Electroencephalography Analysis for Mental Health Classification Using Transformers

Vincenta), Kevin Kedrickb), Bakti Amirul Jabarc), and Felix Indra Kurniadid)

Computer Science Department, Bina Nusantara University, Jakarta, Indonesia

a) Corresponding author: vincent060@binus.ac.id

b) kevin.kedrick@binus.ac.id

c) bakti.jabar@binus.ac.id

d) felix.indra@binus.ac.id

**Abstract.** Mental health issues have become more prevalent in the twenty-first century, with millions of people affected in the world. They pose a serious threat to public health and the accurate classification of these issues is critical for effective treatment and intervention. With the help of machine learning, it is possible to efficiently and accurately classify these mental health issues which could help doctors in treating their patients. In this paper, we compared different transformer-based models to determine whether a patient has a mental health issue or not based on their EEG Signal data. For our experiments, we used the MODMA dataset. We compared three different transformer-based models which utilized a CNN layer before the transformer blocks. The results from our experiments showed that EEG-Deformer was the best model for this task, obtaining an accuracy of 65.45%, while EEGformer performed the worst.

**Keywords:** EEG, Transformers, mental health, EEG classification, machine learning

# INTRODUCTION

In modern society, mental health issues are becoming more prevalent, affecting millions of individuals worldwide. Mental health is a complex and multifaceted aspect of overall well-being, encompassing cognitive, emotional, and social elements. It is not just the absence of mental illness, but also the ability to cope with life's challenges and contribute to one's community. The promotion of mental health is crucial, particularly in the face of adverse conditions such as poverty and discrimination [1]. Factors such as stress, depression, and anxiety can significantly impact mental health, underscoring the importance of maintaining a balance in life activities and responsibilities. The diagnosis of these conditions needs to be as accurate as possible to give effective treatment to the patient. Traditional diagnostic methods for mental health issues typically rely on subjective assessments, resulting in inconsistent diagnoses. Integrating advanced technologies, such as machine learning, with traditional methods has proven to exhibit high potential in diagnosing mental health issues [2].

Electroencephalography (EEG) is a method to record the electrical activity of the brain. This technique has been widely used in many studies to learn how the brain functions, detect abnormalities, and diagnose neurological disorders. The reason why EEG is so widely used in research done on the brain is because of its non-invasiveness, real-time monitoring capability, and high temporal resolution. These advantages make it an ideal candidate for mental health research.

Transformers are a type of neural network that was originally designed for natural language processing tasks. However, they have proven highly adaptable and effective in various other tasks. Transformers excel in capturing information within sequential data. The attention mechanism within a transformer allows for efficient processing of input sequences, making them a good choice for the complex and dynamic nature of EEG data.

# RELATED WORKS

Numerous research studies have investigated the classification of EEG signals using different techniques. A study that used support vector machines (SVM) and empirical mode decomposition (EMD) to categorize interictal and ictal EEGs reported 99% accuracy [3]. In another study, the classification of normal, schizophrenia, and obsessive-compulsive disorder EEGs was accomplished with over 66% accuracy using a three-layered feedforward network with wavelet coefficients [4]. A different study that used convolutional neural networks and machine learning to create a categorization framework reported accuracy rates of 87.2% to 89.4% for five subjects [5]. Additionally, a work that used discrete wavelet transform-based feature extraction for EEG data captured during a challenging cognitive activity achieved an accuracy of more than 98% [6]. Together, these studies show how several approaches can be used to accurately identify EEG signals.

Some research has investigated the categorization of EEG signals. One such work created an atlas for EEG classification, pointing out different types of aberrant patterns [7]. Two studies that used artificial neural networks to classify EEG signals were able to identify between various EEG signal types with good accuracy rates [4], [8]. A thorough analysis of preprocessing, feature extraction, and classification methods was given in a different study [9], emphasizing the application of machine learning and swarm optimization strategies in two research papers, unique classification strategies were suggested. One paper introduced a hybrid technique that included Radial Basis Function Networks with Particle Swarm Optimization, while the other paper achieved high accuracy rates by using a feature extraction scheme based on discrete wavelet transforms [6], [10]. A study was conducted that also showed how well an extreme learning machine and sample entropy combination technique worked for categorizing EEG data. All these studies show how several approaches can be used to classify EEG signals.

Transformer-based models have proven to be useful in EEG analysis in recent studies. Using Transformer networks, two research papers by Pan [11] and Zhang [12] attained great accuracy in the detection and prediction of epileptic seizures. These models demonstrated the promise of Transformers in this field, outperforming conventional CNN and RNN networks. By adding an attention mechanism to the Transformer network, a research paper by Yan significantly enhanced performance by improving the Transformer network's capacity to use time, frequency, and channel information from EEG signals. These results imply that transformer-based models provide a viable method for EEG analysis [11].

# METHODS

## DATASET

This research utilized the MODMA dataset provided by [13]. The dataset used contains 128-electrodes EEG Signal data from 53 participants, which included 24 patients diagnosed with depression and 29 healthy controls. This research used 5 minutes of each patient’s recording for the experiments.

During the preprocessing of the data, this research first performed bandpass filtering on each recording, using 0.1 Hz and 40 Hz as the cutoff frequencies. Next, to standardize the data, this research converted all the recordings to microvolts, followed by performing an exponential moving standardization. After preprocessing, the dataset was then split into an 80/10/10 split in which 80% is used for training, 10% for testing, and the other 10% for validation.

## MODELS

This research employed 3 different transformer architectures named EEG-Conformer, EEGformer and EEG-Deformer. Each of the models used a learning rate of 0.00001. Each model was trained for 5 epochs using a batch size of 1 due to the high computational and memory cost.

### EEG-Conformer

EEG-Conformer is a compact convolutional transformer which aims to encapsulate local and global features in a unified EEG classification framework [14]. The research performed in the original paper proposing this architecture showed that EEG-Conformer could achieve state-of-the-art performances.

This paper used the transformer architecture proposed in [14]. The architecture starts with a convolution layer which then connects with six layers of self-attention layers, each with 10 self-attention heads. A dropout layer is used between each self-attention layer with a dropout rate of 50% to reduce overfitting. The last self-attention layer is connected to a dense layer with a dimension of 19760 to produce the output.

### EEGformer

EEGformer was first proposed in [15]. This model captures the EEG characteristics in a unified manner. [15] claims that the proposed EEGformer model outperforms many state-of-the-art models and can improve the task of classifying EEG signals.

The architecture for the EEGformer model is similar to the one proposed in [15]. However, due to the limited resources this research encountered while performing this experiment, the EEGformer architecture used in this paper only used 1 transformer block which also only used 1 attention head.

### EEG-Deformer

EEG-Deformer is a model that tries to improve the CNN-Transformer architecture that has been utilized in another research such as [14] and [15]. To overcome the limitations of CNN-Transformers, [16] proposed EEG-Deformer which incorporates a Hierarchical Coarse-to-Fine Transformer block and a Dense Information Purification module which utilizes purified temporal information from the EEG data to enhance decoding accuracy. The model was shown to outperform existing state-of-the-art methods or was at least performing as well as such methods.

The EEG-Deformer model in this paper used the same architecture as the one proposed in [16], starting with a CNN layer which then connects to four transformer blocks. Each transformer block used 16 self-attention heads. The last transformer block is then connected to a multilayer perceptron to produce the output.

## PERFORMANCE METRICS

To evaluate the results of our models, we will use accuracy, precision, recall, and F1 score. For precision, recall, and F1 Score we will take the weighted average from each class described in Equations 1-4.

1. Accuracy, Accuracy represents the number of correctly classified data instances over the entire dataset.

Accuracy = (1)

1. Precision, Precision represents the number of true positives over the number of total positives predicted by the model.

Precision = (2)

1. Recall, Recall represents the number of true positives over the number of actual positives in the dataset.

Recall =  (3)

1. F1-Score, The F1-Score is a metric that considers precision and recall. It is the harmonic means of precision and recall.

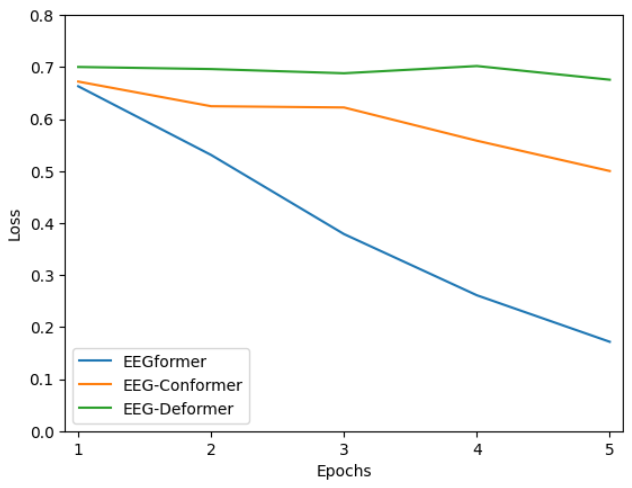
F1 Score = 2 \* (4)

TP (True Positive) and TN (True Negative) denote the total number of true positives and true negatives predicted by the model, respectively. When the model correctly predicts using the provided features that the data belongs to a class, it is considered a true positive, while data correctly predicted as not belonging to a class is considered true negative. Conversely, FP (False Positive) and FN (False Negative) denote the total number of false positives and false negatives, respectively. FP represents the number of incorrect predictions where the model classifies data into an incorrect class, while FN represents the number of incorrect predictions where the model incorrectly classifies data as not belonging to a class.

# RESULTS AND DISCUSSION

Using the three models mentioned, EEG-Conformer, EEGformer, and EEG-Deformer, the results of each model were used to compare which models are better in the task of classifying mental health issues. The models were trained under the same amount of data and the same parameters. Furthermore, the metrics used for evaluation are the same, namely accuracy, precision, recall, and F1 score.

**FIGURE 1** shows the learning curve of the three models used in this experiment. As mentioned before, due to the extremely limited resources that were available to the authors during this experiment, the models did not show promising results. The figure shows that the only model that was successfully trained and showed a consistent downwards trend was the EEGformer model. This downward trend meant that the EEGformer model was able to minimize errors and could converge towards an optimal solution. However, using only an epoch of 5, the model was not able to demonstrate its full ability. The other two models performed even worse, as they were unable to learn from the data and did not work towards a solution at all. The almost linear graph for both the EEG-Conformer and EEG-Deformer shows that the two models were not able to converge towards a solution for the problem.



**FIGURE 1.** Learning curves of each model.

**TABLE 1** shows the results of our experiments. In all performance metrics, the EEG-Deformer model showed the best results, followed by EEG-Conformer, with EEGformer performing the worst out of the three models.

The EEG-Conformer model showed strong performance for detecting mental health issues with a high precision, recall, and F1-score for the mental health issues class. However, the model shows signs of struggles with the healthy control class where all metrics were recorded to be low, which meant that the model was more likely to classify a healthy patient as having a mental health issue. Overall, the accuracy achieved by EEG-Conformer is 0.6364 but it also has an imbalance in performance in which it is very biased towards class 1.

On the other hand, the EEGformer model showed that it was able to work towards a solution for the problem with a more balanced result for both classes with a similar precision, recall, and F1 score for both classes. However, the overall accuracy, precision, recall, and F1-Score were the lowest amongst all the models. When comparing EEGformer and EEG-Conformer, it can be seen that EEGformer has a better performance in scenarios where balanced classification between classes is more important.

EEG-Deformer achieved a rather similar performance to the EEG-Conformer model which was very biased towards the mental health issues class. This model obtained a high precision, recall, and F1-Score for the mental health issues class while obtaining a low performance for the healthy control class. Although it is very biased towards class 1, its overall accuracy was still the best among the other models with an accuracy of 0.6546.

Overall, the three models obtained quite similar results. Showing that it is possible to classify mental health issues based on EEG Signal data using Transformers. However, the low performance shown from the three models indicated that these models are not yet ready to be used in real world scenarios. Further work needs to be done for this method to be a viable way of accurately classifying mental health issues for patients in the real world.

**TABLE 1.** Evaluation of Models for Classification of Mental Health Issues

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Epochs** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| EEG-Conformer | 5 | 0.6364 | 0.6364 | 0.6364 | 0.6364 |
| EEGformer | 5 | 0.6000 | 0.6000 | 0.6000 | 0.6000 |
| EEG-Deformer | 5 | **0.6545** | **0.6659** | **0.6546** | **0.6601** |

# CONCLUSIONS

The EEG-Deformer model showed the highest performance but was very biased towards class 1 which meant that the model was more likely to classify a healthy patient as having mental health issues. The EEG-Conformer model was the second best after the EEG-Deformer model but also suffered from overfitting like the EEG-Deformer model. However, the EEGformer model showed balanced performance among both classes which meant it was more reliable in determining whether a patient was healthy or had mental health issues.

The experiments conducted in this study were very limited due to the use of the dataset containing just 50 data samples. This limited data may have caused the performance issues seen in each model particularly with the class imbalances. Future work should prioritize using a larger and more diverse dataset and increasing the model's complexity while increasing the epochs used during training to allow the model to learn more about it. Besides larger dataset, the fine-tuning of each model is also crucial as it will enhance the overall effectiveness of the model and ensure a more reliable EEG classification to be used in real-world applications.

# References

1. Herrman, H., et al., *Mental Health Promotion.* Elsevier, 2008: p. 450-453.

2. Abd Rahman, R., et al., *Application of Machine Learning Methods in Mental Health Detection: a Systematic Review.* IEEE Access, 2020. **8**: p. 183952-183964.

3. Duan, L., et al. *A Novel EEG Signal Recognition Method Using Modified Optimal Electrodes Recombination Strategy*. in *IScIDE 2018*. 2018. Springer, Cham.

4. Hazarika, N., et al., *Classification of EEG Signals Using the Wavelet Transform.* Signal processing, 1997. **59**(1): p. 61-72.

5. Wu, M., B. Nan, and C. Li. *EEG Classification Based On Deep Learning*. in *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference* 2022. Chongqing, China: IEEE.

6. Amin, H.U., et al., *Feature Extraction and Classification for EEG Signals Using Wavelet Transform and Machine Learning Techniques.* Australasian Physical & Engineering Sciences in Medicine, 2015. **38**: p. 139-149.

7. Lüders, H. and S. Noachtar, *Atlas and classification of electroencephalography*. 2000: Saunders.

8. Lakshmi, N., et al. *Exploring the Performance of EEG Signal Classifiers for Alcoholism*. in *AIDE 2019*. 2020. Singapore: Springer.

9. Houssein, E.H., A.E. Hassanien, and A.A.K. Ismaeel. *EEG Signals Classification for Epileptic Detection: a Review*. in *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing*. 2017. New York, NY, USA: ACM.

10. Cinar, E. and F. Sahin. *A study of Recent Classification Algorithms and a Novel Approach for EEG Data Classification*. in *2010 IEEE International Conference on Systems, Man and Cybernetics*. 2010. Istanbul, Turkey: IEEE.

11. Pan, H., et al., *Epilepsy Seizure Detection Using Transformer.* Highlights in Science, Engineering and Technology, 2022. **1**: p. 325-329.

12. Zhang, X. and H. Li. *Patient-Specific Seizure Prediction from Scalp EEG Using Vision Transformer*. in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference*. 2022. Chongqing, China: IEEE.

13. Cai, H., et al., *A Multi-Modal Open Dataset for Mental-Disorder Analysis.* Scientific Data, 2022. **9**(1): p. 178.

14. Song, Y., et al., *EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022. **31**: p. 710-719.

15. Wan, Z., et al., *EEGformer: a Transformer–Based Brain Activity Classification Method Using EEG Signal.* Frontiers in Neuroscience, 2023. **17**: p. 1-13.

16. Ding, Y., et al., *EEG-Deformer: a Dense Convolutional Transformer for Brain-computer Interfaces.* IEEE Transactions on Neural Networks and Learning Systems, 2024: p. 1-10.