IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 March 2023, accepted 17 April 2023, date of publication 26 April 2023, date of current version 2 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270426

RESEARCH ARTICLE

An End-to-End Deep Learning Model for EEG-Based Major Depressive Disorder Classification

MIN XIA[1], YANGSONG ZHANG[1,2], YIHAN WU[1], AND XIUZHU WANG[1]

[1]School of Computer Science and Technology, Laboratory for Brain Science and Medical Artificial Intelligence, Southwest University of Science and Technology, Mianyang 621010, China

[2]MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu 610054, China

Corresponding author: Yangsong Zhang (zhangysacademy@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62076209, and in part by the House-Level Project of Mianyang Central Hospital under Grant 2021YJ001.

ABSTRACT Major depressive disorder (MDD) is a prevalent mental illness associated with abnormalities in structural and functional brain connectivity, which has become a global public health problem. Early diagnosis is important and challenging for the treatment of MDD. Previous studies have proposed MDD classification methods based on brain connectivity features through functional connectivity and effective connectivity measures. However, artificial selection algorithms to compute brain connectivity features require prior knowledge and experience. Given that the representation learning capabilities of deep learning models and the ability to capture correlations between data of self-attention mechanism, this study proposed an end-to-end integrated DL model for classifying MDD patients and healthy controls based on the resting-state electroencephalography (EEG) data. The model first automatically learned the potential connectivity relationships among EEG channels through a multi-head self-attention mechanism, and then extracted higher-level features through a parallel two-branch convolution neural network module, and finally completed the classification through a fully connected layer. The leave-one-subject-out cross-validation method was utilized to evaluate the effectiveness of the proposed model on a publicly available EEG dataset. Ultimately, the average classification accuracy of the proposed model reached 91.06%, which was better than the comparison methods. The experimental results indicate that this study may provide a novel approach for brain connectivity modeling of MDD detection.

INDEX TERMS Major depression disorder, electroencephalography, convolution neural network, brain connectivity, self-attention mechanism.

I. INTRODUCTION

Major depressive disorder (MDD) is a widespread global psychiatric illness, affecting 322 million people worldwide assessed by the World Health Organization (WHO), which usually manifests by persistent downbeat emotion, loss of interest, and impaired cognitive function [1], [2]. MDD is also characterized by high rates of disability, suicide, and relapse [3]. Early treatment and intervention of depression patients at the onset can prevent their condition from worsening and become a refractory disease. However, traditional MDD detection methods are based on various scales [4], [5], [6], [7], and their diagnosis results may be influenced by some subjective factors, such as the degree of patient cooperation and the experience of doctors, which leads to a low recognition rate. Therefore, we sought a more objective and accurate way to identify MDD.

Advances in brain imaging technology have raised neuro-computational models that use neurological signals to detect psychiatric disorders, such as MDD. Due to its advantages of low cost, high temporal resolution, and non-invasiveness [8], EEG has emerged as a valid and reliable quantitative analysis

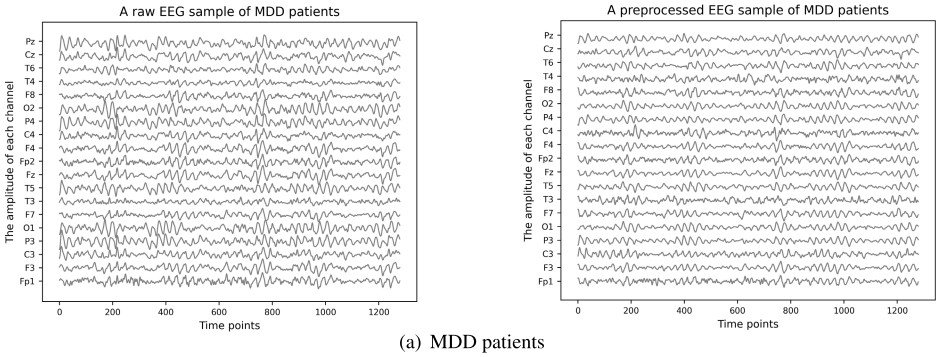
The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

tool for detecting MDD. However, EEG signals are complex and non-stationary, it is difficult to distinguish them by visual observation. Therefore, various EEG-based machine learning (ML) algorithms are proposed for MDD classification. These findings indicate that ML methods, including traditional ML methods and deep learning (DL) methods, have the ability to effectively and correctly detect MDD. For example, Zhu et al. employed feature fusion and hidden layer fusion methods to integrate EEG signals and eye movement signals, and then utilized the support vector machine (SVM) to classify MDD via the leave-one-subject-out cross-validation (LOSOCV) method [9]. The algorithm finally obtained the classification accuracy of 83.42%. Saeedi et al. extracted linear and nonlinear characteristics of EEG signals by frequency band division and two entropy measurement methods, and then applied the genetic algorithm to screen significant features, which were input into an enhanced k-nearest neighbors (E-KNN) classifier [10]. The model reached 98.44% classification accuracy by using the 10-fold CV method. Acharya et al. introduced a 13-layer CNN model for classifying the EEG data of 15 healthy controls (HCs) and 15 MDD subjects [11], and the DL model ultimately achieved the classification accuracies of 93.5% and 96% for the left and right hemispheres, respectively. Seal et al. presented a CNN model named DeprNet for detecting MDD [12]. The model obtained the classification accuracies of 91.4% and 99.37% on the subject-wise split and the record-wise split CV methods, respectively. Song et al. developed a DL model consisting of CNN, long short-term memory (LSTM), and domain discriminator, named LSDD-EEGNet, for MDD classification [13]. The EEG data of 40 depression patients and 40 HCs were divided into training set and test set in the ratio of 7:3 for validating the classification performance of the model, which reached the accuracy of 94.69%. Liu et al. applied the state-of-the-art DL model, named EEGNet [14], to the MODMA dataset containing 24 depression patients and 29 HCs for the recognition of depression [15]. The LOSOCV method was adopted to train the model, which obtained the classification accuracy of 90.98%. By reviewing previous studies, it is found that traditional ML methods require to extract hand-crafted features, which rely on the extensive prior knowledge of the researcher, while DL models can automatically extract high-dimensional nonlinear features. In contrast to the above-mentioned methods that focus on EEG time series, several researchers conducted the classification of MDD based on brain connectivity.

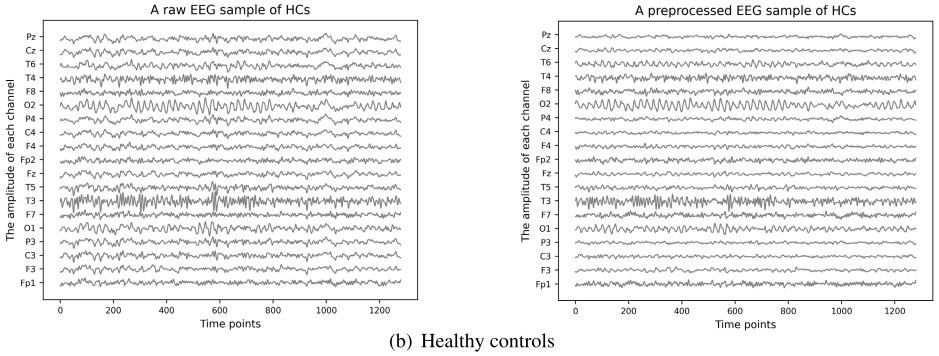
Neuroimaging studies have revealed abnormalities in brain structure and function of MDD [16], [17]. Alterations in brain connectivity and function may affect emotional processes directly related to depressive symptoms and may also negatively affect cognitive function [18]. Consequently, research approaches by modeling and estimating brain connectivity of MDD have been progressively proposed in recent years. For instance, Zhang et al. presented a brain functional network architecture to analyze and diagnose MDD based on the resting-state EEG data of 24 MDD patients and 29 HCs [19]. Network metrics based on the small world and connectivity matrices calculated by the phase lag index of different sub-bands were employed as the extracted features, which were input into the SVM classifier for the classification of MDD. The 10-fold CV method was adopted to evaluate the classification performance of the model, which achieved the best accuracy of 93.31% in the theta frequency band. Movahed et al. presented the SVM with RBF kernel (RBFSVM) method to detect MDD [20]. The sequential backward feature selection (SBFS) algorithm was applied to screen useful features extracted by functional connectivity (FC) measure and other methods. The model obtained an average accuracy of 99% with the 10-fold CV based on the EEG data of 34 MDD patients and 30 HCs. Duan et al. introduced a method that fused inter-hemispheric asymmetry and cross-correlation with the EEG signals [21]. The CNN model with the 10-fold CV method obtained an accuracy of 94.13% on the EEG dataset with 16 MDD patients and 16 HCs. Saeedi et al. utilized a CNN and LSTM fusion model for automatically diagnosing MDD based on the EEG data of 34 MDD patients and 30 HCs [22]. The input data of the model were effective connectivity (EC) features calculated by the direct directed transfer function and generalized partial directed coherence methods. The model was trained with 70% data and tested on 30% data from the EEG dataset, and ultimately achieved the classification accuracy of 99.24%. These studies demonstrate the validity and feasibility of using brain connectivity for the classification of MDD, but there are some limitations existing. Related algorithms to obtain the FC or EC matrices were different among the studies, and the selection of those algorithms was based on the prior knowledge of the researcher. Moreover, the procedure of calculating brain connectivity features was separate from that of classifier training, which may not achieve the best performance. The end-to-end data-driven DL method could provide a new avenue for MDD classification from the brain connectivity perspective.

In this study, we proposed an end-to-end integrated DL architecture to improve the classification accuracy between MDD patients and HCs. The data-driven model could automatically explore the potential connectivity relationships among 19 EEG channels through the multi-head self-attention mechanism, thus avoiding the selection and tedious computation of traditional FC and EC measures. This operation converted the EEG time series signals into connectivity matrices among EEG channels. Subsequently, the high-dimensional features were extracted from different dimensions of the connectivity matrices by the parallel two-branch CNN blocks. Lastly, a dense layer with softmax activation function was employed for MDD patients and HCs classification. The classification performance of the proposed model was verified on the resting-state EEG dataset with 30 MDD patients and 28 HCs using the LOSOCV strategy.

The remainder of this paper is organized as follows. The second section describes the dataset and its preprocessing, the proposed model, and comparison methods. The third section analyzes and discusses the experimental results. The last section gives a summary of this study.



(a) MDD patients



(b) Healthy controls

FIGURE 1. The EEG samples of MDD patients and HCs. The first row shows the raw and preprocessed 19-channel EEG sample of MDD patients, and the second row demonstrates the raw and preprocessed 19-channel EEG signals of HCs.

II. MATERIALS AND METHODS

A. PARTICIPANTS

The publicly available EEG dataset provided by Mumtaz et al. was adopted to evaluate the proposed DL method in this paper [23]. 34 MDD patients(the mean age of 40.3 ± 12.9) and 30 HCs (the mean age of 38.3 ± 15.6) took part in the experiments to acquire EEG data. They were informed of the experimental procedures and signed the consent forms. The MDD patients met the diagnostic criteria with DSM-IV [24], and the HCs were evaluated via psychiatric condition examinations. The experimental setup was approved by the ethics committee of Hospital Universiti Sains Malaysia (HUSM) [23]. Only EEG data from 58 subjects, including 30 MDD patients and 28 HCs, were available for download on the website, which was used for the following experiments.

B. DATA ACQUISITION AND PREPROCESSING

The dataset contained 5-min eye-closed resting-state EEG signals for each subject [23]. According to the international 10-20 system, 19 electrodes (Fp1, Fp2, F7, F8, F3, Fz, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, T5, T6, O1, O2) distributed in different brain regions were adopted to acquire EEG signals, and linked-ear reference was used during EEG data acquisition. The sampling rate of EEG signals was 256 Hz. The notch filter was employed to remove 50 Hz power line noise, and the band-pass filter from 0.1 - 70 Hz was applied for removing artifacts.

Based on the acquired EEG data, a 5 s non-overlapping sliding window was used to divide the raw EEG data of each subject into multiple samples. Each EEG sample contained 1,280 (5 s × 256 Hz) sampling points, where some samples were discarded if there were sampling points with amplitude less than −100 uV or greater than 100 uV [25]. Then, the operations of re-referencing to average reference and Z-score transformation were successively carried out for the remaining samples. Eventually, 57 subjects with a total of 3,163 samples were utilized to verify the proposed method. The EEG samples of MDD patients and HCs are displayed in FIGURE 1. Moreover, because the brain will produce different levels of brain waves in different states [26], the EEG signals were filtered into five sub-bands, namely delta band (0.5 - 4 Hz), theta band (4 - 8 Hz), alpha band (8 - 13 Hz), beta band (13 - 30 Hz), and gamma band (30 - 70 Hz), which were utilized to further study the classification performance of the proposed method on these frequency bands.

C. THE ARCHITECTURE OF PROPOSED MODEL

The architecture of the proposed DL model for the MDD classification is illustrated in FIGURE 2. The model consisted of three modules, where the first module learned the potential connectivity relationships among EEG channels by the multi-head self-attention layer, the second module exploited parallel two-branch CNN blocks to further extract high-level features, and the last module concatenated multiple features for MDD classification via the dense layer with a softmax activation function. The detailed parameter settings of the proposed model are listed in TABLE 1.

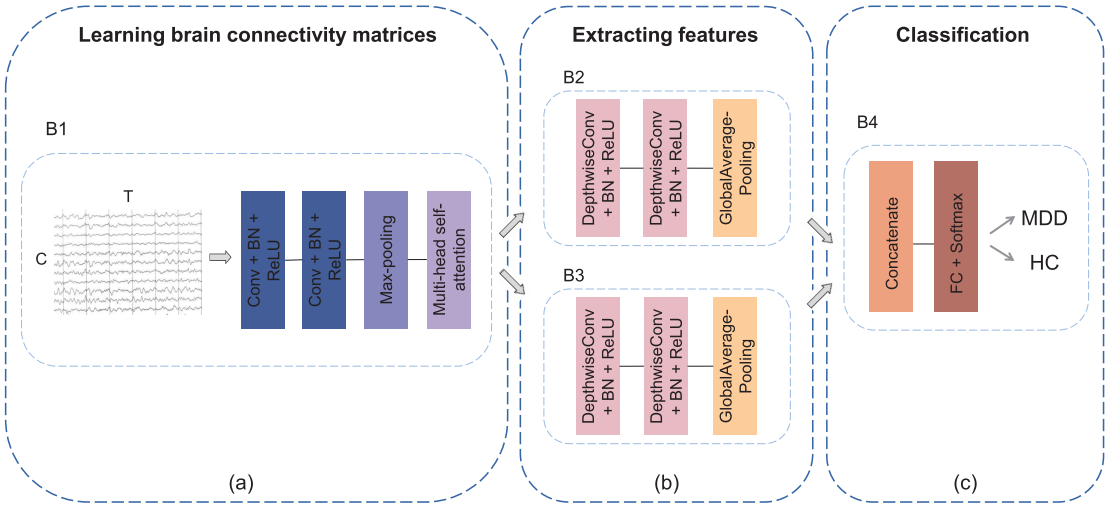


FIGURE 2. The architecture diagram of the proposed model. It consists of three modules: (a) Learning the potential connectivity relationships, (b) Extracting features, and (c) Classification. Where *C* and *T* indicate the numbers of EEG channels and time points, respectively.

1) LEARNING POTENTIAL CONNECTIVITY RELATIONSHIPS AMONG EEG CHANNELS

This subsection describes how the first block (B1 module) automatically learned the potential connectivity relationships among 19 EEG channels, which are represented as the connectivity matrices, and the structural design of the B1 module is shown in FIGURE 2 (a). First, the preprocessed EEG data with the size of (C, T) were fed into the 2D convolutional layer with a convolutional kernel size of $(1, fs/2)$ (fs represents the sampling rate) to perform preliminary feature extraction on the temporal dimension from the EEG data. Due to the similarity of adjacent sampling points of resting-state EEG data, a larger convolutional kernel was employed to obtain features with large differences. Subsequently, the batch normalization (BN) layer and rectified linear unit (ReLU) activation function layer were used, and the extracted multiple feature maps were fed into the convolutional layer with a convolutional kernel size of (1, 1), which was employed for the weighted fusion of the feature maps and reduced the number of feature maps. Similarly, the convolutional layer was followed by the BN layer and ReLU activation function layer. Next, the max-pooling layer was served to reduce the feature dimensions by local down-sampling. Eventually, the feature maps after a series of processing were separately fed into the multi-head self-attention layer to explore the connectivity relationships among 19 EEG channels to form the connectivity matrices. In this paper, the method of obtaining the connectivity matrices was different from the traditional method of calculating the brain connectivity networks. This study adopted a data-driven method, which does not need to manually calculate the connectivity matrices through some FC or EC measure, but directly used the B1 module to automatically mine the potential connectivity relationships among different EEG channels using the multi-head self-attention layer. Because the attention mechanism has the ability to allow the neural network to focus only on part of its input, the 19 channels of EEG signals were set as the object of attention in this study. Besides, the multi-head self-attention improves the performance of the self-attention layer, because it not only has the ability to focus on different locations but also gives the self-attention layer multiple "representation sub-spaces" [27]. Therefore, the multi-head self-attention layer was employed to obtain multiple different connectivity matrices. For the process of the multi-head self-attention layer learning the connectivity matrices, an example of the operation on a feature map is shown in FIGURE 3. The feature map $F \in \mathbb{R}^{C\times S}$ was first weighted by the trainable parameter matrices $W_Q \in \mathbb{R}^{S\times S}$ and $W_K \in \mathbb{R}^{S\times S}$, and then the weighted feature map was reshaped according to the number of multi-head N, where the two reshaped matrices were denoted by $Q \in \mathbb{R}^{N\times C\times D}$ and $K \in \mathbb{R}^{N\times C\times D}$, respectively. Eventually, the dot product of Q and K^T was divided by $\sqrt{D}$ to obtain the weight matrices $W \in \mathbb{R}^{N\times C\times C}$, which can be regarded as the connectivity matrices that reflect the connectivity patterns among different EEG channels. The connectivity matrices are calculated as follows:

$$Q = Reshape(W_Q \cdot F) \tag{1}$$

$$K = Reshape(W_K \cdot F) \tag{2}$$

$$W = \frac{Q \cdot K^T}{\sqrt{D}} \tag{3}$$

where T represents the transpose operation, and $\sqrt{D}$ is the scaling factor, which acts as a modulation so that the inner product was not too large.

2) EXTRACTING HIGH-LEVEL FEATURES

The second module was designed with the parallel two-branch CNN architecture for extracting high-level features from two dimensions of the connectivity matrices, and the detailed architecture is shown in FIGURE 2 (b), which contained the B2 sub-module and B3 sub-module. Each sub-module included two depth-wise convolutional layers,

TABLE 1. Detailed parameter settings of the proposed model.

Block	Layer	Filters	Activation	Options	Output	Parameters
	Input				(None, 19, 1280, 32)	0
	Conv2D	32, (1, fs//2)		padding = 'same'	(None, 19, 1280, 32)	4128
	BatchNormalization				(None, 19, 1280, 32)	128
	Activation		ReLU		(None, 19, 1280, 32)	0
B1	Conv2D	16, (1, 1)		padding = 'same'	(None, 19, 1280, 16)	528
	BatchNormalization				(None, 19, 1280, 16)	64
	Activation		ReLU		(None, 19, 1280, 16)	0
	MaxPooling2D	(1, 4)			(None, 19, 320, 16)	0
	Multi-head self-attention			multi-head = 4	(None, 19, 19, 64)	3276800
	DepthwiseConv2D	(1, 3)		depth = 2, padding = 'same'	(None, 19, 19, 128)	512
	BatchNormalization				(None, 19, 19, 128)	512
	Activation		ReLU		(None, 19, 19, 128)	0
B2	DepthwiseConv2D	(1, 19)		depth = 2	(None, 19, 1, 256)	5120
	BatchNormalization				(None, 19, 1, 256)	1024
	Activation		ReLU		(None, 19, 1, 256)	0
	GlobalAveragePooling2D				(None, 256)	0
	DepthwiseConv2D	(3, 1)		depth = 2, padding = 'same'	(None, 19, 19, 128)	512
	BatchNormalization				(None, 19, 19, 128)	512
	Activation		ReLU		(None, 19, 19, 128)	0
B3	DepthwiseConv2D	(19, 1)		depth = 2	(None, 1, 19, 256)	5120
	BatchNormalization				(None, 1, 19, 256)	1024
	Activation		ReLU		(None, 1, 19, 256)	0
	GlobalAveragePooling2D				(None, 256)	0
classifier	Concatenate (B2, B3)				(None, 512)	0
	Dense	2	Softmax		(None, 2)	1026
Total params: 3,297,010 Trainable params: 3,295,378 Non-trainable params: 1632						

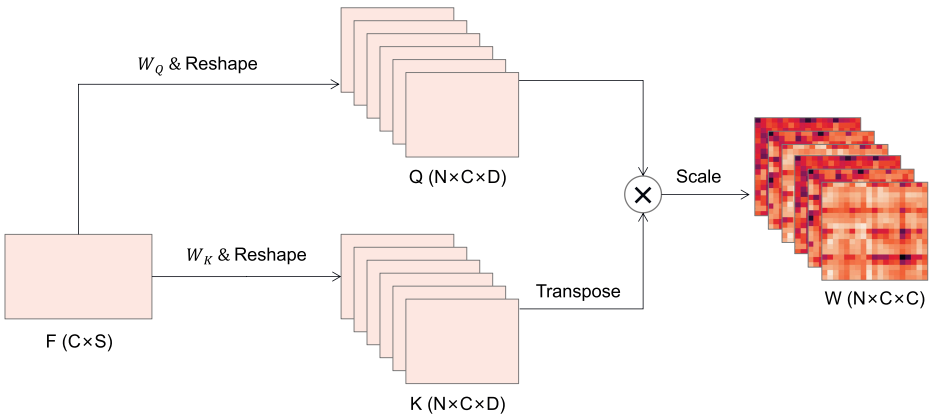


FIGURE 3. The operation process for learning connectivity matrices via a multi-head self-attention layer. *F* and *W* denote the input feature map and the output connectivity matrices, respectively; *Q* and *K* are the matrices weighted by W_Q and W_K, and reshaped according to the number of multi-head *N*, respectively; *C* and *S* represent EEG channels and features, respectively. Note that $S = N * D$.

each of which was followed by a BN layer and a ReLU activation function layer. The number of parameters in the B2 sub-module and B3 sub-module can be greatly reduced by employing a depth-wise convolutional layer instead of a common 2D convolutional layer to separately perform the convolution operation for each feature map. Specifically, in the B2 sub-module and B3 sub-module, the first depth-wise convolutional layer with the kernel size of (1, 3) and (3, 1) was applied to initially extract features from the row/column dimension of the input connectivity matrices, and the second depth-wise convolutional layer with the kernel size of (1, 19) and (19, 1) was adopted to extract features of higher dimensions. At last, the extracted feature maps of the two branches were flattened separately by a global average pooling (GAP) layer. Using the GAP layer instead of the fully connected layer can not only reduce the parameters but also avoid over-fitting [28].

3) CLASSIFICATION

As shown in FIGURE 2 (c), the features extracted from the blocks of B2 and B3 were first concatenated together, and

then input into the dense layer with a softmax activation function for binary classification of MDD patients and HCs.

D. CALCULATION OF FUNCTIONAL AND EFFECTIVE CONNECTIVITY

Using EEG signals to calculate FC, the results can reflect the strong and weak relationships between different brain regions [29], while the calculation results of EC indicate the causal influence of one brain region on another, thus it has directional information flow between cortical regions [30]. In this paper, three types of brain connectivity measures were chosen to compute the different connectivity relationships among EEG channels as comparison methods.

- Phase locking value (PLV) is a phase synchronization approach proposed by Lachaux et al [31], which is commonly utilized to calculate FC matrix based on the instantaneous phases of multi-channel EEG signals [32]. It is defined as:

$$PLV = \left| n^{-1} \sum_{t=1}^{n} e^{i(\phi_{xt} - \phi_{yt})} \right| \tag{4}$$

  where n represents the time point, ϕ_{xt} and ϕ_{yt} represent the phase angle of signal x and y at time point t, respectively. The value range of PLV matrix is [0, 1], and the larger the value is, the stronger the phase synchronization degree between two signals is.
- Coherence represents the linear relationship between two EEG signals at a specific frequency [33]. It is also a FC measure used to quantify the connection strength between each pair of electrodes, and it can be computed as:

$$Coh_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f) \cdot P_{yy}(f)} \tag{5}$$

  where P_{xy}, P_{xx} and P_{yy} represent the cross-spectral density, the self-spectral density of signal x and y, respectively. The value range of $Coh_{xy}(f)$ is [0, 1], and the larger the value is, the greater the correlation degree of the two signals at frequency f is.
- Transfer entropy (TE) is employed to measure the EC relationships of EEG signals as a conditional mutual information. It is a non-parametric statistic that measures the amount of directed information transferred between two random processes [34]. Concretely, the TE of process X to process Y refers to the reduction of the uncertainty of current Y when the past X is obtained from the given past Y, so it is calculated as:

$$TE_{X \to Y}(t) = H\left(Y_t \mid Y_{t-1}\right) - H\left(Y_t \mid Y_{t-1}, X_{t-1}\right) \tag{6}$$

  where the amount of information is measured by Shannon entropy H. Concretely, $H\left(Y_t \mid Y_{t-1}\right)$ expresses the conditional entropy between current Y_t and past Y_{t-1}, and $H\left(Y_t \mid Y_{t-1}, X_{t-1}\right)$ indicates the information of current Y_t obtained from past Y_{t-1} and X_{t-1}.

The visualization of connectivity matrices from the multi-head self-attention layer, PLV, Coherence, and TE measures in the full frequency band is displayed in FIGURE 4. The first and second rows respectively demonstrate a connectivity matrix of four different methods for MDD patients and HCs. It can be noted that significant variability in brain connectivity between MDD patients and HCs.

E. BASELINE METHODS

In this paper, three DL models were selected as the baseline methods, which were utilized to compare the classification performance with the proposed method.

- Since the FC matrices computed by the PLV and coherence measures were symmetric, the parallel two-branch DL model presented in this paper was not applicable to these FC matrices. Therefore, an appropriate model was selected from the research on MDD detection for feature extraction and classification based on the two FC matrices. The comparison model was composed of two convolutional layers, a max-pooling layer, a dense layer, and a softmax layer, which was proposed by Li et al [35]. The parameters of the model were set as those in the original paper. Each convolutional layer had 32 convolutional kernels with a size of (3, 3) and a stride of 1. ReLU was utilized as the activation function of the two convolutional layers. The input of each convolutional layer was filled with one unit to maintain the spatial resolution after convolution. The max-pooling with a pool size of (2, 2) and a stride of 2 was followed by a dense layer with 512 hidden units and a ReLU activation function. The final layer was a softmax layer for CNN classification. The dense and softmax layers were preceded by a dropout with a probability of 0.5.
- The hybrid model composed of CNN and LSTM was proposed by Ay et al., which was termed as CNN-LSTM model [36]. The CNN-LSTM model consisted of 11 layers, where the first two, fifth and sixth layers were convolutional layers with a ReLU activation function, with the numbers of filters as 64, 128, 128, and 32, with different filter sizes of 5, 3, 13, and 7, respectively. A max-pooling layer with a kernel size of 2 and a stride of 2 was employed for the third layer. The fourth and tenth layers adopted dropout layers with a dropout rate of 0.2. The LSTM layer with a unit size of 32 was applied for the seventh layer. The ninth and eleventh layers were dense layers with a unit size of 64 and 2, with ReLU and softmax activation function after a flatten layer of the eighth layer.
- DeprNet was a CNN model for depression classification proposed by Seal et al., which included five convolutional blocks, a flatten layer, and three fully connected layers [12]. Each convolutional block consisted of a convolutional layer, a BN layer, and a max-pooling layer. The number of filters used for the five convolutional layers was 128, 64, 64, 32 and 32, respectively. The filter sizes used in the first three, fourth, and fifth convolutional layers were (1,5), (1,3), and (1,2), respectively. The pool size of five max-pooling layers was equal to (1, 2). The stride of all convolutional layers and max-pooling layers was equal to (1, 2). The number

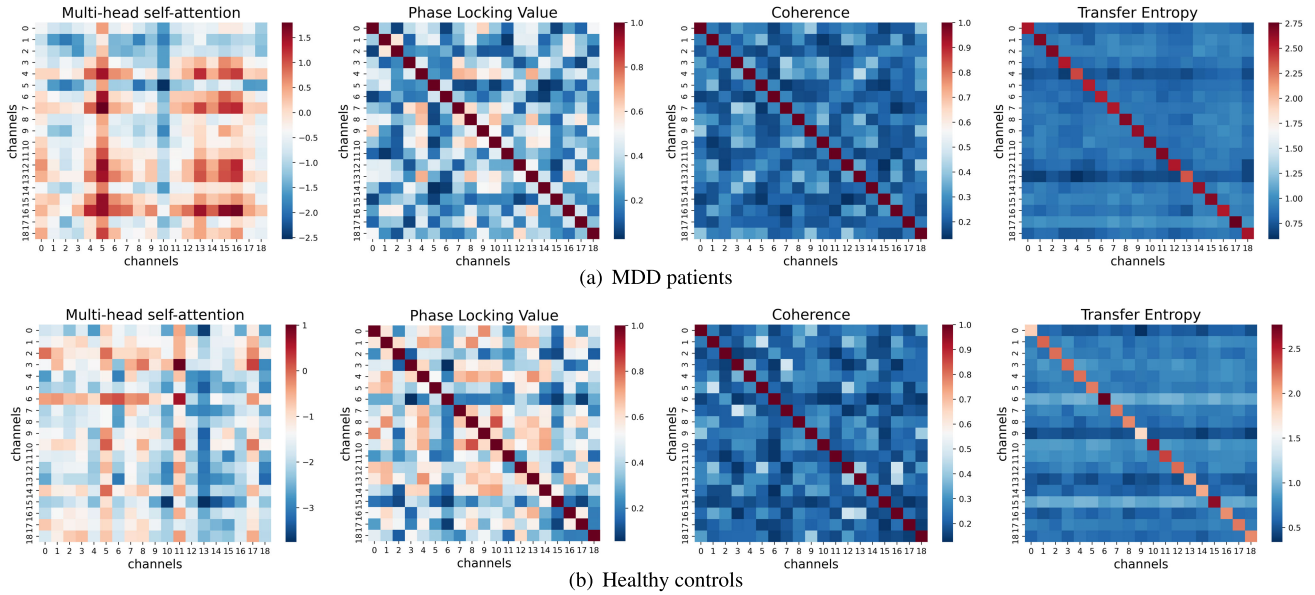


FIGURE 4. The visualization of connectivity matrices by multi-head self-attention and three brain connectivity measures in full frequency band from MDD patients and HCs. The first and second rows show the visualization of a connectivity matrix of MDD patients and HCs, respectively. The depth of color indicates the strength of connection among 19 channels.

of neurons in the three fully connected layers was 8, 4, and 2, respectively. The last fully connected layer employed the softmax activation function, while the LeakyReLU activation function was used in other layers.

F. MODEL IMPLEMENTATION AND EXPERIMENTAL EVALUATION

Disease classification differs from researches in other fields in that it is not rational to divide the data of a subject into both training and testing sets in a supervised learning scenario. Therefore, this study employed the LOSOCV method to divide the training set and test set at the subject level. Specifically, for each validation loop, the EEG data of one subject was used as the test data, and the EEG data of the remaining subjects were utilized as the training data. The whole process was repeated 57 times until the data from each subject served as the testing data once. Furthermore, in order to reduce the bias of experimental results, all experiments were performed 5 times in each LOSO loop, and the average evaluation metrics were taken as the final experimental results for each test subject. All experiments were implemented with Python using Keras libraries. All models were trained with the batch size and the number of epoch being set as 32 and 30, respectively. The binary cross-entropy loss function and Adam optimizer with a learning rate of 10^{-4} were utilized for model training.

In addition, the sensitivity, specificity, and modified accuracy were adopted as evaluation metrics to evaluate the classification performance of the models. Sensitivity is the ability of a classifier to correctly identify positive samples, while specificity is the ability of a classifier to correctly identify negative samples. Modified accuracy is the ability of a classifier to classify all samples correctly. The three evaluation metrics are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Modified\_accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{9}$$

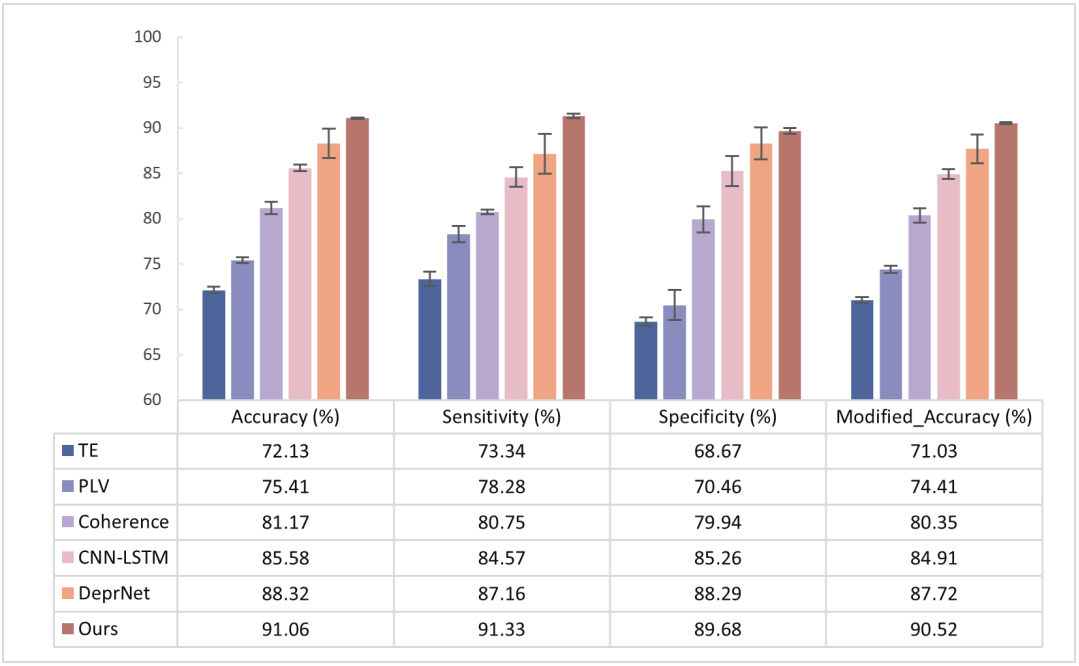
where TP, FN, TN and FP indicate the number of true positive samples, false negative samples, true negative samples and false positive samples, respectively.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the classification performance of the proposed end-to-end DL model, experiments were conducted from different perspectives in this section. First, the classification ability of the proposed model and the baseline models in the full frequency band were compared. Subsequently, ablation experiments were conducted to verify the effectiveness and rationality of the proposed model architecture. Conclusively, the subject-dependent experiments in the full frequency band and experiments to exploring the classification performance of five sub-bands were further implemented.

A. RESULTS AND COMPARISON

In this study, EEG data containing 3,163 samples obtained from 57 subjects were used for discriminating between MDD and HCs. The LOSOCV strategy was employed to evaluate the proposed method and the five comparison methods mentioned above. The classification results of the full frequency band are shown in FIGURE 5. It can be seen that the proposed model yields better classification performance in four evaluation metrics compared with the other five models,



	Accuracy (%)	Sensitivity (%)	Specificity (%)	Modified_Accuracy (%)
TE	72.13	73.34	68.67	71.03
PLV	75.41	78.28	70.46	74.41
Coherence	81.17	80.75	79.94	80.35
CNN-LSTM	85.58	84.57	85.26	84.91
DeprNet	88.32	87.16	88.29	87.72
Ours	91.06	91.33	89.68	90.52

FIGURE 5. The bar chart indicates the classification performance of proposed method and five comparison methods in the full frequency band using LOSOCV method. The error bar indicates the standard deviation.

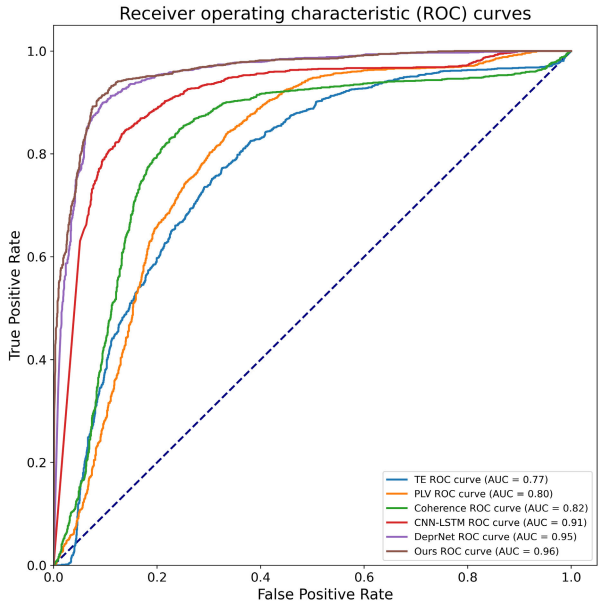


FIGURE 6. The receiver operating characteristic (ROC) curves of the proposed approach and five comparison methods obtained in the full frequency band.

TABLE 2. The classification performance of proposed method and five comparison methods in the full frequency band using 10-fold cv method. (mean ± std).

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
TE	92.60 ± 0.27	95.66 ± 0.80	89.55 ± 1.22
PLV	97.54 ± 0.20	98.26 ± 0.38	96.87 ± 0.61
Coherence	98.13 ± 0.14	98.45 ± 0.39	97.81 ± 0.23
CNN-LSTM	88.04 ± 4.59	**99.24 ± 0.17**	77.15 ± 9.30
DeprNet	96.26 ± 0.50	96.32 ± 0.67	96.22 ± 1.00
Ours	**98.45 ± 0.23**	98.55 ± 0.14	**98.35 ± 0.39**

where the classification accuracy, sensitivity, specificity and modified accuracy are 91.06%, 91.33%, 89.68% and 90.52%, respectively. The proposed model outperforms DeprNet in terms of the classification accuracy by 2.31%. Besides, our model has the smallest standard deviation in all classification metrics, which indicates that the proposed method is more robust in classify MDD using EEG data as compared with the baseline methods, and has the minimum standard deviation of 0.08% in the classification accuracy. The reason for the better performance of the proposed method may be that our model could effectively extracted the features of potential connectivity relationships among different EEG channels. In contrast, the comparison models had some limitations for the extraction of temporal and spatial features of EEG signals. First, the CNN-LSTM and DeprNet models did not appropriately consider spatial information of EEG signals. Second, the connectivity matrices calculated by TE, PLV and coherence measures may not be able to represent the underlying connectivity patterns among the multi-channel EEG data, which could result in the suboptimal features extracted by these comparison models, thereby leading to the loss of important information for following classification.

This paper also employed the receiver operating characteristic (ROC) curve to describe the process by which the performance of the model changes with the threshold change. The area under the ROC curve is denoted by AUC, whose value ranges from [0, 1]. The larger the value is, the more discriminative the model is. FIGURE 6 visually shows the

TABLE 3. The classification performance of the proposed method and comparison methods in five sub-bands. (mean ± std).

	Model \ Band	Delta	Theta	Alpha	Beta	Gamma
Accuracy (%)	TE	77.75 ± 0.55	65.66 ± 0.55	69.60 ± 0.60	66.74 ± 1.07	83.05 ± 0.73
	PLV	70.30 ± 0.49	72.34 ± 0.35	74.50 ± 0.78	73.53 ± 0.55	77.91 ± 0.61
	Coherence	68.95 ± 0.83	75.27 ± 0.64	78.07 ± 0.77	71.16 ± 0.54	73.25 ± 1.02
	CNN-LSTM	82.42 ± 0.88	**77.33 ± 1.32**	**80.13 ± 0.85**	85.06 ± 1.08	85.83 ± 0.50
	DeprNet	87.38 ± 0.62	71.82 ± 0.74	72.79 ± 1.58	84.83 ± 1.01	86.80 ± 0.99
	Ours	**90.61 ± 0.39**	75.11 ± 0.33	78.47 ± 0.55	**87.88 ± 0.41**	**88.28 ± 0.16**
Sensitivity (%)	TE	87.40 ± 0.91	**82.10 ± 0.76**	75.62 ± 0.88	71.21 ± 1.58	83.92 ± 1.47
	PLV	77.07 ± 0.95	76.69 ± 0.79	75.88 ± 0.95	72.64 ± 0.83	79.57 ± 0.68
	Coherence	71.53 ± 1.01	78.45 ± 0.90	**79.36 ± 0.60**	70.46 ± 0.49	72.91 ± 1.28
	CNN-LSTM	84.70 ± 1.36	77.28 ± 1.11	77.24 ± 0.65	83.47 ± 0.86	86.14 ± 1.43
	DeprNet	88.08 ± 1.43	69.51 ± 0.67	69.77 ± 2.04	81.87 ± 1.38	87.78 ± 2.34
	Ours	**92.50 ± 0.33**	76.93 ± 0.48	78.02 ± 0.53	**85.34 ± 0.43**	**89.79 ± 0.38**
Specificity (%)	TE	65.02 ± 0.51	45.06 ± 0.61	60.78 ± 0.80	59.05 ± 1.12	82.39 ± 1.82
	PLV	59.45 ± 0.53	65.57 ± 1.07	71.21 ± 1.05	72.49 ± 0.64	73.77 ± 0.93
	Coherence	63.14 ± 1.08	69.83 ± 0.81	74.92 ± 1.10	70.23 ± 0.98	70.97 ± 1.13
	CNN-LSTM	78.14 ± 1.20	**76.31 ± 2.31**	**81.74 ± 1.59**	85.68 ± 1.82	85.16 ± 0.73
	DeprNet	85.07 ± 1.32	71.35 ± 1.11	74.04 ± 1.52	86.68 ± 1.54	85.59 ± 0.88
	Ours	**87.45 ± 0.51**	69.68 ± 0.60	77.19 ± 1.38	**89.76 ± 0.83**	**86.83 ± 0.22**
Modified_Accuracy (%)	TE	76.31 ± 0.59	63.76 ± 0.62	68.27 ± 1.15	65.19 ± 0.77	83.16 ± 0.33
	PLV	68.35 ± 0.54	71.19 ± 0.40	73.57 ± 0.84	72.56 ± 0.52	76.70 ± 0.60
	Coherence	67.38 ± 0.86	74.18 ± 0.63	77.16 ± 0.79	70.34 ± 0.51	71.95 ± 1.09
	CNN-LSTM	81.45 ± 0.89	**76.80 ± 1.32**	**79.47 ± 0.90**	84.57 ± 1.17	85.65 ± 0.48
	DeprNet	86.59 ± 0.70	70.42 ± 0.68	71.88 ± 1.58	84.25 ± 1.02	86.70 ± 1.00
	Ours	**90.00 ± 0.37**	73.34 ± 0.36	77.61 ± 0.55	**87.52 ± 0.53**	**88.33 ± 0.17**

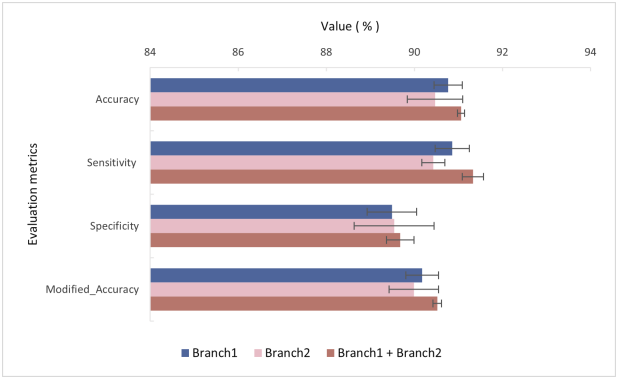


FIGURE 7. The bar chart illustrates the classification results of ablation experiments on the four evaluation metrics, including accuracy, sensitivity, specificity, and modified accuracy. The error bar indicates the standard deviation.

ROC curves of the proposed end-to-end DL method and the five comparison methods in the full frequency band, and the AUC values of each model are presented. The experimental results indicate that the proposed DL model has better recognition ability of MDD patients and HCs, and its AUC value is 0.96.

B. ABLATION STUDY

Since this study designed a parallel two-branch model to extract features from two different dimensions of the connectivity matrices learned from the multi-head self-attention layer, the ablation experiments were performed to sequentially remove each branch to verify the rationality. It can be seen from FIGURE 7 that the model with a parallel two-branch has better classification performance and smaller standard deviation than the model with only a single branch on the four evaluation metrics. It further suggests that the parallel two-branch model could extract more effective features from two different dimensions, which leads to better performance and robustness of the model.

C. FURTHER STUDY

In order to further verify the ability of the proposed model to classify MDD patients and HCs, this paper designed subject-dependent experiments based on the EEG data of MDD patients and HCs as previous studies [37], [38]. Concretely, the proposed model and the five comparison models were carried out in five rounds of 10-fold CV experiments. The 3163 samples of MDD patients and HCs obtained after preprocessing were randomly divided into ten folds, of which seven folds contained 316 samples, and the remaining three folds contained 317 samples. The mean evaluation metrics of 5-round 10-fold CV were taken as the final classification results. As can be seen from TABLE 2, the proposed model yielded better classification performance compared with the other five models, and its classification accuracy, sensitivity, and specificity are 98.45%, 98.55%, and 98.35%, respectively.

Additionally, the EEG data were further decomposed into five sub-bands to evaluate the classification performance of the proposed model and comparison models.

The classification results of the four evaluation metrics, including accuracy, sensitivity, specificity, and modified accuracy are listed in TABLE 3. The results demonstrate that the classification performance of the proposed method is superior to that of the other five methods in delta, beta, and gamma bands, where it exceeds the accuracy of DeprNet with the second highest accuracy by more than 3% in the delta band. Furthermore, the method achieved a maximum sensitivity of 92.50% in the theta band and a maximum specificity of 89.76% in the beta band. It indicates that the proposed method has a relatively low probability of both missing and misdiagnosing depression. Except for the delta band, which reached the largest classification accuracy among the five sub-bands, beta and gamma bands also have better classification performance, which may be related to depression in neural mechanisms. Previous studies have reported that beta and low gamma power are associated with inattention in depression patients [39], and delta oscillations are more pronounced in pathological states caused by brain tissue damage, such as depression [40]. Therefore, the features in delta, beta, and gamma bands could be more discriminative for classifying MDD, and provide favorable biomarkers for the identification of MDD.

D. DISCUSSION

Although the performance of the proposed DL model was satisfactory as compared with the other five baseline models, several limitations still need to be discussed. First, only one public dataset was utilized to evaluate the proposed model, thus extra cohort datasets should be acquired to further validate the effectiveness of the proposed model, such as the MODMA dataset [41]. Second, although the proposed method achieved better performance on the angle of the spatial connectivity relationships, we need to refine the model that could integrate the temporal information with the spatial information, which may further enhance the detection performance for MDD. Besides, the multi-head self-attention mechanism adopted in the current study was just one type of attention mechanism, and other ones can be explored in our future study [42], [43], [44]. Lastly, the interpretability of the model is important for DL model applications, we did not discuss this aspect in the current study, which is worthy of investigating in the following study.

IV. CONCLUSION

This study presented a novel end-to-end DL model for MDD classification. The proposed model can learn the potential connectivity relationships among 19 EEG channels in a data-driven fashion by a multi-head self-attention mechanism directly from the EEG data. The parallel two-branch CNN modules were leveraged to extract high-level features from multiple perspectives on the learned connectivity matrices, which were finally input into a dense layer to complete the classification. Based on the EEG signals of 30 MDD patients and 28 HCs, the highest average accuracy of 91.06% was obtained in the full frequency band by the proposed model with the LOSOCV strategy. Besides, ROC curves, ablation experiments, and subject-dependent experiments were conducted to assess the validity and feasibility of the proposed model. Lastly, EEG data were divided into five sub-bands for further experiments, the proposed method outperformed the comparison methods in terms of classification performance in the delta, beta, and gamma sub-bands. These results indicate that the proposed model could be an alternative method to detect MDD patients, which could be beneficial for early intervention and treatment of MDD.

ACKNOWLEDGMENT

The authors sincerely thank all anonymous reviewers for their insightful suggestions and comments.

REFERENCES

[1] W. Depression, *Other Common Mental Disorders: Global Health Estimates*, vol. 24. Geneva, Switzerland: World Health Organization, 2017.

[2] C. Otte, S. M. Gold, B. W. Penninx, C. M. Pariante, A. Etkin, M. Fava, D. C. Mohr, and A. F. Schatzberg, "Major depressive disorder," *Nature Rev. Disease Primers*, vol. 2, no. 1, pp. 1–20, 2016.

[3] E. G. Cuevas, R. Wu, K. Manning, L. Wang, and D. Steffens, "Functional brain networks of trait and state anxiety in late-life depression," *Amer. J. Geriatric Psychiatry*, vol. 29, no. 4, pp. S52–S53, Apr. 2021.

[4] J. R. Lave, R. G. Frank, H. C. Schulberg, and M. S. Kamlet, "Cost-effectiveness of treatments for major depression in primary care practice," *Arch. Gen. Psychiatry*, vol. 55, no. 7, pp. 645–651, 1998.

[5] J. B. Williams, "A structured interview guide for the Hamilton depression rating scale," *Arch. Gen. Psychiatry*, vol. 45, no. 8, pp. 742–747, 1988.

[6] A. T. Beck, R. A. Steer, and M. G. Carbin, "Psychometric properties of the beck depression inventory: Twenty-five years of evaluation," *Clin. Psychol. Rev.*, vol. 8, no. 1, pp. 77–100, Jan. 1988.

[7] *Diagnostic and Statistical Manual of Mental Disorders*, vol. 21, American Psychiatric Association, Washington, DC, USA, 2013.

[8] A. Biasiucci, B. Franceschiello, and M. M. Murray, "Electroencephalography," *Curr. Biol.*, vol. 29, no. 3, pp. R80–R85, 2019, doi: 10.1016/j.cub.2018.11.052.

[9] J. Zhu, Y. Wang, R. La, J. Zhan, J. Niu, S. Zeng, and X. Hu, "Multimodal mild depression recognition based on EEG-EM synchronization acquisition network," *IEEE Access*, vol. 7, pp. 28196–28210, 2019.

[10] M. Saeedi, A. Saeedi, and A. Maghsoudi, "Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 1007–1018, Sep. 2020.

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

[12] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A deep convolution neural network framework for detecting depression using EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[13] X. Song, D. Yan, L. Zhao, and L. Yang, "LSDD-EEGNet: An efficient end-to-end framework for EEG-based depression detection," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103612.

[14] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[15] B. Liu, H. Chang, K. Peng, and X. Wang, "An end-to-end depression recognition method based on EEGNet," *Frontiers Psychiatry*, vol. 13, Mar. 2022, Art. no. 864393.

[16] W. C. Drevets, J. L. Price, and M. L. Furey, "Brain structural and functional abnormalities in mood disorders: Implications for neurocircuitry models of depression," *Brain Struct. Function*, vol. 213, nos. 1–2, pp. 93–118, Sep. 2008.

[17] R. H. Kaiser, J. R. Andrews-Hanna, T. D. Wager, and D. A. Pizzagalli, "Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity," *JAMA Psychiatry*, vol. 72, no. 6, pp. 603–611, 2015.

[18] K. M. Albert, G. G. Potter, B. D. Boyd, H. Kang, and W. D. Taylor, "Brain network functional connectivity and cognitive performance in major depressive disorder," *J. Psychiatric Res.*, vol. 110, pp. 51–56, Mar. 2019.
[19] B. Zhang, G. Yan, Z. Yang, Y. Su, J. Wang, and T. Lei, "Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 215–229, 2021.
[20] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis," *J. Neurosci. Methods*, vol. 358, Jul. 2021, Art. no. 109209.
[21] L. Duan, H. Duan, Y. Qiao, S. Sha, S. Qi, X. Zhang, J. Huang, X. Huang, and C. Wang, "Machine learning approaches for MDD detection and emotion decoding using EEG signals," *Frontiers Hum. Neurosci.*, vol. 14, p. 284, Sep. 2020.
[22] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach," *Cognit. Neurodyn.*, vol. 15, no. 2, pp. 239–252, Apr. 2021.
[23] W. Mumtaz, L. Xia, M. A. M. Yasin, S. S. A. Ali, and A. S. Malik, "A wavelet-based technique to predict treatment outcome for major depressive disorder," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171409.
[24] C. C. Bell, "DSM-IV: Diagnostic and statistical manual of mental disorders," *J. Amer. Med. Assoc.*, vol. 272, no. 10, pp. 828–829, 1994.
[25] Y. Lin, P. Du, H. Sun, Y. Liang, Z. Wang, Y. Cui, K. Chen, Y. Xia, D. Yao, L. Yu, and D. Guo, "Identifying refractory epilepsy without structural abnormalities by fusing the common spatial patterns of functional and effective EEG networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 708–717, 2021.
[26] E. Başar and B. Güntekin, "Review of delta, theta, alpha, beta, and gamma response oscillations in neuropsychiatric disorders," *Suppl. Clin. Neurophysiol.*, vol. 62, pp. 303–341, Jan. 2013.
[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
[28] Q. Xin, S. Hu, S. Liu, L. Zhao, and Y.-D. Zhang, "An attention-based wavelet convolution neural network for epilepsy EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 957–966, 2022.
[29] K. J. Friston, "Functional and effective connectivity: A review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, Jan. 2011.
[30] J. Cao, Y. Zhao, X. Shan, H. Wei, Y. Guo, L. Chen, J. A. Erkoyuncu, and P. G. Sarrigiannis, "Brain functional and effective connectivity based on electroencephalography recordings: A review," *Hum. Brain Mapping*, vol. 43, no. 2, pp. 860–879, Feb. 2022.
[31] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Hum. Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.
[32] C. Hua, H. Wang, H. Wang, S. Lu, C. Liu, and S. M. Khalid, "A novel method of building functional brain network using deep learning algorithm with application in proficiency detection," *Int. J. Neural Syst.*, vol. 29, no. 1, Feb. 2019, Art. no. 1850015.
[33] L. H. Koopmans, *The Spectral Analysis of Time Series*. Amsterdam, The Netherlands: Elsevier, 1995.
[34] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, Jul. 2000.
[35] X. Li, R. La, Y. Wang, B. Hu, and X. Zhang, "A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography," *Frontiers Neurosci.*, vol. 14, p. 192, Apr. 2020.
[36] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil, and U. R. Acharya, "Automated depression detection using deep representation and sequence learning with EEG signals," *J. Med. Syst.*, vol. 43, no. 7, p. 205, Jul. 2019.
[37] W. Dang, Z. Gao, X. Sun, R. Li, Q. Cai, and C. Grebogi, "Multilayer brain network combined with deep convolutional neural network for detecting major depressive disorder," *Nonlinear Dyn.*, vol. 102, no. 2, pp. 667–677, Oct. 2020.
[38] D. M. Khan, K. Masroor, M. F. M. Jailani, N. Yahya, M. Z. Yusoff, and S. M. Khan, "Development of wavelet coherence EEG as a biomarker for diagnosis of major depressive disorder," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4315–4325, Mar. 2022.
[39] S.-C. Roh, E.-J. Park, M. Shim, and S.-H. Lee, "EEG beta and low gamma power correlates with inattention in patients with major depressive disorder," *J. Affect. Disorders*, vol. 204, pp. 124–130, Nov. 2016.
[40] G. G. Knyazev, "EEG delta oscillations as a correlate of basic homeostatic and motivational processes," *Neurosci. Biobehavioral Rev.*, vol. 36, no. 1, pp. 677–695, Jan. 2012.
[41] H. Cai, "A multi-modal open dataset for mental-disorder analysis," *Sci. Data*, vol. 9, no. 1, p. 178, Apr. 2022.
[42] Z. Gao, J. Xie, Q. Wang, and P. Li, "Global second-order pooling convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3019–3028.
[43] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3183–3192.
[44] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 173–190.

MIN XIA is currently pursuing the master's degree with the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China. Her current research interests include machine learning and the classification of diseases based on EEG.

YANGSONG ZHANG is currently with the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China. His current research interests include brain–machine interaction, machine learning, deep learning, brain information processing, medical image processing, and biomedical signal processing.

YIHAN WU is currently pursuing the master's degree with the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China. His current research interests include machine learning and emotion recognition based on EEG.

XIUZHU WANG is currently with the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China. Her research interests include data mining, big data technology, and intelligent medicine.