Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 12, No. 3, 2025

ADHD-Conformer: EEG-Based Classification and Detection of ADHD

Junming Bai, Bingxu Hou, Jiansheng Wu

Science and technology Liaoning University, Anshan, Liaoning, China

Abstract: Attention Deficit Hyperactivity Disorder (ADHD) is one of the most prevalent neurodevelopmental disorders among children, often leading to cognitive, emotional, and social impairments. Traditional diagnostic methods rely heavily on behavioral observation and subjective questionnaires, lacking objective physiological indicators. This study proposes a novel brain-computer interface (BCI) assisted diagnostic method based on EEG signal classification using a deep learning hybrid model named ADHD-Conformer. The model combines convolutional neural networks (CNNs) for local feature extraction and transformer-based architectures for capturing global temporal dependencies. Experiments on open-source datasets demonstrate that our approach achieves superior classification performance, with accuracy reaching 99.01%, thus proving the feasibility of deep learning-assisted ADHD diagnosis.

Keywords: ADHD; Brain-Computer Interface; EEG; Deep Learning; Conformer; CNN; Transformer.

1. Introduction

ADHD affects approximately 5–7% of children globally and remains one of the leading causes of academic failure, behavioral issues, and social dysfunction. Current clinical diagnosis is based on interviews and questionnaires, which are highly subjective and prone to misdiagnosis. Given the overlap of ADHD symptoms with other neurological disorders, an urgent need arises for objective, biomarker-based diagnostic systems.

EEG (Electroencephalography) signals, as direct reflections of neuronal activity, provide a promising alternative. The advent of Brain-Computer Interface (BCI) technology enables non-invasive brain activity monitoring, paving the way for automated, physiological signal-based ADHD diagnosis. This study presents an innovative deep learning approach that enhances both the accuracy and generalization of ADHD classification from EEG data.

2. Related Work

Previous studies have employed various deep learning models, such as ADHD-Net, which leverages depthwise separable convolutions and LSTM networks. However, LSTM struggles with long-term dependency modeling and introduces heavy computational overhead. In contrast, the Conformer architecture, which fuses CNN and Transformer modules, offers improved feature representation and computational efficiency.

3. Materials and Methods

3.1. Datasets

Two publicly available EEG datasets were used:

IEEE DataPort ADHD/Control Dataset: 144 participants with 56-channel EEG signals (including ADHD, ADD, and healthy control groups).

Deep Learning EEG Dataset: Covers multiple cognitive tasks with 64-channel EEG recordings.

Preprocessing included bandpass filtering (4–30Hz), Z-score normalization, and time-window alignment (padding/cropping to 385 time points). A 5-fold stratified cross-validation scheme was adopted for robust evaluation.

3.2. Evaluation Metrics

We evaluated the model performance using accuracy (Acc), recall, loss, and F1-score.

4. Model Architecture

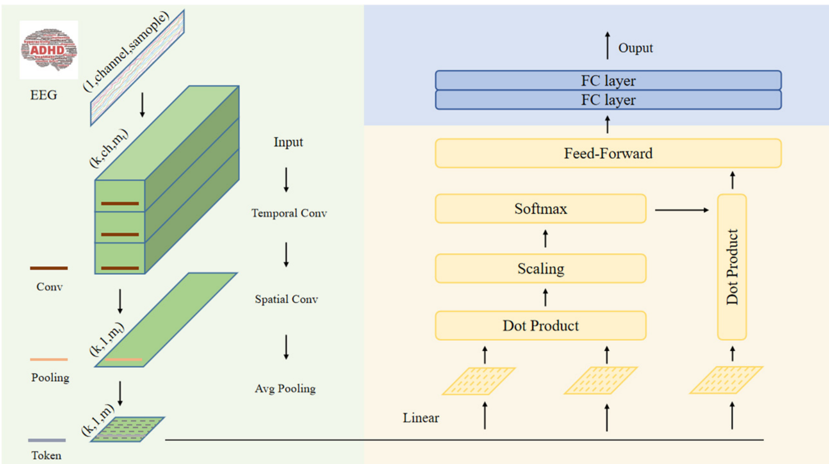

Figure 1. ADHD-Conformer Research Workflow Diagram.

The proposed ADHD-Conformer is a hybrid deep neural network designed specifically for modeling the spatiotemporal characteristics of EEG signals in the context of ADHD diagnosis. The architecture integrates both local and global feature extraction mechanisms, consisting of four main components: Input Preprocessing, Local Feature Extraction (CNN), Global Temporal Modeling (Transformer), and a Classification Module. The overall structure is illustrated in the model diagram (figure to be inserted)[3].

4.1. Input Representation and Preprocessing

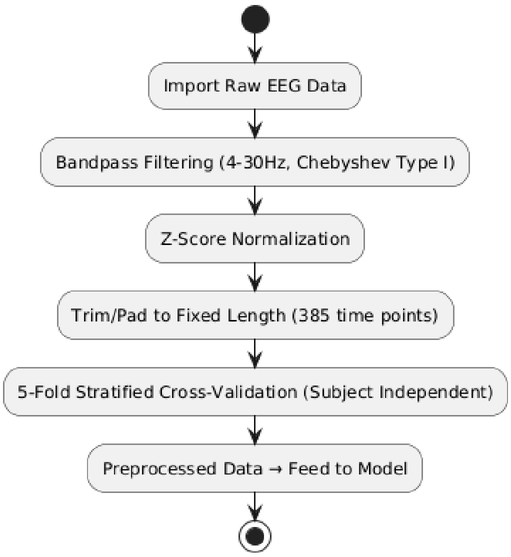

Figure 2. EEG Data Preprocessing Workflow.

The input to the model is a preprocessed EEG tensor with a shape of [B,385,56], where B is the batch size, 385 is the number of time points, and 56 represents the EEG channels. To prepare for convolutional and attention-based operations, the signal is first segmented into patches. Each patch consists of 25 time points, resulting in 39 patches per trial, allowing the model to focus on localized temporal dynamics[2].

4.2. Local Feature Extraction Module (CNN)

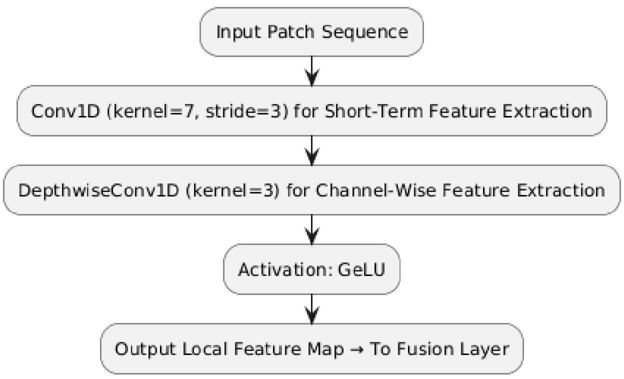

Figure 3. CNN Module Architecture Diagram.

This module employs 1D Convolutional Neural Networks to capture localized frequency-domain features of EEG signals. It consists of two primary layers:

Conv1D (kernel=7, stride=3): Extracts short-term temporal features, such as θ (4–8 Hz) and α (8–13 Hz) wave patterns commonly observed in ADHD.

Depthwise Conv1D (kernel=3): Enhances channel-wise feature specialization while reducing computational cost.

To improve nonlinearity and gradient flow, the Gaussian Error Linear Unit (GeLU) activation function is applied after convolution. Compared to ReLU, GeLU offers smoother transitions and mitigates the "dying neuron" problem, enhancing the model's sensitivity to subtle EEG variations.

4.3. Global Temporal Modeling Module (Transformer)

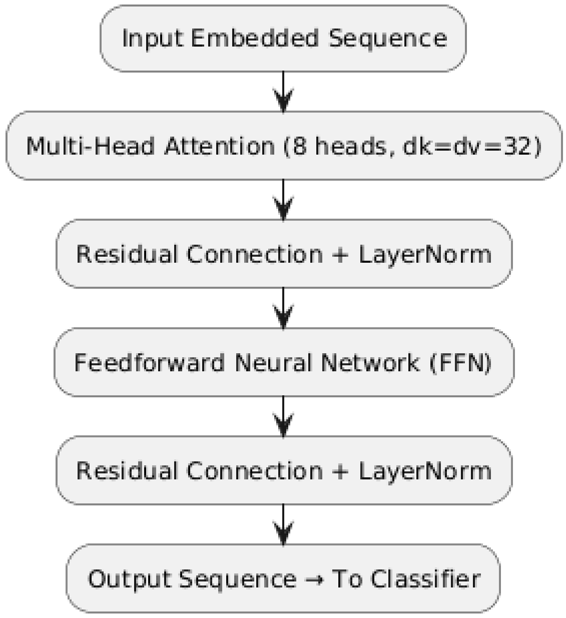

Figure 4. Transformer Module Architecture Diagram

To capture long-range temporal dependencies in EEG signals, the model incorporates an improved Transformer architecture with the following components:

Multi-head Self-Attention (8 heads, $d_k = d_v = 32$); Learns interactions between EEG channels and time steps in parallel, enabling rich contextual modeling.

Relative Positional Encoding: Enhances the model's ability to learn sequence order relationships, which is crucial for handling EEG dynamics over time.

Feedforward Layers and Residual Connections: Integrated with Layer Normalization to stabilize training and facilitate deeper network construction[4].

This module is particularly effective at modeling EEG signal patterns such as cross-frequency interactions (e.g., θ–β coupling), which are indicative of ADHD-related neural dynamics.

4.4. Feature Fusion and Classification Module

The outputs of the CNN and Transformer modules are projected to the same dimension and concatenated. The fused feature vector is then passed through a classification head, which consists of:

Feature Concatenation,Fully Connected Layer.

Softmax Activation: Outputs probabilities for the three target classes: ADHD-Inattentive (ADHD-I), ADHD-Hyperactive/Impulsive (ADHD-H), and Control.

This fusion mechanism enables the model to integrate local frequency sensitivity with global temporal coherence, significantly improving the overall classification performance.

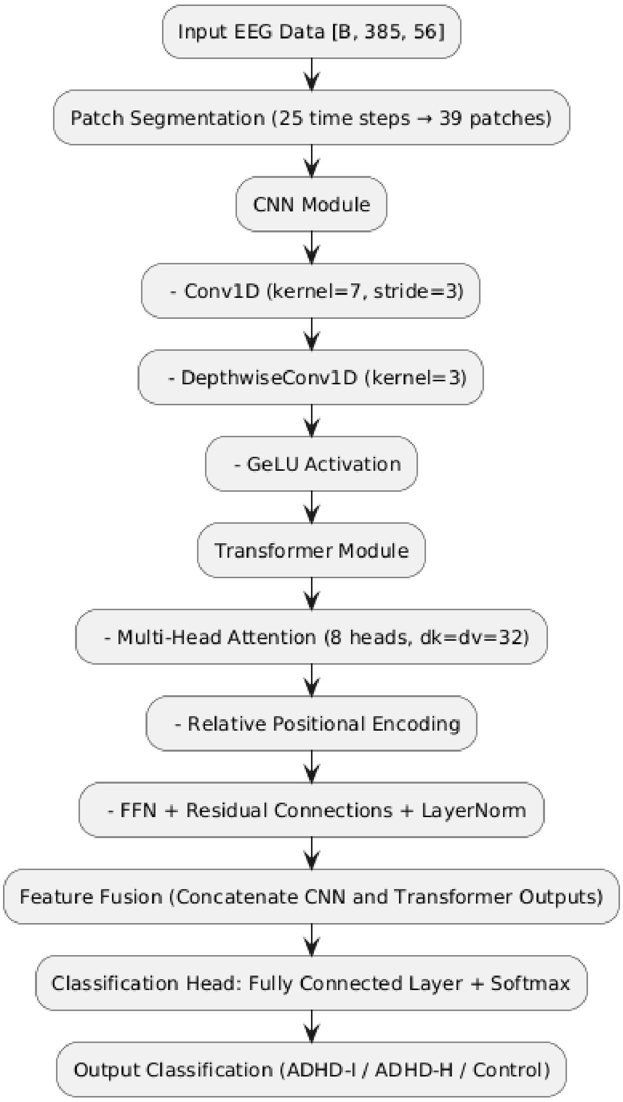

Figure 5. ADHD-Conformer Model Architecture Diagram.

4.5. Summary of Model Advantages

Local–Global Joint Modeling: CNNs capture short-term variations, while Transformers model long-term dependencies, providing a comprehensive view of EEG features.

Lightweight and Scalable: The use of patching and depthwise convolutions reduces parameter count, making the model suitable for deployment on edge devices and real-time applications.

Generalization Across Datasets: The architecture adapts well to various EEG configurations and tasks, demonstrating strong cross-dataset performance and transferability.

5. Innovations

Multimodal Fusion Architecture: Simultaneously captures local frequency features and global temporal dynamics.

GeLU Activation: Replaces ReLU to mitigate dead neuron problems and improve training stability.

Cross-Dataset Generalization: Training on diverse datasets improves the model's robustness and diagnostic reliability.

6. Experimental Results

The model achieves stable performance on the Deep Learning EEG dataset with 99.01% classification accuracy. Metrics such as recall and F1-score exhibit consistent improvement across validation and test sets. Loss curves show convergence, confirming model effectiveness.

All experiments were conducted using PyTorch 1.11.3 on an NVIDIA RTX 3090 GPU, enabling efficient training and deployment.

7. Conclusion

This study presents a novel ADHD diagnosis method based on EEG signals and a hybrid Conformer architecture. The results show significant improvements in diagnostic accuracy, model robustness, and training convergence. The proposed model offers a powerful, objective tool to assist clinicians in diagnosing ADHD. Future work will explore model compression and real-time deployment on portable BCI devices.

Acknowledgments

This work was supported by the innovation and entrepreneurship training program for college students of University of Science and Technology Liaoning 2024.

References

[1] Arns M, Conners C K, Kraemer H C. A decade of EEG Theta/Beta Ratio research in ADHD: a meta-analysis. Journal of Attention Disorders, 2013, 17(5): 374–383.

[2] Li Y, Li X, Zhang Q, et al. EEG-based diagnosis of ADHD using hybrid CNN-LSTM model. IEEE Access, 2020, 8: 108766–108774.

[3] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 5998–6008.