MDDNet: EEG-based Transformer with Domain Adversarial Learning for Major Depression Disorder Diagnosis

Shaozhe Liu[1] , **Leike An**[2] , **Ziyu Jia**[3,*]
[1]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China
[2]College of Engineering, Peking University, Beijing, China
[3]Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China
liushaozhe@pku.org.cn, anleike721@pku.edu.cn, jia.ziyu@outlook.com

Abstract

Early identification and intervention for depression is crucial in achieving optimal therapeutic outcomes. Despite its prevalence, depression remains insufficiently comprehended in both clinical and research contexts. Recent progress in deep learning strategies, involving convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has facilitated the diagnostic capability of Major Depressive Disorder (MDD). However, CNNs and RNNs are limited in extracting essential brain spatiotemporal data during the classification process of MDD. Moreover, cross-subject variability poses an additional challenge in MDD classification. The proposed MDDNet overcomes the limitations of CNNs and RNNs by incorporating a spatial-temporal transformer architecture that can extract crucial brain spatiotemporal features from EEG signals. Additionally, domain adversarial learning techniques are employed to improve the model's generalization ability across cross-subject variations in EEG data. The model is validated on an MDD diagnosis dataset, where it achieves exceptional classification performance, demonstrating its state-of-the-art diagnostic capability.

1 Introduction

Major depressive disorder is a serious medical condition that has a detrimental impact on an individual's emotions, thoughts, and behaviors. Projections suggest that by 2040, depression will surpass all other causes to become the leading cause of disability worldwide [Reddy, 2013]. Among the different types of depressive disorders, major depression is the most prevalent. It presents as one of the most complex mental health challenges humankind faces today. Its incidence has been steadily rising over the past few decades. Early identification and timely intervention with appropriate treatment options can significantly enhance the chances of recovery, prevent relapses, and lessen the emotional and economic burden of this illness. Analyzing EEG signals is indeed an effective tool in diagnosing and managing major depressive disorder. The signals recorded through the electrodes on the scalp reflect the electrical activity and brain waves of the brain. These signals can provide valuable insight into the communication of different nerve cells and networks within the brain. By analyzing these signals, researchers and clinicians can identify brain wave patterns that are associated with depression and use that information to design appropriate treatment plans. Figure 1 suggests that individuals with MDD have reduced neural activity in certain areas of the brain, which can contribute to symptoms of depression, such as low mood, fatigue, and difficulty concentrating. EEG signals provide a non-invasive way to measure brain activity and can be used to identify biomarkers that can aid in the diagnosis and treatment of MDD [Drevets *et al.*, 1997].

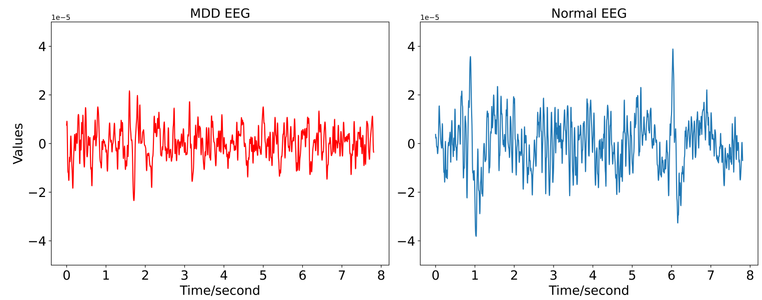

Figure 1: Examples of EEG from normal and depressed participants

Traditional clinical diagnosis methods for MDD using EEG signals have relied on constructing handcrafted features such as amplitude, frequency, and wavelet coefficients [Hosseinifard *et al.*, 2013; Xu *et al.*, 2016]. These methods often require prior knowledge of MDD diagnosis and lack the ability to fully capture the complex spatiotemporal patterns of EEG signals. This can lead to decreased accuracy in MDD classification, particularly when dealing with large and complex data sets. On the other hand, deep learning can be used to automatically learn complex features from EEG signals, enabling the detection of more subtle differences between normal and MDD patients. Deep learning models are designed to learn hierarchies of increasingly abstract features from raw data, allowing them to uncover hidden patterns that may be difficult to detect using handcrafted features. Despite these achievements, the challenges still remain for MDD diagnosis.

The first challenge encountered in MDD diagnosis pertains to the insufficient consideration of spatiotemporal characteristics in EEG signals. EEG signals represent the electrical activity of the brain and are affected by the functional connectivity between brain regions. Therefore, it is important to incorporate the spatial connections between brain regions when analyzing EEG signals. CNN-based approaches [Xu *et al.*, 2022; Acharya *et al.*, 2018] are widely adopted for effective learning of data representations from EEG channels, enabling extraction of spatially invariant and localized features. For example, DeprNet [Seal *et al.*, 2021] uses the ConvNet architecture and 5 stacked CNN layers to capture temporal features. InceptionNet [Lei *et al.*, 2022] uses kernels of different sizes to capture features in different terms and uses channel-wise attention in high-level layers to learn the channel importance. In contrast to conventional approaches, RNN-based methods are adept at capturing the temporal information inherent in EEG trials, which is critical for modeling the temporal dependencies of the EEG signal. Some methods [Yildirim, 2018; Oh *et al.*, 2018] employ LSTM networks to extract spatiotemporal features separately. Also, [Ay *et al.*, 2019] provides sequence learning with LSTM block and enables the representation of input signals through CNN layers. Nonetheless, it is noteworthy that CNNs may suffer from losing some temporal information in the EEG time-series data, thereby limiting their ability to capture a wide range of internal relationships within the signal. Simultaneously, RNNs only take into account the preceding states and the current situation, thereby potentially missing out on the entire temporal sequence of actions embedded in the EEG signals. In recent years, transformer architecture has gained significant attention for its ability to process sequential data, such as natural language or time series data. The transformer models the dependencies between the input sequence and captures long-range dependencies through self-attention mechanisms [Vaswani *et al.*, 2017]. Moreover, the transformer architecture is capable of selecting relevant channels in an unbiased and automated manner [Battaglia *et al.*, 2018]. This could potentially improve the accuracy and efficiency of automated systems for MDD diagnosis and other neurological disorders.

An additional challenge in EEG signals analysis lies in the cross-subject variability problem, also known as inter-subject variability. This refers to the fact that EEG signals can vary significantly across different individuals due to differences in brain morphology, neural connectivity, and other factors. The cross-subject variability problem is commonly recognized as a domain shift issue [Ganin *et al.*, 2016], which means that the statistical properties of the data differ between the training and testing domains. This can result in the model performing poorly on new, unseen data, especially when the model has only been trained on a limited set of subjects. Domain adversarial learning(DAL) can be a useful approach in MDD diagnosis using EEG signals. In EEG-based MDD diagnosis, DAL refers to building models that can generalize well across different patient populations. By capturing the common patterns across multiple domains, DAL can help improve the model's robustness and adaptability to new domains.

This paper presents a new method for MDD diagnosis that effectively tackles the previously mentioned challenges. The proposed method relies on two principal modules: the Transformer encoder module and the Domain adversarial learning module. The Transformer encoder module uses spatial and temporal transformer blocks to capture spatial and temporal information from EEG signals, allowing the model to extract critical high-level features that aid in the diagnosis of MDD. The Domain adversarial learning module addresses the common issue of cross-subject variability in EEG analysis by utilizing domain-labeled data and exhibiting increased generalization abilities toward new subjects while also improving robustness across a range of experimental environments. The primary contributions of this research are the following:

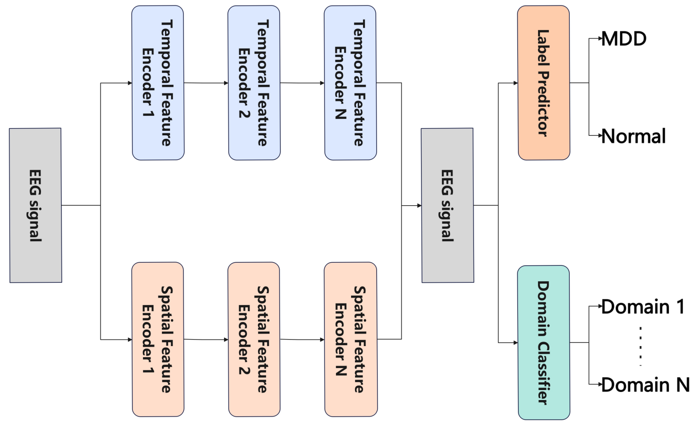

Figure 2: The overview of our method. (1) Encode spatiotemporal features related to MDD diagnosis with independent modules. (2) Integrate a Domain Adversarial Learning module for extracting subject-invariant features.

- To the best of our knowledge, it is the first attempt to integrate the Transformer and Domain Adversarial Learning in a unified framework for MDD diagnosis.
- We design the Transformer encoder to capture the spatial and temporal characteristics of the EEG signals in an effective manner.
- We apply Domain Adversarial Learning to extract subject-invariant features and further improve the generalization of the proposed model.
- The proposed MDDNet achieves the state-of-the-art performance on MDD diagnosis datasets in subject-independent cross-validation.

2 Methodology

The overall framework of MDDNet is depicted in Fig. 2. The framework comprises three primary modules, namely the temporal feature encoder, spatial feature encoder, and Domain Adversarial Learning module. The temporal feature encoder is designed to capture the temporal information contained in EEG signals. This module utilizes a multi-head transformer to track the temporal evolution at different instances in time. The spatial feature encoder uses an SE transformer to capture spatial features and to understand the inter-channel relationships and their relative significance. The Do-

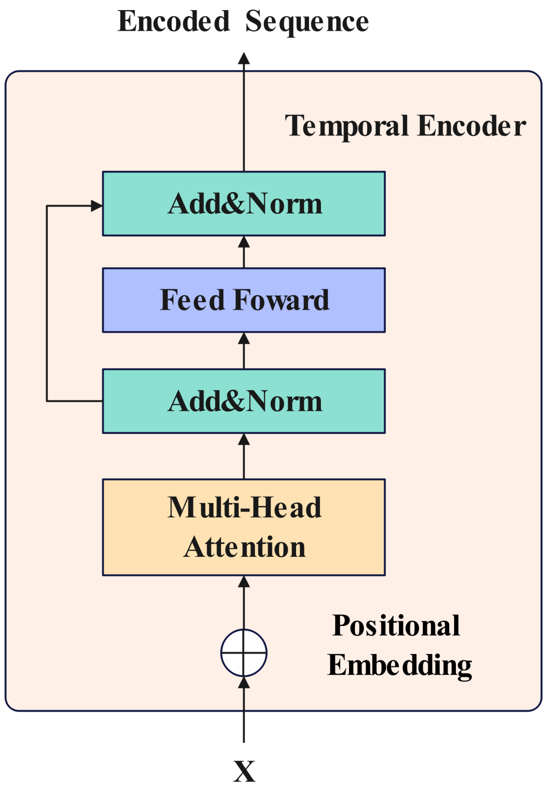

Figure 3: The structure of the temporal encoder module.

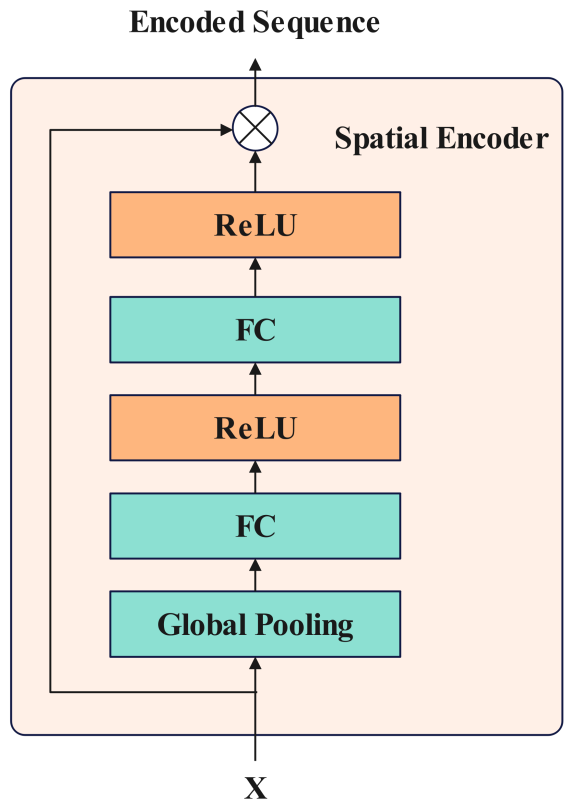

Figure 4: The structure of the spatial encoder module.

main Adversarial Learning module is utilized to enhance the model's generalization abilities.

2.1 Problem definition

We reformulate the MDD diagnosis problem into a binary classification task, wherein a subject's original EEG recording, denoted as $X \in R^{C \times T}$, comprises C channels and T sampling points. The corresponding label for each X, denoted as $y_i \in \{0, 1\}$, signifies whether the subject is healthy or has MDD. Our aim is to learn a classifier f, which takes an EEG sample X as input and produces a binary output $y = f(X)$ that represents the probability of the subject being an MDD patient.

2.2 Temporal Feature Encoder

The EEG is a time-varying recording of brain activity in the form of wave signals. The amplitude of these signals represents the magnitude of brain activity at any given point in time. Multi-head self-attention mechanism allows the encoder to capture temporal dependencies in the EEG signals. This mechanism computes a weighted sum of the input signals, where the weights are learned based on the relationships between different time points in the signals. This process allows the encoder to learn which parts of the signal are most important for the classification task.

The standard transformer encoder module depicted in Figure 3 comprises various multi-head self-attention mechanisms, followed by feed-forward neural networks, which facilitate the learning of complex temporal relationships within the EEG signals. The sequential processing of each layer ensures the gradual accumulation of information, leading to the formation of higher-level features that aid in achieving accurate classification. Lastly, the fully connected layer serves to correlate the learned EEG features with the pre-defined diagnostic categories. The temporal feature encoder can be defined as:

$$Q, K, V = \mathrm{Linear}(X), X \in R^{C \times T} \tag{1}$$

$$\mathrm{Attention_t}(Q, K, V) = \mathrm{softmax}(\frac{Q^T K}{\sqrt{d_t}})V^T \tag{2}$$

$$\mathrm{head_t} = \mathrm{Attention_t}(QW_i{}^Q, KW_i{}^K, VW_i{}^V) \tag{3}$$

$$\mathrm{MultiHead} = \mathrm{Concat}(\mathrm{head_1}, \ldots \mathrm{head_h})\, W^O \tag{4}$$

where $Q^T K \in R^{T \times T}$ represent the attention map in Equation 2. $W_i{}^Q$, $W_i{}^K$, and $W_i{}^V$ denote the linear mapping to obtain the query sequence, key sequence, and value sequence of each head. W^o denotes the head projection matrix. To enhance the non-linear learning ability of the model, two fully-connected layers with residual and the activation function ELU is connected behind the multi-head attention [Vaswani *et al.*, 2017].

2.3 Spatial Feature Encoder

In addition to temporal features, multi-channel EEG signals offers crucial spatial features primarily concerned with the interrelationships and significance of diverse channels. As demonstrated in [Khosla *et al.*, 2022], the inter-channel relationships mirror the functional topology of the brain, aiding in the accurate differentiation of individuals with MDD from healthy controls. To better capture the aforementioned spatial features in EEG signals, we incorporate an additional transformer encoder layer in the spatial feature encoder, as presented in Figure 4. The module incorporates the SE attention layer [Hu *et al.*, 2020] to learn the channel interdependence and the weight assigned to each channel. The concept behind the SE module is generating a vector $V \in R^C$ that characterizes the importance of the feature map originating from each of the C channels within the input tensor. The vector $V = [v_1, v_2, \ldots, v_c]$ denotes the weight ascribed to each feature map, where v_c refers to the importance assigned to the c^{th} feature map. The spatial feature encoder operates as follows:

(1) Initially, the module performs global information encapsulation of each channel through average pooling. The dimension is consequently reduced through the utilization of global average pooling, which is implemented as follows:

$$f_c = \frac{1}{T}\sum_{t=1}^{T} v_c(t) \tag{5}$$

where f_c denotes the encoded feature map of EEG signals from channel c and we obtain the feature vector $F^s = [f_1, f_2, \ldots, f_c]$.

(2) To adaptively assign weights to feature maps, F^s is subsequently excitation by two fully connected parameterized layers. After the adaptive recalibration, we obtain the unioned attention map representing the importance of each channel from both modalities:

$$\mathrm{Attention_s}(X) = \mathrm{Sigmoid}\left(W_s\,\mathrm{Relu}\left(W_s^{'}F^s\right)\right) \tag{6}$$

$$X_T = \sum_{c=1}^{C} \mathrm{Attention_s}\left(X\right) \otimes X \tag{7}$$

where W_s, $W_s^{'}$ are employed to impart varying levels of significance to channels derived from distinct modalities. To achieve this objective, channel-wise multiplication is conducted to produce the attention map, which offers insights into the interrelationship among individual EEG channels. This facilitates an emphasis by our model on spatial features that have shown to be efficacious while filtering out ineffective ones.

Having captured both temporal and spatial features, it is of critical importance to consider their respective contributions to the diagnosis of MDD. As a result, we combine the encoded temporal features with the encoded spatial features in order to generate a comprehensive encoding representation.

2.4 Domain Adversarial Learning Module

Our model incorporates a transfer learning method termed DAL [Ma *et al.*, 2019; Ganin *et al.*, 2016]. In DAL, an adversarial relationship is established between the tasks of discriminating MDD patients and indiscriminating domains, facilitating the extraction of domain-invariant features.

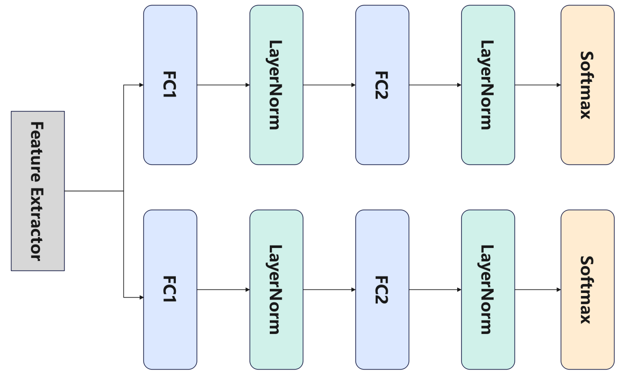

Figure 5: The architecture of DAL module. Specifically, the label predictor and domain classifier components of DAL architecture are equipped with two fully connected layers, both incorporating the softmax function.

The DAL technique is composed of three primary elements, namely a feature extractor ($\mathcal{G}_f$), a label predictor ($\mathcal{G}_l$), and a domain classifier ($\mathcal{G}_d$). Our model employs the Temporal and Spatial feature Encoder as the feature extractor $\mathcal{G}_f$, which operates by converting input data into a feature space that is insensitive to domain shifts. The feature extractor can be defined as:

$$\mathcal{G}_f\left(X;\theta_f\right) = \mathrm{Attention_t}\left(X\right) \oplus \mathrm{Attention_s}\left(X\right) \tag{8}$$

$$F^{st} = \mathcal{G}_f\left(X;\theta_f\right) \tag{9}$$

where X refers to the raw EEG data input, θ_f represents the model's trainable parameters, and F^{st} corresponds to the feature vector embedded with both temporal and spatial information.

The extracted sequence is subsequently inputted into both the label predictor and the domain classifier, both of which are defined as:

$$\hat{y_i} = \mathrm{Softmax}\,\mathcal{G}_l\left({F^{st}}_i;\theta_y\right) \tag{10}$$

$$\hat{d_i} = \mathrm{Softmax}(\mathcal{G}_d({F^{st}}_i;\theta_d)) \tag{11}$$

where ${F^{st}}_i$ refers to the embedded sequence originating from the i^{th} sample. The predicted results of $\mathcal{G}_l$ and $\mathcal{G}_d$ denoted as $\hat{y_i}$ and $\hat{d_i}$ respectively are both multi-class classifiers. The loss functions can be defined as follows:

$$\mathcal{L}_y = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C_y} y_{i,j} log\hat{y_{i,j}} \tag{12}$$

$$\mathcal{L}_d = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C_d} d_{i,j} log\hat{d_{i,j}} \tag{13}$$

where $\mathcal{L}_y$ is the cross entropy loss function of the MDD diagnose task, N denotes the number of samples, C_y and C_d denote the number of classes, and the number of domains. y is the true label and d is the true domain.

In order to establish an adversarial correlation between the label predictor and the domain classifier, the Gradient Reversal Layer (GRL) approach proposed by [Ganin *et al.*, 2016] is utilized in our study. This layer is placed between the feature extractor $\mathcal{G}_f$ and the domain classifier $\mathcal{G}_d$ to enable the amalgamation of feature learning and domain generalization within a cohesive framework that supports backpropagation algorithms. Specifically, we design the overall loss function and optimization routine of our model in the subsequent manner:

$$\mathcal{L}_{DG} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C_y} y_{i,j} log \hat{y_{i,j}} + \lambda \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C_d} d_{i,j} log \hat{d_{i,j}} \tag{14}$$

$$\left(\hat{\theta_f}, \hat{\theta_y}\right) = \arg\min_{\theta_f, \theta_y} \mathcal{L}\left(\theta_f, \theta_y, \hat{\theta_d}\right) \tag{15}$$

$$\left(\hat{\theta_d}\right) = \arg\max_{\theta_d} \mathcal{L}\left(\hat{\theta_f}, \hat{\theta_y}, \theta_d\right) \tag{16}$$

where $\mathcal{L}_{DG}$ is the total loss function of the proposed model, λ is a hyperparameter for the GRL.

The optimization of the loss function can facilitate the realization of the feature extractor $\mathcal{G}_f$ goal, which is to identify a subject-invariant feature space. In order to accomplish this objective, the parameters θ_y and θ_d are leveraged to minimize the losses of $\mathcal{G}_y$ and $\mathcal{G}_d$, respectively. Notably, our optimization strategy involves the simultaneous minimization of the loss of $\mathcal{G}_y$ and maximization of the loss of $\mathcal{G}_d$, with θ_y being optimized to achieve the aforementioned goal.

3 Experiment

3.1 Dataset

The MDD diagnosis dataset in [Mumtaz, 2016] is selected as the experimental dataset for this study. The dataset contains EEG recordings of 30 MDD patients and 30 healthy subjects in a resting state with their eyes open or closed. A standard 10-20 system is used to place 19 EEG electrodes, namely Fp1, F3, C3, P3, O1, F7, T3, T5, Fz, Fp2, F4, C4, P4, O2, F8, T4, T6, Cz, Pz. The EEG signals are sampled at a rate of 256 Hz and filtered using a 0.5 Hz high-pass and 80 Hz low-pass filter to eliminate unwanted noise. The recordings are divided into 4-second-long samples with a sliding window stride of 1 second, resulting in a total of 16,024 EEG samples, including 7,542 positive and 8,482 negative samples.

3.2 Baseline and Experiment Setting

We compare our model with the following approaches:

- **EEGNet** [Lawhern *et al.*, 2018]: A classical compact fully convolutional network with depth-wise and separable convolutions for EEG classification using CNN.
- **ConvNet** [Schirrmeister *et al.*, 2017]: The ConvNet employs four convolution-pooling blocks that give it a strong ability to extract features from raw EEG signals.

Parameters	Value	Parameters	Value
Learning Rate	0.001	Batch Size	50
Optimizer	Adam	Epochs	100
Dropout Rate	0.2	λ	$2(1+e^{-10})^{-1}$

Table 1: Training Settings

Model	ACC(%)	Recall(%)	Precision(%)
EEGNet	85.0	87.3	85.2
ConvNet	83.7	87.6	82.8
MMCNN	89.7	91.3	89.7
CNN+LSTM	86.2	90.5	87.2
MDDNet	92.3	94.1	91.4

Table 2: Overall performance comparison with other MDD diagnosis methods.

- **MMCNN** [Jia *et al.*, 2021]: An end-to-end deep learning model consists of 5 parallel EEG Inception Networks. Each EIN consists of an EEG Inception block, a Residual block, and a Squeeze and Excitation block.
- **CNN+LSTM** [Sharma *et al.*, 2021]: The model utilizes a 1D-CNN and Bi-LSTM to learn temporal features of short-period and long-period, respectively.

In order to prevent data leakage, a 10-fold subject-independent cross-validation methodology is employed for evaluating all models. Specifically, for each fold of the cross-validation, EEG samples of three MDD patients and three healthy control subjects are used as the test set, while samples from the remaining subjects are used for training. Notably, all samples pertaining to one subject were included either solely in the training set or solely in the test set, to ensure subject independence.

We implement our model using PyTorch and the source code will be public on GitHub. Our method is implemented using Python 3.7 and the PyTorch library on a Geforce 2080Ti GPU. The training settings in our experiments are given in Table 1.

We use common binary classification metrics including accuracy, recall, and precision for performance evaluation.

3.3 Comparison and Analysis of Experiment Results

Classification Performance and Hyperparameters Tuning

The present study reports a comparative analysis of the proposed MDDNet versus several baseline models, as shown in Table 2. Our findings indicate that the MDDNet outperforms all other baseline models, achieving an accuracy of 92.3%, Recall of 94.1%, and Precision of 91.4%. Specifically, the ConvNet and EEGNet models, which rely on CNN-based deep networks for EEG classification, demonstrate strong performance due to their adept handling of convolutions of various scales to extract spatial and temporal features. However, their limitation in capturing global dependencies hinders their overall efficacy. The CNN+LSTM model extends

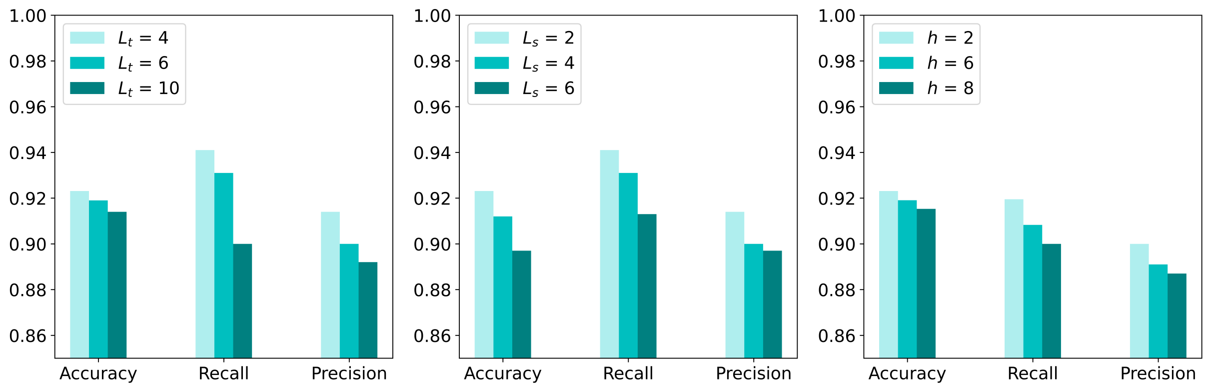

Figure 6: Hyperparameters tuning process. Three figures show the tuning of the number of temporal encoders L_t, the number of spatial encoders L_s, and the number of attention heads h.

the approach by introducing LSTM to account for temporal dependencies. However, it still falls short in capturing the spatiotemporal information critical to MDD diagnosis. Overall, our findings suggest that the proposed MDDNet model is superior to all the baseline models in fully capturing and utilizing spatiotemporal information essential for MDD diagnosis.

In Figure 6, the MDDNet underwent fine-tuning on critical hyperparameters, namely, the quantity of temporal and spatial encoders and temporal attention heads. Despite these variations, the performance metrics of Accuracy, Recall, and Precision persist at an elevated level above 89%, with a nominal decline of no more than 3%. The obtained outcomes indicate the model's insensitivity towards variations in certain key hyperparameters.

Ablation Experiments

Ablation experiments are undertaken to demonstrate the efficacy of the amalgamation of heterogeneous modules within the MDDNet framework. Models used for comparison in the ablation experiment are as follows:

- **Variant 1**: Spatial encoder module + Temporal encoder module.
- **Variant 2**: Spatial encoder module + DAL.
- **Variant 2**: Temporal encoder module + DAL.

According to the results presented in Figure 7, it can be inferred that the MDDNet outperforms its counterpart variants. This can be attributed to the inclusion of a spatial and temporal feature encoder within the MDDNet architecture, enabling it to extract more informative features from varying MDD diagnostic modalities, thus validating its effectiveness. Additionally, the performance of the model is observed to improve with the incorporation of the DAL module, thereby corroborating the efficacy of the DAL approach in extracting subject-invariant features.

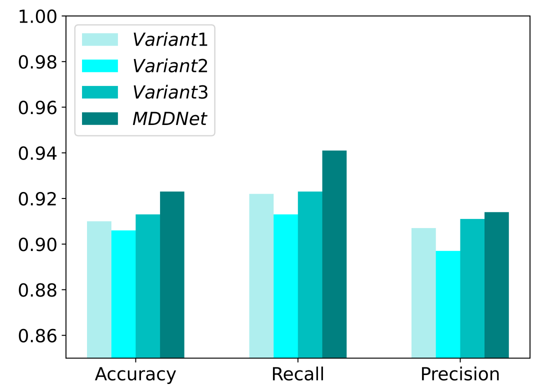

Figure 7: Ablation experiments results.

4 Conclusion

The present study presents a novel MDD diagnostic approach named MMDNet, which incorporates effective techniques for modeling the spatiotemporal dynamics of EEG signals while also accounting for inter-subject variability. Specifically, an EEG-based transformer module is introduced to encode the most pertinent spatial-temporal features for MDD diagnosis. To extract subject-invariant features, a unified framework is developed by amalgamating the DAL approach and the Transformer module. Experimental evaluations conducted on our proposed MDDNet model demonstrate its superior performance, surpassing the state-of-the-art MDD diagnostic approaches. Moreover, our proposed approach provides a generalizable framework for the analysis of multivariate physiological time-series data.

Acknowledgments

This project is funded by China Postdoctoral Science Foundation through Grant No. 2023M733738.

References

[Acharya *et al.*, 2018] U. R. Acharya, Shu L. Oh, Yuki Hagiwara, Jen H. Tan, and Hojjat Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in biology and medicine*, 100:270–278, 2018.

[Ay *et al.*, 2019] Betul Ay, Ozal Yildirim, Muhammed Talo, Ulas B. Baloglu, Galip Aydin, Subha D. Puthankattil, and

U. R. Acharya. Automated depression detection using deep representation and sequence learning with eeg signals. *Journal of medical systems*, 43(7):205–12, 2019.

[Battaglia *et al.*, 2018] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. 2018.

[Drevets *et al.*, 1997] Wayne C. Drevets, Joseph L. Price, Joseph R. Simpson, Richard D. Todd, Theodore Reich, Michael Vannier, and Marcus E. Raichle. Subgenual prefrontal cortex abnormalities in mood disorders. *Nature (London)*, 386(6627):824–827, 1997.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Franois Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[Hosseinifard *et al.*, 2013] B. Hosseinifard, M. H. Moradi, and R. Rostami. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal. *Elsevier*, (3), 2013.

[Hu *et al.*, 2020] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(8):2011–2023, 2020.

[Jia *et al.*, 2021] Ziyu Jia, Youfang Lin, Jing Wang, Kaixin Yang, Tianhang Liu, and Xinwang Zhang. *MMCNN: A Multi-branch Multi-scale Convolutional Neural Network for Motor Imagery Classification*. 2021.

[Khosla *et al.*, 2022] Ashima Khosla, Padmavati Khandnor, and Trilok Chand. Automated diagnosis of depression from eeg signals using traditional and deep learning approaches: A comparative analysis. *Biocybernetics and biomedical engineering*, 42(1):108–142, 2022.

[Lawhern *et al.*, 2018] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: A compact convolutional network for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013.1–056013.17, 2018.

[Lei *et al.*, 2022] Y. Lei, A. N. Belkacem, X. Wang, S. Sha, C. Wang, and C. Chen. A convolutional neural network-based diagnostic method using resting-state electroencephalograph signals for major depressive and bipolar disorders. *Biomedical Signal Processing and Control*, 72:103370–, 2022.

[Ma *et al.*, 2019] Bo-Qun Ma, He Li, Yun Luo, and Bao-Liang Lu. Depersonalized cross-subject vigilance estimation with adversarial domain generalization. pages 1–8. IEEE, 2019.

[Mumtaz, 2016] Wajid Mumtaz. MDD Patients and Healthy Controls EEG Data (New). 2016.

[Oh *et al.*, 2018] Shu L. Oh, Eddie Y. K. Ng, Ru S. Tan, and U. R. Acharya. Automated diagnosis of arrhythmia using combination of cnn and lstm techniques with variable length heart beats. *Computers in biology and medicine*, 102:278–287, 2018.

[Reddy, 2013] Gvp Reddy. Depression – the global crisis. *Indian Journal of Psychological Medicine,34,3(2013-01-14)*, 34(3):201, 2013.

[Schirrmeister *et al.*, 2017] R. T. Schirrmeister, J. T. Springenberg, Ldj Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. 2017.

[Seal *et al.*, 2021] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, and O. Krejcar. Deprnet: A deep convolution neural network framework for detecting depression using eeg. *IEEE Transactions on Instrumentation and Measurement*, PP(99):1–1, 2021.

[Sharma *et al.*, 2021] Geetanjali Sharma, Abhishek Parashar, and Amit M. Joshi. Dephnn: A novel hybrid neural network for electroencephalogram (eeg)-based screening of depression. *Biomedical signal processing and control*, 66:102393, 2021.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. volume 2017-, pages 5999–6009, 2017.

[Xu *et al.*, 2016] Fangzhou Xu, Weidong Zhou, Yilin Zhen, Qi Yuan, and Qi Wu. Using fractal and local binary pattern features for classification of ecog motor imagery tasks obtained from the right brain hemisphere. *International journal of neural systems*, 26(6):1650022, 2016.

[Xu *et al.*, 2022] Min Xu, Chen Tang, Nian Hong, and Zhenkun Lei. Mdd-net: A generalized network for speckle removal with structure protection and shape preservation for various kinds of espi fringe patterns. *Optics and lasers in engineering*, 154:107017, 2022.

[Yildirim, 2018] Ozal Yildirim. A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. *Computers in biology and medicine*, 96:189–202, 2018.