NEUROPHOTONICS

RESEARCH PAPER

Functional near-infrared spectroscopy-based diagnosis support system for distinguishing between mild and severe depression using machine learning approaches

Zhiyong Huang,[a,*,†] Man Liu,[a] Hui Yang,[b,*,†] Mengyao Wang,[a] Yunlan Zhao,[a] Xiao Han,[a] Huan Chen,[c] and Yaju Feng[c]

[a]Chongqing University, School of Microelectronics and Communication Engineering, Chongqing, China
[b]Chongqing University Central Hospital, Chongqing Emergency Medical Center, Chongqing, China
[c]Chongqing Mental Health Center, Department of Clinical Psychology, Chongqing, China

ABSTRACT. **Significance:** Early diagnosis of depression is crucial for effective treatment. Our study utilizes functional near-infrared spectroscopy (fNIRS) and machine learning to accurately classify mild and severe depression, providing an objective auxiliary diagnostic tool for mental health workers.

Aim: Develop prediction models to distinguish between severe and mild depression using fNIRS data.

Approach: We collected the fNIRS data from 140 subjects and applied a complete ensemble empirical mode decomposition with an adaptive noise-wavelet threshold combined denoising method (CEEMDAN-WPT) to remove noise during the verbal fluency task. The temporal features (TF) and correlation features (CF) from 18 prefrontal lobe channels of subjects were extracted as predictors. Using recursive feature elimination with cross-validation, we identified optimal TF or CF and examined their role in distinguishing between severe and mild depression. Machine learning algorithms were used for classification.

Results: The combination of TF and CF as inputs for the prediction model yielded higher classification accuracy than using either TF or CF alone. Among the prediction models, the SVM-based model demonstrates excellent performance in nested cross-validation, achieving an accuracy rate of 92.8%.

Conclusions: The proposed model can effectively distinguish mild depression from severe depression.

Keywords: severity of depression; functional near-infrared spectroscopy; machine learning; CEEMDAN-WPT

Paper 23076GRR received Sep. 12, 2023; revised Mar. 14, 2024; accepted Mar. 28, 2024; published Apr. 24, 2024.

1 Introduction

Depression is a prevalent psychological disorder characterized by persistent feelings of sadness and loss of interest in activities, significantly impacting the lives and work of individuals during the onset of the illness. Recognized as a pressing concern by the WHO's Mental Health Gap Action Program (mhGAP),[1] ~3.8% of the global population suffers from this condition.

*Address all correspondence to Zhiyong Huang, zyhuang@cqu.edu.cn; Hui Yang, yanghui0925@126.com

†Zhiyong Huang and Hui Yang are equally responsible for this study.

The severity of depression, as assessed by the Hamilton Depression Rating Scale (HAMD),[2] can be roughly categorized as mild, moderate, and severe. However, the diagnosis process often involves subjective evaluations, such as clinical interviews and professional medical assessments, which may impede the timely and accurate diagnosis of severe depression. Moreover, due to the shortage of mental health workers, only 9.2% of individuals with depression receive appropriate treatment, whereas some individuals with anxiety disorders or mild depression are mistakenly prescribed antidepressant medications.[3] Research[4] has indicated that antidepressants are ineffective in treating mild depression. Thus, there is an urgent need to develop an objective and efficient approach to quickly differentiate between mild and major depression for early diagnosis and effective treatment evaluation.

Several studies[5,6] have found evidence of structural and functional changes in various brain regions among individuals with severe depression. These changes include abnormal activity in prefrontal, limbic, thalamic, and cortical areas. In prefrontal regions, increased depression severity is associated with decreased neural activity in the low-frequency range and increased activity in the high-frequency range.[7] In addition, severe depression is characterized by extremely low mood, which may be accompanied by severe sleep problems, loss of appetite, and loss of interest in daily activities. Even for mild depression, the changes in the brain and physiology may be relatively mild but can still lead to problems with mood, concentration, and cognition.[8,9] Depression severity has been found to be significantly associated with cognitive performance in situational memory, executive functioning, and processing speed, and as depression severity increases, these cognitive abilities decline.[10] Impaired cognitive function in depressed patients results in reduced regional cerebral blood flow (rCBF) in the frontal, temporal, and anterior cingulate gyrus during cognitive tasks.[11] Functional near-infrared spectroscopy (fNIRS) serves as a non-invasive, portable tool for monitoring hemodynamic responses in the cerebral cortex, commonly utilized to observe hemodynamic changes in psychiatric patients undertaking cognitive tasks due to its simplicity, safety, and resistance to interference. When performing a cognitive task, the brain experiences an oversupply of rCBF, leading to the hemodynamic response that can be measured through fNIRS.[12] Researchers[13–15] found that values of oxyhemoglobin concentration in the prefrontal lobe measured by the fNIRS device were negatively correlated with HAMD scores, which were associated with depression severity, suggesting that changes in oxygen-hemoglobin are a potential biomarker for recognizing depression severity.

In recent years, several researchers[16,17] utilized machine learning techniques to construct prediction models that differentiate between severe depression and healthy individuals based on fNIRS data, obtaining favorable classification performance. Although these studies[13,18–20] suggest that machine learning prediction models developed based on fNIRS data serve as effective analytical tools for identifying patients with severe depression, few studies have focused on discriminating depression severity. Ramasubbu et al.[21] developed a predictive model for distinguishing between patients with mild depression and severe depression but did not achieve high accuracy. Richter et al.[22] proposed a diagnostic system (25 clinical anxiety or depression patients, 76 healthy control participants) and achieved an accuracy of 81.69% in distinguishing the mixed anxiety/depression group from the control group, but it only achieved a success rate of 50.66% in differentiating the anxiety group from the depression group.

Inspired by previous researches,[14,15,23] we found that fNIRS measurement can be instrumental in distinguishing the severity of depression. In this study, we proposed a predictive model based on fNIRS data to differentiate between patients with mild and severe depression. This model provides mental health professionals with an objective and effective diagnosis method for determining the severity of depression.

2 Materials

2.1 Participants

In this study, 140 patients with drug-naïve, first-episode depression were recruited with their permission and informed consent. All subjects had no other neurological disorders (stroke, brain tumor, severe concussion, migraine, etc.), cardiovascular (myocardial infarction, arrhythmia, etc.) diseases, and no obvious impairment in vision and hearing. The experiment was conducted under the guidance of a professional psychiatrist. No data identifying the participants were

Table 1 Subject information of our data.

	Severe depression	Mild depression
Number	82	58
Gender	54 females, 28 males	39 females, 19 males
Age (years)	32.48 ± 17.71	36.65 ± 19.97

recorded during the entire research process. A senior professional psychiatrist conducted interviews and observations, utilizing the HAMD[2] scale to evaluate depressive symptom severity. According to the diagnosis of the psychiatrist, among these 140 first-episode depression patients, 58 individuals were assessed as mildly depressed, whereas the remaining 82 subjects were classified as having severe depression. More details of the participants are shown in Table 1.

2.2 Activation Task (Verbal Fluency Task)

The verbal fluency task (VFT) is a cognitive task commonly used in fNIRS research, which primarily reflects executive function and has been correlated with a number of basic neurocognitive activities (e.g., working memory, motivation, and attention).[24] In China, the Chinese version of the VFT has been widely used in the diagnosis of psychiatric disorders as a sensitive indicator for assessing deficits in domains of cognition and executive function that depend on activation in prefrontal regions.[25] Before the VFT task, the participants were instructed to stay in a separate, quiet room and sit on a chair, ensuring that their eyes were directly in front of the monitor at a distance of 60 cm. Participants were asked to look at the "+" symbol on the monitor and keep their heads still. At the beginning of each experiment, during a 30-s rest period before the task, a guiding voice prompt was played: "Check begins, please repeat the reading 12345." Concurrently, the fNIRS data were recorded. Next, during a 60-s task period, the participants were required to complete three tasks involving the formation of words using the three characters "zhong," "ri," and "lan," respectively. Each word formation task had a time limit of 20 s. Finally, during a 30-s resting period after the task completion, the instructional audio prompt stating, "Please repeat reading 12345 until the end of the check" was played and the fNIRS data acquisition was terminated. The entire VFT process is shown in Fig. 1(b).

2.3 fNIRS Data Acquisition

In the experiments, an fNIRS instrument (ETG-4100 Optical Topography System) manufactured by Hitachi Medical is utilized to obtain data. This device utilizes near-infrared light at two specific wavelengths (659 ± 20 and 830 ± 20 nm) to penetrate the scalp, skull, and cerebrospinal fluid, enabling irradiation of the cerebral cortex. The instrument consists of 17 transmitters and 16 detectors, forming a total of 52 channels, including 18 in the prefrontal lobes and 17 channels each in the right and left frontal lobes. The distance between each pair of transmitters and detectors was 3.0 cm, with a sampling frequency of the instrument was 10 Hz. During the VFT, neural activity leads to changes in CBF, causing changes in brain oxygen-hemoglobin (ΔHbO) and deoxygen-hemoglobin (ΔHbR).[6] According to Lambert-Beer,[26] changes in ΔHbO and ΔHbR can be deduced from the NIRS signals measured during the task. These changes provide a more accurate reflection of the activation state of the cerebral cortex in relation to cognitive tasks. The entire process of data collection in the experiment is shown in Fig. 1.

3 Methods

To predict the severity of depression, we developed a prediction model for mild and severe depression. The model involves data preprocessing, feature extraction, feature selection, feature fusion, and classification prediction, as shown in Fig. 2.

3.1 Data Pre-Processing

We utilized the HOMER2 toolbox in MATLAB 2018b to process NIRS data and obtained the change in ΔHbO and ΔHbR using the Beer-Lambert law.[27] These hemodynamic parameters are

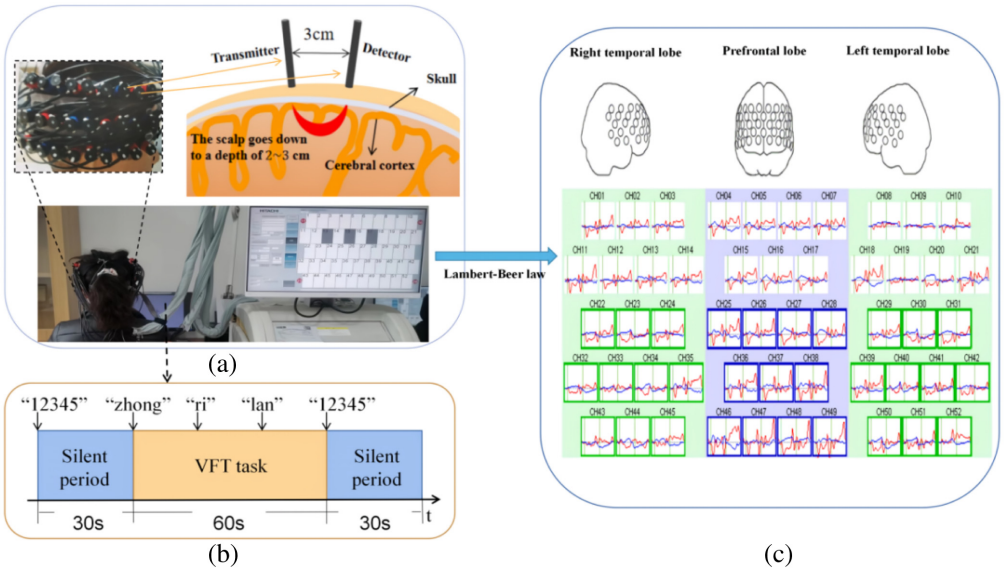

Fig. 1 The experimental data collection process. (a) Details of the probe for the fNIRS device. (b) The VFT consisted of three stages: a 30-s pre-task baseline period, a 60-s VFT task period, and a 30-s post-task baseline period. (c) The ΔHbO and ΔHbR curves for 17 channels in the right temporal lobe, 18 channels in the prefrontal lobe, and 17 channels in the left temporal lobe.

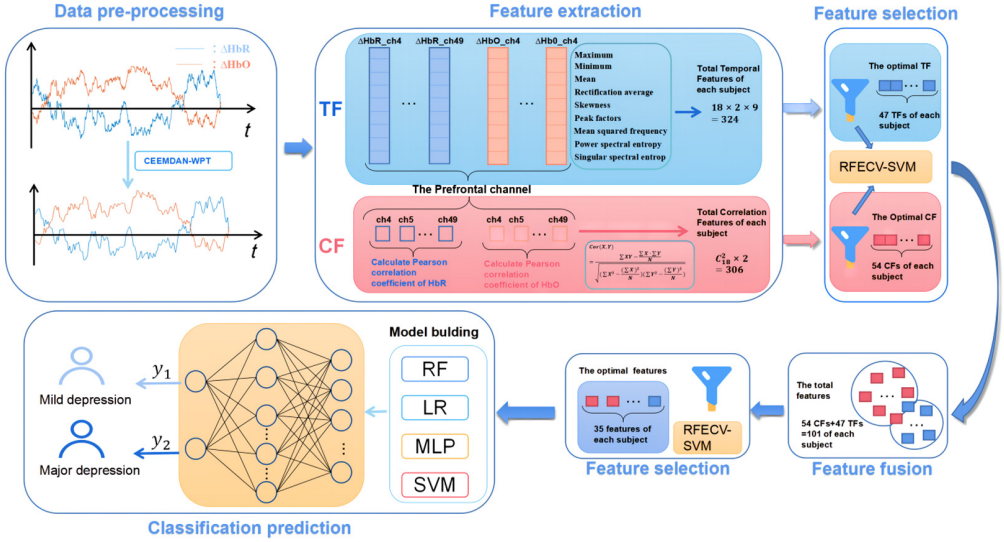

Fig. 2 The overview of the model.

recognized as sensitive indicators for investigating psychiatric disorders.[28] However, the raw hemodynamic signals often contain various types of noise that can affect the feature extraction process. Therefore, preprocessing of the fNIRS data is essential to remove noise before further analysis.[29] We utilized a polynomial regression model to estimate linear or nonlinear trends and then subtracted this trend from the original hemoglobin concentration signal to obtain detrended data. The temporal derivative distribution repair method[30] was employed for motion artifact correction. In addition, a third-order Chebyshev type II filter[31] with a cutoff frequency of 0.01 Hz and a stopband frequency of 0.2 Hz was used to remove most physiological noise [e.g., respiration (0.2 to 0.4 Hz) and heartbeat (0.5 to 2.0 Hz)]. However, various artifacts, such as ambient light, electrical interference from the instrumentation, and positional movements, still contaminated the effective frequency band. In previous studies,[32,33] empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD) methods have been used to

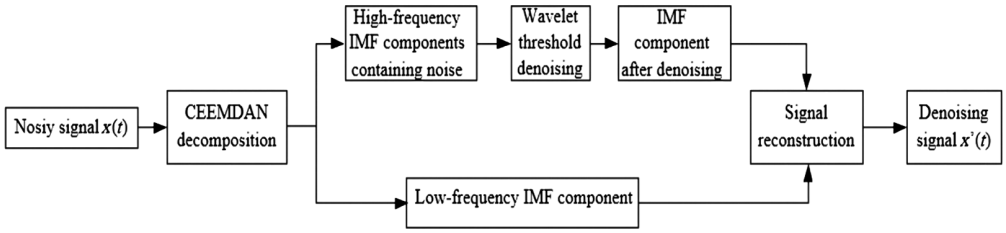

Fig. 3 The overall process of the CEEMDAN-WPT.

eliminate motion artifacts and baseline drifts in physiological signals. In this study, to enhance the signal-to-noise ratio (SNR) and facilitate subsequent feature extraction, we applied a complete EEMD with adaptive noise-wavelet threshold (CEEMDAN-WPT) denoising method to eliminate noise contaminating the effective frequency band of fNIRS data. The flowchart of the CEEMDAN-WPT algorithm is shown in Fig. 3, and the specific implementation steps are detailed below.

1. Suppose $E_i(\cdot)$ represent the i'th intrinsic mode function (IMF) component operation calculated by the EMD algorithm. White noise that conforms to the standard distribution is added to the original signal, resulting in I initial signal $x^i(t)$:

$$x^i(t) = x(t) + \varepsilon_i \omega^i(t). \tag{1}$$

   Here, $x(t)$ represents the original signal, ε_i is the white noise amplification factor, and $\omega^i(t)$ denotes the white noise following a standard Gaussian distribution.

2. Perform EMD decomposition on all initial signals to obtain the first modal component:

$$\overline{\mathrm{IMF}_1(t)} = \frac{1}{I}\sum_{i=1}^{I} \mathrm{IMF}_1^i(t). \tag{2}$$

   Then, calculate the first-order remaining components:

$$r_1(t) = x(t) - \overline{\mathrm{IMF}_1(t)}. \tag{3}$$

3. Add $E_1(\omega^i(n))$ to $r_1(t)$, obtaining the new signal $r_1(t) + \varepsilon_1 E_1(\omega^i(n))$, and perform EMD decomposition again to calculate the second residual component:

$$\overline{\mathrm{IMF}_2(t)} = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t)) + \varepsilon_1 E_1(\omega^i(n)). \tag{4}$$

   Then, calculate the second-order residual components:

$$r_2(t) = r_1(t) - \overline{\mathrm{IMF}_2(t)}. \tag{5}$$

4. Repeat the above steps until the number of poles of the obtained residual signal is less than two, indicating the end of the decomposition algorithm. At this stage, the original signal $x(t)$ has been decomposed into

$$x(t) = \sum_{k=1}^{K} \overline{\mathrm{IMF}_k} + R(t). \tag{6}$$

5. Carry out wavelet threshold denoising on the noisy high-frequency component IMF_k:

$$\begin{cases} y = W(\mathrm{IMF}_k) \\ \tilde{y} = \tilde{D}(y, \lambda) \\ \overline{\mathrm{IMF}_k} = W(\tilde{y}) \end{cases}. \tag{7}$$

   In the equation, W and $\overline{W}$ are the wavelet packet change and its inverse operation, respectively. $\tilde{D}$ and λ are the threshold function and threshold magnitude, respectively, and $\overline{\mathrm{IMF}_k}$ denotes the high-frequency component after denoising processing.

6. The denoising high-frequency component $\overline{\mathrm{IMF}_k}$ and the low-frequency component IMF are reconstructed to obtain the denoising signal $x'(t)$.

3.2 Methodology for the Effective Feature and Identified Model

In our study, we employed a methodology to extract key features for effective classification of patients with mild and severe depression. This involved identifying critical features from the data obtained from channels. Subsequently, we established a recognition model using these key features to classify patients with severe depression and those with mild depression. Figure 4 shows a detailed description of feature selection and classification model construction process in this study. To ensure an objective assessment of the classification performance, the analysis was conducted exclusively on the training set within the blue-bordered box, whereas the final selected features were validated on the test set within the orange-bordered box. For comprehensive model evaluation, we utilized nested cross-validation (black-bordered box) to obtain model metrics. The principle of nested cross-validation, as shown in Fig. 4, involves two loops: an outer loop and an inner loop. The inner loop consists of cross-validation and search for the best hyperparameters of the model, such as grid search, which determines the optimal hyperparameters for the outer loop. Meanwhile, the outer loop provides training data to the inner loop while retaining some data for testing the models within the inner loop. The entire process was repeated for all possible fold combinations within the outer loop, providing a robust estimation of the model performance. This approach prevents information leakage and overfitting.

3.2.1 *Feature extraction*

Temporal feature. Considering that the actual task time for VFT is only 60 s, we extracted the temporal features (TF) within this time period. These features encompass time domain characteristics, frequency domain attributes, and information entropy. The time domain features consist of maximum, minimum, mean, rectification average, skewness, and peak factors. Mean squared frequency was selected as frequency domain feature. Detailed information about the selected time and frequency domain features can be found in Table S1 in the Supplementary

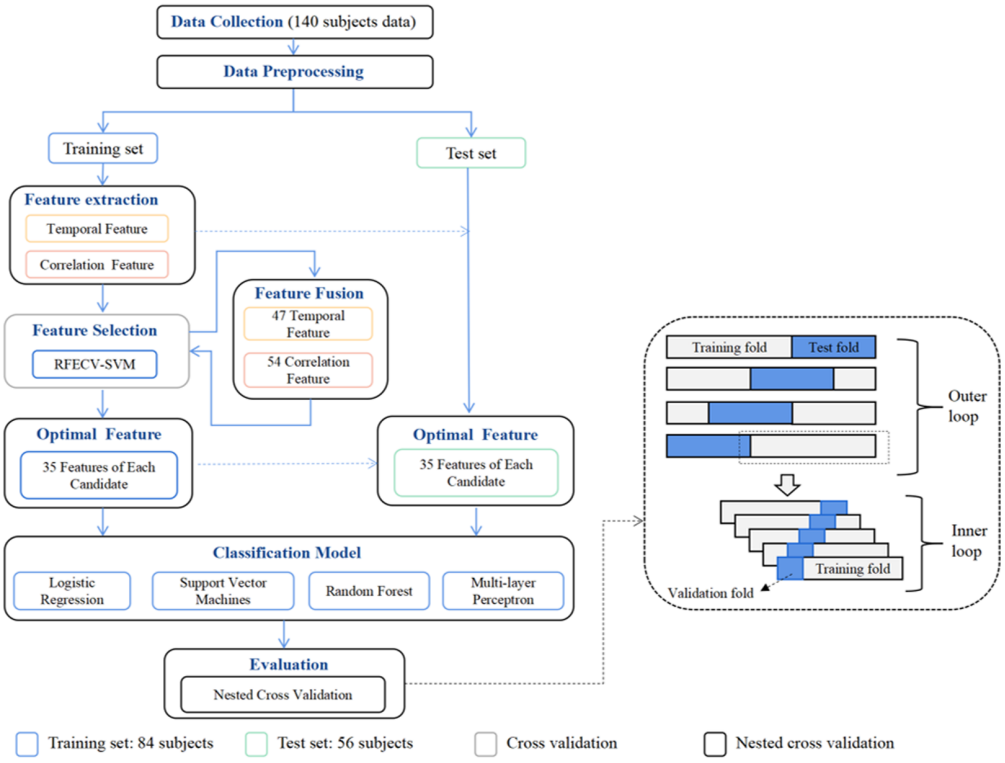

Fig. 4 The procedure for feature selection and model training.

Material. Power spectral entropy and singular spectral entropy serve as the information entropy features, calculated as follows:

$$H_f = -\sum_{k=0}^{k-1}\left(S_k \sum_{k=1}^{N} S_k\right) \log\left(S_k \sum_{k=1}^{N} S_k\right). \tag{8}$$

In the equation, H_f represents the power spectral entropy, and S_k denotes the energy distribution of the signal in the frequency domain:

$$H_t = -\sum_{i=1}^{l}\left(\frac{\lambda_i}{\sum_{i=1}^{l}\lambda_i}\right) \log\left(\frac{\lambda_i}{\sum_{i=1}^{l}\lambda_i}\right), \tag{9}$$

where H_t is the singular spectral entropy and λ_i represents the singular value spectrum obtained through the singular value decomposition of the signal.

Correlation feature. In our study, the correlation among channels was extracted as the correlation features (CF). These were obtained by calculating the Pearson correlation coefficient[19] of the change in ΔHbO and ΔHbR between 18 frontal channels of each participant. The equation for calculating the CF is as follows:

$$\mathrm{Cor}(X,Y) = \frac{\sum XY - \frac{\sum X \cdot \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right)}}, \tag{10}$$

where X and Y represent the concentration change of hemoglobin in different channels of the participant, respectively.

Therefore, a total of $18 \times 2 \times 9 = 324$ TF and $C_{18}^2 \times 2 = 306$ CF can be extracted from the change curves of ΔHbO and ΔHbR in the 18 frontal channels of each participant.

3.2.2 *Feature selection*

In the classification task, when the sample size is small and the number of features is large, it can lead to the problem of dimensionality curse and reduce the accuracy. To address this issue, recursive feature elimination with cross-validation (RFECV) method[34] can be utilized to select the optimum features from TF and CF. The RFECV method utilizes RFE[35] to obtain the importance of each feature and find the optimal number of features based on cross-validation accuracy. Compared to dimension reduction methods, such as PCA,[36] the RFECV method can eliminate redundant feature information and determine the most impactful features for optimal classification performance. In our prediction model, we applied REFCV method twice. First, the optimal 47 TF and 54 CF were screened separately. Then, after fusing the optimal CF and TF selected in the first step, we further identified the best 35 features for distinguishing between mild and severe depression.

3.2.3 *Classification models*

Given the limited number of samples in depression data, we employed machine learning techniques to construct the classification prediction. Four classical algorithms were utilized, including two supervised learning methods [logistic regression (LR)[37] and support vector machines (SVM)[38]], an ensemble learning algorithm [random forest (RF)],[39] and an artificial neural networks [multi-layer perceptron (MLP)].[40] According to the participants, the dataset was split into two parts: a training set comprising 60% (49 severe depression and 35 mild depression) and a testing set comprising 40% (33 severe depression and 23 mild depression). To identify the optimal prediction model for each pattern, a grid search[41] based on hyperparameter turning[42] was employed to determine the model parameter with the highest accuracy.

3.3 Model Metrics

The denoising performance of the CEEMDAN-WPT algorithm is evaluated using two metrics: SNR and root mean square error (RMSE). SNR quantifies the relationship between signal and noise intensity, where a high SNR indicates stronger useful signals compared to noise, resulting in clearer and more reliable signals. RMSE measures the average deviation between predicted values and actual values. The equations for calculating SNR and RMSE are as follows:

$$\mathrm{SNR} = 10\ \lg\left(\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}[x(n)-y(n)]^2}\right), \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[x(n)-y(n)]^2}, \tag{12}$$

where $x(n)$ represents the original signal, n is the the signal length, and $y(n)$ denotes the denoised signal.

For the classification models, we calculated the area under the ROC curve (AUC) for different models to compare their performance differences. In addition, several other metrics, including specificity, sensitivity, accuracy, and $F1$ score, were calculated. Supplementary Material provide detailed calculation methods for these metrics.

4 Results and Analysis

4.1 Result on Denoising

As shown in Fig. 5, the original signal is decomposed into five high-frequency components and five low-frequency components. The high-frequency components denoised by wavelet packet threshold and low-frequency components are reconstructed into a new signal, which is visibly smoother. Figure 6 shown the spectrum of the original signal and the signal processed using the CEEMDAN-WPT method. In Fig. 6(a), it can be observed that even after filtering, a small amount of noise remains mixed within the effective frequency range. However, our proposed method effectively removes this noise while retaining the information relevant to neural activity. In Table 2, compared to EMD,[43] the CEEMDAN-WPT method significantly improves SNR and reduces the RMSE. Compared to wavelet filtering,[44] although there is a slight difference in RMSE, the SNR is improved by 4.1 dB.

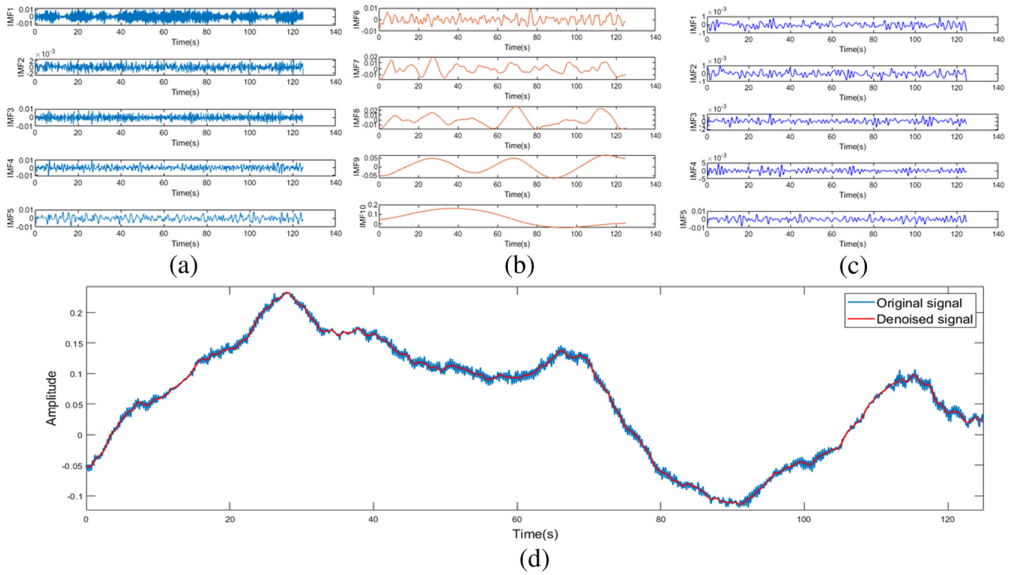

Fig. 5 Noise reduction signal comparison. (a) IMF1 through IMF5 represent the high-frequency components, (b) IMF6 through IMF10 represent the low-frequency components; (c) the high-frequency components after denoising. (d) Comparison diagram of the original signal and the denoised signal.

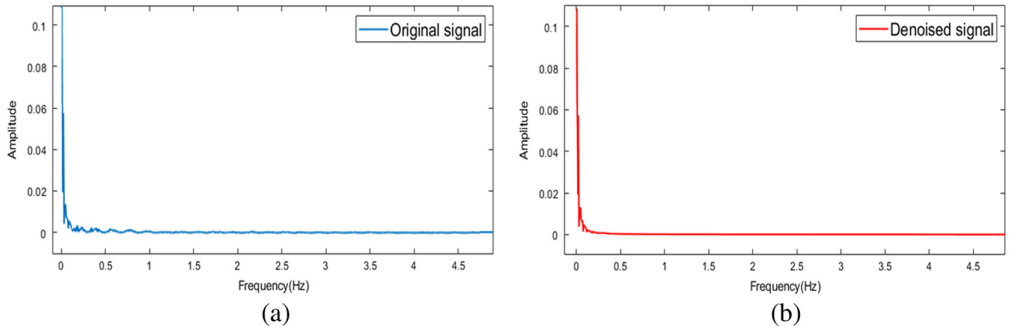

Fig. 6 The spectrum of the original signal and the denoised signal; (a) represents the one-sided spectrum of the original signal and (b) illustrates the one-sided spectrum of the denoised signal.

Table 2 Performance comparison of three algorithms.

Method	SNR	RMSE
EMD	5.8842	0.1675
Wavelet filtering	16.8702	0.0297
CEEMDAN-WPT	20.9767	0.0258

4.2 Result on Feature Selection

Through the RFECV algorithm, we identified 35 optimal features that are related to the severity of depression. As shown in Fig. 7, we plotted some violin figures regarding the features. It can be

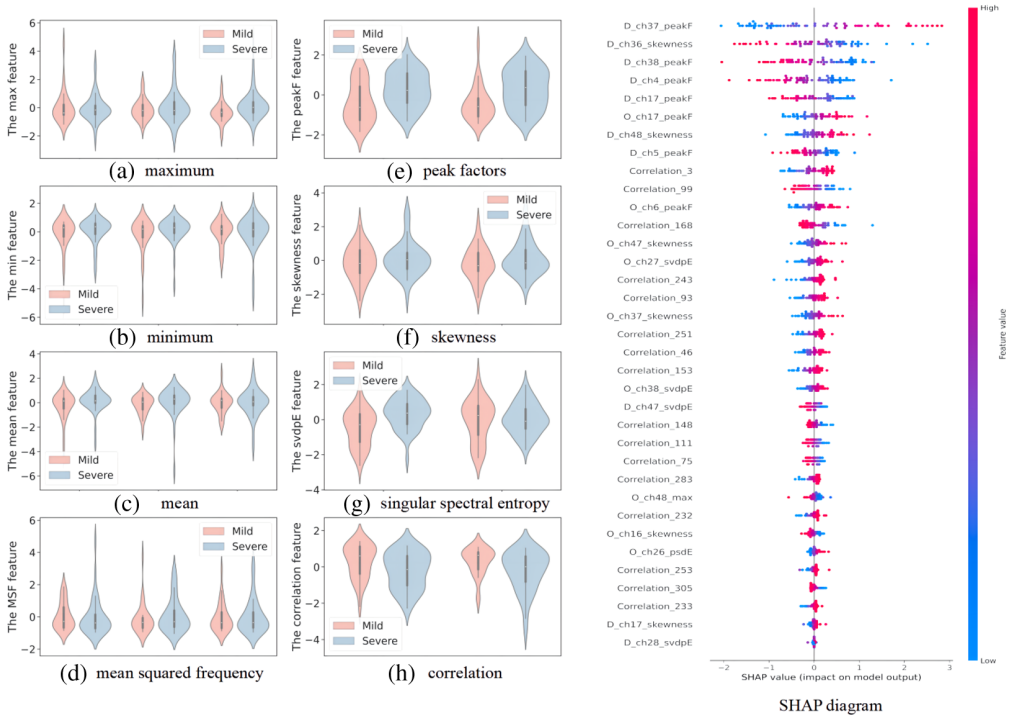

Fig. 7 Violin and SHAP diagram about features. (a) through (d) represent the maximum, minimum, average, and root mean square frequency, respectively. (e) through (h) correspond to the peak factor, skewness, singular spectrum entropy, and correlation coefficient, respectively. (a) through (d) visualize the features that were excluded, while (e) through (h) showcase features selected by RFECV. The SHAP diagram ranks the importance of 35 optimal features, depicting the distribution of their impact on the prediction results along the *X*-axis. Each point represents a sample, and the color represents the feature value. For instance, the first row indicates that high peak factor has a positive impact on the prediction results, whereas low peak factor has a negative impact.

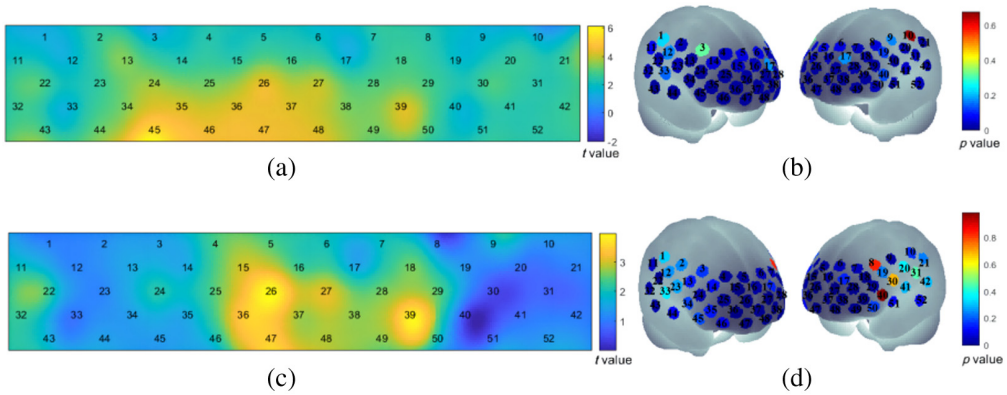

Fig. 8 Analysis of channel activation status and activation intensity in mild and severe depression. (a) The heat map of *t*-values representing the brain channel activations during the VFT, where the blue regions indicate non-activated channels. (b) Dark blue channels ($p < 0.05$) indicate activated channels. (c) The heat map of *t*-values compares the intensity of activation between the mild depression and severe depression groups, with blue regions indicating no significant differences between the two groups. (d) Dark blue channels ($p < 0.05$) represent significant differences between the mild depression and severe depression groups.

seen from the figure that there are significant differences between mild depression and severe depression subjects in terms of peak factor, skewness, singular spectral entropy, and correlation in certain channels. Conversely, the differences in maximum, minimum, mean, and mean square frequency are relatively small. The SHAP diagram in Fig. 7 demonstrated that peak factor and skewness in certain channels have a significant impact on the prediction results, whereas the influence of maximum, minimum, average, and mean square frequency is relatively minor.

4.3 Activation Analysis

Individual-level statistical analysis was performed on preprocessed fNIRS data using the general linear model to calculate beta values representing brain activation intensity during the VFT task. The activation levels of individual channels were evaluated using a one-sample *t*-test, and the heat map of channel activation *t*-values is presented in Fig. 8(a). Channels, such as ch1, ch3, ch9, ch10, ch11, ch12, ch17, ch19, ch33, and ch40, did not show significant activation during the VFT task, whereas the remaining channels exhibited significant activation. In Fig. 8(a), 52 channels were mapped onto the brain, with non-activated channels mainly located in the left and right temporal lobes. To compare brain activation intensity between the severe depression and mild depression groups, a paired *t*-test was conducted. The results indicated that, during the VFT task, the activation levels in the mild depression group were higher than those in the severe depression group. Notably, there was a significant difference in channel activation intensity in the frontal lobe region, as depicted in Figs. 8(c) and 8(d). The Supplementary Material contain detailed results of both the one-sample and paired sample *t*-tests.

4.4 Result on Classification Model

We evaluated the TF, the CF, and the fusion of both as inputs for the four classification models and validated the results on the test set, as shown in Table 3. Across all models, the fusion of temporal and CFs consistently outperformed individual temporal or CFs. Furthermore, in LR, MLP, and SVM models, the AUC of the fusion features was greater than 0.9, indicating its ability to better distinguish between mild and severe depression. Notably, the fusion features of temporal and correlation fusion features showed outstanding performance in the SVM classification model, with an accuracy of 92.8%, a sensitivity of 91.6%, a specificity of 93.7%, and an $F1$-score of 92.7%.

5 Discussion

Cerebral blood flow (CBF) is associated with brain activity and metabolism. During cognitive tasks, rCBF increases[12] while significant changes in CBF occur in patients with depression,

Table 3 The results of classification models.

Method		Classification metrics				
Model	Feature	AUC	Accuracy	Sensitivity	Specificity	*F*1-score
RF	TF	0.757	0.678	0.333	0.937	0.619
	CF	0.645	0.642	0.333	0.875	0.590
	TCF	0.739	0.714	0.50	0.875	0.688
LR	TF	0.770	0.714	0.50	0.875	0.688
	CF	0.911	0.821	0.833	0.812	0.819
	TCF	0.927	0.857	0.666	1.0	0.844
MLP	TF	0.916	0.821	0.583	1.0	0.801
	CF	0.906	0.821	0.750	0.875	0.815
	TCF	0.968	0.892	0.750	1.0	0.885
SVM	TF	0.942	0.857	0.750	0.937	0.850
	CF	0.895	0.821	0.750	0.875	0.815
	TCF	0.963	0.928	0.916	0.937	0.927

Abbreviation: RF, random forest; LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine; TF, temporal features; CF, correlation features; TCF, temporal and correlation fusion features.

particularly in the prefrontal region. Specifically, rCBF is lower in patients with mild depression compared to the normal control group, whereas it is higher in patients with moderate to severe depression compared to the rCBF in most cortical regions of the normal control group.[45,46] The changes in rCBF can lead to hemodynamic responses, which can be measured using an fNIRS device, making it an important tool for studying depression. However, the fNIRS device is sensitive to motion artifacts and noise, posing a challenge in filter out noise while retaining depression-related information. Some researchers[47] attempted to use wavelet filtering to remove the motion artifacts from fNIRS data and compared it with other methods for correcting motion artifacts, such as principal component analysis, spline interpolation, and Kalman filtering. The results indicated that wavelet filtering had better performance. However, in this paper, the CEEMADN-WPT method was utilized to remove noise from fNIRS data, which is a more powerful technique. Compared to signals filtered by wavelet filtering, the signal filtered by CEEMADN-WPT had higher SNR (as shown in Table 2). With this denoising method, we effectively reduced the noise mixed in the valid frequency band in fNIRS data, thus preserving information related to depression. Shallow tissues, such as the scalp, skull, and meninges, are abundant in capillaries. The concentration of hemoglobin within these capillaries fluctuates due to respiration, heartbeat, and task-related autonomous neural activity. When near-infrared light penetrates these shallow tissues, changes in hemoglobin concentration lead to variations in fNIRS light attenuation, referred to as shallow physiological noise. The "short-separation" method (channels where the emitter-detector distance is below 1.0 cm) is recognized as an effective approach of removing shallow noise. This method involves using additional short-separation fNIRS channels to record shallow physiological noise and then subtract it from the signal. The absence of physiological noise data during our data collection process has posed a certain limitation on our preprocessing of fNIRS data. In future studies, we plan to collect both fNIRS data and physiological noise data to better capture data related to brain neural activity.

In this study, temporal and channel CFs extracted from fNIRS data of the prefrontal lobe of patients with severe and mild depression were compared. Our findings suggest that using TFs as inputs leads to higher accuracy compared to using only CFs. Furthermore, the fusion of temporal and CFs obtained the highest accuracy with an AUC of 96.3% on SVM models. These results

indicate that studying the functional connectivity of the brain alone is insufficient in distinguishing between patients with mild and severe depression. By combining the temporal and CFs of fNIRS data, the classification accuracy can be improved. Through feature visualization, significant differences can be observed between patients with mild depression and those with severe depression in certain TFs, such as peak factor, skewness, and singular spectral entropy. In addition, the peak factor and skewness of ΔHbO and ΔHbR curves play a crucial role in the output of the model. The research results indicate that temporal and channel CFs extracted from ΔHbO and ΔHbR are important for distinguishing the severity of depression, and these features have the potential to serve as biomarkers for distinguishing between mild and severe depression, which is consistent with previous research results.[13]

The findings of our research have several potential applications in clinical services. The diagnostic support system based on fNIRS that we propose provides a non-invasive, objective, and quantitative method for assessing the severity of depression, which can serve as a supplement to traditional diagnostic methods, such as clinical interviews or self-report questionnaires. By combining fNIRS technology and machine learning algorithms, our model can extract valuable information from brain activity patterns that are difficult to detect using traditional methods. These results can serve as an auxiliary tool for psychiatrists, helping them accurately assess the severity of depression in patients and develop more personalized treatment plans. In addition, our research outcomes can be utilized for monitoring treatment progress. By regularly conducting fNIRS tests, clinicians can objectively understand changes in patient's conditions and make timely adjustments to treatment plans for better therapeutic efficacy.

It is important to note that this study has several limitations. First, the small sample size of 140 participants, including 82 cases of severe depression and 58 cases of mild depression, may have limited the precision of our model. A larger dataset is needed to improve the effectiveness of the model. The participants in the study were drug-naive, first-episode depression patients receiving outpatient treatment at the Mental Health Center in Chongqing City. Therefore, our findings are not influenced by medication, or previous treatments, as these data were not included in the analyzed dataset. However, our study did not consider potential differences among the participants, such as long-term smoking, alcohol consumption, the number of past depressive episodes, or a family history of mental illness, all of which could have had an impact on the severity of depression. In addition, the diagnostic data were obtained from a single psychiatric research center, and the generalizability of our model to other hospitals has not been validated. Furthermore, due to the small sample size, patients with moderate depression were not included in the study, necessitating further research in this area. Future work will involve collecting a larger sample size to investigate the severity of depression and analyzing clinical information among participants. We also plan to develop a deep learning-based predictive model for classifying mild, moderate, and severe depression and embed the model into the software used by hospitals. By connecting to the fNIRS database and integrating clinical information from patients, we aim to obtain more accurate and personalized diagnostic results through big data analysis, providing psychiatrics with a more reliable and efficient tool to support their decision-making in depression treatment.

6 Conclusions

This paper proposes a depression severity prediction model based on the temporal and CFs extracted from fNIRS data, achieving high classification accuracy for mild and severe depression. As an objective auxiliary tool, this model can improve the diagnostic efficiency of depression severity and assist doctors in clinical diagnosis, which is of great significance for the treatment of depressed patients.

Disclosures

The authors declare that they have no conflict of interests.

Code and Data Availability

The data and materials that support the findings of this study are available upon request. Due to privacy and confidentiality concerns, the raw data cannot be publicly shared. However, interested

researchers can contact Zhiyong Huang at zyhuang@cqu.edu.cn to request access to the data used in this study. The code used to generate the results and figures is available in a Github repository: https://github.com/liumanliu/fNIRS_depression.

Ethical Approval

The experiment was conducted under the supervision of a professional psychiatrist. No identifiable data of the participants were recorded during the whole study. The experiment proposal was approved by the Ethics Committee of Chongqing Mental Health Center (2020-037).

Authors' Contributions

Man Liu and Zhiyong Huang wrote the main manuscript text, whereas the experiment was conducted under the supervision of professional psychiatrist Hui Yang. Yunlan Zhao and Xiao Han reviewed the manuscript. Huan Chen and Yaju Feng contributed equally to this work. They jointly took the management and coordination responsibility for the research activity planning and execution.

Acknowledgments

The work was supported by grants from the Natural Science Foundation of Chongqing (Grant No. cstc2020jcyj-msxmX0641), the Fundamental Research Funds for the Central Universities (Grant No. 2020CDJ-LHZZ-025), and the Medical Research Project of Chongqing Health Commission (Grant No. 2020FYYX031).

References

1. World Health Organization (WHO), "Depression," (2023). https://www.who.int/news-room/fact-sheets/detail/depression (accessed 12 February 2023).
2. M. Zimmerman et al., "Severity classification on the Hamilton depression rating scale," *J. Affect. Disord.* **150**(2), 384–388 (2013).
3. H. Baumeister, "Inappropriate prescriptions of antidepressant drugs in patients with subthreshold to mild depression: time for the evidence to become practice," *J. Affect. Disord.* **139**(3), 240–243 (2012).
4. P. A. Vöhringer and S. N. Ghaemi, "Solving the antidepressant efficacy question: effect sizes in major depressive disorder," *Clin. Ther.* **33**(12), B49–B61 (2011).
5. S. Wang et al., "Aberrant patterns of brain cerebral blood flow in Chinese Han first-episode drug-naive depressive patients with and without a family history of depression," *Oncotarget* **8**(45), 79906 (2017).
6. H. S. Mayberg, "Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment," *Br. Med. Bull.* **65**(1), 193–207 (2003).
7. J. Xiao et al., "Decoding depression severity from intracranial neural activity," *Biol. Psychiatry* **94**(6), 445–453 (2023).
8. Y. Lu et al., "A regressional analysis of maladaptive rumination, illness perception and negative emotional outcomes in Asian patients suffering from depressive disorder," *Asian J. Psychiatry* **12**, 69–76 (2014).
9. G. Y. Lim et al., "Prevalence of depression in the community from 30 countries between 1994 and 2014," *Sci. Rep.* **8**(1), 2861 (2018).
10. L. M. McDermott and K. P. Ebmeier, "A meta-analysis of depression severity and cognitive function," *J. Affect. Disord.* **119**(1–3), 1–8 (2009).
11. K. Oda et al., "Regional cerebral blood flow in depressed patients with white matter magnetic resonance hyperintensity," *Biol. Psychiatry* **53**(2), 150–156 (2003).
12. P. Pinti et al., "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience," *Ann. N. Y. Acad. Sci.* **1464**(1), 5–29 (2020).
13. C. S. H. Ho et al., "Diagnostic and predictive applications of functional near-infrared spectroscopy for major depressive disorder: a systematic review," *Front. Psychiatry* **11**, 378 (2020).
14. M. Kawano et al., "Correlation between frontal lobe oxy-hemoglobin and severity of depression assessed using near-infrared spectroscopy," *J. Affect. Disord.* **205**, 154–158 (2016).
15. T. Noda et al., "Frontal and right temporal activations correlate negatively with depression severity during verbal fluency task: a multi-channel near-infrared spectroscopy study," *J. Psychiatric Res.* **46**(7), 905–912 (2012).
16. H. Song et al., "Automatic depression discrimination on FNIRS by using general linear model and SVM," in *7th Int. Conf. Biomed. Eng. and Inf.*, IEEE, pp. 278–282 (2014).
17. J. Chao et al., "fNIRS evidence for distinguishing patients with major depression and healthy controls," *IEEE Trans. Neural Syst. Rehabilit. Eng.* **29**, 2211–2221 (2021).

18. Z. Li et al., "Identifying neuroimaging biomarkers of major depressive disorder from cortical hemodynamic responses using machine learning approaches," *EBioMedicine* **79**, 104027 (2022).
19. Q. Yu et al., "GNN-based depression recognition using spatio-temporal information: a fNIRS study," *IEEE J. Biomed. Health Inf.* **26**(10), 4925–4935 (2022).
20. R. Wang et al., "Depression analysis and recognition based on functional near-infrared spectroscopy," *IEEE J. Biomed. Health Inf.* **25**(12), 4289–4299 (2021).
21. R. Ramasubbu et al., "Accuracy of automated classification of major depressive disorder as a function of symptom severity," *NeuroImage Clin.* **12**, 320–331 (2016).
22. T. Richter et al., "Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders," *J. Psychiatric Res.* **141**, 199–205 (2021).
23. S. Y. Dong et al., "Prefrontal functional connectivity during the verbal fluency task in patients with major depressive disorder: a functional near-infrared spectroscopy study," *Front. Psychiatry* **12**, 659814 (2021).
24. J. Huang et al., "Increased prefrontal activation during verbal fluency task after repetitive transcranial magnetic stimulation treatment in depression: a functional near-infrared spectroscopy study," *Front. Psychiatry* **13**, 876136 (2022).
25. T. Yang et al., "The fNIRS evaluation of frontal and temporal lobe cortical activation in Chinese first-episode medication-naïve and recurrent depression during a verbal fluency task," *Front. Psychiatry* **14**, 1132666 (2023).
26. F. Scholkmann et al., "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *NeuroImage* **85**, 6–27 (2014).
27. S. F. Husain et al., "Cortical haemodynamic response measured by functional near infrared spectroscopy during a verbal fluency task in patients with major depression and borderline personality disorder," *EBioMedicine* **51**, 102586 (2020).
28. F. Herold et al., "Applications of functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise-cognition science: a systematic, methodology-focused review," *J. Clin. Med.* **7**(12), 466 (2018).
29. M. A. Yücel et al., "Best practices for fNIRS publications," *Neurophotonics* **8**(1), 012101 (2021).
30. F. A. Fishburn et al., "Temporal derivative distribution repair (TDDR): a motion correction method for fNIRS," *NeuroImage* **184**, 171–179 (2019).
31. A. Faress and T. Chau, "Towards a multimodal brain-computer interface: combining fNIRS and fTCD measurements to enable higher classification accuracy," *NeuroImage* **77**, 186–194 (2013).
32. G. Li et al., "Baseline wander removal for ECG signals based on improved EMD," in *15th IEEE Int. Conf. Signal Process. (ICSP)*, IEEE, Vol. 1, pp. 484–487 (2020).
33. J. Jenitta and A. Rajeswari, "Denoising of ECG signal based on improved adaptive filter with EMD and EEMD," in *IEEE Conf. Inf. Commun. Technol.*, IEEE, pp. 957–962 (2013).
34. C. Wang, Z. Xiao, and J. Wu, "Functional connectivity-based classification of autism and control using SVM-RFECV on RS-fMRI data," *Phys. Med.-Eur. J. Med. Phys.* **65**, 99–105 (2019).
35. I. Guyon et al., "Gene selection for cancer classification using support vector machines," *Mach. Learn.* **46**, 389–422 (2002).
36. I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**(2065), 20150202 (2016).
37. S. Sperandei, "Understanding logistic regression analysis," *Biochem. Med.* **24**(1), 12–18 (2014).
38. J. Cervantes et al., "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing* **408**, 189–215 (2020).
39. P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscip. Rev. Data Mining Knowl. Discov.* **9**(3), e1301 (2019).
40. F. Rossi and B. Conan-Guez, "Functional multi-layer perceptron: a non-linear tool for functional data analysis," *Neural Netw.* **18**(1), 45–60 (2005).
41. S. M. LaValle, M. S. Branicky, and S. R. Lindemann, "On the relationship between classical grid search and probabilistic roadmaps," *Int. J. Rob. Res.* **23**(7–8), 673–692 (2004).
42. L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing* **415**, 295–316 (2020).
43. Y. Gu et al., "Empirical mode decomposition-based motion artefact correction method for functional near-infrared spectroscopy," *J. Biomed. Opt.* **21**(1), 015002 (2016).
44. B. Molavi and G. A. Dumont, "Wavelet-based motion artefact removal for functional near-infrared spectroscopy," *Physiol. Meas.* **33**(2), 259 (2012).
45. T. Chithiramohan et al., "Investigating the association between depression and cerebral haemodynamics: a systematic review and meta-analysis," *J. Affect. Disord.* **299**, 144–158 (2022).
46. M. T. Abou-Saleh et al., "Single photon emission tomography with 99m Tc-HMPAO in Arab patients with depression," *J. Affect. Disord.* **55**(2–3), 115–123 (1999).
47. S. Brigadoi et al., "Motion artefacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data," *NeuroImage* **85**, 181–191 (2014).

Zhiyong Huang is a professor and doctoral supervisor at the School of Microelectronics and Communication Engineering, Chongqing University, China. He serves as the principal investigator for the "Active Health and Aging Technology Response" project under the Chinese National Support Program. He has also led several projects, including the National Key R&D Program, Key Projects of Technological Innovation and Application Development in Chongqing Municipality, Chongqing Natural Science Foundation, and Key Collaborative Projects in Sports Science Research in Chongqing Municipality.

Biographies of the other authors are not available.