG-TACT architecture to support heterogeneous EEG acquisition conditions and to solve multiple tasks towards a clinical foundational model [47]. Lastly, although attention maps and learned filters provide useful model-level interpretability, they should not be interpreted as direct causal neurophysiological explanations. Hence, a thorough validation on model interpretability should be conducted, even including transformer variants and calibration- or uncertainty-aware model responses [48].

Author Contributions: Conceptualization, J.D.-P., A.G.-R. and D.C.-P.; methodology, A.A.-M. and J.G.-G.; validation, J.D.-P. and A.G.-R.; formal analysis, J.D.-P., A.G.-R., D.C.-P. and A.A.-M.; investigation, J.D.-P., A.G.-R. and D.C.-P.; resources, J.G.-G. and D.C.-P.; writing—original draft preparation, J.D.-P. and A.G.-R.; writing—review and editing, D.C.-P. and A.A.-M.; visualization, D.C.-P.; supervision, A.A.-M.; project administration, D.C.-P.; funding acquisition, A.A.-M., D.C.-P. and J.G.-G. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the financial support provided by the research program "Alianza Científica con Enfoque Comunitario para Mitigar Brechas de Atención y Manejo de Trastornos Mentales Relacionados con Impulsividad en Colombia – ACEMATE" through grant 111091991908.

Data Availability Statement: The EEG data for ADHD/control children used in this study are publicly available through IEEE DataPort under the title *EEG Data for ADHD/Control Children*, used in the validation of the current study, is publicly available at https://ieee-dataport.org/open-access/eeg-data-adhd-control-children.

Acknowledgments: The authors used Grammarly only to support English language editing during the preparation of this manuscript. All scientific content, interpretations, and conclusions were reviewed, revised, and approved by the authors, who take full responsibility for the final version of the publication.

Conflicts of Interest: The authors declare that they have no conflicts of interest. The funding institution was not involved in the study design, data collection, data analysis, interpretation of the results, manuscript preparation, or the decision to submit the work for publication.

References

1. Pino, A.; Papatheodorou, N.; Kouroupetroglou, G.; Giannopoulos, P.A.; Makris, G.; Papageorgiou, C. Hand Dexterity Evaluation Grounded on Cursor Trajectory Investigation in Children with ADHD Using a Mouse and a Joystick. *Technologies* **2025**, *13*. https://doi.org/10.3390/technologies13030099.
2. Salari, N.; Ghasemi, H.; Abdoli, N.; Rahmani, A.; Shiri, M.H.; Hashemian, A.H.; Akbari, H.; Mohammadi, M. The Global Prevalence of ADHD in Children and Adolescents: A Systematic Review and Meta-Analysis. *Italian Journal of Pediatrics* **2023**, *49*, 48. https://doi.org/10.1186/s13052-023-01456-1.
3. Faraone, S.V.; Banaschewski, T.; Coghill, D.; Zheng, Y.; Biederman, J.; Bellgrove, M.A.; Newcorn, J.H.; Gignac, M.; Al Saud, N.M.; Manor, I.; et al. The World Federation of ADHD International Consensus Statement: 208 Evidence-Based Conclusions About the Disorder. *Neuroscience & Biobehavioral Reviews* **2021**, *128*, 789–818. https://doi.org/10.1016/j.neubiorev.2021.01.022.
4. Dhiabi, R.; Walha, R.; Drira, F. A Neural Approach to ADHD Detection in Children: Enhanced EEG Analysis with Wavelet-Transformer Synergy. In Proceedings of the Proceedings of the 18th International Conference on Agents and Artificial Intelligence (ICAART 2026) - Volume 3. SCITEPRESS, 2026, pp. 2899–2906. https://doi.org/10.5220/0014334800004052.
5. Peterson, B.S.; Trampush, J.; Brown, M.; Maglione, M.; Bolshakova, M.; Rozelle, M.; Miles, J.; Pakdaman, S.; Yagyu, S.; Motala, A.; et al. Tools for the Diagnosis of ADHD in Children and Adolescents: A Systematic Review. *Pediatrics* **2024**, *153*, e2024065854. https://doi.org/10.1542/peds.2024-065854.
6. Knyazhansky, M.; Shrot, T. ADHD Diagnostic Tools Across Ages: Traditional and Digital Approaches. *Frontiers in Psychiatry* **2025**, *16*, 1668070. https://doi.org/10.3389/fpsyt.2025.1668070.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

7. Chan, H.K.; Rowe, R.; Carroll, D. Factors Associated with Parent-Teacher Hyperactivity/Inattention Screening Discrepancy: Findings from a UK National Sample. *PLOS ONE* **2024**, *19*, e0299980. https://doi.org/10.1371/journal.pone.0299980.
8. Gondek, T. Diagnosing ADHD in adults: Diagnostic tools and differential diagnosis. *European Psychiatry* **2021**, *64*, S72–S73. https://doi.org/10.1192/j.eurpsy.2021.226.
9. Fallahpour, B.; Dastjerdi, G.; Akbarian, E.; Emarati, A.; Sadr, Z.; Dastgheib, S.A.; Shahbazi, A.; Bahrami, R.; Golshan-Tafti, M.; Shiri, A.; et al. Artificial Intelligence in ADHD Diagnosis: A Comprehensive Review of Machine Learning Applications, Clinical Validation Challenges, and Implementation Barriers in Precision Medicine. *Egyptian Pediatric Association Gazette* **2026**, *74*, 36. https://doi.org/10.1186/s43054-026-00530-7.
10. Bian, J.; Liu, X.; Wang, C. Executive Function and Brain Region Development in ADHD: Mechanisms and Interventions in the Prefrontal Cortex and Related Circuits. *Advances in Precision Medicine* **2025**, *10*, 15–21. https://doi.org/10.18063/APM.V10I1.681.
11. Wang, C.; Wang, S.; Sun, L.; Sui, J. Abnormal MRI Features in Children with ADHD: A Narrative Review of Large-Scale Studies. *Brain Sciences* **2026**, *16*, 104. https://doi.org/10.3390/brainsci16010104.
12. Tian, L.; Zheng, H.; Zhang, K.; Qiu, J.; Song, X.; Li, S.; Zeng, Z.; Ran, B.; Deng, X.; Cai, J. Structural or/and Functional MRI-Based Machine Learning Techniques for Attention-Deficit/Hyperactivity Disorder Diagnosis: A Systematic Review and Meta-Analysis. *Journal of Affective Disorders* **2024**, *355*, 459–469. https://doi.org/10.1016/j.jad.2024.03.111.
13. Quintero López, C.; Gil Vera, V.D.; Ruiz Quintero, M.J. Diagnosis of ADHD in Children with EEG and Machine Learning: Systematic Review and Meta-Analysis. *Clínica y Salud* **2025**, *36*, 109–121. https://doi.org/10.5093/clh2025a16.
14. Mao, Y.; Qi, X.; He, L.; Wang, S.; Wang, Z.; Wang, F. Advanced Machine Learning Techniques Reveal Multidimensional EEG Abnormalities in Children with ADHD: A Framework for Automatic Diagnosis. *Frontiers in Psychiatry* **2025**, *16*, 1475936. https://doi.org/10.3389/fpsyt.2025.1475936.
15. López, C.Q.; Vera, V.D.G.; Quintero, M.J.R. Diagnosis of ADHD in children with EEG and machine learning: Systematic review and meta-analysis. *Clinical and Health* **2025**, *36*, 109–121. https://doi.org/10.5093/clh2025a16.
16. Kim, J.W.; Kim, B.N.; Kim, J.I.; Yang, C.M.; Kwon, J. Electroencephalogram (EEG) Based Prediction of Attention Deficit Hyperactivity Disorder (ADHD) Using Machine Learning. *Neuropsychiatric Disease and Treatment* **2025**, *21*, 271–279. https://doi.org/10.2147/NDT.S509094.
17. Maya-Piedrahita, M.C.; Herrera-Gomez, P.M.; Berrio-Mesa, L.; Cardenas-Pena, D.A.; Orozco-Gutierrez, A.A. Supported Diagnosis of Attention Deficit and Hyperactivity Disorder from EEG Based on Interpretable Kernels for Hidden Markov Models. *International Journal of Neural Systems* **2022**, *32*, 2250008. https://doi.org/10.1142/S0129065722500083.
18. Belhadi, A.; Yazidi, A.; Lind, P.G.; Djenouri, Y. EEG Data Classification: Review and Taxonomy. *ACM Transactions on Computing for Healthcare* **2025**, *6*, 45. https://doi.org/10.1145/3742795.
19. Lyu, R. Deep Learning Approaches for EEG-Based Healthcare Applications: A Comprehensive Review. *Frontiers in Human Neuroscience* **2026**, *19*, 1689073. https://doi.org/10.3389/fnhum.2025.1689073.
20. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Human Brain Mapping* **2017**, *38*, 5391–5420. https://doi.org/10.1002/hbm.23730.
21. 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. *Journal of Neural Engineering* **2018**, *15*, 056013. https://doi.org/10.1088/1741-2552/aace8c.
22. Song, Y.; Zheng, Q.; Liu, B.; Gao, X. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2023**, *31*, 710–719. https://doi.org/10.1109/TNSRE.2022.3230250.
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, 2017, Vol. 30. https://doi.org/10.48550/arXiv.1706.03762.
24. Zhao, W.; Jiang, X.; Zhang, B.; Xiao, S.; Weng, S. CTNet: A Convolutional Transformer Network for EEG-Based Motor Imagery Classification. *Scientific Reports* **2024**, *14*, 20237. https://doi.org/10.1038/s41598-024-71118-7.
25. Nasrabadi, A.M.; Allahverdy, A.; Samavati, M.; Mohammadi, M.R. EEG data for ADHD/control children. IEEE DataPort, 2020. https://doi.org/10.21227/rzfh-zn36.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

26. Ekhlasi, A.; Motie Nasrabadi, A.; Mohammadi, M. Analysis of Effective Connectivity Strength in Children with Attention Deficit Hyperactivity Disorder Using Phase Transfer Entropy. *Iranian Journal of Psychiatry* **2021**, *16*, 374–382. https://doi.org/10.18502/ijps.v16i4.7224.
27. Li, L.; Guo, X.; Yang, Z.; Zhao, Y.; Liu, X.; Yang, J.; Chen, Y.; Peng, X.; Han, L. ADHD Detection from EEG Signals Using GCN Based on Multi-Domain Features. *Frontiers in Neuroscience* **2025**, *19*, 1561994. https://doi.org/10.3389/fnins.2025.1561994.
28. Attallah, O. ADHD-AID: Aiding Tool for Detecting Children's Attention Deficit Hyperactivity Disorder via EEG-Based Multi-Resolution Analysis and Feature Selection. *Biomimetics* **2024**, *9*, 188. https://doi.org/10.3390/biomimetics9030188.
29. Jovanović, V.; Petrušić, I.; Ković, V.; Savić, A.M. The Practical Implications of Re-Referencing in ERP Studies: The Case of N400 in the Picture–Word Verification Task. *Diagnostics* **2025**, *15*, 156. https://doi.org/10.3390/diagnostics15020156.
30. Kaur, G.; Aggarwal, H.; Goel, N. Artificial Intelligence Driven Neuropsychiatry: A Systematic Review of Electroencephalography-Based Computational Techniques for Major Depressive Disorder Prediction. *Neuroscience* **2025**, *581*, 179–207. https://doi.org/10.1016/j.neuroscience.2025.07.010.
31. Ilse, M.; Tomczak, J.; Welling, M. Attention-based Deep Multiple Instance Learning. In Proceedings of the Proceedings of the 35th International Conference on Machine Learning; Dy, J.; Krause, A., Eds. PMLR, 10–15 Jul 2018, Vol. 80, *Proceedings of Machine Learning Research*, pp. 2127–2136.
32. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, 2009.
33. Liu, K.; Yang, M.; Yu, Z.; Wang, G.; Wu, W. FBMSNet: A Filter-Bank Multi-Scale Convolutional Neural Network for EEG-Based Motor Imagery Decoding. *IEEE Transactions on Biomedical Engineering* **2023**, *70*, 436–445. https://doi.org/10.1109/TBME.2022.3193277.
34. Salami, A.; Andreu-Perez, J.; Gillmeister, H. EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification. *IEEE Access* **2022**, *10*, 36672–36685. https://doi.org/10.1109/ACCESS.2022.3161489.
35. Zhang, S.; Yu, S.; Cui, X.; Liang, L.; Li, X. Neural Oscillation Features of ADHD Symptoms in Children: EEG Evidence From Resting State and Oddball Task. *Journal of Attention Disorders* **2026**, *30*, 552–565. https://doi.org/10.1177/10870547251405008.
36. Jain, S.; Wallace, B.C. Attention is not Explanation. In Proceedings of the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota, 2019; Vol. 1, pp. 3543–3556. https://doi.org/10.18653/v1/N19-1357.
37. Wiegreffe, S.; Pinter, Y. Attention is not not Explanation. In Proceedings of the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 2019; pp. 11–20. https://doi.org/10.18653/v1/D19-1002.
38. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.
39. Kapoor, S.; Narayanan, A. Leakage and the Reproducibility Crisis in Machine-Learning-Based Science. *Patterns* **2023**, *4*, 100804. https://doi.org/10.1016/j.patter.2023.100804.
40. Wang, C.; Wang, X.; Jing, X.; Yokoi, H.; Huang, W.; Zhu, M.; Chen, S.; Li, G. Towards High-Accuracy Classifying Attention-Deficit/Hyperactivity Disorders Using CNN-LSTM Model. *Journal of Neural Engineering* **2022**, *19*, 046015. https://doi.org/10.1088/1741-2552/ac7c3b.
41. Alim, A.; Imtiaz, M.H. Automatic Identification of Children with ADHD from EEG Brain Waves. *Signals* **2023**, *4*, 193–205. https://doi.org/10.3390/signals4010013.
42. Alsharif, N.; Al-Adhaileh, M.H.; Al-Yaari, M. Diagnosis of attention deficit hyperactivity disorder: A deep learning approach. *AIMS Mathematics* **2024**, *9*, 10580–10608. https://doi.org/10.3934/math.2024517.
43. Salazar-Dubois, D.V.; Álvarez-Meza, A.M.; Castellanos-Dominguez, G. T-GARNet: A Transformer and Multi-Scale Gaussian Kernel Connectivity Network with Alpha-Rényi Regularization for EEG-Based ADHD Detection. *Mathematics* **2025**, *13*, 4026. https://doi.org/10.3390/math13244026.
44. Rainio, O.; Teuho, J.; Klén, R. Evaluation Metrics and Statistical Tests for Machine Learning. *Scientific Reports* **2024**, *14*, 6086. https://doi.org/10.1038/s41598-024-56706-x.
45. Pugnana, A.; Perini, L.; Davis, J.; Ruggieri, S. Deep Neural Network Benchmarks for Selective Classification. *Journal of Data-centric Machine Learning Research* **2024**, *1*, 1–58.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2026 doi:10.20944/preprints202605.1894.v1

46. Tay, Y.; Dehghani, M.; Bahri, D.; Metzler, D. Efficient Transformers: A Survey. *ACM Computing Surveys* **2022**, *55*, 1–28. https://doi.org/10.1145/3530811.
47. Kuruppu, G.; Wagh, N.; Kremen, V.; Varatharajah, Y. EEG Foundation Models: A Critical Review of Current Progress and Future Directions. *Journal of Neural Engineering* **2026**, *23*, 021001. https://doi.org/10.1088/1741-2552/ae4455.
48. Gawlikowski, J.; Tassi, C.R.N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. A Survey of Uncertainty in Deep Neural Networks. *Artificial Intelligence Review* **2023**, *56*, 1513–1589. https://doi.org/10.1007/s10462-023-10562-9.

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.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.