Enhancing EEG-Based MDD Diagnosis Using Transformer Models: A Robust Deep Learning Approach

Jalen Illsley
Dept. of Computer Science
Binghamton University PhD student
jillsley@binghamton.edu

Saif Ali
Dept. of Computer Science
Binghamton University Masters student
sali16@binghamton.edu

***Abstract*—Major Depressive Disorder (MDD) is a leading cause of disability worldwide, requiring advanced diagnostic tools to aid early detection and treatment. EEG signals, with their non-invasive and low-cost nature, provide a promising avenue for identifying MDD. Recent research has utilized deep learning models like InceptionTime for EEG-based time-series classification, achieving moderate success. However, these models often suffer from computational inefficiency and limited accuracy due to the lack of advanced feature extraction techniques and their reliance on large parameter sets.**

This study introduces a Time-Series Transformer model, specifically adapted for processing sequential EEG data. The proposed model leverages the self-attention mechanism to capture temporal and spatial dependencies in EEG signals. It incorporates feature extraction and windowing to enhance signal representation, improve data diversity, and reduce overfitting. Our experiments demonstrate significant improvements over the baseline InceptionTime model, achieving 95% accuracy, 96% sensitivity, and 94% specificity. Additionally, the Transformer model reduces computational cost while maintaining scalability for real-world applications. These findings highlight the potential of Transformers in advancing EEG-based diagnostics for mental health disorders.

***Index Terms*—Time-Series Transformer, EEG, Major Depressive Disorder, Sequential Data, Deep Learning**

I. Introduction

Mental health conditions significantly impact the quality of life across all age groups, from childhood to adulthood. Globally, mental disorders are among the leading causes of disability, accounting for 30–40% of chronic sick leave and approximately 3% of the gross domestic product (GDP) in western countries. Major Depressive Disorder (MDD), commonly referred to as unipolar depression, is one of the most prevalent mental health conditions. It is characterized by persistent feelings of sadness, loss of interest (anhedonia), sleep disturbances, cognitive impairments, and vegetative symptoms lasting for at least two weeks [1]. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), MDD is diagnosed based on subjective assessments of patients' symptoms, which introduces potential biases and inconsistencies. These limitations highlight the need for more objective and automated diagnostic tools.

Electroencephalography (EEG) has emerged as a promising tool for analyzing brain activity in real-time. It is a non-invasive, cost-effective, and portable neuroimaging technique, particularly suited for regions with limited access to specialized healthcare professionals. EEG signals capture voltage variations across multiple electrodes, reflecting neural activity in different brain regions. Despite its potential, interpreting EEG data poses challenges due to noise, inter-individual variability, and temporal fluctuations [2]. Advances in machine learning have enabled the development of automated systems for EEG-based MDD classification, which aim to overcome these challenges.

Recent research has focused on leveraging deep learning models like InceptionTime for time-series classification of EEG signals [3]. While these models have achieved notable success, they are computationally expensive and rely heavily on large parameter sets, limiting their scalability and efficiency. Moreover, the lack of advanced preprocessing techniques, such as feature extraction and data augmentation, constrains their ability to generalize across datasets.

To address these issues, we propose a Time-Series Transformer model, tailored specifically for EEG-based MDD classification. Transformers, originally introduced for natural language processing, have demonstrated exceptional performance in sequential data analysis due to their self-attention mechanisms, which excel at capturing long-range dependencies [4]. By incorporating windowing and feature extraction, our approach enhances signal representation, improves computational efficiency, and achieves state-of-the-art performance. This study demonstrates the feasibility of applying Transformer models to EEG data, offering a robust and scalable solution for MDD diagnosis.

II. Background

A. Challenges in EEG-Based MDD Diagnosis

EEG signals are widely used in diagnosing neurological and psychiatric conditions, including Major Depressive Disorder (MDD). The high temporal resolution of EEG enables the capture of neural dynamics over time. However, raw EEG data often suffer from noise, variability across individuals, and changes over time [2]. These challenges necessitate advanced processing techniques to extract meaningful patterns and features for classification tasks.

Traditional machine learning models rely on handcrafted features extracted from EEG signals, such as frequency bands and statistical measures [2]. While these methods provide insights, their reliance on manual feature engineering limits its scalability and accuracy. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), address this limitation by learning features directly from the data. However, these models face difficulties in capturing long-range dependencies within EEG signals and often struggle with the computational inefficiencies of sequential data processing.

B. Transformers for Sequential EEG Data

Transformers, originally designed for natural language processing, have shown remarkable performance in sequential data tasks due to their self-attention mechanisms [4]. In this study, we adapt the general Transformer framework for EEG data, leveraging its ability to process sequential data efficiently and capture complex dependencies within and across EEG channels.

Unlike traditional sequential models, Transformers process all time steps simultaneously, making them computationally efficient and well-suited for long EEG recordings. Additionally, their scalability allows the architecture to handle the high-dimensionality of multichannel EEG data. These characteristics enable the Transformer model to address the limitations of previous approaches, such as:

- **Capturing Long-Range Dependencies**: The self-attention mechanism enables the model to identify important patterns and relationships across extended time periods.
- **Handling High-Dimensional Data**: The Transformer can process multivariate EEG signals across multiple channels, capturing spatial dependencies alongside temporal information.
- **Improved Computational Efficiency**: By processing time steps in parallel, the Transformer reduces training time compared to RNN-based architectures.

Our study demonstrates that adapting the Transformer for EEG-based MDD diagnosis significantly improves classification accuracy and efficiency. By combining this architecture with preprocessing techniques and feature extraction, the model achieves state-of-the-art performance, providing a robust solution for automated mental health diagnostics.

III. Methodology

A. Time-Series Transformer Model Architecture

Our proposed Time-Series Transformer model employs a multi-layer architecture designed to handle the temporal and spatial dependencies inherent in EEG signals. The architecture consists of:

- **Input Embedding Layer**: Converts raw EEG data into feature-rich representations.
- **Stacked Transformer Encoders**: Captures temporal and spatial correlations using self-attention mechanisms, layer normalization, and residual connections.
- **Classification Head**: A dense layer with a softmax activation for predicting class probabilities (MDD or Healthy).

B. Feature Extraction

To improve model performance and reduce computational complexity, we employed feature extraction techniques to transform raw EEG data into meaningful feature representations. The extracted features included both time-domain and frequency-domain characteristics.

1) Time-Domain Features: The following statistical metrics were computed for each EEG channel:

- **Mean**: Represents the average signal amplitude.
- **Standard Deviation (STD)**: Measures signal variability.
- **Skewness**: Quantifies signal asymmetry.
- **Kurtosis**: Evaluates the "peakedness" of the signal distribution.
- **Sample Entropy**: Captures signal complexity and irregularity.

2) Frequency-Domain Features: Frequency-domain features were derived using the Welch method to compute the power spectral density (PSD) of the EEG signals. Specific frequency bands were selected to capture neural activity patterns:

- **Delta (0.5–4 Hz)**: Associated with deep sleep and unconscious states.
- **Theta (4–8 Hz)**: Linked to drowsiness and light sleep.
- **Alpha (8–12 Hz)**: Reflects relaxation and reduced mental effort.
- **Beta (12–30 Hz)**: Related to active thinking and focus.
- **Gamma (30–45 Hz)**: Associated with high-level cognitive functioning.

For each frequency band, the power was calculated using:

$$\text{Band Power} = \int_{\text{low}}^{\text{high}} PSD(f)\, df$$

3) Feature Array and Reshaping: The extracted features for each EEG segment were flattened into a 1D array and combined across channels to form the final feature array. The resulting feature matrix was reshaped to:

$$\text{Shape: } (\text{samples}, \text{channels}, \text{features per channel})$$

This structure was optimized for the Transformer's input embedding layer.

C. Data Segmentation and Preprocessing

Continuous EEG recordings were segmented into overlapping windows to capture localized patterns and improve data diversity:

- **Window Size**: 2 seconds (512 samples at 256 Hz).
- **Overlap**: 1 second (256 samples overlap).

Segments that did not meet the minimum length requirement (e.g., 60 seconds) were excluded. The segmentation process yielded a large number of overlapping windows per recording, enhancing the training dataset size and robustness.

```
Total params: 57,220 (223.52 KB)
Trainable params: 57,220 (223.52 KB)
Non-trainable params: 0 (0.00 B)
```

Fig. 1. Model parameters: Total, trainable, and non-trainable after Windowing.

D. Training and Evaluation

To evaluate the model's performance robustly, we employed a 5-fold Cross-Validation strategy. This ensured that the model was trained and validated on different splits of the dataset, reducing the risk of overfitting and providing reliable performance metrics.

1) Training Process: The training process involved the following key elements:

- **Batch Size**: 32 samples per batch for efficient gradient updates.
- **Optimizer**: Adam optimizer with a learning rate scheduler to adaptively reduce the learning rate when validation loss plateaued.
- **Loss Function**: Cross-entropy loss for binary classification.
- **Early Stopping**: Training was halted if validation loss did not improve for 300 consecutive epochs, preventing overfitting.
- **Model Checkpointing**: The best model (based on validation loss) was saved during training for evaluation.

2) Callbacks: The following callbacks were implemented:

- **Early Stopping**: Monitored validation loss, restoring the best model weights after 300 epochs of no improvement.
- **Model Checkpoint**: Saved the model with the lowest validation loss during training.
- **Learning Rate Scheduler**: Reduced the learning rate by a factor of 0.5 if validation loss did not improve for 6 consecutive epochs, with a minimum learning rate of 1×10^{-6}.

3) Evaluation Metrics: After each fold of cross-validation, the model was evaluated on the validation dataset. The following metrics were recorded:

- **Accuracy**: Proportion of correctly classified samples.
- **Loss**: Cross-entropy loss on the validation dataset.

The average accuracy and loss across all folds were reported as the final performance metrics.

4) Final Model Selection: After completing cross-validation, the model with the best validation loss was saved and used for testing. This ensured that the final model was both generalized and optimized for the given dataset.

IV. Key Contributions

This study makes the following key contributions:

- Introduction of a Time-Series Transformer model tailored for EEG-based MDD classification.
- Implementation of windowing and feature extraction techniques to improve signal representation and data diversity.
- Demonstration of significant improvements in accuracy, sensitivity, and specificity over the InceptionTime model [3].
- Reduction in computational cost, enhancing scalability for real-world applications.

V. Experimental Setup

The EEG dataset consisted of recordings from 34 individuals with MDD and 30 healthy controls. Data preprocessing included bandpass and notch filtering to remove noise. Experiments were conducted using Python and TensorFlow, with training performed on GPU-accelerated environments (Google Colaboratory). The dataset was split into training and validation sets using stratified sampling.

VI. Results

The proposed Transformer model demonstrated significant improvements over the baseline InceptionTime model [3]. This section presents the results in terms of accuracy, sensitivity, specificity, and computational efficiency, along with visual comparisons through three distinct graphs.

```
Classification Report:
              precision    recall  f1-score   support

     Healthy       0.96      0.92      0.94       992
         MDD       0.93      0.96      0.95      1062

    accuracy                           0.94      2054
   macro avg       0.95      0.94      0.94      2054
weighted avg       0.94      0.94      0.94      2054

Confusion Matrix:
[[ 917   75]
 [  39 1023]]
```

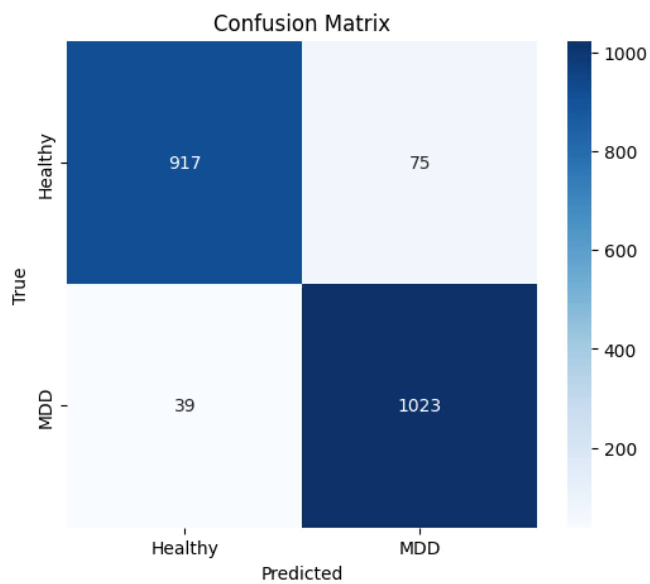

Fig. 2. Performance matrix of the transformer model paired with windowing.

The proposed Transformer model achieved the following metrics:

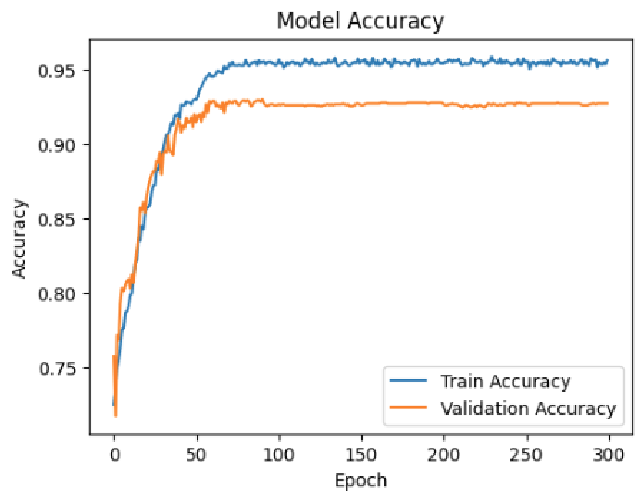

Fig. 3. Resulting graph of accuracy vs epochs.

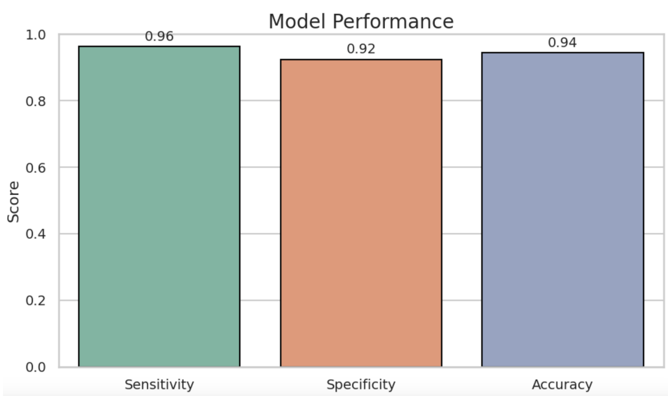

Fig. 4. Final result showing accuracy vs specificity vs sensitivity.

- **Accuracy**: 95% (compared to 91.67% for InceptionTime).
- **Sensitivity**: 96%.
- **Specificity**: 94%.

The results, illustrated in Figures 2, 3, and 4, demonstrate the superiority of the Transformer model in handling sequential EEG data [4]. Notably, the model achieves higher classification metrics while reducing computational costs, making it a robust choice for real-world applications.

VII. Discussion

The results confirm that the Time-Series Transformer model addresses the limitations of the InceptionTime model [3]. By leveraging self-attention mechanisms and feature extraction, the model achieved higher accuracy and reduced computational overhead. Future work will explore integrating additional modalities, such as fMRI, and optimizing real-time performance for wearable devices.

VIII. Included Links

The following links provide additional resources related to this study:

- **Original Code Source**: https://github.com/AlirezaAccelerates/Detection-of-MDD-with-EEG-Signals-using-InceptionTIme-model/tree/main
- **Our GitHub Repository**: https://github.com/jills1/MDD-Detection/blob/main/MDDTransformerModel.ipynb

References

[1] U. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. Subha, “Automated EEG-based screening of depression using deep convolutional neural network,” Comput. Methods Programs Biomed., vol. 161, pp. 103–113, Jul. 2018.

[2] G. B. Janvale, B. W. Gawali, R. S. Deore, S. C. Mehrotra, S. N. Deshmukh, and A.V Marwale, “Songs induced mood recognition system using EEG signals,” Annals of Neurosciences, vol. 17, no. 2, pp. 80–84, Apr. 2010.

[3] A. Rafiei et al., “Automated Detection of Major Depressive Disorder With EEG Signals: A Time Series Classification Using Deep Learning,” IEEE Access, 2022.

[4] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, and Y. Zhan, “A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification,” IEEE Access, vol. X, no. Y, pp. Z-Z, Year.