

Classifying and Scoring Major Depressive Disorders by Residual Neural Networks on Specific Frequencies and Brain Regions

Cheng Kang[✉], Graduate Student Member, IEEE, Daniel Novak[✉], Xujing Yao[✉], Jiayong Xie, and Yong Hu[✉], Senior Member, IEEE

Abstract—Major Depressive Disorder (MDD) - can be evaluated by advanced neurocomputing and traditional machine learning techniques. This study aims to develop an automatic system based on a Brain-Computer Interface (BCI) to classify and score depressive patients by specific frequency bands and electrodes. In this study, two Residual Neural Networks (ResNets) based on electroencephalogram (EEG) monitoring are presented for classifying depression (classifier) and for scoring depressive severity (regression). Significant frequency bands and specific brain regions are selected to improve the performance of the ResNets. The algorithm, which is estimated by 10-fold cross-validation, attained an average accuracy rate ranging from 0.371 to 0.571 and achieved average Root-Mean-Square Error (RMSE) from 7.25 to 8.41. After using the beta frequency band and 16 specific EEG channels, we obtained the best-classifying accuracy at 0.871 and the smallest RMSE at 2.80. It was discovered that signals extracted from the beta band are more distinctive in depression classification, and these selected channels tend to perform better on scoring depressive severity. Our study also uncovered the different brain architectural connections by relying on phase coherence analysis. Increased delta deactivation accompanied by strong beta activation is the main feature of depression when the depression symptom is becoming more severe. We can therefore conclude that the model developed here is acceptable for classifying depression and for scoring depressive severity. Our model

can offer physicians a model that consists of topological dependency, quantified semantic depressive symptoms and clinical features by using EEG signals. These selected brain regions and significant beta frequency bands can improve the performance of the BCI system for detecting depression and scoring depressive severity.

Index Terms—Depressive severity, mental disorder monitoring, BCI, classifying and scoring, deep learning.

I. INTRODUCTION

MDD is a mental illness which is often accompanied by a high risk of suicidal thoughts [1]. Depressed individuals are often misdiagnosed by physicians, which leads to a range of problems, including self-medication, substance abuse, inappropriate treatment, social isolation, and impaired performance in education or at work [2], [3]. Cognitive behavioural therapy is the best way to treat mild depression, and for severe depression, currently, the combination of psychotherapy and antidepressant drugs is the most effective treatment [4], [5], [6]. Improper treatments would lead to future relapse and prolonged discontinuation symptoms [7].

Depression is widely categorized as non-depressed, mild, moderate, and severe, according to the severity of the depressive symptoms [8]. However, a descriptive study has shown that the rate of misdiagnosis of MDD is as high as 65.9% [3]. This means that the primary accuracy rate is less than 35% [3]. Failure to correctly diagnose MDD is caused by inadequate training of clinicians, as well as reasons that sufferers are not given appropriate appointments, medical examinations and proper treatments at the early stage [3], [9]. Existing tools for diagnosing MDD tend not to be used by clinical psychologists and physicians because these complex approaches have three main challenges: (1) they are time-consuming and need to be administrated by well-trained engineers or by professional clinicians [10], [11]; (2) they cannot classify depressive severity; (3) there is no visualization result provided, for example, brain topological maps.

To overcome these three above challenges, we firstly assume that the delta and beta brain activities are connected to depression as previous studies [12], [13], [14], [15], [16] noted. Therefore, in an attempt to achieve early detection

Manuscript received 22 February 2023; revised 15 June 2023; accepted 28 June 2023. Date of publication 7 July 2023; date of current version 26 July 2023. This work was supported in part by the Research Centre for Informatics under Grant CZ.02.1.01/0.0/0.0/16_019/0000765 and in part by the Student Grant Agency of the Czech Technical University in Prague under Grant SGS22/165/OHK3/3T/13. (Corresponding authors: Xujing Yao; Jiayong Xie; Yong Hu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Shenzhen Mental Health Center.

Cheng Kang and Daniel Novak are with the Department of Cybernetics and Robotics, Faculty of Electrical Engineering, Czech Technical University in Prague, 16636 Prague, Czech Republic (e-mail: kangchen@fel.cvut.cz; xnovakd1@fel.cvut.cz).

Xujing Yao is with the School of Informatics, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: xy147@leicester.ac.uk).

Jiayong Xie is with the Department of Nephrology, Xinghua City People's Hospital, Taizhou, Jiangsu 225700, China (e-mail: 1000007211@ujs.edu.cn).

Yong Hu is with the Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong (e-mail: yhud@hku.hk).

Digital Object Identifier 10.1109/TNSRE.2023.3293051

of depression, we have made an analysis of the delta and beta activities, as well as the corresponding brain networks which were provided as the visualization result. In our study, the Phase Synchrony Index (PSI) [12], [13], [14], [15], [16] was calculated to construct the brain functional networks. The electrodes and frequency bands were chosen based on different PSIs between depressive and healthy groups. Secondly, one classifier relying on ResNet [17] was designed to process selected EEG signals and detect depression. Additionally, we proposed one regression model relying on ResNet to score depressive severity. Both these two optimized ResNets on EEGs aims to accelerate the computing and diagnosing. This BCI system is expected to be used as a complementary tool to detect depression and monitor depressive severity, as well as a tool for evaluating conventional treatments in hospitals and clinics.

To detect depression from the healthy and, to score the depressive severity based on some specific psychological scales, we proposed this BCI system. The contributions of this paper are as follows: (1) We present the central-parietal increased delta deactivation accompanied by strong beta activation in the severe depression group under working memory tasks. (2) We propose one classification ResNet with the use of specific frequencies and brain regions which achieves better and more practical results in detecting the depressive from the healthy. (3) We also propose one regression ResNet with specific frequencies and brain regions for scoring the depressive severity based on two professional psychologists' score labels. The codes and corresponding documentation can be found here: <https://github.com/ChengKang520/Classifying-and-Scoring-MDD-BCI>.

II. RELATED WORKS

It is important to detect depression at the early stage, and not to delay proper treatments. Undetected depression can lead to long-term suffering and even suicide. Machine learning approaches can make early detection.

In recent years, medical images and methods relying on electrophysiological signals have been developed. Most of them have focused on extracting brain networks and using diagnostics models. Figure 1 illustrates the experimental flow. Our previous studies found that delta and beta brain activities are different from activities which are observed in the control group [12], [13], [14]. Based on these findings and inferences, a particularly designed system is illustrated in Figure 1. From A1 to A4, A1 presents the calculation of the PSI between two EEG signals. The formulas for calculating PSI can be denoted as:

$$\begin{aligned} \Delta\theta_{trialk}^{n \rightarrow m} &= \theta_{trialk}^n - \theta_{trialk}^m \\ r^{n \rightarrow m} &= \frac{\sqrt{\left\{ \sum_{trialk=1}^N \sin(\Delta\theta_{trialk}^{n \rightarrow m}) \right\}^2 + \left\{ \sum_{trialk=1}^N \cos(\Delta\theta_{trialk}^{n \rightarrow m}) \right\}^2}}{N} \end{aligned} \quad (1)$$

$$lag^{n \rightarrow m} = \arctan \left\{ \frac{\sum_{trialk=1}^N \sin(\Delta\theta_{trialk}^{n \rightarrow m})}{\sum_{trialk=1}^N \cos(\Delta\theta_{trialk}^{n \rightarrow m})} \right\} \quad (2)$$

$$= \arctan \left\{ \frac{\sum_{trialk=1}^N \sin(\Delta\theta_{trialk}^{n \rightarrow m})}{\sum_{trialk=1}^N \cos(\Delta\theta_{trialk}^{n \rightarrow m})} \right\} \quad (3)$$

where $\Delta\theta_{trialk}^{n \rightarrow m}$ is the difference between the angles of two electrodes (θ_{trialk}^n and θ_{trialk}^m) under the k -th trial. N is the number of total trials, and $r^{n \rightarrow m}$ is the mean value. We also denote it as PSI . Lastly, $lag^{n \rightarrow m}$ is the averaged angle of N trials. The entire procedure for constructing brain functional connection networks is presented in our previous study [14]. Moreover, A2 shows the significant features that we have detected from functional brain networks during working memory tasks - beta frequency band and 16 selected electrodes out of 64. After these pre-processing steps, A3 shows the use of ResNet architectures [17] and lists out the strategies for classifying and scoring depression. A4 shows two outputs which consist of the detection result for depression and the score of grading depressive severity.

A. Brain Regions and Extraction of Functional Networks

Methods relying on functional or structural brain networks are at the core of many mental health diagnosis methods, such as brain-connected networks for identifying bipolar disorders [18], [19] and schizophrenia [20], especially for detecting depression by EEGs [21], [22], [23]. To improve detection accuracy, researchers have focused on extracting useful information in the first pre-processing step.

In the first stage, when constructing functional brain networks or selecting obvious brain regions, some indexes were computed to estimate the connections or spectrums of brain regions. For example, during the resting state, one study used Adaboost classifiers to identify Cognitive Emotion Regulation Strategies (CERSs) with the help of spectral coherence [24]. Those spectrums of left and right frontal-prefrontal regions show obvious advantages to estimate the depression symptom under resting state [25]. The absolute power of the theta wave indicated a valid characteristic for discriminating depression. Relying on this finding, researchers used K-Nearest Neighbor (KNN) with 10-fold cross-validation to test the classification performance [21]. After the calculation of the relative wavelet energy and various entropy features by the Decomposed Discrete Wavelet Transform (DWT) coefficients on the EEG signals, a feed-forward Artificial Neural Networks (ANNs) was used to classify depression conditions [23]. A feature-level fusion approach was used to find powerful features, and then, traditional machine learning classifiers were utilized to detect depression by multimodal EEG data [22].

The brain networks have the attribute to show some well-known cognitive patterns, such as the abnormal cognitive control network of depressive patients [26], and they also can present the electrophysiological brain connections of some frequency bands (delta, theta, alpha and beta) [14]. According to brain oscillations in different frequencies, the PSI [12], [13], [14], [15], [16] was calculated to construct the brain functional networks. The PSI can reflect the degree of synchronization between two EEG channels, and after the calculation of correlation coefficients based on PSIs, an online clustering approach was used to construct convergent brain networks

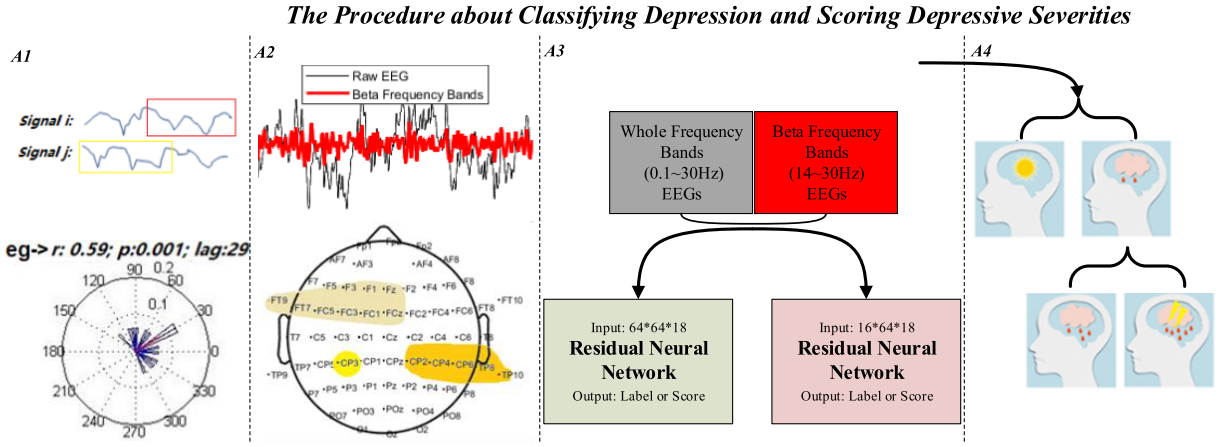


Fig. 1. The framework of this study. The entire procedure about classifying depression and scoring depressive severity (A1 → A2 → A3 → A4).

as described in previous studies [12], [13], [14], [15], [16]. Consequently, the Morlet's wavelet was utilized to calculate the time-frequency domain and the angle:

$$\begin{aligned} \varphi_{trialk}^n(f, t) &= \frac{1}{\sqrt{\pi}\delta_t} \exp\left(\frac{-t^2}{2\delta_t^2}\right) \exp(j2\pi ft) \end{aligned} \quad (4)$$

$$\begin{aligned} \Delta\theta_{trialk}^{n \rightarrow m} &= \text{angle}\{\exp(i[\varphi_{trialk}^n(f, t)])\} - \text{angle}\{\exp(i[\varphi_{trialk}^m(f, t)])\} \end{aligned} \quad (5)$$

where $\varphi_{trialk}^i(f, t)$ is the Morlet's wavelet at frequency domain f , and δ_t is the standard deviation of the Gaussian function $\varphi_{trialk}^n(f, t)$. When relying on the EEGLAB in the MATLAB environment, the wavelet cycles and the lowest time-frequency window are selected referring to our previous studies [12], [13], [14].

B. Artificial Neural Networks Utilization

Classifying depression by machine learning approaches mainly contributes to the short time-consuming diagnosis. With the use of machine learning methods (Support Vector Machine (SVM), Adaptive Boosting (AdaBoost) and Random Forest (RF)), the most popular basic clinical techniques are Magnetic Resonance Imagings (MRIs) and EEGs. Some particularly selected channels were utilized before training tasks, as large amounts of irrelevant information will slow down training and make the model susceptible to overfitting [27]. The common flow of depression detection systems can be divided into the following three steps.

1) Step 1: Psychological Paradigm: One study reported that adaptive dual n-back Working Memory (WM) training can reduce subclinical anxiety and depression symptomology in adolescents [28]. During the learning processes of the WM capacity, WM was found to moderate the relation between the Brain Derived Neurotrophic Factor (BDNF) and psychotherapy outcome for depression [29]. These studies indicate that WM plays an important role in reflecting depression severity. Therefore, there are two reasons why we chose n-back tasks as the paradigm. (1) because most emotional tasks cannot control the affection intensity, we design this n-back paradigm and

evaluate the WM capacity of participants by adjusting the n-back task (0-back is the base, and from 1 to 2,3,...,n, the difficulty of these n-back tasks will become higher); (2) to develop potential rehabilitation methods by using WM training for our future work.

2) Step 2: Feature Extraction: Commonly, significant neuroimaging regions [26] and significant electrophysiological areas [8] should be extracted and fed into machine learning models, as well as some selected EEG channels contributed to depression classification during eye-close resting states or tasks completing states.

3) Step 3: Classifying and Scoring Depressive Symptoms: In recent five years, researchers used traditional machine learning approaches with EEGs to identify depressed subjects, for examples, ANN [23], logistic regression [30], SVM [31], bagged tree [32] and Convolution Neural Network (CNN) [8]. For classifying depression subjects, a deep CNN [8], a combined architecture of CNN and Long Short Term Memory (LSTM) [33], HybridEEGNet [34], and Spiking Neural Network (SNN) [35] were used for discriminating depression under various cognitive or resting state tasks. Deep learning methods, especially CNN architectures, can automatically extract important features and can score depression severity within several psychological tasks. But for LSTM, one time-series model, which focuses on processing long time-series signals, it requires participants to complete long and huge continuous tasks.

For scoring the severity of depressive symptoms, one study used fMRI images and kernel partial least squares regression model, and the authors applied RMSE to evaluate the performance of their models [36]. Finally, we propose an approach based on beta EEG and particularly selected sixteen channels for classifying depression and scoring depressive symptoms. ResNets were selected because they can avoid gradient vanishing and go deeper with fewer parameters [17].

III. METHODOLOGY

A. Participants and EEGs Recording

The EEG signals were obtained from Shenzhen University, and Shenzhen Kangning Hospital, in Shenzhen, China. This study was approved by the ethics committee of Shenzhen

Mental Health Center. This dataset consists of 52 healthy undergraduate dextrorotational students (6:4 males to females; $Mean \pm SD = 20.4 \pm 9.7$ age distribution) and 48 depressed patients (6:4 males to females; $Mean \pm SD = 34.3 \pm 12.1$ age distribution). Strict selection and assessment procedures were employed as detailed elsewhere [12]. In both healthy and depressive groups, no medication was taken, and no personal or family history of psychiatric or neurological diseases was found before experiments. Based on using the DSM-IV Axis I Disorders, Clinician Version (SCID-CV) [37], the 17-item Hamilton Depression Rating Scale (HAMD), the depressed scores come from two professional clinical psychologists. The accepted depressive patients should be tested before the experiment by these two psychologists. If when two scores of one patient from these two psychologists could not be verified as the same level (one is mild, another is severe), this patient will be rejected to participate in this experiment. On the contrary, the average score will be computed to be the final score label. The depressive severity based on the HAMD and SCID-CV was classified, and the 17-item HAMD cut-off points were also defined as follows: $> 24 = severe$; $17-23 = moderate$; $8-16 = mild$; and $none(non - depressed) = 0 - 7$ [38]. The difference between moderate depression and mild depression is small, and after pre-processing of EEG data, we found the data distribution is imbalance. To avoid these two potential risks that would affect the performance of our proposed models, we eventually reduced the number of categories to three. Based on these previous studies, in this system, three groups were selected: healthy controls ($non - depressed : 0 - 7$), depressed with low scores ($Score : 8 - 23$) and depressed with high scores ($Score : > 24$).

B. Working Memory Experiments

Following our previous studies [12], [14], this n-back experiment was developed under the E-Prime 5.0 environment. The letter variant version of the n-back tasks was used in the experiment. 0-back tasks were set as the baseline, 1-back 2-back tasks were set as the WM load. These volunteers observed and responded to the black letter stimuli on the screen with a white background, and at the same time, they should press two buttons including the index finger for matching stimulus and the middle finger for mismatching stimulus. Under the 0-back tasks, participants were required to identify a single prespecified letter "X" from the screen by pressing the matching button. Meanwhile, if they recognized a particular letter which matched the letter presented 2 trials before in the 2-back tasks, they should press the matching button. Presented letters were randomly selected from English consonants. This experiment was divided into three segments, and every segment has three tasks with a separately random arrangement of 0-back, 1-back and 2-back. We designed three tasks: 0-back, 1-back and 2-back. In each experiment segments, we randomly arranged the sequence of these three n-back tasks, because a predictable and fixed sequence design will affect the performance of participants when they implementing these WM tasks. Random arrangement of 0-back, 1-back and 2-back can avoid this potential risk. We set the duration of

each given task as 75 seconds. All tasks should consist of a pseudo-random sequence of 30 consonants (10 targets and 20 nontargets). To avoid incorrect manipulation, as well as to provide enough time for reaction, letters were presented for 0.5 seconds and then disappeared in the following 2 seconds. Meanwhile, during every two parts, there are 45 seconds for participants to take a break. The behavioural performance was also recorded, for example, the reaction time and the response accuracy rate. Particularly, incorrect responses were excluded during the EEG analysis. After they convinced that there is no questions and every detail is clear, the warm-up tasks for guiding participants before the formal experiment would be ended.

C. Preprocessing of EEGs Before Training

All procedures including EEG recording and preprocessing have been detailed in the previous studies [12]. Briefly, after (1) the removal of eye movements, (2) 0.16-30 Hz (24dB/Octave) band-pass filtering, (3) artefact rejection and (4) baseline correction, the phase coherence calculation should be completed before the training tasks, because this study aims to develop an automatic system which can classify depression and score depressive severity by selected frequency bands and electrodes. Brain connection maps were constructed using the phase coherence method, as described in our previous studies [12], [13], [14], [15], [16]. The inputs are EEG signals from 64 or 16 channels during one type of task, and there are three different task types including 0-back, 1-back and 2-back.

D. Residual Neural Networks

In Figure 2, there are 64 channels that record EEGs, meanwhile, 2.5 seconds of EEG signals were collected. After that, a down-sampling process makes the data length from 2500 points to 1250 points. Thus, after discarding 98 points of the tail, we set the size of the input as $64 \times 64 \times 18$ in the first model. Then, two residual neural networks are used to train this EEG data (0-back, 1-back and 2-back tasks). Moreover, 16 selected electrodes based on the phase synchronization method were used in the second training phase. Therefore, we set the size of the input as $16 \times 64 \times 18$ for further training. The whole size of this EEG data is $22.5M$ sampling points ($[48 \text{ depressive patients} + 52 \text{ healthy controls}] \times 60 \text{ trials} \times 3 \text{ tasks [0-back, 1-back and 2-back]} \times 2.5 \text{ seconds} \times 500 \text{ sampling rates} = 22.5M$). After we tested the CNNs with 6 residual blocks, the performance tends to be best, as the size of its parameters is $0.85M$ (properly selecting the size of the model's parameters can avoid overfitting, as well as insufficient fitting).

As the misdiagnosis rate of MDDs is widely recognized as 65.9% [3], we set the threshold of detecting rate as 70%. There are 60 trials that one participant should implement, and the depressive probability of one participant is the division product: (the number of trials whose predicted probability from the model is above 70%) $\div$ (the total number of trials). Thus, if the predicted probability of one subject on one trial being depressive is above 70%, the system will classify him or her as 100% depressive on this trial. Finally, during one trial,

The Structure of the Residual Neural Network

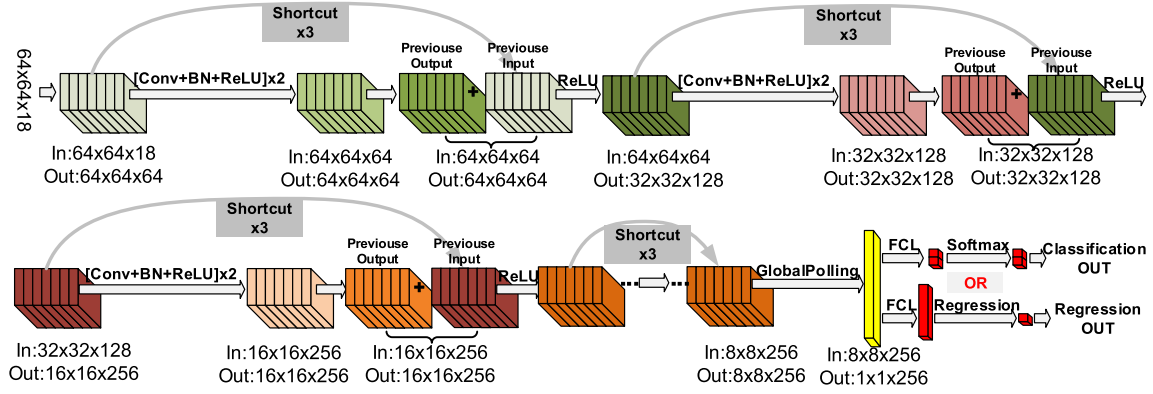


Fig. 2. The structure of the constructed residual neural network. The input size is $64 \times 64 \times 18$ or $16 \times 64 \times 18$. Conv+BN+ReLU means the processing of convolution (Conv), batch normalization (BN) and rectified linear unit (ReLU). FCL is the fully connected layer. The shortcut is purely forward plus. $\times 3$ means this block should be repeated triple times.

if there are 33 subjects out of a total of 40 subjects whose probabilities of the ResNet classifier are equal to or greater than 70%, the accuracy rate of the model is 0.825 (33/40). Moreover, the second ResNet regression model outputs the score of depressive severity (as referring to the SCID-CV system and the HAMD score).

IV. RESULT

A. Memory Load Comparison of Behavioural Results

Table I shows the significant level between the low and the high depressed group in terms of response accuracy rate and reaction time during three different working memory tasks (0-back, 1-back and 2-back). During the 0-back task, there is no significant difference ($P = 0.061$) between MDDs with low scores and the MDDs with high scores in terms of the response accuracy rate. But for the reaction time, the difference is significant ($P = 0.017$). In the 1-back task, both the response accuracy rate and the reaction time show a significant level ($P < 0.01$). When implementing the 2-back task, the MDDs with low scores demonstrated a significant difference in response accuracy rate ($P < 0.01$).

B. The Connections Comparison

We used the 0-back task as the “rest-state”, and the 2-back task as the WM load. Thus, the PSI decrease refers to the weak neuronal activity in the corresponding regions and the inhibition trends returning to the “rest-state”. The PSI increase means the activation of the neuronal activity in the corresponding brain regions and strong WM related mechanisms.

Depending on two WM tasks (0-back and 2-back), the number of significantly connected pairs is presented in Figure 3. For the PSI decrease, connections in the whole delta frequency components are the most significant different part among the three groups. The depressed group with high scores demonstrates the dominant whole theta frequency connections, but other frequency bands show no significant difference. When considering the PSI increase, the depressed group with high scores has the fewest delta, theta and alpha-connected pairs. Both two depressed groups demonstrate greater numbers of

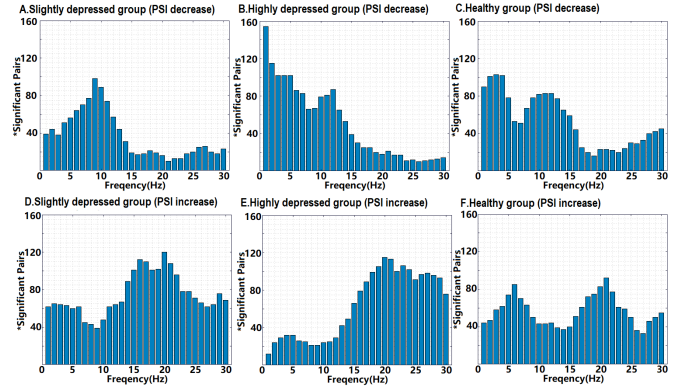


Fig. 3. The number of the significant pairs in terms of the comparison between 2-back and 0-back tasks.

the whole beta connections, but the depressed group with low scores presents stronger connections in delta, theta and alpha bands when compared with the depressed group with high scores. After the comparison of the product (the number of significant pairs $\times$ the corresponding PSI values), which could represent significant PSI whole levels among three groups, t-value results deriving from the two-sample t-test are demonstrated in Figure 4 ($P < 0.01$). We marked the most significant frequency component in every histogram. Apart from Figure 4B, which shows that the slightly obvious frequency part locates on delta bands ($P < 0.05$), the last three histograms show that the beta frequency activation represents the most significant difference.

C. Clusters Between These Three Groups

According to the PSI connections comparison, as shown in Figure 5A, the PSI decrease in the depressive group with lower scores contributes to few electrode connections, and it also presents the flat distribution in the beta frequency band in Figure 3. However, for the PSI increase in Figure 5B, the control group cannot generate one cluster, and the depressive group with low scores tends to gather the connected pairs in the left parietal and the left central regions (as shown in Cluster A). Accordingly, between the depressive group

TABLE I

THE COMPARISON OF REACTION TIME AND RESPONSE ACCURACY RATES BETWEEN TWO DIFFERENT MEMORY LOADS (AVERAGE $\pm$ VARIANCE) IN TWO DEPRESSIVE GROUPS

Memory Load	0-back		1-back		2-back	
	Low scores	High scores	Low scores	High scores	Low scores	High scores
Response Accuracy Rate	98.9 $\pm$ 1.4	96.3 $\pm$ 2.2	92.8 $\pm$ 4.7	86.3 $\pm$ 3.6	84.9 $\pm$ 5.3	75.5 $\pm$ 7.6
Reaction Time	545 $\pm$ 53	561 $\pm$ 47	701 $\pm$ 147	751 $\pm$ 129	769 $\pm$ 176	791 $\pm$ 183
Statistics (the low and the high)	P value		P value		P value	
	Accuracy Rate	Reaction Time	Accuracy Rate	Reaction Time	Accuracy Rate	Reaction Time
	P = 0.061	P = 0.017	P < 0.01	P < 0.01	P < 0.01	P = 0.053

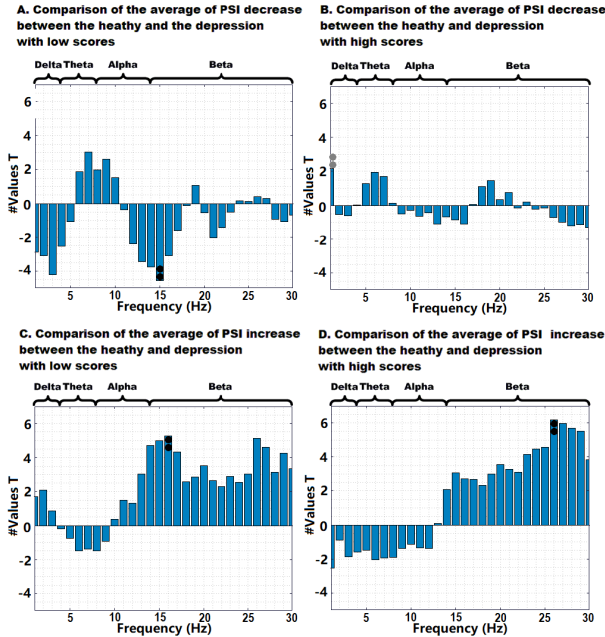


Fig. 4. The t values (significant level) of the comparison between the depression group and the healthy control group.

TABLE II

CLASSIFICATION (ACCURACY) AND SCORING DEPRESSION (RMSE) RESULTS (MEAN AND VARIANCE) USING WHOLE FREQUENCY BANDS (DELTA, THETA, ALPHA AND BETA)

Accuracy Rate (>0.70)	0-back	1-back	2-back	Best Result
	0.457 $\pm$ 0.004	0.429 $\pm$ 0.010	0.514 $\pm$ 0.027	0.734 in 2-back
Score Difference (RMSE)	0-back	1-back	2-back	Best Result
	8.38 $\pm$ 10.39	8.41 $\pm$ 12.43	7.73 $\pm$ 10.40	3.22 in 2-back

with high scores and the control group (the down panel in Figure 5C), in terms of the PSI decrease, the control group obtains fewer connected pairs mainly in the left frontal and the whole parietal areas (as shown in Cluster C), but the depressed group shows almost the whole cerebral connections apart from the occipital areas (as shown in Cluster B). Regarding the PSI increase, the depressed group with high scores presents the compact connecting pattern presenting the left frontal-central and right central-parietal regions, as well as the left frontal-temporal and the right temporal-parietal areas (as shown in Cluster D).

D. The Result of Classifying and Scoring MDD Patients

After these above pre-processing steps, there are no more than 60 trials for each subject, because we removed

TABLE III

CLASSIFICATION (ACCURACY) AND SCORING DEPRESSION (RMSE) RESULTS (MEAN AND VARIANCE) USING BETA FREQUENCY BANDS

Accuracy Rate (>0.70)	0-back	1-back	2-back	Best Result
	0.514 $\pm$ 0.047	0.429 $\pm$ 0.051	0.371 $\pm$ 0.047	0.783 in 0-back
Score Difference (RMSE)	0-back	1-back	2-back	Best Result
	7.97 $\pm$ 5.08	7.59 $\pm$ 2.29	8.05 $\pm$ 1.96	4.10 in 0-back

TABLE IV

CLASSIFICATION (ACCURACY) AND SCORING DEPRESSION (RMSE) RESULTS (MEAN AND VARIANCE) USING WHOLE FREQUENCY BANDS (DELTA, THETA, ALPHA AND BETA) AND SELECTED EEG CHANNELS

Accuracy Rate (>0.70)	0-back	1-back	2-back	Best
	0.514 $\pm$ 0.047	0.514 $\pm$ 0.057	0.571 $\pm$ 0.020	0.714 in 1-backs
Score Difference (RMSE)	0-back	1-back	2-back	Best
	7.77 $\pm$ 9.67	7.25 $\pm$ 4.81	7.37 $\pm$ 4.61	2.88 in 1-back

TABLE V

BY SCALING THE SIZE OF PROPOSED RESNETs, THE BELOW SHOWS THE CLASSIFICATION (ACCURACY) AND SCORING (RMSE) RESULTS USING BETA FREQUENCY BAND AND SELECTED EEG CHANNELS

ResNet (Size: 2.4M)				
Accuracy Rate (>0.70)	0-back	1-back	2-back	Best
	0.452 $\pm$ 0.0913	0.409 $\pm$ 0.222	0.414 $\pm$ 0.367	0.833 (0-back)
Score Difference (RMSE)	0-back	1-back	2-back	Best
	8.12 $\pm$ 11.45	8.07 $\pm$ 11.83	7.74 $\pm$ 13.36	3.02 (0-back)
Default ResNet (Size: 4.6M)				
Accuracy Rate (>0.70)	0-back	1-back	2-back	Best
	0.429 $\pm$ 0.051	0.514 $\pm$ 0.016	0.457 $\pm$ 0.055	0.871 (2-back)
Score Difference (RMSE)	0-back	1-back	2-back	Best
	7.97 $\pm$ 12.76	7.83 $\pm$ 10.93	7.59 $\pm$ 14.68	2.80 (2-back)

substandard trials. The result of the whole frequency band is shown in Table II. Results of beta frequency bands are presented in Table III. In the second model, we extended the system for classifying depression and scoring depressive severity by extracting the beta frequency band and 16 significant electrodes. The online clustering step based on PSIs generates Cluster A and Cluster D, and the most frequently connected electrodes - Fz, F1, F3, FCz, FC1, FC3, FC5, FT7, FT9, T7, CP3, CP2, CP4, CP6, TP8 and TP10 - both in Cluster A and Cluster D can contribute to improve the performance of classifying depression and scoring depressive severity. To avoid random result, we used 10-fold computing method to choose the best result. For example, in Table IV, the classification accuracy rate can reach 0.714 when using the whole frequency bands. But as shown in Table V when relying on the beta frequency bands, the accuracy rate can even reach 0.871. Eventually, based on 2-back tasks in beta frequency bands with the contribution of particularly selected channels, in terms of 10-fold testing, 0.871 is the maximum value.

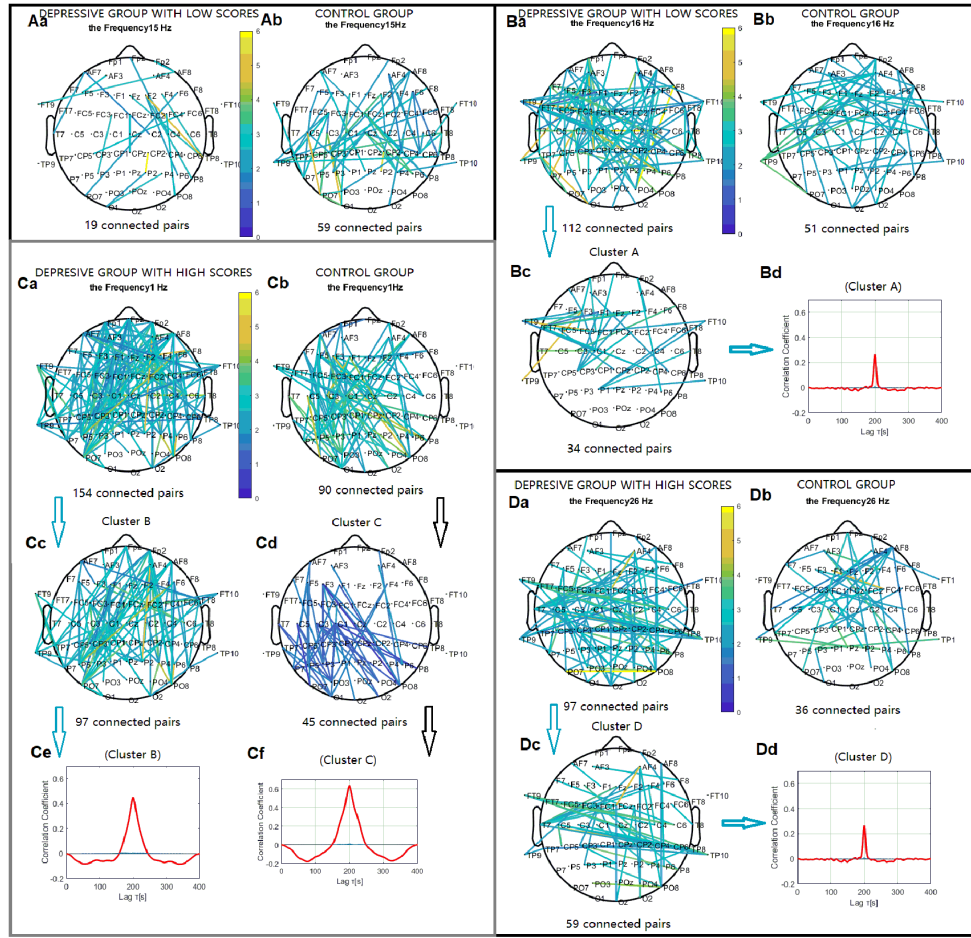


Fig. 5. Clustering of some significantly increased and decreased phase synchronization indices mainly in beta bands for both the two depression groups and control groups. Lines in the up panel (panel A and B) respectively represent the significant PSI decrease and increase during the 2-back condition. Relative to that during the 0-back condition ($p < 0.05$) between the depressed group with low scores and the control group. Connections in the down panel (panel C and D) respectively represent significant PSI decrease and increase between the depressed group with high scores and the control group. (Bc, Cc, Cd and Dc) Cluster A, B, C and D identified in the control group and two depressed groups respectively were significant using a control of family-wise error rate at the level of $= 0.01$. Bd, Ce, Cf and Dd are correlation coefficient of phase synchronization within corresponding clusters. The gray panel C means that the significant level is slightly obvious.

Regarding scoring depressive severity, although the smallest RMSE result is 2.8 in 2-back in Table V when depending on beta frequency bands and particularly selected channels, the whole performance of scoring the depressive severity in Table V is weaker than that in Table IV.

V. DISCUSSION

In this study, deactivation means that the rest-state takes the dominant role, and activation presents the processing of the working memory. We found that the low depressive group presents weaker delta deactivations but stronger beta activations, while the high depressive group shows more obviously deactivated delta connections and more activated beta connections. Moreover, there are beta right central parietal functional connections appearing in depression patients when the depressive severity becomes severe. In addition, the beta frequency bands contribute to classifying depressive patients from healthy controls, and particularly selected channels tend to easily distinguish depressive patients. Relying on beta frequency bands can increase the possibilities for scoring the

depressive severity, and these selected channels also show obvious scoring advantages under the beta frequency band.

A. Possible Inducing Reason for Getting Depression

As the depressive symptoms becoming more severe, depression patients show more obvious delta deactivations and beta activations, but no evidence presents the obvious theta and alpha activities. This is a coincidence that subjects who were infected by Human Herpesvirus 6 (HHV-6) could not show a relationship with theta and alpha EEG oscillations [39]. Moreover, for patients suffering the HHV-6 infection, after medical treatment and 14 days of improvement, their theta/delta EEG oscillations will become slower [40], which means theta/delta activities become weaker [41]. Human Betaherpesvirus 6B (HHV-6B) infection can increase the potential risk of mental disorders [42], especially depression [43]. Then, we can conclude that there is a potential relationship between HHV-6 and depression. Our following research would focus on how serious HHV-6 can induce depression.

B. Topological Analysis

The approach of topological networks provides the comparison of different cognitive patterns. On the one hand, the phase coherence analysis shows that the depressive group tends to weaken the low-frequency WM activation, especially in the delta and theta frequency bands. When depressive symptoms are becoming severe from moderately depressed to severely depressed, the above finding becomes more obvious. On the other hand, in Figure 4 C and D, the beta WM activation of highly depressed patients present a significant difference when compared to beta WM activities of the slightly depressed group. Thus, the depressive group would gain stronger beta activations than the healthy controls, and highly depressive patients show more risk of suffering from this imbalance. Moreover, slightly depressive patients present a lack of delta and theta WM deactivation, and on the contrary, the highly depressive group shows redundant delta and theta WM deactivation. During the implementation of WM tasks, the depressive group reported reduced frontal-midline theta power and increased occipital upper alpha power during WM encoding [44], and this similar research perhaps provides evidence that depressive patients present abnormal brain activities in all frequency bands. The beta-frequency topological structure (Cluster D in Figure 5) of the highly depressive patients shows the extra central-parietal WM activation when compared with that (Cluster A in Figure 5) of the slightly depressive patients. This may correspond to the findings that only MDDs are characterized by unique EEG oscillations in beta frequencies. EEG beta oscillations are dominant in relation to delta, theta, and alpha when compared with healthy subjects [45], [46], and moreover, the high beta coherence is relative to connections within and between Dorsolateral Prefrontal Cortex (DLPFC) or temporal regions [25].

The increased delta deactivation during WM tasks represents the low WM loads, and this could be related to the resting recovery mechanism from cognitive maintenance. Considering the difference between Cluster B and Cluster C (Figure 5, panel C), as well as the pairs increase in the delta band (Figure 3), in terms of the condition when depressive symptoms are becoming serious, the same climbing trend of the delta deactivation was also seen in a neuromodulation therapy study [47]. Although connections of the highly depressive group in the PSI decrease show no obvious significance, the whole cerebral delta connections (Cluster B) indicate that they need more brain areas to implement WM deactivation than the control group (Cluster C). The research [47] also found that beta and gamma power increases at the Left-Dorsolateral Prefrontal Cortex (L-DLPFC) were correlated with an improvement in depressive symptoms. Increased attentional processes were proved to be connected to oscillations of the beta and gamma bands [48], and this might proof that Cluster A and Cluster D appearing in beta oscillations could adjust the attention processing of depressive subjects. When compared with Figure 3D, decreased pairs of the alpha activation in Figure 3E present the similar evidence that greater reductions of upper alpha and gamma power during WM maintenance were relative to high depressive severity [44].

TABLE VI
COMPARISON WITH EXISTING METHODS ON CLASSIFYING DEPRESSION WITH EEGs

References	Subjects	Cross validation EEGs (Scenario)	Method + Feature	Accuracy
Hanshu Cai et al (2020)[22]	MDD = 86, HC = 92	10-fold	KNN + EEGs (Fp1, Fpz, Fp2)	Highest at 86.98%
Xiaowei Zhang et al (2020)[49]	MDD = 81, HC = 89	10-fold	CNN + EEGs + demographic	Average at 75.29%
Xiaowei Li et al (2019)[50]	MDD = 24, HC = 24	24-fold	CNN + EEGs (all frequencies)	80.74% for mild
The proposed method	MDD = 48, HC = 52	10-fold	ResNet + EEGs (beta bands 16 electrodes)	Max: 87.1% and Average at 45.7%

C. Contribution of Frequency and Topological Selection for Classifying and Scoring Depressive Patients

When distinguishing depressive patients from healthy controls by a ResNet classifier in this study, the strategy depending on the single beta frequency band shows that the accuracy rate is higher than using the whole four frequency bands. Furthermore, for scoring the depressive severity in the depression group, this system presents a suitable approach to quantizing the grade of depressive severity. This finding may indirectly suggest that the beta frequency has the advantage to identify depression patients when implementing WM tasks [14]. Additionally, beta frequency cerebral activities can offer a tool to detect depression, but they cannot help to improve the performance of scoring depressive severity.

However, in the beta band, the scoring result presents wider variances. The whole frequency bands should be considered to score depressive severity only. The reason why the average accuracy rate is not high is that there are only two psychologists who diagnosed the patients and provided the results. This will cause the instability of data, especially when using the probably misdiagnosed subjects to test deep learning models.

D. State of the Art for Classifying Depressive Patients

Table VI shows the significant advantage of this proposed method, and the highest accuracy rate of detecting depression can reach 87.1%. However, the whole performance of scoring the depressive severity in V is weaker than that in IV which uses the whole frequency bands. We infer that the reason is the quality of data and the robustness of the proposed model is not strong enough. In terms of the average accuracy rate, the limitation of this proposed method is that it cannot provide stable results, and this approach relies on the psychological paradigms (n-back) which only can represent the brain function of the working memory.

E. State of the Art for Scoring Depressive Severities

Scoring of depressive severity is addressed in two studies based on MRI-related images with Partial Least Squares Regression (PLSR) and Relevance Vector Regression (RVR) [38,75]. Table VII shows that under the leave-one-out cross-validation, the minimum RMSE can reach 2.50 [51], which means the RVR+MRI method can precisely grade the depressive severity within 2.50 error. In this study, the proposed

TABLE VII

COMPARISON WITH EXISTING METHODS ON SCORING DEPRESSIVE SEVERITIES WITH EEGs

References	Subjects	Cross validation	Method + Feature	RMSE
Images (Scenario)				
Kosuke Yoshida et al (2017)[36]	MDD = 58, HC = 65	leave-one-out	PLS + sMRI	9.56
Benson et al (2012)[51]	MDD = 30, HC = 0	leave-one-out	RVR + MRI	2.50
EEGs (Scenario)				
The proposed method	MDD = 48, HC = 52	10-fold	ResNet + EEGs (beta bands 16 electrodes)	2.80

method shows a minimum RMSE of 2.80 under 10-fold cross-validation.

VI. CONCLUSION AND FUTURE WORK

In this study, we proposed a BCI system including two models based on the ResNet architecture (2) to detect depression and (1) to score the depressive severity by using 16 particularly selected channels and beta frequency EEG signals. The ResNet classifier is mainly for detecting depressive subjects from healthy controls, and the ResNet regression model is aiming to grade the depressive severity. Specifically, the coherence analysis provides the significant frequency bands, as well as the identified brain functional networks of depressive patients. We proved that the beta frequency can contribute to detecting depression and scoring depressive severity. Particularly selected EEG channels to show a significant advantage for classifying depression.

Future works will mainly focus on (1) the construction of further advanced ANNs, (2) the EEG data acquisition and selection of depressive patients, (2) more proper experiments design, and (4) the estimation of the antidepressant drug treatment. We will also aim to figure out whether there is a strong connection between inducing factors of depression and HHV-6.

REFERENCES

- [1] G. V. P. Reddy, "Depression—The global crisis," *Indian J. Psychol. Med.*, vol. 34, no. 3, p. 201, 2013.
- [2] B. F. Grant et al., "Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the national epidemiologic survey on alcohol and related conditions," *Arch. Gen. Psychiatry*, vol. 29, no. 7, pp. 807–816, 2006.
- [3] M. Vermani, M. Marcus, and M. A. Katzman, "Rates of detection of mood and anxiety disorders in primary care: A descriptive, cross-sectional study," *Primary Care Companion CNS Disorders*, vol. 13, no. 2, p. 27211, Apr. 2011.
- [4] J. C. Fournier et al., "Initial severity and depressive relapse in cognitive behavioral therapy and antidepressant medications: An individual patient data meta-analysis," *Cognit. Therapy Res.*, vol. 46, no. 3, pp. 517–531, Jun. 2022.
- [5] I. Kirsch, B. J. Deacon, T. B. Huedo-Medina, A. Scoboria, T. J. Moore, and B. T. Johnson, "Initial severity and antidepressant benefits: A meta-analysis of data submitted to the food and drug administration," *PLoS Med.*, vol. 5, no. 2, p. e45, Feb. 2008.
- [6] H. S. Sharma et al., "The 2021 yearbook of neurorestoratology," *J. Neurorestoratology*, vol. 10, no. 3, Sep. 2022, Art. no. 100008.
- [7] J. Davies and J. Read, "A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: Are guidelines evidence-based?" *Addictive Behaviors*, vol. 97, pp. 111–121, Oct. 2019.
- [8] 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.
- [9] G. Andrews and L. Peters, "The psychometric properties of the composite international diagnostic interview," *Social Psychiatry Psychiatric Epidemiology*, vol. 33, no. 2, pp. 80–88, Jan. 1998.
- [10] M. Buoli, B. M. Cesana, J. L. Barkin, G. Tacchini, and A. C. Altamura, "Validity of a clinical diagnosis of bipolar disorder among participants in a multicenter study using the mini-international neuropsychiatric interview," *Bipolar Disorders*, vol. 20, no. 3, p. 284, May 2018.
- [11] B. K. Natamba et al., "Reliability and validity of the center for epidemiologic studies-depression scale in screening for depression among HIV-infected and-uninfected pregnant women attending antenatal services in northern Uganda: A cross-sectional study," *BMC Psychiatry*, vol. 14, no. 1, pp. 1–8, Dec. 2014.
- [12] C. Kang, Y. Li, D. Novak, Y. Zhang, Q. Zhou, and Y. Hu, "Brain networks of maintenance, inhibition and disinhibition during working memory," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1518–1527, Jul. 2020.
- [13] Y. Li, C. Kang, X. Qu, Y. Zhou, W. Wang, and Y. Hu, "Depression-related brain connectivity analyzed by EEG event-related phase synchrony measure," *Frontiers Human Neurosci.*, vol. 10, p. 477, Sep. 2016.
- [14] Y. Li et al., "Beta oscillations in major depression—signalling a new cortical circuit for central executive function," *Sci. Rep.*, vol. 7, no. 1, p. 18021, Dec. 2017.
- [15] H. Mizuhara, L.-Q. Wang, K. Kobayashi, and Y. Yamaguchi, "Long-range EEG phase synchronization during an arithmetic task indexes a coherent cortical network simultaneously measured by fMRI," *NeuroImage*, vol. 27, no. 3, pp. 553–563, Sep. 2005.
- [16] H. Mizuhara and Y. Yamaguchi, "Human cortical circuits for central executive function emerge by theta phase synchronization," *NeuroImage*, vol. 36, no. 1, pp. 232–244, May 2007.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [18] J. Jeganathan, A. Perry, D. S. Bassett, G. Roberts, P. B. Mitchell, and M. Breakspear, "Fronto-limbic dysconnectivity leads to impaired brain network controllability in young people with bipolar disorder and those at high genetic risk," *NeuroImage, Clin.*, vol. 19, pp. 71–81, Jan. 2018.
- [19] S. M. Strakowski, C. M. Adler, S. K. Holland, N. P. Mills, M. P. DelBello, and J. C. Eliassen, "Abnormal fMRI brain activation in euthymic bipolar disorder patients during a counting stroop interference task," *Amer. J. Psychiatry*, vol. 162, no. 9, pp. 1697–1705, Sep. 2005.
- [20] M. E. Lynall et al., "Functional connectivity and brain networks in schizophrenia," *J. Neurosci.*, vol. 30, no. 28, pp. 9477–9487, 2010.
- [21] H. Cai et al., "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, pp. 1–13, Jan. 2018.
- [22] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal EEG data for depression recognition," *Inf. Fusion*, vol. 59, pp. 127–138, Jul. 2020.
- [23] S. D. Puthankattil and P. K. Joseph, "Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy," *J. Mech. Med. Biol.*, vol. 12, no. 4, Sep. 2012, Art. no. 1240019.
- [24] S. Aydin, "Cross-validated adaboost classification of emotion regulation strategies identified by spectral coherence in resting-state," *Neuroinformatics*, vol. 20, pp. 627–639, Sep. 2021.
- [25] A. F. Leuchter, I. A. Cook, A. M. Hunter, C. Cai, and S. Horvath, "Resting-state quantitative electroencephalography reveals increased neurophysiologic connectivity in depression," *PLoS ONE*, vol. 7, no. 2, Feb. 2012, Art. no. e32508.
- [26] B.-J. Li, K. Friston, M. Mody, H.-N. Wang, H.-B. Lu, and D.-W. Hu, "A brain network model for depression: From symptom understanding to disease intervention," *CNS Neurosci. Therapeutics*, vol. 24, no. 11, pp. 1004–1019, Nov. 2018.
- [27] Y. Cui, M. Jia, T. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9260–9269.
- [28] P. Beloe and N. Derakshan, "Adaptive working memory training can reduce anxiety and depression vulnerability in adolescents," *Develop. Sci.*, vol. 23, no. 4, p. e12831, Jul. 2020.

- [29] S. J. E. Bruijnicks et al., "Working memory moderates the relation between the brain-derived neurotrophic factor (BDNF) and psychotherapy outcome for depression," *J. Psychiatric Res.*, vol. 130, pp. 424–432, Nov. 2020.
- [30] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, Mar. 2013.
- [31] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, Jun. 2017.
- [32] G. M. Bairy et al., "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features," *J. Med. Imag. Health Informat.*, vol. 7, no. 8, pp. 1857–1862, Dec. 2017.
- [33] 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. Neurodynamics*, vol. 15, pp. 239–252, Jul. 2020.
- [34] Z. Wan, J. Huang, H. Zhang, H. Zhou, J. Yang, and N. Zhong, "HybridEEGNet: A convolutional neural network for EEG feature learning and depression discrimination," *IEEE Access*, vol. 8, pp. 30332–30342, 2020.
- [35] D. Shah, G. Y. Wang, M. Dobarjeh, Z. Dobarjeh, and N. Kasabov, "Deep learning of EEG data in the NeuCube brain-inspired spiking neural network architecture for a better understanding of depression," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2019, pp. 195–206.
- [36] K. Yoshida et al., "Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0179638.
- [37] M. B. First and M. Gibbon, "The structured clinical interview for DSM-IV axis I disorders (SCID-I) and the structured clinical interview for DSM-IV axis II disorders (SCID-II)," in *Comprehensive Handbook of Psychological Assessment. Personality Assessment*, vol. 2. Hoboken, NJ, USA: Wiley, 2004. [Online]. Available: <https://psycnet.apa.org/record/2004-12821-011>
- [38] J. C. Tolentino and S. L. Schmidt, "DSM-5 criteria and depression severity: Implications for clinical practice," *Frontiers Psychiatry*, vol. 9, p. 450, Oct. 2018.
- [39] M. Tanaka, Y. Shigihara, M. Funakura, E. Kanai, and Y. Watanabe, "Fatigue-associated alterations of cognitive function and electroencephalographic power densities," *PLoS ONE*, vol. 7, no. 4, Apr. 2012, Art. no. e34774.
- [40] A. Yassin et al., "Clinical, radiological, and electroencephalographic features of HHV-6 encephalitis following hematopoietic stem cell transplantation," *Ann. Med. Surgery*, vol. 60, pp. 81–86, Dec. 2020.
- [41] F. Zhang et al., "Therapeutic effects of subthalamic nucleus deep brain stimulation on anxiety and depression in Parkinson's disease patients," *J. Neurorestoratology*, vol. 10, no. 1, pp. 31–42, 2022.
- [42] B. K. Prusty et al., "Active HHV-6 infection of cerebellar purkinje cells in mood disorders," *Frontiers Microbiology*, vol. 9, p. 1955, Aug. 2018.
- [43] N. Kobayashi et al., "Human herpesvirus 6B greatly increases risk of depression by activating hypothalamic-pituitary-adrenal axis during latent phase of infection," *iScience*, vol. 23, Jun. 2020, Art. no. 101187.
- [44] O. W. Murphy, K. E. Hoy, D. Wong, N. W. Bailey, P. B. Fitzgerald, and R. A. Segrave, "Individuals with depression display abnormal modulation of neural oscillatory activity during working memory encoding and maintenance," *Biol. Psychol.*, vol. 148, Nov. 2019, Art. no. 107766.
- [45] A. A. Fingelkurts and A. A. Fingelkurts, "Altered structure of dynamic electroencephalogram oscillatory pattern in major depression," *Biol. Psychiatry*, vol. 77, no. 12, pp. 1050–1060, Jun. 2015.
- [46] A. A. Fingelkurts, A. A. Fingelkurts, H. Rytälä, K. Suominen, E. Isometsä, and S. Kähkönen, "Composition of brain oscillations in ongoing EEG during major depression disorder," *Neurosci. Res.*, vol. 56, no. 2, pp. 133–144, Oct. 2006.
- [47] Y. Pathak, O. Salami, S. Baillet, Z. Li, and C. R. Butson, "Longitudinal changes in depressive circuitry in response to neuromodulation therapy," *Frontiers Neural Circuits*, vol. 10, p. 50, Jul. 2016.
- [48] K. Benchenane, P. H. Tiesinga, and F. P. Battaglia, "Oscillations in the prefrontal cortex: A gateway to memory and attention," *Current Opinion Neurobiol.*, vol. 21, no. 3, pp. 475–485, Jun. 2011.
- [49] X. Zhang, J. Li, K. Hou, B. Hu, J. Shen, and J. Pan, "EEG-based depression detection using convolutional neural network with demographic attention mechanism," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 128–133.
- [50] X. Li et al., "EEG-based mild depression recognition using convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 57, no. 6, pp. 1341–1352, Jun. 2019.
- [51] B. Mwangi, K. Matthews, and J. D. Steele, "Prediction of illness severity in patients with major depression using structural MR brain scans," *J. Magn. Reson. Imag.*, vol. 35, no. 1, pp. 64–71, Jan. 2012.