MULTIMODAL CONVOLUTIONAL TRANSFORMER (MCT-DD): DEPRESSION DIAGNOSIS THROUGH JOINT TASK ANALYSIS

Firoz N.[1], Beresteneva O.G.[2], Aksyonov S.V.[3]
[1]Tomsk State University, (Institute of Applied Mathematics and Computer Science, PhD. 3rd year Student), e-mail: nedafiroz1910@gmail.com
[2]Tomsk Polytechnic University (Department of Information Technology, Professor), e-mail: ogb6@yandex.ru
[3]Tomsk Polytechnic University (Department of Information Technology, Associate Professor), e-mail: axyonov@tpu.ru

Abstract

A new deep learning method, Multimodal Convolutional Transformer, analyzes EEG and genetic data to diagnose MDD. This approach achieved high accuracy (97,16 %) and surpasses other methods for early MDD detection, potentially aiding healthcare professionals.

Keywords: Major Depressive Disorder, EEG, Genetics, Deep Learning, Convolutional Neural Networks, Transformers.

Introduction

The worldwide prevalence of Major Depressive Disorder (MDD) affects millions, highlighting the critical need for novel diagnostic and therapeutic strategies [1]. Despite extensive efforts, the availability of non-invasive and accurate diagnostic tools for MDD remains limited. Analyzing Electroencephalography (EEG) signals presents a promising avenue in addressing this challenge, as EEG records the brain's electri-cal activity and waves via electrodes on the scalp. This method offers unique advantages, including high frequency resolution and affordability and portability compared to alternatives like MRI [2].

Utilizing such a non-invasive approach, coupled with automated deep learning techniques, and initiating timely treatment courses, holds significant promise [3]. EEG, as a non-invasive modality, allows for the measurement of electrical activity originating from various brain regions [4]. These signals can reveal the occurrence and localization of operational abnormalities inside the brain. Identifying deviations in brain physiology during depressive states holds potential for early disease detection [5].

Biological markers are objective features that are measured and assessed to serve as Markers of typical biological functions, disease advancement, or reactions to treatment interventions [6]. Despite dec-ades of research aimed at understanding and treating Major Depressive Disorder (MDD), the quest for non-invasive and quantitative diagnostic tests remains unfulfilled. Moreover, there are currently no approved biomarkers established as clinical diagnostic criteria for MDD patients, with clinical diagnosis largely de-pendent on subjective assessments of depressive behavior and clinical examinations [7, 8]. Therefore, there is an urgent need to develop a biomarker-based system for discerning MDD, which can aid in predicting disease progression and guiding treatment decisions during the early stages of the disorder.

To address this gap, we employed Deep Learning (DL) techniques to classify Major Depressive Disorder (MDD) utilizing both EEG and gene expression data, showcasing the potential of deep learning in advancing depression diagnostics. Previous methodologies typically focused on data modalities such as audio, text, video, EEG, or gene expression. For instance, in [9], hybrid transformer model was introduced, while in

[10], temporal convolutional transformer model was introduced for joint diagnostic task using text, audio, video, and EEG data, while in [11] the authors introduced transformer model over text data only. Other models like hybrid CNN-LSTM model performed well in only text-based models [12, 13]. In contrast to prior studies, our innovative approach integrates EEG and gene expression data, supported by robust meth-odologies, representing a unique contribution to depression detection amidst the exploration of multimodal data within a unified Multimodal Convolutional Transformer for Depression Detection (MCT-DD) for Joint Task, akin to previous methodologies [10, 14].

Methodolgy

In our study, we employed the MODMA dataset [15], which encompasses data from electroencephalography (EEG) signals and audio recordings of individuals diagnosed with clinical depression. This multimodal dataset was curated and published by Lanzhou University in 2020. Following the methodology outlined by Zhao et al. [16], the MODMA dataset was organized into 1321 segments. Subsequently, these

segments were partitioned into a training set comprising 971 segments and a test set containing 350 seg-ments.

The GSE98793 dataset, sourced from the Gene Expression Omnibus database (GEO), provided pub-licly accessible transcriptomic data for individuals with Major Depressive Disorder (MDD) and healthy controls (HCs). This dataset included high-resolution gene expression profiles obtained from whole blood samples of 128 MDD patients and 64 HCs. Gene expression was measured using the Affymetrix Human Genome U133-Plus 2.0 gene expression microarray. Diagnosis of MDD patients was based on the identifi-cation of at least two depressive episodes meeting DSM-IV or ICD-10 criteria, assessed using the semi-structured Schedule for Clinical Assessment in Neuropsychiatry (SCAN). Detailed demographic and clini-cal information regarding the GSE98793 dataset can be found publicly on the website.

EEG Data Preprocessing: Before feeding in the deep learning model, we pre-process the EEG signal as described in [18].

Genetic Data Preprocessing: We utilized significant gene data from the prominent genes in MDD patients compared to healthy controls, as suggested by [17]. The bioinformatics analysis conducted in [17] provided transcriptomic data with a large number of genes (features). In general, the transcriptomic data is huge surpassing the sample size, possibly resulting in overfitting during classification tasks. To address this issue, we applied Principal Component Analysis (PCA) for feature selection, reducing data dimensionality while retaining significant information. The significant genes identified in [17] were exclusively used in our model. Through an integrative analysis of bioinformatics and machine learning methods in [17], 10 primary MDD-associated biomarkers were pinpointed: NRG1, CEACAM8, CLEC12B, DEFA4, HP, LCN2, OLFM4, SERPING1, TCN1, and THBS1.

Model description

We used Convolutional and transformer models in hybrid form. Given that sequence models entail lengthier computation due to sequential processing, whereas Transformer models necessitate less execution time owing to parallelized processing [9]. Thus, we propose Multimodal Convolutional Transformer Depression detection (MCT-DD) model for joint task repression. Fig. 1 below depicts the proposed model workflow.

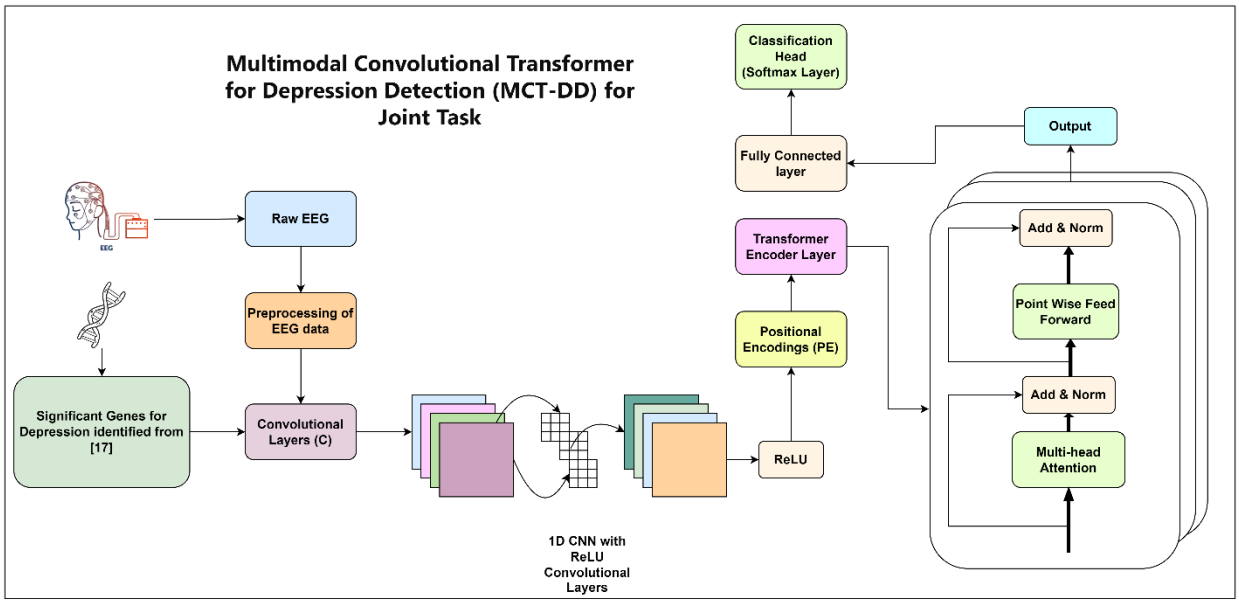

Fig. 1. Workflow of Multimodal Convolutional Transformer for Depression Detection (MCT-DD) for Joint Task

Further we discuss the two segments of the proposed architecture using CNN and Transformer Models.

CNN Layer

The input features from the EEG signal and audio modalities, respectively are fed to Convolutional Layers (C), it undergoes a spatial filtering operation to extract local patterns or features from the input feature set. The convolutional layer applies a set of learnable filters (also called kernels) to the input. Each filter slides across the input signal and computes a dot product between its weights and the values in the receptive field. This dot product represents a localized feature map, capturing spatial patterns within the input multimodal data. As the filters slide across the input signal, they detect various spatial patterns or features present in different

parts of the EEG signal and audio features. These features represent characteristics such as amplitude variations, frequency components, or temporal dynamics within the EEG signal and audio features.

We have X = the input EEG signal with dimensions channels × time × electrodes.

channels = the number of EEG channels (e.g., number of electrodes).

time = the length of the time series data (e.g., number of time samples).

electrodes = the number of electrodes.

F = number of filters (or kernels) used in the convolutional layer.

K = size of the filter/kernel (also referred to as the kernel size).

S = stride (step size) of the filter as it slides/spatially convolves across the input.

ReLU = Rectified Linear Unit activation function.

Then Convolutional Equations for EEG data signals are presented in equation 1.

$$Conv(x)_{i,j,f} = Re\left(\sum_{c=1}^{c}\sum_{k=1}^{k}\sum_{e=1}^{E} W_{k,ec,t^*} x_{i+k-1,j+e-1,c} + b_f\right) \quad (1)$$

where:

$Conv(X)_{i,j,f}$= the value at position (i,j) of the f^{th} feature map produced by the convolutional layer.

$X_{i,j,c}$= value at position (i,j) of the c^{th} channel of the input EEG signal.

$W_{k,e,c,f}$ = weight (or parameter) associated with the k^{th} element of the e^{th} filter for the c^{th} input channel and the f^{th} output feature map.

b_f = the bias term associated with the f^{th} feature map.

C = number of input channels (number of EEG electrodes).

E = the number of elements in the filter/kernel (kernel size)

i and j = iterate over the spatial dimensions of the output feature map.

The sums are computed over all input channels, kernel elements, and electrodes.

After the convolution operation, a non-linear activation function (e.g., ReLU) is applied elementwise to the feature maps. This introduces non-linearity to the model, enabling it to learn complex relationships between the input and the target output. Average pooling was applied after the convolutional operation. Pooling reduces the spatial dimensions of the feature maps while retaining the most salient information, aiding in spatial hierarchies, and reducing computational complexity.

The output of the convolutional layer consists of a stack of feature maps, each representing a different aspect or pattern extracted from the input signals.

These feature maps serve as input to subsequent layers in the Multimodal Convolutional Transformer for Depression Detection (MCT-DD) for Joint Task architecture.

Transformer Encoder

The output from the convolutional layer is fed into a Transformer model via positional encodings to provide positional information to the model. Further it undergoes a transformation through attention mechanism followed by fully connected neural networks.

Here X = output from the convolutional layer, which consists of feature maps representing spatial patterns extracted from the input EEG signals.

N = the number of positions (or time steps) in the input.

FFN(·) = the position-wise feedforward neural network.

The self attention mechanism is described as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where: Q, K, V = the query, key, and value matrices, respectively, obtained by linear transformations of X.

d_k = the dimensionality of the key vectors.

While the Multi-head Attention is represented as:

$$Multihead\ (X) = Concat\ (head_1, head_2\ , \ldots\ldots\ldots head_h) \cdot W^0 \quad (3)$$

where:

$head_i$ = Attention (XW_i^Q ,XW_i^K, XW_i^Vrepresents the i^{th} attention head).

W_i^Q ,W_i^K, W_i^V = the weight matrices for the query, key, and value projections for the i^{th} head.

W^0 = the weight matrix for the output projection.

The output is further processed by a fully connected layer, with Softmax function for classification into depressed or healthy controls.

A similar task was applied on genetic model using the significant genes in the dataset [17]. The input and output of both EEG based data and gene-based data are processed individually to test the performance of model.

Results

In the results section we compare our model with other existing deep learning models that were essentially high performing in [12, 13]. The results are illustrated in Table 1. Table 2 shows the hyperparameters used in our model.

Table 1

Performance of the Proposed Model

Model Name	Accuracy		F1		Precision		Recall	
CNN-LSTM-BiLSTM	EEG (MODMA dataset)	Gene model	EEG (MODMA dataset)	Gene model	EEG (MODMA dataset)	Gene model	EEG (MODMA dataset)	Gene Model
	0.88	0.94	0.89	0.95	0.85	0.96	0.88	0.94
CNN-LSTM	0.85	0.92	0.85	0.94	0.86	0.92	0.87	0.93
Our Proposed Convolutional Transformer Model	0.95	0.97	0.95	0.97	0.96	0.98	0.95	0.97

Table 2

Hyperparameter settings of the Convolutional Transformer proposed model

Layer Name	Parameter Settings
Optimizer	Adam
Learning Rate	0.001
Batch Size	256
Epochs	100
Regularization (L2)	0.001
Dropout Rate	0.2
Key Dimension in Multi-Head Attention	16
Units in Dense Layers	512, 16, 256, 256
Activation Function in Dense Layers	ReLU
Units in Output Layer	1
Activation Function in Output Layer	Softmax
Early Stopping	Yes

Our Convolutional Transformer model (MCT-DD) for Joint Task outperforms CNN-LSTM-BiLSTM [12] and CNN-LSTM [13] architectures in classifying MDD using EEG and gene expression data, achieving accuracies of 95 % and 97,16 %, respectively. The higher precision, recall, and F1 scores obtained by our

model on comparison with existing models indicate its ability to effectively distinguish between MDD subjects and healthy controls with fewer false positives and false negatives. These results highlight the efficacy of leveraging a unified Convolutional Transformer architecture for joint analysis of EEG and gene expression data, facilitating improved feature extraction and modeling of complex interactions within the data. By capturing both spatial and temporal dependencies in EEG signals while integrating gene expression information, our model demonstrates enhanced discriminatory power for MDD classification.

Conclusion

In conclusion, Major Depressive Disorder (MDD) presents a significant global health challenge, highlighting the need for innovative diagnostic tools, such as our Multimodal Convolutional Transformer model (MCT-DD), which integrates EEG and gene expression data, showcasing promising potential for enhancing depression diagnostics. While the model shows promise but lacks clinical validation and also it cannot predict MDD type or severity. Integrating EEG and genetic data in the (MCT-DD) model was hindered by dataset differences; future research may focus on collecting data from the same subjects to enhance diagnostic accuracy for complex diseases like MDD.

References

1. Shiryaeva O.S., Surikova Ya.A., Kondrashenkova S.V. Psychological support for personality in extreme conditions of life: monograph. – Petropavlovsk-Kamchatsky: KamSU, – 2013. – 193 p.
2. Saeedi A., Maghsoudi A., & Rahatabad F.N. Depression Diagnosis and Drug Response Prediction via Recurrent Neural Networks and Transformers Utilizing EEG Signals. arXiv preprint – arXiv:2303.06033, 2023.
3. Tasci G., Loh H.W., Barua P.D., Baygin M., Tasci B., Dogan S., ... & Acharya U.R. Automated accurate detection of depression using twin Pascal's triangles lattice pattern with EEG Signals // Knowledge-Based Systems. – 2023. – Vol. 260. – P. 110190.
4. Sharma M., Achuth P.V., Deb D., Puthankattil S.D., & Acharya U.R. An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals // Cognitive Systems Research. – 2018. – Vol. 52. – P. 508–520.
5. Li X., Hu B., Sun S., & Cai H. EEG-based mild depressive detection using feature selection methods and classifiers // Computer methods and programs in biomedicine. – 2016. – Vol. 136. – P. 151–161.
6. Aronson J.K., & Ferner R.E. Biomarkers—a general review // Current protocols in pharmacology. – 2017. – Vol. 76(1). – P. 9–23.
7. Lakhan S.E., Vieira K., & Hamlat E. Biomarkers in psychiatry: drawbacks and potential for misuse // International Archives of Medicine. – 2010. – Vol. 3(1), – P. 1.
8. Gururajan A., Clarke G., Dinan T.G., & Cryan J.F. Molecular biomarkers of depression. Neuroscience & Biobehavioral Reviews. – 2016. – Vol. 64. – P. 101–133.
9. Zhang Y., He Y., Rong L., & Ding Y. A hybrid model for depression detection with transformer and bi-directional long short-term memory // 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). – 2022. – P. 2727–2734.
10. Zheng W., Yan L., & Wang F.Y. Two birds with one stone: Knowledge-embedded temporal convolutional transformer for depression detection and emotion recognition // IEEE Transactions on Affective Computing, 2023.
11. Ilias L., Mouzakitis S., & Askounis D. Calibration of transformer-based models for identifying stress and depression in social media // IEEE Transactions on Computational Social Systems, 2023.
12. Firoz N., Beresteneva O. G., Vladimirovich A.S., Tahsin M.S., & Tafannum, F. Automated text-based depression detection using hybrid ConvLSTM and Bi-LSTM model // 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) – 2023. – P. 734–740.
13. Wani M.A., ELAffendi M.A., Shakil K.A., Imran A.S., & Abd El-Latif A.A. Depression screening in humans with AI and deep learning techniques // IEEE transactions on computational social systems, 2022.
14. Kamal M.S., Northcote A., Chowdhury L., Dey N., Crespo R.G., & Herrera-Viedma E. Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes // IEEE Transactions on Instrumentation and Measurement. – 2021. – Vol. 70. – P. 1–7.
15. Cai H., Gao Y., Sun S., Li N., Tian F., Xiao H., ... & Zhao Q. Modma dataset: A multi-modal open dataset for mental-disorder analysis. arXiv – 2020. arXiv preprint – arXiv:2002.09283.
16. Zhao Y., Liang Z., Du J., Zhang L., Liu C., & Zhao L. Multi-head attention-based long short-term memory for depression detection from speech // Frontiers in Neurorobotics. – 2021. – Vol. 15. – P. 684037.
17. Bouzid A., Almidani A., Zubrikhina M., Kamzanova A., Ilce B.Y., Zholdassova M., ... & Hamoudi R. Integrative bioinformatics and artificial intelligence analyses of transcriptomics data identified genes associated with major depressive disorders including NRG1 // Neurobiology of Stress. – 2023. – Vol. 26. – P. 100555.
18. Qayyum A., Razzak I., Tanveer M., Mazher M., & Alhaqbani B. «High-density electroencephalography and speech signal based deep framework for clinical depression diagnosis» // IEEE/ACM Transactions on Computational Biology and Bioinformatics. – 2023.
19. Hua J., Xiong Z., Lowey J., Suh E., & Dougherty E.R. «Optimal number of features as a function of sample size for various classification rules» // Bioinformatics. – 2005. – Vol. 21(8). – P. 1509–1515.