Ensemble Deep Learning Framework Combining CNN, Transformer, and XGBoost for EEG-Based ADHD Detection

Catherine Joy[1], Rahul Krishnan[2], Albert Rajan[3] [1]Postdoctoral Researcher, Lincoln University College, Malaysia;
[2]Associate Professor, Sree Buddha College of Engineering, Pattoor, Alappuzha; [3]Associate Professor, Karunya Institute of Technology and Sciences, Coimbatore, India; pdf.catherine@lincoln.edu.my, rahulkrish1990@gmail.com, albertrajan.a@gmail.com

Abstract

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental diseases affecting children and adolescents, frequently causing difficulties with attention, impulse control, and cognitive performance. Early and correct ADHD diagnosis is critical for successful intervention and treatment planning. EEG recordings are valuable biosignals in determining the presence of this disorder. However, previous studies using conventional machine learning algorithms have failed to capture the complexity and nuances of these bio-signals, leading to less than optimal results. The current study describes an ensemble-based deep learning system for identifying ADHD using EEG signals. The proposed approach uses convolutional neural networks, transformer networks, and extreme gradient boosting methods to automate ADHD diagnosis based on EEG data. The system makes use of EEG signals collected from 19 channels, which were pre-processed and normalized before being used in the model. The convolutional neural network was used to extract local features from EEG data, while the transformer network captured global information from encoded input signals. Finally, the extreme gradient boosting classifier was utilized to divide the collected features into two groups: ADHD and normal. The model's performance was assessed on a benchmark EEG dataset for ADHD classification. The model had an accuracy of 98.94%, 99% precision, recall, and F1-score, suggesting good performance in identifying the two types of EEG signals. The confusion matrix showed few classification mistakes, and a comparison of findings demonstrated that the suggested CNN-Transformer-XGBoost framework outperformed existing machine learning and deep learning models in distinguishing ADHD from normal EEG signals. Thus, the suggested framework can be used as a computer-aided diagnostic system to help doctors identify ADHD patients by analyzing EEG records.

Keywords—*ADHD Detection, XGBoost, Transformer Network, Deep Learning, Brain Signal Analysis, EEG, CNN..*

1. Introduction

ADHD, which stands for Attention Deficit Hyperactivity Disorder, is a neurodevelopmental disorder that is associated with characteristics of inattention, hyperactivity, and impulsivity. This condition develops at an early stage of development and continues until adulthood, seriously affecting the academic performance, social conduct, and quality of life of affected individuals. In spite of the fact that the traditional way of detecting ADHD involves a subjective evaluation of behavior and interviews, there have been reports of objective neurophysiological tests for accurate and reliable diagnosis of the condition. EEG is one such test that is proven to be very effective; in past research, there have been noticeable spectral differences between healthy individuals and those with ADHD. Recently, state-of-the-art techniques such as convolutional neural networks (CNN) have been used to extract features automatically from EEG data with good classification results. Though CNN exhibits excellent results in extracting local features through convolutional filters, the self-attention architecture of the vision transformer is used to model global context. But most of these models tend to perform poorly when both local and global features are extracted together and each of these may not have the ability to leverage information from the classification head.

The proposed model is a combination of CNN-Transformer-XGBoost for ADHD detection in EEG signals. The CNN is employed for extracting local discriminative features, and the Transformer encoder is used for capturing the global context. In addition, the combination of deep features extracted by the CNN and Transformer models is classified via XGBoost, which improves the classifier performance through gradient-boosted trees and ensemble learning.

The primary contribution of this study is the development of a hybrid CNN-Transformer-XGBoost architecture for automated EEG-based ADHD detection, as well as thorough experimental findings demonstrating the efficacy of the proposed method for distinguishing ADHD participants from healthy ones. Using 19 multichannel EEG recordings, the created model achieves 98.94% classification accuracy. Other evaluation criteria, including as precision, recall, F1-score, confusion matrix, and receiver operating characteristic analysis, are also taken into account for a comprehensive performance assessment. This paper is broken down into six pieces. Section 2 discusses related efforts on detecting ADHD using EEG waves. The suggested CNN-Transformer-XGBoost framework is described in depth in Section 3. Section 4 explains the experimental design and the validation dataset. Furthermore, Section 5 offers the results and discussion, and Section 6 closes this work and suggests possible next directions.

2. Related Work

A number of studies have been made in order to identify ADHD on the basis of EEG. Some authors focus on traditional signal processing and machine learning methods, while others propose deep learning concepts. Among others, Catherine Joy et al. [1] analyzed several entropy-based features extracted from EEG and classified them using artificial neural networks. The authors concluded that nonlinear entropy features contain discriminant pattern information for ADHD classification. Another work, conducted by Tor et al. [2], described an automated method based on signal decomposition and nonlinear techniques, which allows detecting both conduct disorder and ADHD with higher sensitivity.

In their study, Ahire et al. [3] evaluated conventional machine learning algorithms for ADHD detection on the basis of EEG features extracted from raw signals. Similarly to previous works, the authors proved that features obtained using classical signal processing techniques are relevant for ADHD classification. The use of traditional feature engineering was limited due to a lack of ability to fully model complex nonlinear patterns. On the other hand, Pedrollo et al. [4] applied the Random Forest classifier to frequency domain features, extracted from resting-state EEG signals. The authors identified several discriminant power ratios related to frontal and central electrodes located at the head.

In his recent work, Bai et al. [5] proposed the ADHD-Conformer – a hybrid convolutive-attention neural network, which was inspired by Conformer models used in automatic speech recognition. The architecture is characterized by superior performance due to the combination of deep convolution layers with multi-head attention mechanisms. The authors claimed that the developed approach achieved accurate results in identifying distinct patterns of ADHD patients compared to healthy controls. Ahire [6] suggested an attention-driven deep learning approach for child EEG recordings with special focus on self-attention to identify ADHD-related time frames. Li et al. [7] proposed to represent EEG channels as a graph and perform a graph convolution over multiple domains to encode spatial relations between electrodes that traditional CNNs cannot capture. Similarly to our approach, Sarker et al. [8] designed an ADHD classifier based on a Transformer encoder and an XGBoost classifier, claiming that gradient-boosted trees' application on the top of a deep neural network's embeddings improves the performance of an end-to-end model.

Besides specific contributions to ADHD classification, the present research is based on several foundational works. First of all, Vaswani et al. [9] introduced the Transformer architecture and the scaled dot-product self-attention operation for modeling sequential data. Chen and Guestrin [10] proposed an XGBoost algorithm, a regularized gradient boosting framework that can be applied to regression and classification tasks. Last but not least, LeCun et al. [11] and Krizhevsky et al. [12] designed CNN architectures and Kingma and Ba [13] suggested an Adam optimization algorithm that we use to train our CNN–Transformer branch.

Furthermore, the selection of preprocessing procedures is essential to the effectiveness of ADHD classification models. García-Ponsoda et al. [14] found that EEG preprocessing and temporal segmentation have a significant impact on ADHD classifier performance. They investigated the effect of several segmentation procedures and concluded that windowing, normalization, and artifact rejection have a significant impact on classifier performance. In turn, Hossain et al. [15] proposed a multi-band spatial feature augmentation strategy for EEG-based ADHD classification. Overall, our findings emphasize the need of developing sophisticated feature encoding techniques for EEG signals.

Summing up, the analysis of related research revealed that (i) nonlinear and entropy-based features tend to yield a better performance than linear ones, (ii) attention-based encoder architectures are efficient for modeling long-range temporal dependencies in EEG, (iii) the use of CNNs and XGBoost in an end-to-end manner increases the performance of traditional machine learning pipelines. However, most of the prior research employs CNNs or Transformers separately before applying ensemble classification. So, in this research work, we would be combining the architecture of CNN and Transformer to extract local and global features along with their application using XGBoost to enhance ADHD detection performance.

3. Proposed Methodology

The current study offers an ensemble-based deep learning technique that uses Convolutional Neural Networks (CNNs), Transformer networks, and the Extreme Gradient Boosting (XGBoost) algorithm to automate ADHD diagnosis using electroencephalogram data. The methodology involves five stages, which include acquiring EEG data, preprocessing, extracting deep features using CNNs and Transformers, fusing the learned features, and classifying the EEG segments using XGBoost. The proposed framework distinguishes between ADHD patients and the control group using multi-channel EEG recordings. The obtained EEG signals are first preprocessed and normalized to guarantee that the distribution of input data is consistent. After normalization, the signals are partitioned into equal-length segments. The segmented data are then encoded using CNNs and Transformers independently. The obtained features are concatenated to produce a fused feature vector. The feature vector is then fed into a dense layer to generate the final classification results. Finally, the XGBoost classifier predicts whether an EEG segment falls in the ADHD or Control category. Figure 1 depicts the general design of the proposed framework.

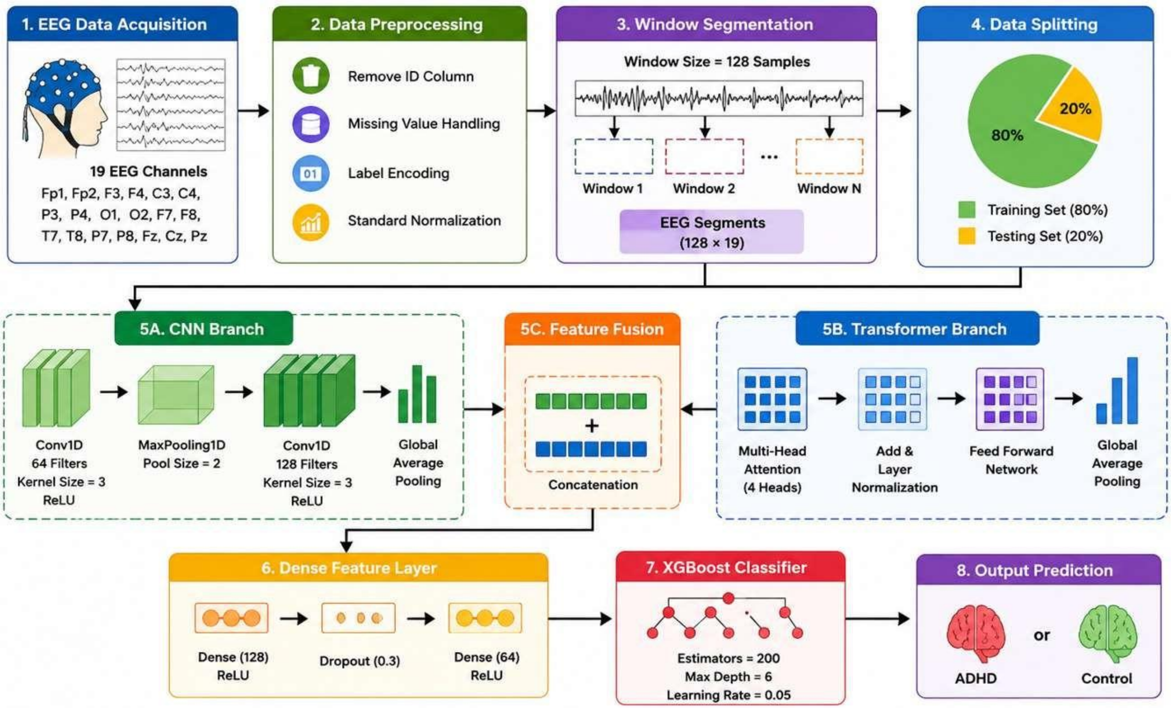

Fig. 1. Architecture of the proposed CNN–Transformer–XGBoost framework for EEG-based ADHD detection.

3.1 EEG Data Preprocessing

Since raw EEG data may have different amplitudes and statistical properties for different channels, it is necessary to perform preprocessing steps to make the data more consistent and suitable for learning. As a result, the following steps were taken:

1. The attributes containing subject identifiers were removed.
2. The values of the EEG channels were converted to numbers.
3. The missing values were imputed with the mean.
4. The features were normalized using the Standard Scaler.
5. The data was split into segments of 128 samples each.

The normalized feature vector is calculated using the formula below:

$X\_norm = (X - \mu) / \sigma$, where X is the initial feature vector, μ is the mean, and σ is the standard deviation. This transformation is needed to prevent features with large variances (e.g., EEG channels with high amplitudes) from overpowering others during learning.

3.2 CNN-Based Feature Extraction

The preprocessed and segmented EEG windows are first fed into a one-dimensional convolutional neural network (CNN) branch, which is intended to extract local temporal-spatial properties from the input data. The CNN consists of two stacked 1D convolutional layers with 64 and 128 filters, respectively, each with a kernel size of 3 and employing the ReLU function. After each convolution process, a MaxPooling1D is used to down sample the input data and minimize the dimensionality of the feature space. The network is quite deep, allowing the model to learn complicated temporal-spatial patterns from the provided EEG data. Particularly, the first few convolutional layers can be responsible for detecting low-level waveform patterns, whereas the later layers can capture high-level features. Finally, it is noteworthy that each layer is equipped with batch normalization, which accelerates the training process and improves the model performance.

3.3 Transformer-Based Feature Extraction

A transformer module may capture long-range temporal-spatial dependencies in EEG data, whereas CNNs are primarily geared for learning local patterns. The transformer architecture includes a multi-head self-attention mechanism that allows the model to balance the importance of distinct points in the input data relative to one another. In mathematical terms, the transformer attention is calculated as

$$\text{Attention } Q,K,V = \text{Softmax}(QK^T / \sqrt{d_k})\, V ,$$

where Q , K and V are the query, key, and value projection matrices, respectively, and d_k is the key dimension. In our case, each head has a dimension of 32, so that the four self-attention heads can jointly capture prominent patterns in the data. After the attention operation, layer normalization and average pooling are applied to obtain relevant high-level features, which encode global features in the EEG segments.

3.4 Feature Fusion and Dense Representation

The CNN and Transformer branch outputs are combined together to produce the fused feature representation as shown below: Fusion = CNN Features $\oplus$ Transformer Features.

The fused features are then further processed with fully connected dense layers and ReLU activation. Following that, a 0.3-rate dropout layer is added to boost the network's generalization performance and reduce overfitting. The last dense layer produces a 64-dimensional deep feature vector, which is used for classification.

3.5 XGBoost Classification

The XGBoost classifier, an ensemble learning technique that solves the classification problem using gradient boosted decision trees, receives the extracted feature vector. The objective function below provides the mathematical formulation of XGBoost. $Obj = \Sigma L(y_i, \hat{y}_i)$
$+ \Sigma\Omega(f_k)$, where L is the loss function that measures the error between the true label y_i and the predicted probability $\hat{y}_i$ and Ω is the regularization term that penalizes the complexity of the model to prevent overfitting. Using the deep features, XGBoost learns the decision boundaries and classifies the input EEG signals into either the ADHD or Control category.

4 Dataset Description

An EEG-based ADHD dataset comprising recordings from ADHD and healthy control patients was used to test the proposed architecture. The dataset includes 19 EEG channels (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, and Pz) that
correspond to typical electrode placements. 2,166,383 EEG samples from the ADHD and Control classes made up the initial dataset. The EEG recordings were divided into fixed-length windows of 128 samples, yielding 16,924 EEG segments, to enable deep learning-based analysis. There were 9,430 segments in the ADHD class and 7,494 segments in the Control class. An 80:20 split of the dataset was used to create training and testing subsets.

4.1 Hyperparameter Configuration

The CNN–Transformer architecture was trained using optimized hyperparameters to ensure effective learning and convergence.

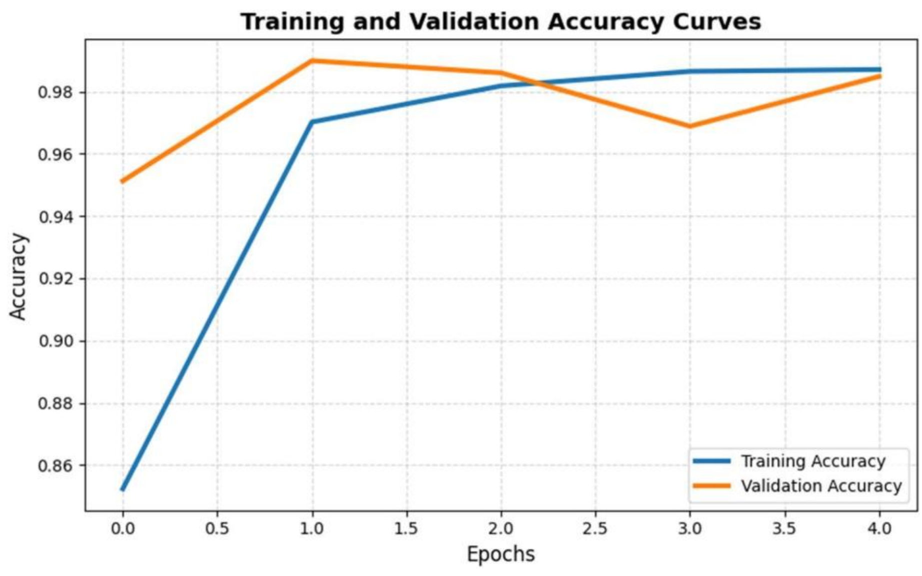

Fig. 2. Training and validation accuracy curves of the CNN–Transformer model across epochs.

The CNN-Transformer model can learn the most discriminative features for ADHD classification, as measured by the training and validation accuracy improving quickly during the first few epochs and then stabilizing at a pretty high level beyond 96%. Furthermore, the two lines of accuracy (for training and validation) are extremely close to one another and approach about 98.5%, indicating that the suggested model performs well in generalization without significantly overfitting.

4.2 Evaluation Metrics

A number of commonly used classification metrics, including accuracy, precision, recall, F1-score, confusion matrix, and receiver operating characteristic (ROC) analysis, were used to assess the effectiveness of the suggested model. These metrics offer a thorough evaluation of the classifier's ability to differentiate between individuals with ADHD and healthy control participants.

Accuracy determines overall classification performance by calculating the proportion of correctly identified samples out of a total of samples.

Its definition is: Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision measures the proportion of accurately predicted positive samples out of all positive samples, demonstrating the classifier's positive prediction reliability.

It is calculated as: Precision = TP / (TP + FP)

Recall, also known as sensitivity, is the fraction of actual positive samples accurately identified by the classifier. It is expressed as follows: Recall = TP / (TP + FN).

Finally, the F1-score is calculated as the harmonic mean of Precision and Recall, providing a balanced measure of classification performance that accounts for both false positives and false negatives. Where TP denotes true positive predictions, TN denotes true negative predictions, FP denotes false positive predictions, and FN denotes false negative predictions.

4.3 Comparative Performance Analysis

The performance of the suggested framework was assessed by contrasting it with both deep learning and traditional machine learning models.

Table 4. Performance Comparison of ADHD Detection Models

Model	SVM	Random Forest	XGBoost	CNN	CNN-BiLSTM	Transformer	CNN-Transformer	Proposed CNN-Transformer-XGBoost

Accuracy (%)	90.21	92.48	94.36	95.82	97.13	97.56	98.12	98.94

The results show that the framework has the highest classification accuracy across all models, indicating that CNN, Transformer, and XGBoost can effectively classify ADHD.

5. Results and Discussion

The ability of the proposed CNN–Transformer–XGBoost model to distinguish ADHD individuals from healthy controls was evaluated on the EEG-based ADHD dataset. Standard classification metrics, such as accuracy, precision, recall, confusion matrix, F1-score, and ROC analysis, were used to assess the model.

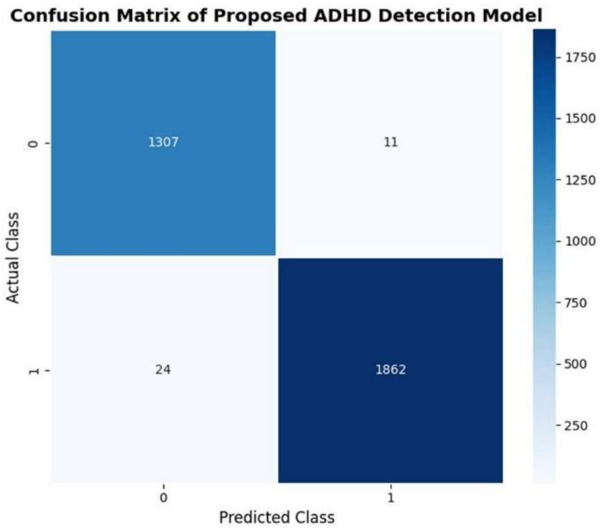

Figure 3: Confusion matrix resulting from the proposed framework.

The confusion matrix results prove that the proposed CNN–Transformer–XGBoost approach can achieve a good classification performance on ADHD and Control. Most of the samples are correctly classified, and only a few samples are misclassified. This result shows that the proposed method has a strong ability for classifying ADHD EEG data.

6. Conclusion

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental disorder that must be identified accurately and swiftly. In this study, a novel ensemble deep learning architecture including the Convolutional Neural Network (CNN), Transformer, and XGBoost algorithms was built and evaluated for ADHD classification. The system made use of CNNs' capacity to extract local features, Transformers' global dependency modeling, and XGBoost's excellent classification.

A multichannel EEG dataset with 19 channels was used to evaluate the suggested model. After preprocessing, normalizing, and dividing the EEG signals into time frames, features were retrieved and identified using the CNN-Transformer framework and the XGBoost algorithm. The overall accuracy of the suggested model was 98.94%, with precision, recall, and an F1-score of 99%. To evaluate the performance of the created framework, a confusion matrix and ROC analysis were used. The results showed that the suggested model efficiently distinguished ADHD patients from healthy controls.

Compared to typical machine learning and deep learning models, the CNN-Transformer-XGBoost architecture performed much better at classifying ADHD. The framework substantially reduced the possibility of misclassification and increased the accuracy of diagnosis by leveraging the strengths of several algorithms in local and global feature learning of EEG data.

References:

[1] R. Catherine Joy, S. Thomas George, A. Albert Rajan, and M. S. P. Subathra, "Detection of ADHD From EEG Signals Using Different Entropy Measures and ANN," Clinical EEG and Neuroscience, vol. 53, no. 1, pp. 12–23, 2022.

[2] H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth, J. K. E. Wei, V. Jahmunah, S. L. Oh, U. R. Acharya, and D. S. S. Fung, "Automated Detection of Conduct Disorder and Attention Deficit Hyperactivity Disorder Using Decomposition and Nonlinear Techniques with EEG Signals," Computer Methods and Programs in Biomedicine, vol. 200, 105941, 2021.

[3] N. Ahire, R. N. Awale, and A. Wagh, "Electroencephalogram (EEG) Based Prediction of Attention Deficit Hyperactivity Disorder (ADHD) Using Machine Learning," Applied Neuropsychology: Adult, vol. 32, no. 4, pp. 966–977, 2025.

[4] G. R. Pedrollo, L. B. Bagesteiro, A. R. Franco, and A. Balbinot, "ADHD Diagnosis Through Resting-State EEG Frequency Analysis With Random Forest," in Proc. IEEE EMBC, 2024, pp. 1–4.

[5] J. Bai, B. Hou, and J. Wu, "ADHD-Conformer: EEG-Based Classification and Detection of ADHD," Frontiers in Computing and Intelligent Systems, vol. 12, no. 3, pp. 97–99, 2025.

[6] N. K. Ahire, "Attention-Driven Deep Learning Framework for EEG Analysis in ADHD Detection," Applied Neuropsychology: Child, 2025.

[7] L. Li, X. Guo, Z. Yang, Y. Zhao, X. Liu, J. Yang, Y. Chen, X. Peng, and L. Han, "ADHD Detection From EEG Signals Using GCN Based on Multi-Domain Features," Frontiers in Neuroscience, vol. 19, 1561994, 2025.

[8] S. R. Sarker, S. Mehjabin, M. M. Piper, R. Rahman, F. U. Islam, and M. G. R. Alam, "A Hybrid Approach to Attention Deficit Hyperactivity Disorder Detection Leveraging Transformer and XGBoost Models Using XSparseFormerNet," Scientific Reports, vol. 15, 41039, 2025.

[9] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.

[10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016,
pp. 785–794.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[12] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc.

International Conference on Learning Representations (ICLR), 2015.
[14] 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, 2024.
[15] M. B. Hossain, M. A. I. Himel, M. A. Rahim, S. Mahmood, A. S. M. Miah, and J. Shin, “Classification of ADHD and Healthy Children Using EEG Based Multi-Band Spatial Features Enhancement,” 2025.