ISNNI: 3007-3391 HEC Category: Y

Journal of Computers and Intelligent Systems
Vol. 3, No. 4, 2025, pages 353 - 365
Journal Homepage:
https://journals.iub.edu.pk/index.php/JCIS/

A Multi-Branch EEGNet–Transformer Hybrid Network for Automated ADHD Screening from Multichannel EEG

Abdul Haseeb Wajid[1], Ayesha Qureshi[2], Ali Samad[3],
[1]Affilition of Author 1, The Islamia University Bahawalpur
[2]Affilition of Author 2, The Islamia University Bahawalpur
[3]Affilition of Author 3, The Islamia University Bahawalpur

ARTICLE INFO

Article History:

Received: December 01, 2025
Revised: December 15, 2025
Accepted: December 20, 2025
Available Online: December 30, 2025

Keywords:

ADHD 1
Electroencephalography 2
EEGNet 3
Transformer 4
Attention 5
Deep learning
Data augmentation
Clinical decision support

Classification Codes:

Funding:

This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

Attention-Deficit/Hyperactivity Disorder (ADHD) diagnosis remains dependent on subjective instruments and interviews, motivating objective screening tools. Electroencephalography (EEG) is non-invasive and cost-effective, but automated ADHD classification is challenged by noisy recordings, inter-subject variability, and complex spatiotemporal structure. We propose MB-EEGNet-T, a compact multi-branch hybrid model that fuses an EEGNet-like convolutional branch for local temporal–spatial feature extraction with a temporal Transformer branch that models longer-range dependencies using multi-head self-attention. Using an open-access 19-channel pediatric EEG dataset of 61 ADHD and 60 control participants [1], we construct overlapping 128-sample segments (≈1 s at 128 Hz) with 50% overlap. Preprocessing applies per-channel standardization and a 0.5–45 Hz Butterworth band-pass filter; training is regularized via physiologically plausible augmentations (temporal shift, amplitude scaling, additive Gaussian noise, and stochastic channel dropout). On a stratified segment-level split, MB-EEGNet-T achieves 92.2% accuracy, 0.922 F1-score, and 0.976 ROC-AUC on a held-out test set, improving over an EEGNet-only baseline. The model uses 61k parameters (<0.25 MB), supporting efficient inference. We discuss evaluation pitfalls (e.g., subject leakage) and outline steps toward subject-independent validation and clinical decision support.

Corresponding Author's Email:

Citation:

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental condition that involves persistent inattention and/or hyperactivity-impulsivity, and impairs development and functioning. The condition usually onsets in early childhood and may continue through adulthood, leading to educational, social and occupational dysfunction. While diagnostic criteria are clear, clinical diagnosis still relies heavily on interviews, rating scales, and observations, which are subjective and may lead to delayed and variable diagnosis, especially across clinics, cultures and age ranges.

EEG reflects population activity in the brain with sub-millisecond resolution, and has been investigated for ADHD biomarkers for decades. ADHD has been linked to changes in oscillatory activity and electrophysiological subtypes, such as spectral power and connectivity in conventional bands (delta to gamma). A recent review in clinical neurophysiology

reviews how EEG features might reflect attentional control and arousal regulation [32]. However, EEG signals are noisy, non-stationary and prone to artifacts, and ADHD effects are weak and can be spread across channels and time. This makes automated classification more difficult, and motivates models that can capture local features (e.g., bursts of rhythmic activity) as well as global features (e.g., longer-range temporal context or coordination across channels).

Traditional machine learning techniques typically rely on engineering features such as band powers, ratios, entropy, fractal dimension, functional/effective connectivity and so on, then training classifiers like support vector machines (SVMs) or random forests. These approaches can be interpretable, but require elaborate feature engineering and can be sensitive to preprocessing and segmentation.

Recently, deep learning has been applied to learn representations from EEG. EEGNet [30] is a popular small convolutional network that performs temporal convolution and spatial channel mixing using depth wise separable convolutions. On the other hand, Transformers [28] apply self-attention to sequences of data and have been modified to decode and classify clinical EEG data (in hybrid CNN–Transformer networks such as EEG Conformer [31] and EEGformer [33]). Hybrid models are appealing due to the strong inductive biases of convolutions for local time-frequency features and the potential of attention to model long-range temporal dependencies.

Here, we present MB-EEGNet-T, a hybrid multi-branch network with an EEGNet-like convolutional branch and a temporal Transformer branch. The branches learn different representations that are subsequently combined for classification. We test the architecture on an open dataset of pediatric epilepsy [1] with a segment-based protocol and discuss limitations and recommendations for future subject-independent validations [25], [26].

A. Contributions

The main contributions of this paper are:

1) A lightweight, multi-branch EEGNet–Transformer architecture (MB-EEGNet-T) with 61k parameters;

2) An end-to-end pipeline for segmentation, filtering, and physiologically plausible data augmentation;

3) Segment-level evaluation on an open pediatric dataset with comprehensive reporting (training curves, confusion matrix, ROC-AUC), plus discussion of validity threats and recommended evaluation protocols.

B. Paper Organization

Section II reviews related work on EEG-based ADHD classification, Transformer-based EEG modeling, and evaluation protocols. Section III describes the dataset, preprocessing, augmentation, and MB-EEGNet-T architecture. Section IV presents experiments and results. Section V discusses implications, limitations, and future work. Section VI concludes.

2. Related Work

We summaries the literature on ADHD classification using EEG and on modeling EEG using Transformers. As performance results depend heavily on data splits and preprocessing, we focus on evaluation procedures.

A. Classical EEG Features for ADHD Screening

Traditionally ADHD EEG analysis has relied heavily on features. For instance, initial studies used nonlinear features and shallow neural networks to classify ADHD and control groups [27]. Connectivity measures have been especially common: for instance, directed phase transfer entropy has been used to quantify brain connectivity in ADHD children [17] and phase-based connectivity biomarkers combined with graph measures have been reported to achieve high accuracy in some cases [14]. Effective connectivity measures (linear and nonlinear) have also been investigated in large-scale research [5], [36]. Such methods can provide interpretability, but care is required when dealing with artifacts, segmentation and cross-subject transfer.

B. Deep Learning Models for EEG-Based ADHD Detection

Recent deep learning methods for ADHD detection involve CNNs applied to raw EEG segments, or some form of representation such as a spectrogram or connectivity map. Dubreuil-Vall et al. showed CNNs using event-related spectral EEG can distinguish adult ADHD from controls [2]. BIOCYBERNETICS AND BIOMEDICAL ENGINEERING by Karabiber Cura et al. suggested building EEG feature maps and training deep models to detect ADHD [8]. Health Information Science and Systems proposed a simple 2-layer CNN that achieved high performance when using all 19 channels [13]. Other reports have provided explainability and multi-task descriptions (ADHD/CD-NET) [11], attention-

based residual connectivity and autoencoder-based deep feature learning [18]. But many of the reported accuracies are greater than 95%, which may relate to protocol-specific data sets and perhaps, in some cases, segment-based evaluation that could lead to subject leakage.

There have also been studies on connectivity-based deep learning. Neuroinformatics suggested to combine Pearson correlation and phase-locking value connectivity maps and train an attention-based CNN, and report high accuracies on the IEEE Data Port dataset [9]. A 2025 Frontiers in Neuroscience paper combined time-frequency features and a fused functional connectivity matrix with a graph convolutional network (GCN), achieving average accuracies of over 96% on two datasets [19]. Such methods indicate spatial channel relations and connectivity may contain cues that are highly discriminative, but they also point to the importance of rigorous validation to achieve generalization across subjects.

C. Transformers for EEG and Hybrid Architectures

Transformers have been applied to EEG since self-attention can capture long-term dependencies in time series. EEG Conformer uses convolutional stems and Transformer blocks with an emphasis on interpretability through attention visualization [31]. EEGformer proposes region and synchronous attention based on EEG characteristics [33]. A recent review discusses Transformer designs for motor imagery, seizure detection, and emotion recognition and provides a summary of design strategies that combine convolution and attention [24]. These works motivate our hybrid approach: convolution offers a strong inductive bias to capture local oscillatory patterns and attention to capture long-range temporal information.

D. Preprocessing, Augmentation, and Evaluation Protocols

Preprocessing and segmentation can significantly affect outcomes. García-Ponsoda et al. measured the impact of preprocessing, temporal segmentation, and reproducibility in ADHD classification, with sensitivity analysis [7]. EEG augmentation is used to prevent overfitting in deep networks; a methodological article surveyed augmentation methods and their effects on deep learning for EEG classification [3], while a review article provides a taxonomy of augmentation approaches and applications [4]. Meanwhile, the EEG community is increasingly alert to evaluation issues: random segment-wise splitting can leak subject identity and normalization using the entire dataset can leak test-set statistics. Recent reviews of deep learning studies on translational EEG document the common sources of leakage and suggest using subject-wise partitions and training-only preprocessing [25]. Meanwhile, a 2025 cross-subject analysis demonstrates that partition choice has a large impact on supervised EEG deep learning performance [26].

E. Summary of Prior ADHD EEG Classifiers

Table I presents some example ADHD EEG classification studies and their results. These numbers cannot be directly compared due to dataset, task, preprocessing, and evaluation differences; we present them to show trends in methodology, and to emphasize the design of evaluation.

TABLE I EEG-based ADHD classification studies and reported results.

Study	Input / Features	Model	Dataset	Validation	Reported performance
Ahmadi Moghadam et al. (2024) [9]	Connectivity maps (PCC+PLV)	Att-CNN	IEEE DataPort [1]	Not stated in abstract	Acc 98.88%, Prec 98.41%, Rec 98.19% (theta band)
Li et al. (2025) [19]	Time/Freq + fused connectivity	GCN + CNN/LSTM	Two EEG datasets	Average over splits	Acc 97.29% / 96.67%
Rodriguez et al. (2025) [22]	Entropy-based channel selection	EEG classifier	IEEE DataPort [1]	5-fold CV (as stated)	Acc 99.29%
Amar et al. (2025) [21]	Raw EEG segments	Multi-headed DNN	IEEE DataPort [1]	Patient-independent split	Acc 89.87%

Karabiber Cura et al. (2024) [8]	EEG feature maps	Deep CNN	Pediatric EEG	Not stated in abstract	Reported high accuracy (see paper)
Hassan & Singhal (2024) [13]	Raw 19-ch EEG	2-layer CNN	Pediatric EEG	Not stated	Acc 100% (as reported)
Bansal et al. (2025) [18]	Autoencoder + ResNet	ResNet + attention	IEEE DataPort [1]	Not stated in abstract	Reported improved diagnosis (see paper)

3. Materials and Methods

A. Dataset and Problem Formulation

We use the open-access IEEE 19-channel EEG dataset of children with ADHD and healthy controls [1]. EEG was recorded at 128 Hz during a visual attention task. Let $X \in \mathbb{R}^{L\times C}$denote a multichannel EEG segment of length Lsamples and Cchannels, and let $y \in \{0,1\}$denote the class label (0 = Control, 1 = ADHD). The goal is to learn a function $f_\theta(X)$that outputs class probabilities $p_\theta(y \mid X)$by minimizing the empirical risk over training segments.

Given Ntraining segments $\{(X_i, y_i)\}_{i=1}^{N}$, we minimize the categorical cross-entropy loss:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k\in\{0,1\}} \mathbb{I}[\,y_i = k]\log p_\theta(y = k \mid 7X_i). \qquad (1)$$

B. Preprocessing

EEG preprocessing aims to reduce artifacts and normalize scale while preserving diagnostic information. We apply four steps: (i) missing value imputation, (ii) per-channel standardization, (iii) segmentation into overlapping windows, and (iv) band-pass filtering.

1. Imputation: NaNs are replaced with 0.0. In practice, alternative artifact handling (e.g., artifact subspace reconstruction) may further improve robustness but is not used in this baseline pipeline.

2. Standardization: each channel $c$is standardized to zero mean and unit variance using z-scores:

$$\hat{x}(t,c) = \frac{x(t,c) - \mu_c}{\sigma_c + \varepsilon}. \qquad (2)$$

3. Segmentation: a sliding window of length $L = 128$samples and overlap $\rho = 0.5$creates segments with step $S = L(1-\rho) = 64$. Segment $i$spans $[iS, iS + L - 1]$. Segments are treated as independent examples for training and evaluation. We recommend segmenting within participant boundaries to avoid mixed-label windows.

4. Band-pass filtering: we apply a 4th-order Butterworth band-pass filter from 0.5 to 45 Hz, implemented via forward–backward filtering to produce zero-phase distortion. The Butterworth magnitude response is maximally flat in the passband and is commonly used for EEG preprocessing. The digital filter is designed by bilinear transformation from an analog prototype.

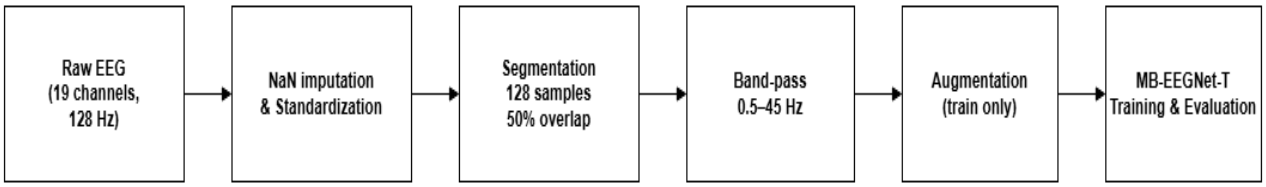

Figure. 1. Overview of preprocessing and modeling.

C. Data Augmentation

To increase robustness to nuisance variability, we augment only the training set using physiologically plausible transformations surveyed in EEG augmentation literature [3], [4]. The four transformations are: time shifting, amplitude scaling, additive Gaussian noise, and stochastic channel dropout. Formally, given a segment X, we generate an augmented segment $\tilde{X}$via a composition of random operators:

$$\tilde{X} = \mathcal{A}(X) = \mathcal{D}_{ch}\left(\mathcal{N}\left(\mathcal{S}_{amp}\left(\mathcal{S}_{time}(X)\right)\right)\right). \qquad (3)$$

where $\mathcal{S}_{time}$applies a circular shift along time, $\mathcal{S}_{amp}$scales amplitude, $\mathcal{N}$adds Gaussian noise, and $\mathcal{D}_{ch}$randomly zeros a subset of channels. In our experiments, the training set is doubled (augmentation factor $2\times$).

D. MB-EEGNet-T Architecture

MB-EEGNet-T processes an input segment $X \in \mathbb{R}^{L\times C\times 1}$in two parallel branches and fuses their feature vectors for classification. The architecture is designed to remain lightweight while capturing both local and global temporal dependencies.

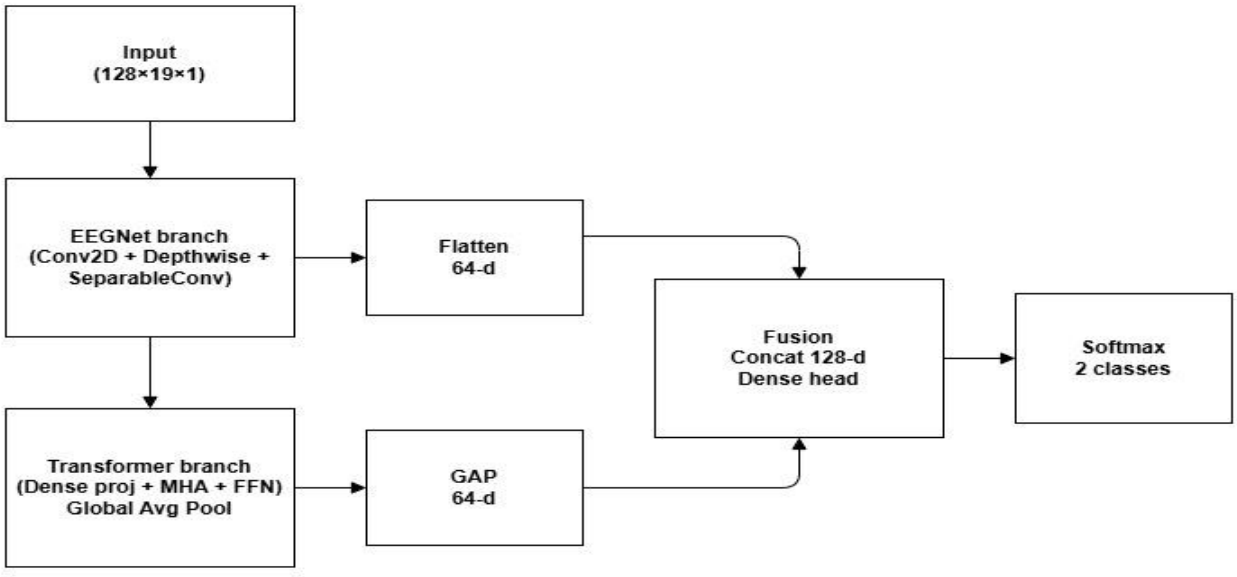

Figure. 2. Block diagram of MB-EEGNet-T.

1) EEGNet-inspired convolutional branch

The first branch follows EEGNet [30]. A temporal Conv2D with kernel $(K\times 1)$learns F_1temporal filters. A depthwise convolution with kernel $(1\times C)$learns spatial filters across channels with depth multiplier D. Batch normalization and ELU activation stabilize training, while average pooling and spatial dropout reduce

dimensionality and overfitting. A SeparableConv2D block provides further temporal filtering with parameter efficiency. The output is flattened to a feature vector $f_{\mathrm{CNN}} \in \mathbb{R}^{64}$.

2) Temporal Transformer branch

The second branch reshapes the input to a sequence $Z \in \mathbb{R}^{L\times C}$and projects each time step to a d_{model}-dimensional embedding. We apply one Transformer encoder block consisting of multi-head self-attention (MHA), a feed-forward network, residual connections, and layer normalization. For head h, attention is computed as:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V. \qquad (4)$$

The concatenated outputs of multiple heads are linearly projected, followed by a position-wise feed-forward network. Global average pooling over the Ltime steps yields $f_{\mathrm{TX}} \in \mathbb{R}^{64}$. Transformer-based EEG modeling has demonstrated strong performance across tasks and supports interpretability through attention visualization [24], [31], [33].

3) Feature fusion and classifier

The two feature vectors are concatenated: $f = [f_{\mathrm{CNN}}; f_{\mathrm{TX}}] \in \mathbb{R}^{128}$. A small MLP classifier with ELU activations and dropout produces logits and softmax probabilities over the two classes.

TABLE II Key hyperparameters for preprocessing and MB-EEGNet-T.

Parameter	Setting
Sampling rate	128 Hz
Band-pass filter	0.5–45 Hz, 4th-order Butterworth
Segment length / overlap	128 samples / 50%
EEGNet temporal kernel K	64
F1 / D / F2	8 / 2 / 16
Transformer d_model / heads	64 / 4
Dropout	0.5 (CNN and head), 0.1 (attention)
Optimizer	Adam with exponential decay
Batch size	32
Epochs / early stopping	80 max / patience 20
Trainable parameters	61,186

E. Training and Evaluation

Segments are split into train/validation/test sets using stratified sampling (70/15/15). Augmentation is applied only to training segments. We train with categorical cross-entropy and report accuracy, precision, recall, F1-score, and ROC-AUC. Precision and recall are defined as: $\mathrm{Recall} = \frac{TP}{TP+FN}$, $F_1 = \frac{2\cdot\mathrm{Precision}\cdot\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$.

ROC-AUC is computed from predicted probabilities by integrating the ROC curve. To support reproducibility and avoid optimistic estimates, we discuss subject-wise evaluation and training-only preprocessing in Section V [25], [26].

4. Experimental Results

A. Experimental Setup

Experiments were conducted in TensorFlow/Keras (v2.18). GPU acceleration was available (Tesla P100, 16 GB). We fixed random seeds ($\mathrm{SEED} = 42$) for NumPy and TensorFlow. The training process used model checkpointing (best validation loss) and early stopping with patience 20 epochs.

During import and initialization, the runtime emitted CUDA plugin registration warnings (e.g., cuFFT/cuDNN/cuBLAS already registered) and repeated protobuf MessageFactory AttributeError messages. These messages are commonly encountered in some notebook environments due to library version mismatches; training and evaluation proceeded normally once TensorFlow initialized.

B. Data Characteristics

The original data set includes 1,207,069 samples labeled as ADHD and 959,314 as control (55.7% ADHD). The test data, reserved for segment-level evaluation, consisted of 2,248 ADHD segments and 2,830 control segments (stratified split after windowing).

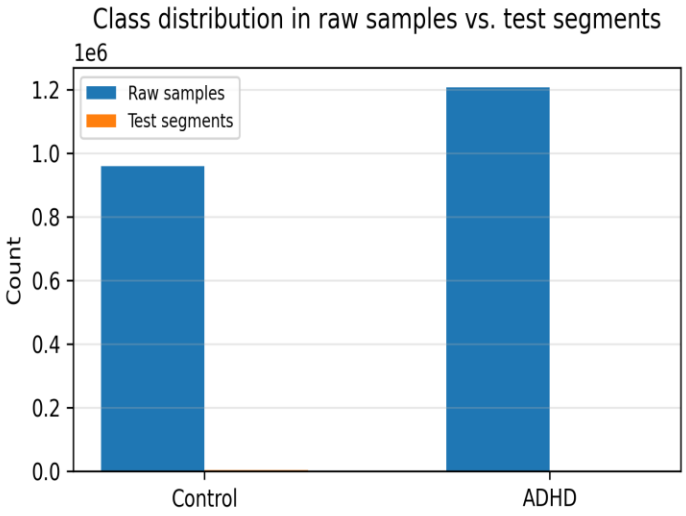

Figure. 3. Class distribution in raw samples versus the test segment partition.

C. Training Dynamics

The loss on the validation set drops quickly in the early epochs and achieves its lowest value at epoch 6. There is little improvement and a slight decline in subsequent epochs, so training stops at epoch 26. This suggests fast convergence and that strong regularization (dropout, augmentation, learning-rate decay) is necessary but more capacity may be needed with more varied data or per-subject validation to determine subject-specific performance.

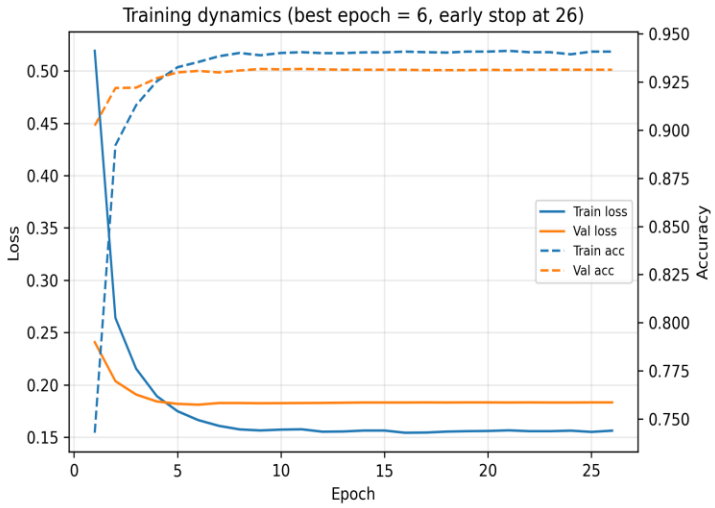

Figure. 4. Training and validation loss/accuracy across epochs 1–26.

D. Test-Set Results

On the held-out test set (5,078 segments), MB-EEGNet-T achieved loss 0.205, accuracy 0.922, precision 0.922, recall 0.922, F1-score 0.922, and ROC-AUC 0.976. Table III reports overall metrics, while Table IV lists per-class precision, recall, and F1-score.

TABLE III Overall segment-level test performance for MB-EEGNet-T.

Metric	Value
Loss	0.2051
Accuracy	0.9222
Precision (micro)	0.9222
Recall (micro)	0.9222
F1-score (micro)	0.9222
ROC-AUC	0.9759

TABLE IV Per-class classification report on the test set (segments).

Class	Precision	Recall	F1-score	Support
Control	0.9250	0.9364	0.9306	2830
ADHD	0.9187	0.9044	0.9115	2248
Macro avg	0.9218	0.9204	0.9210	5078
Weighted avg	0.9222	0.9222	0.9221	5078

Fig. 5 shows the confusion matrix. The model correctly classifies 2650/2830 control segments and 2034/2248 ADHD segments. The lower recall for ADHD suggests that some ADHD segments resemble control patterns under this protocol, motivating further exploration of subject-level variability and task effects.

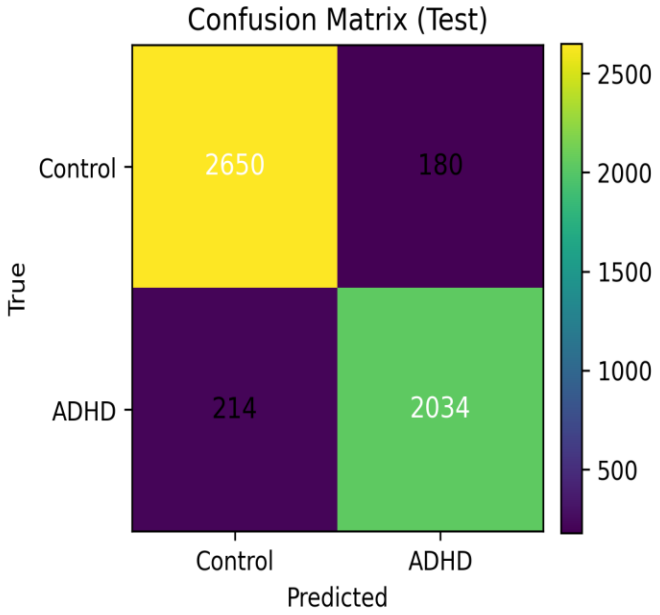

Figure. 5. Confusion matrix on the test partition.

The ROC curve is generated from predicted probabilities. In this draft, Fig. 6 is provided as a placeholder visualization; the exact curve should be regenerated directly from y_pred_proba saved by the evaluation script. The reported AUC of 0.9759 indicates strong ranking performance across thresholds.

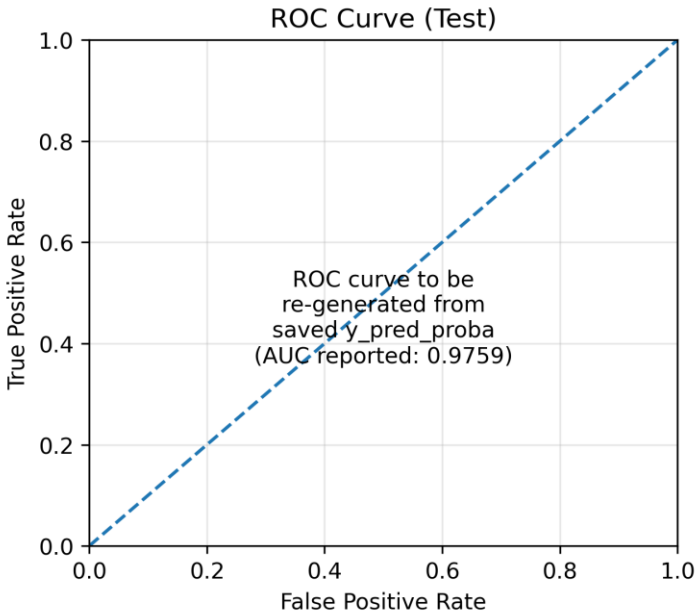

Figure. 6. ROC curve placeholder (AUC reported: 0.9759).

E. Baseline Comparison and Parameter Efficiency

We benchmark MB-EEGNet-T against a baseline (EEGNet-only) model that eliminates the Transformer branch, leaving the convolutional branch and head. With exactly the same preprocessing and segment-level split, the baseline model has 88.9% accuracy and 0.967 ROC-AUC, compared to 92.2% accuracy and 0.976 AUC with the MB-EEGNet-T model. This result supports our hypothesis that the attention-based temporal modelling captures information in addition to local features.

The complete model of MB-EEGNet-T has 61,266 parameters (61,186 of them trainable), or roughly 239 KB of 32-bit floats. This is a reasonable number of parameters for deployment compared to many deep networks.

TABLE V Comparison to an EEGNet-only baseline on the same protocol.

Model	Accuracy	F1-score	ROC-AUC
EEGNet-only (CNN branch)	0.889	0.889	0.967
MB-EEGNet-T (CNN + Transformer)	0.922	0.922	0.976

F. Additional Analysis: Decision Threshold and Screening Trade-offs

Screening tests typically place emphasis on sensitivity (recall for a condition) to avoid missing cases, but not specificity. While we estimate using argmax classification, the softmax probabilities can be thresholded on ADHD probability to control the proportion of false positives and false negatives. Threshholds should be tuned on a validation set reflecting the target prevalence and cost structure, and probability calibration evaluated, in future clinical studies.

5. Discussion

MB-EEGNet-T leverages inductive biases from convolutions with attention-based temporal processing and delivers good performance on segment-level pediatric data. The improvement over baseline EEGNet suggests the Transformer branch provides additional capacity to capture more long-term dependencies in the one-second segments. It is a compact model that enables fast inference and could be deployed at the edge.

But care must be taken with absolute results. High accuracies are common in ADHD EEG studies, and as shown in Table I, the results between studies can be quite variable. This can be due to variations in paradigm (resting-state vs thinking paradigms), preprocessing and feature extraction, but more importantly, the protocol used to validate the results.

A. Interpretation and Potential Mechanisms

The convolutional branch should learn temporal filters similar to band-pass filters and blend channel information through depthwise spatial convolutions, which can pick up local spatial features. The Transformer branch can learn attention patterns for non-contiguous time events, which may capture stimulus locked features or longer-term temporal structures. Future work on interpretability may explore attention maps and gradient-based saliency maps to understand the attention allocation across channels and temporal segments, and to compare these patterns with known electrophysiological patterns in ADHD [32].

B. Limitations and Threats to Validity

1) Segment-level splitting: Randomly dividing segments can cause subject identity leakage if segments from multiple subjects are included in the training and testing splits. This can bias performance relative to subject-independent testing. A recent study shows leakage in EEG deep learning and suggests using subject-wise partitions and nested designs [25]. Another study in 2025 also demonstrates that data partitioning has a significant impact on supervised cross-subject classification accuracy [26].

2) Preprocessing only on training data: Centering and scaling parameters should be estimated on the training data and then applied to the validation and test data. Statistics computed on the entire data set may be slightly leaky. Re-implementations should compute μ_c and σ_c on training subjects.

3) Segmentation at subject boundaries: Sliding windows should not cross subject boundaries. If the subject ID is known, this should be done for each subject separately.

4) Generalizability: The dataset is collected from a particular visual attention task. This should be validated (and potentially adapted) in other tasks (resting-state, auditory) and clinical conditions.

C. Ethical and Clinical Considerations

Our approach is on automated screening, not diagnosis. Prospective validation and clinical supervision are required for a clinical decision support system. Potential biases and fairness across groups should be assessed and model predictions should include confidence intervals to guide clinicians. Given the privacy concerns around pediatric EEG data, privacy preserving deployment and data security are crucial.

D. Future Work

In future work, we will rigorously assess the performance of MB-EEGNet-T with strict subject-independent testing procedures (such as leave-one-subject-out) and other datasets. We will also investigate multi-scale temporal modeling (longer pooling and hierarchical attention), connectivity-based attention and self-supervised pretraining of EEG features.

6. Conclusion

We proposed MB-EEGNet-T, a lightweight multi-branch EEGNet-Transformer fusion network for automatic ADHD diagnosis based on multichannel EEG. On an open pediatric dataset, our technique attained 92.2% segment-level accuracy and 0.976 ROC-AUC with 61k parameters. Our findings suggest that combining convolutional feature learning and attention-based temporal modeling can be effective. We highlighted the need for future work to focus on subject-independent validation and only train-time preprocessing to evaluate clinical generalizability.

7. References

[1] A. Motie Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi, “The Open-Access IEEE 19 Channel EEG Dataset of 61 ADHD and 60 Healthy Children,” IEEE DataPort. doi:10.21227/rzfh-zn36.

[2] L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon, “Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG,” Frontiers in Neuroscience, vol. 14, art. 251, 2020, doi:10.3389/fnins.2020.00251.

[3] E. Lashgari, D. Liang, and U. Maoz, “Data augmentation for deep-learning-based electroencephalography,” Journal of Neuroscience Methods, vol. 346, art. 108885, 2020, doi:10.1016/j.jneumeth.2020.108885.

[4] C. He, J. Liu, Y. Zhu, and W. Du, “Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review,” Frontiers in Human Neuroscience, vol. 15, 2021, doi:10.3389/fnhum.2021.765525.

[5] N. Talebi and A. M. Nasrabadi, “Investigating the discrimination of linear and nonlinear effective connectivity patterns of EEG signals in children with attention-deficit/hyperactivity disorder and typically developing children,” Computers in Biology and Medicine, vol. 148, art. 105791, 2022, doi:10.1016/j.compbiomed.2022.105791.

[6] D. Zhou, Z. Liao, and R. Chen, “Deep learning enabled diagnosis of children’s ADHD based on the big data of video screen long-range EEG,” Journal of Healthcare Engineering, 2022, Art. no. 5222136, doi:10.1155/2022/5222136.

[7] S. García-Ponsoda, A. Maté, and J. Trujillo, “Refining ADHD diagnosis with EEG: The impact of preprocessing and temporal segmentation on classification accuracy,” Computers in Biology and Medicine, vol. 183, art. 109305, 2024, doi:10.1016/j.compbiomed.2024.109305.

[8] O. Karabiber Cura, A. Akan, and S. Kocaaslan Atli, “Detection of Attention Deficit Hyperactivity Disorder based on EEG feature maps and deep learning,” Biocybernetics and Biomedical Engineering, vol. 44, no. 3, pp. 450–460, 2024, doi:10.1016/j.bbe.2024.07.003.

[9] E. Ahmadi Moghadam, F. Abedinzadeh Torghabeh, S. A. Hosseini, and M. H. Moattar, “Improved ADHD Diagnosis Using EEG Connectivity and Deep Learning through Combining Pearson Correlation Coefficient and Phase-Locking Value,” Neuroinformatics, vol. 22, pp. 521–537, 2024, doi:10.1007/s12021-024-09685-3.

[10] H. Jahani and A. A. Safaei, “Efficient Deep Learning Approach for Diagnosis of Attention-Deficit/Hyperactivity Disorder in Children Based on EEG Signals,” Cognitive Computation, 2024, doi:10.1007/s12559-024-10302-3.

[11] H. W. Loh, A. J. Ooi, A. B. B. A. Aidil, N. Dey, and U. R. Acharya, “ADHD/CD-NET: automated EEG-based characterization of ADHD and CD using explainable deep neural network technique,” Cognitive Neurodynamics, vol. 18, pp. 1609–1625, 2024, doi:10.1007/s11571-023-10028-2.

[12] O. Attallah, “ADHD-AID: Aiding Tool for Detecting Children’s Attention Deficit Hyperactivity Disorder via EEG-Based Multi-Resolution Analysis and Feature Selection,” Biomimetics, vol. 9, no. 3, art. 188, 2024, doi:10.3390/biomimetics9030188.

[13] U. Hassan and A. Singhal, “Convolutional neural network framework for EEG-based ADHD diagnosis in children,” Health Information Science and Systems, vol. 12, art. 44, 2024, doi:10.1007/s13755-024-00305-7.

[14] F. Abedinzadeh Torghabeh, S. A. Hosseini, and Y. Modaresnia, “Potential biomarker for early detection of ADHD using phase-based brain connectivity and graph theory,” Physical and Engineering Sciences in Medicine, vol. 46, no. 4, pp. 1447–1465, 2023, doi:10.1007/s13246-023-01310-y.

[15] A. Alim and M. H. Imtiaz, “Automatic Identification of Children with ADHD from EEG Brain Waves: A Deep Learning Approach,” Signals, vol. 4, no. 1, pp. 193–205, 2023, doi:10.3390/signals4010010.

[16] F. Esas and F. Latifoğlu, “Detection of ADHD from EEG signals using new hybrid decomposition and deep learning techniques,” Journal of Neural Engineering, vol. 20, 2023, doi:10.1088/1741-2552/acc902.

[17] R. E. Tavasoli, A. Sadeghian, A. Faghihzadeh, S. Bakhshi, and A. Ahmadlou, “Analysis of EEG brain connectivity of children with ADHD using directed phase transfer entropy,” Biomedical Engineering / Biomedizinische Technik, 2023, doi:10.1515/bmt-2022-0100.

[18] J. Bansal, G. Gangwar, M. Aljaidi, A. Alkoradees, and G. Singh, “EEG-Based ADHD classification using autoencoder feature extraction and ResNet with double augmented attention mechanism,” Brain Sciences, 2025, Art. no. 95, doi:10.3390/brainsci15010095.

[19] L. Li et al., “ADHD detection from EEG signals using GCN based on multi-domain features,” Frontiers in Neuroscience, 2025, doi:10.3389/fnins.2025.1561994.

[20] Y. Mao, X. Qi, L. He, S. Wang, Z. Wang, and F. Wang, “Advanced machine learning techniques reveal multidimensional EEG abnormalities in children with ADHD: A framework for automatic diagnosis,” Frontiers in Psychiatry, 2025, doi:10.3389/fpsyt.2025.1475936.

[21] Z. Amar, A. Adigun, V. Rosas, J. L. Little, and V. J. Lawhern, “Comparative study of multi-headed and baseline deep learning models for ADHD classification from EEG signals,” Physical and Engineering Sciences in Medicine, 2025, doi:10.1007/s13246-025-01609-y.

[22] M. E. Rodriguez, M. Zhang, and D. Matthews, “Entropy difference-based EEG channel selection technique for automated detection of ADHD,” PLOS ONE, 2025, Art. no. e0319487, doi:10.1371/journal.pone.0319487.

[23] S. R. Sarker et al., “Hyperdimensional computing boosting EEG-based diagnosis of ADHD: An accurate classifier and interpretable framework,” Scientific Reports, 2025, doi:10.1038/s41598-025-24919-3.

[24] E. Vafaei and M. Hosseini, “Transformers in EEG analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification,” Sensors, vol. 25, no. 5, Art. no. 1293, 2025, doi:10.3390/s25051293.

[25] G. Brookshire et al., “Data leakage in deep learning studies of translational EEG,” Frontiers in Neuroscience, 2024, doi:10.3389/fnins.2024.1373515.

[26] F. Del Pup et al., “The role of data partitioning on the performance of EEG-based deep learning models in supervised cross-subject analysis: A preliminary study,” Computers in Biology and Medicine, 2025, Art. no. 110608, doi:10.1016/j.compbiomed.2025.110608.

[27] D. C. Lohani, V. Chawla, and B. Rana, “A systematic literature review of machine learning techniques for the detection of attention-deficit/hyperactivity disorder using MRI and/or EEG data,” Neuroscience, 2025, doi:10.1016/j.neuroscience.2025.02.019.

[28] A. Vaswani et al., “Attention Is All You Need,” in Proc. NeurIPS, 2017.

[29] R. T. Schirrmeister et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” Human Brain Mapping, vol. 38, pp. 5391–5420, 2017, doi:10.1002/hbm.23730.

[30] V. J. Lawhern et al., “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces,” Journal of Neural Engineering, vol. 15, 056013, 2018, doi:10.1088/1741-2552/aace8c.

[31] Y. Song, Q. Zheng, B. Liu, and X. Gao, “EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31, pp. 710–719, 2023, doi:10.1109/TNSRE.2022.3230250.

[32] R. J. Barry, A. R. Clarke, and S. J. Johnstone, “A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography,” Clinical Neurophysiology, vol. 114, no. 2, pp. 171–183, 2003, doi:10.1016/S1388-2457(02)00362-0.

[33] Z. Wan et al., “EEGformer: A transformer–based brain activity classification method using EEG signal,” Frontiers in Neuroscience, vol. 17, 2023, doi:10.3389/fnins.2023.1148855.