

# Classification of EEG signals in depression by fusing temporal convolution and feature recalibration

SUN Fanglin, ZHAI Fengwen\*, JIN Jing

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

\*Corresponding author: ZHAI Fengwen (zhai@mail.lzjtu.cn)

Received: March 17, 2024

Revised: May 21, 2024

Accepted: June 7, 2024

**Abstract:** Aiming at the problem of insufficient feature extraction in single scale neural network model and the problem that convolutional neural network cannot process sequential tasks in the classification of EEG signals in depression, a hybrid model (BFTCNet) of dual-branch convolutional neural network (Bi\_CNN) and temporal convolutional network (TCN) based on feature recalibration (FR) was proposed to classify EEG signals of depressed patients and healthy controls. Firstly, Bi\_CNN module was used to extract the mixed EEG features between different frequency bands and different channels. Secondly, FR module was used to enhance the features extracted by Bi\_CNN. Finally, TCN with dilated causal convolution was used for the sequence learning to capture the temporal dependency between features. In this study, 128 EEG channels of resting-state (closed-eye) EEG data from the public dataset MODMA were used as experimental data, including 29 healthy controls and 24 depression patients. The performance of the model was evaluated by the 10-fold cross validation method. The proposed BFTCNet achieves a classification accuracy of 95.98%,  $F1$  score value of 95.47%, sensitivity and specificity of 94.21% and 97.50%, respectively. Compared with the single-scale network model EEGNet-8,2, the classification accuracy and  $F1$  value are improved by 1.5% and 1.48%, respectively. Meanwhile, the ablation experiment proved that each sub-module had its contribution to the improvement of the model's classification ability.

**Key words:** multi-channel EEG signal; dual-branch convolutional neural network; feature recalibration; temporal convolutional network

## 0 Introduction

Depression is a common mental disorder characterized by depression, loss of interest, and a sense of worthlessness, which may lead to suicidal thoughts or behaviors<sup>[1]</sup>. Currently, the population of depressed patients is growing rapidly worldwide and showing a trend of younger age<sup>[2]</sup>. Therefore, it is necessary to develop a reliable and objective evaluation method to effectively detect depression.

Electroencephalography (EEG) has been widely explored for the detection of nervous system diseases, such as depression, epilepsy, Alzheimer's disease and so on, due to its high sensitivity, relatively low cost, and convenient recording<sup>[3]</sup>. EEG-based techniques for depression recognition have also been extensively studied. For example, the EEG signals of patients with depression have significant differences in amplitude, energy, and other indicators<sup>[4]</sup>. Grin-Yatsenko et al.<sup>[5]</sup> also found increased activity in theta, alpha, and beta bands in the occipital-parietal region of the brain in

patients with depression.

EEG signals are nonlinear and non-stationary in essence. It is time-consuming and tedious to understand EEG signals by manual inspection when a large amount of EEG information needs to be studied<sup>[6]</sup>. As a result, computer-aided detection methods are beginning to emerge. The application of deep learning models in EEG signal analysis has attracted the attention of many researchers. Acharya et al.<sup>[7]</sup> proposed a convolutional neural network (CNN) model with a 13-layer structure. It can learn input features adaptively, so that the accuracy of depression recognition of EEG signals in the left and right hemispheres reaches 93.5% and 96.0%, respectively. Uyulan et al.<sup>[8]</sup> used EEG data of four frequency bands (delta, theta, alpha, beta) to differentiate depressed patients from healthy controls by simulating CNN structures with different depths, and found that the classification effects of MobileNet architecture in the left and right hemispheres were 89.33% and 92.66%, respectively. Lawhern et al.<sup>[9]</sup> proposed a CNN model commonly used in EEG signals

called EEGNet, which can perform well in different EEG datasets. Yan et al.<sup>[10]</sup> and Liu et al.<sup>[11]</sup> distinguished EEG signals from depressed patients and healthy controls by the EEGNet structure. It was found that EEGNet achieved the highest classification accuracy of 94.27% and 90.98%, respectively. At the same time, they proved the effectiveness of EEGNet classification of EEG signals in depression.

In addition, more and more scholars have realized that EEG has temporal correlation, so extracting EEG temporal features is helpful to improve the classification ability of networks. However, the state of a single CNN network structure only depends on the input of the current moment, ignoring the influence of the previous moment. So a single CNN is not suitable for dealing with temporal problems. Liu et al.<sup>[12]</sup> proposed a method based on brain maps with frequency domain and spatial features, using CNN and gated recurrent unit (GRU) to extract sequence features, and the classification accuracy of this method was 89.63%. While the network state of the recurrent network model (such as GRU) at each moment depends on the training results of its previous moment, so parallel computation cannot be realized, resulting in low computational efficiency. The temporal convolution networks (TCN) proposed by Bai et al.<sup>[13]</sup> performed well in temporal modeling and classification tasks.

With the features of good parallelism, flexible receptive field, and stable gradient, TCN model has been attracting more and more attention. Wang et al.<sup>[14]</sup> combined the extracted EEG amplitude and frequency characteristics with TCN for depression recognition of EEG signals and achieved 86.87% classification accuracy. Ingolfsson et al.<sup>[15]</sup> applied a TCN structure after the feature extraction layer of EEGNet and achieved 83.84% classification accuracy in classifying EEG signals based on motor imagery brain-machine interfaces.

Above all, in the process of reading the relevant literature and reproducing the network model, it is found that the above models only use single-scale network models to capture feature information from different frequency band data and ignore the potential correlation among channels, suffering from the problem of inadequate feature extraction.

To address the above problems, we used a parallel dual-branch convolutional network structure to enrich the EEG features extracted by the model firstly, and then further extracted its temporal features to obtain an effective EEG classification model for depression. A

hybrid model, BFTCNet, based on feature recalibration-based Bi\_CNN and temporal convolutional network to classify EEG signals from depressed patients and healthy controls was proposed. The main contributions of this study are as follows.

1) A dual-branch convolutional feature extraction network Bi\_CNN was designed to capture the different frequency and channel spatial features from the basic EEG signals. At the same time, the integration of Bi\_CNN module and FR module made the network performance more efficient, which helped the BFTCNet model to pay more attention to the significant information in EEG data, thus strengthening the feature representation extracted by Bi\_CNN.

2) TCN was introduced and combined with Bi\_CNN to construct BFTCNet, which not only considered the features' relationship between different channels and had the ability of recurrent network to process temporal issues, but also retained the parallel computation ability of CNN network.

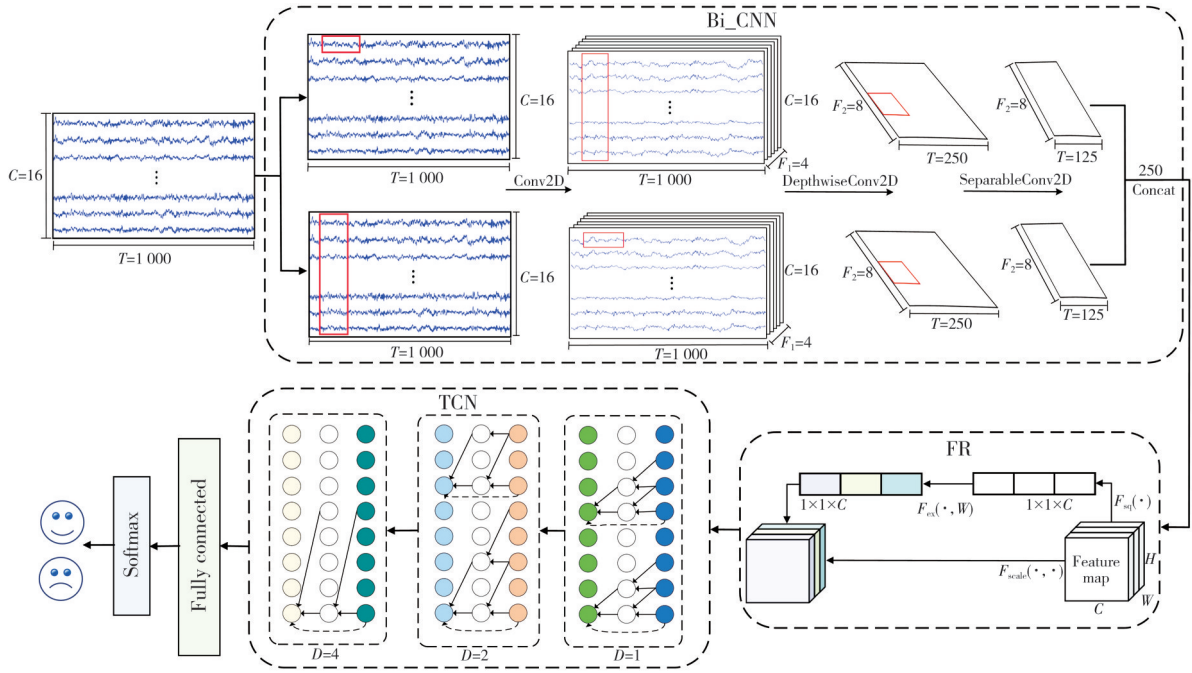
3) The contribution of each sub module to the overall classification capability of the proposed model was progressively validated by the ablation experiment. In the meanwhile, t-SNE dimensional reduction visualization method was used to visually present the classification effects of the middle network layers of the model.

4) Comparing experiments on the public dataset MODMA showed that the classification accuracy,  $F1$  value, sensitivity, and specificity index values of the model were improved than the compared models.

## 1 Proposed method

The hybrid model BFTCNet was proposed for the classification of EEG signals in depression. The model is mainly composed of a dual-branch convolutional module, a feature recalibration module, and a temporal convolutional module, as shown in Fig. 1, which is divided into four modules: feature extraction (Bi\_CNN), feature recalibration (FR), TCN, classification.  $C$  represents the number of electrode channels,  $T$  represents the number of sampling points,  $F_1$  represents the number of time filters,  $F_2$  represents the number of point-by-point filters, and  $D$  represents the size of the expansion factor in the dilated causal convolution.

Firstly, Bi\_CNN is designed to extract fusion features from different electrode channels and multiple frequency bands from EEG signals to obtain rich representation of spatio-temporal domain information.



**Fig. 1 Overall architecture of BFTCNet**

Secondly, the FR module is employed to automatically highlight the most valuable information in the spatio-temporal features by optimizing the feature weights, and at the same time, to enhance the recognition accuracy by leveraging the complementary information of different types of features.

Thirdly, TCN ensures that the shallow network obtains a wider receptive field by dilated convolution, extracts the state and degree of EEG signal change over time, and further extracts advanced time characteristics.

Finally, the classification decision is done by a fully connected layer with a softmax activation function.

### 1.1 Feature extraction module (Bi\_CNN)

In order to explore different types of frequency band and channel characteristics, Bi\_CNN implements two convolution branches with different kernel sizes. Each branch is designed based on the network structure of EEGNet<sup>[9]</sup>, which differs from EEGNet in that the network design process is based on the relevant EEG neurophysiological characteristics. In this method the selection of convolution kernel size is related to the data sampling frequency and channel number in the EEG signal, and different CNN kernel sizes are used to extract low-frequency and high-frequency features.

As shown in the left part of Fig. 2, the input data dimension is  $(n, C, T)$ , and the reshape operation casts the dimension to  $(n, C, T, 1)$ , where  $n$  is the number of samples,  $C$  is the number of electrode channels in EEG data set,  $T$  is the number of data sampling points, which is equal

to the sampling time multiplied by the data sampling frequency (250 Hz). The network structure of each branch consists of three different convolution layers and two maximal pooling layers, where each convolution layer includes a batch normalization layer and uses gaussian error linear units (GELU) as the activation function. The GELU activation function is chosen because it allows some negative weights to pass. These negative weights play a role in the feature recalibration process in the FR module in the next step.

In the left branch, Conv2D (4, (1, 250)) indicates that four convolution kernels with a size of (1, 250) are used, which means that the features are first extracted on different frequency bands. Here, in order to extract frequency features greater than or equal to 1 Hz in the EEG signal data, the width of the convolution kernel is set to the same size as the sampling frequency (250 Hz).

In the right branch, Conv2D (8, (16, 1)) indicates that 8 convolution kernels with a size of (16, 1) are used, and 16 is the number of electrode channels, that is, the aggregated signal feature representation formed by the weighted combination of 16 channels is extracted firstly. Subsequently, different frequency band features were extracted, and different spatial features were mixed together.

Finally, the feature fusion of the two branches is completed by tensor splicing. In order to reduce the risk of overfitting, dropout is applied after the splicing of the two branch is completed. This section corresponds to the part in the first rectangular box in Fig.1.

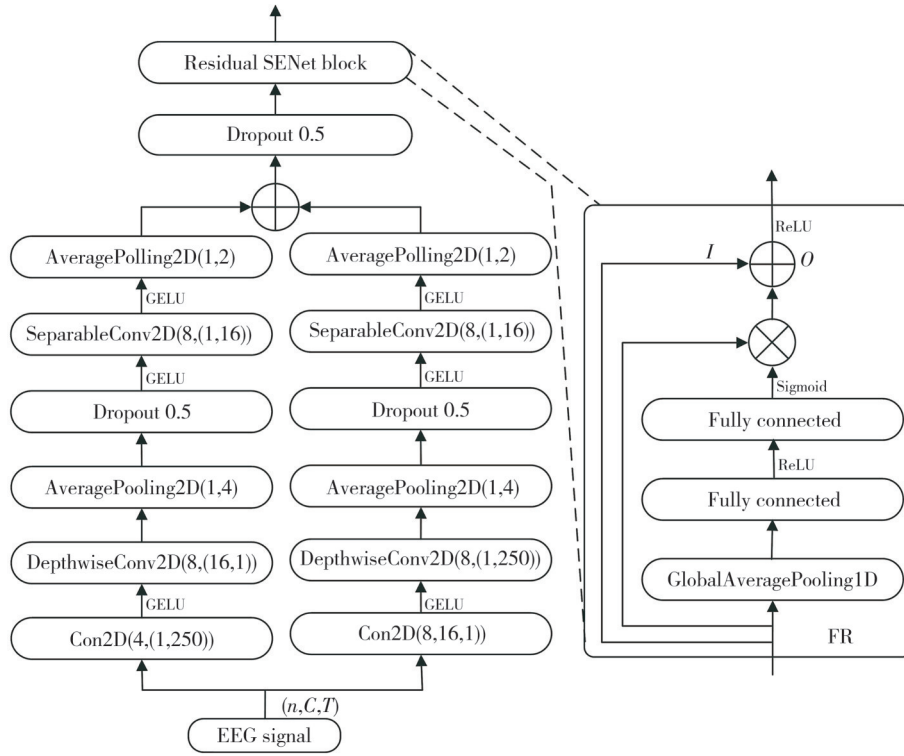


Fig. 2 Bi\_CNN and FR structure

## 1.2 Feature recalibration (FR)

Each convolution layer of the CNN architecture shares weights, which means that equal attentions are paid to each part of the EEG signal. In order to learn the global dependencies in the EEG signal and focus on the relevant parts that play an important role in the classification task, an attention mechanism based feature recalibration (FR) module was designed to incorporate with the former Bi\_CNN module. FR aims to recalibrate the features learned by Bi\_CNN to improve the classification performance, and the process is implemented through the SENet<sup>[16]</sup>, as shown in the right part of Fig.2.

The squeeze operation performs a global average pooling according to the input features  $F \in \mathbf{R}^{N \times d}$  to obtain  $s = \{s_1, s_2, \dots, s_N\}$ , denoting  $d$  data points in the features  $F_i \in \mathbf{R}^d$ , where  $1 \leq i \leq N$ . Through the previous processing, the feature information of each channel is compressed into a specific value to obtain a feature mapping with a global sensory field.

The excitation operation captures the channel correlation, and the first fully connected layer is followed by a rectified linear unit (ReLU) activation function to perform dimensionality reduction, and the second fully connected layer is followed by a sigmoid activation function to perform dimensionality increase. Compared with using only one full connection layer, double full connection layer can better fit the correlation of features between channels.

Finally, the skip connection is added to combine the original input  $I$  with the correction features  $O$  obtained from the network. The calculation process is

$$X = I + O \in \mathbf{R}^{N \times d}. \quad (1)$$

FR operation completes the operation of assigning weights to each feature. In the back propagation, the ReLU and sigmoid activation function are used to update the weight parameters, and finally to strengthen the features that play a greater role in the classification and weaken the features that play a smaller role. This section corresponds to the part in the second rectangular box in Fig.1.

## 1.3 Time convolutional network (TCN)

Although CNN has excellent feature extraction capabilities, it is still difficult to maintain time consistency. TCN is introduced for time series modeling tasks after extracting features, as shown in Fig.3. The TCN residual block has two branches. The first branch includes two dilated causal convolution layers, two-layer batch normalization normalization and two-layer nonlinear transformation layers, namely exponential linear unit (ELU). Further optimization is done based on the original TCN<sup>[13]</sup>, such as the use of batch normalization between convolution layers instead of weight normalization, which is done because the batch normalization is directly to the output of neurons for batch normalization. The process of the object is the output of each layer of network, and it will make EEG

data more stable. The reason for using the ELU activation function instead of the ReLU activation function is that, compared to ReLU, ELU has negative values, which can make the output mean approach 0, thereby achieving the effect of making the gradients close to the natural gradients and reducing the computational load. And the regularization technique (Dropout) is used to reduce the risk of overfitting, and the second is the skip connection of TCN input. In order to ensure that the input and output have the same tensor ratio between the addition operations, the original input is added to the output feature map with the convolution layer.

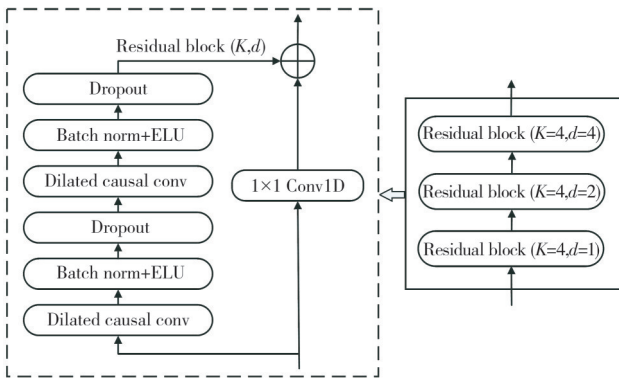


Fig. 3 TCN structure

The most special feature of the dilated causal convolution structure is that its specific form is determined by the filter and the expansion factor. The expansion factor of layer  $i$  is calculated by

$$d = 2^i. \tag{2}$$

It can be seen that the expansion factor increases exponentially with the depth of the network depth. The filter size ( $K$ ) selected in this paper is 4, and the expansion factor size is 1, 2, and 4, respectively. This section corresponds to the part in the third rectangular box in Fig.1.

### 1.4 Classification

As shown in Fig.1, the TCN module uses the high-level time features extracted by the TCN in the time series and sends them to the fully connected (FC) layer with softmax classifier. Finally, the function is used to calculate the classification probability of the corresponding category, and the category with the largest probability value is taken as the final category. The definition of the softmax function is

$$p = \text{soft max} (z_i) = \frac{e^{z_i}}{\sum_{B=1}^B e^{z_B}}, \tag{3}$$

where  $p$  is the probability value that the model assigns to the  $i$ -th output label for the input tensor  $z_i$ ,  $B$  indicates

the number of categories of the classification.  $Z_B$  indicates the output of category  $B$ .

## 2 Experiment

### 2.1 Dataset

MODMA (Multi-modal open dataset for mental-disorder analysis ) is used as the experimental dataset, which is released in 2020 by the Gansu Provincial Wearable Equipment Key Laboratory of Lanzhou University<sup>[17]</sup>. The dataset contains EEG data of clinical depression patients and EEG data of normal control groups. Throughout the course of the experiment, 128 channel resting state EEG dataset in MODMA is used, which contains 53 subjects, including 24 depressed patients and 29 healthy people. All depressed patients in the dataset met the diagnostic criteria for depression of the DSMIV-based “Diagnostic and Statistical Manual of Mental Disorders”<sup>[18]</sup>. The sampling frequency of EEG signal in this dataset is 250 Hz, the reference electrode is Cz electrode, and the data recording time of each subject (the person under test) is about 5 min.

### 2.2 Data processing

The data processing stage is implemented with the help of the open-source Python toolkit MNE<sup>[20]</sup>, which is used to explore, visualize, and analyze human neurophysiological data, such as magnetoencephalography (MEG), electroencephalography (EEG), etc. The data processing of raw EEG data consists of six main steps.

**Step 1** Inspired by the research work<sup>[20]</sup>, the effective electrodes are selected. Considering the time performance and calculation cost, 16 tested electrodes (Fp1/2, F3/4, C3/4, P3/4, O1/2, F7/8, T3/4, T5/6) are chosen from 128 electrodes. The position information of the selected electrodes is shown in Fig.4, where Cz is the reference electrode.

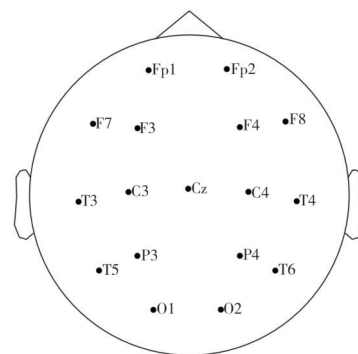


Fig. 4 16 EEG electrode positions

**Step 2** The notch filter is used to remove the power frequency, which can eliminate power line interference at 50 Hz.

**Step 3** The band-pass FIR filter is used to filter out high-frequency noise and low-frequency drift between 1 Hz and 40 Hz, leaving only the information frequency band needed for research.

**Step 4** A window with the size of 4 s is used to segment up to 300 s of EEG data<sup>[21]</sup>, that is, the data sampling point is 1 000, and there is no overlap between adjacent windows.

**Step 5** To fix broken segments, the bad segment is repaired in this step using Autoreject, an automatic artifact suppression algorithm proposed by Jas et al.<sup>[22]</sup>, which is used to generate clean physiological data.

**Step 6** The datasets need to undergo normalization processing before inputting into the network model, and the normalization processing formula is

$$z = \frac{X_{i,j} - \mu}{\sigma}, \quad (4)$$

where  $X_{i,j}$  represents the feature vector,  $\mu$  and  $\sigma$  the average and the standard deviation of the feature vectors, respectively.

### 2.3 Evaluation indicators

The evaluation metrics used in this study are classification accuracy ( $Acc$ ), precision ( $Pre$ ), sensitivity ( $Sen$ ), specificity ( $Spe$ ), and  $F1$  value. The calculation formulas are

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (5)$$

$$Pre = \frac{TP}{TP + FP}, \quad (6)$$

$$Sen = \frac{TP}{TP + FN}, \quad (7)$$

$$Spe = \frac{TN}{TN + FP}, \quad (8)$$

$$F1 = \frac{2TP}{2TP + FP + FN}, \quad (9)$$

where  $TP$  represents true positive,  $TN$  represents true negative,  $FN$  represents false negative, and  $FP$  represents false positive.

### 2.4 Experimental environment and parameter settings

The experimental environment is built on the remote cloud service platform provided by AutoDL.com. The environment is configured as follows: Platinum 8255C processor, 45 GB CPU memory, 11 GB GPU video memory, and RTX 2080Ti video card. The operating system is windows 10 Professional edition 64 bits, the

simulation software is PyCharm 2021, and the compilation environment is python 3.8. All experiments are implementations in the Keras framework TensorFlow backend.

The experimental parameter settings are shown in Table 1.

**Table 1 Experimental parameter settings**

Parameter	Value
Batch size	16
Epochs	100
Optimizer	Adam
Learning rate	$10^{-3}$
Dropout rate of Bi_CNN	0.5
Dropout rate of TCN	0.2

### 2.5 Training and testing methods

Cross-validation is an effective way to improve the credibility of the model. A 10-fold cross-validation method was used to split the training set and the test set. There were two categories of samples, those with depression were labeled 1 and those with normal controls were labeled 0. During the experiment, all samples were randomly divided into 10 sub-samples with the same number, each sub-sample was used as the test set in turn, and the other 9 sub-samples were used as the training set to train model. The classification accuracy, precision, sensitivity, specificity, and  $F1$  score of the model were calculated at the end of each fold, and finally mean and standard deviation of each metric on the test set of the 10-fold model were taken as the final evaluation metric of the model.

## 3 Results and discussion

### 3.1 Ablation experiments

The model BFTCNet proposed in this paper has three functional components: dual-branch convolution module (Bi\_CNN), feature recalibration module (FR), and temporal convolution network module (TCN). In order to clarify the independent influence of each component of the BFTCNet model on the overall performance of the model, this section designs experiments on the disassembly of BFTCNet model components in four seconds slice size EEG data. They are the classification results after removing the feature re-correction module (W/O FR), removing the time convolution network module (W/O TCN), and removing the feature re-correction module, and the time convolution network (W/O FR+TCN) from the BFTCNet, as shown in Table 2.

**Table 2 Ablation experimental study**

Model	Acc/%	Pre/%	F1/%	Sen/%	Spe/%
EEGNet-8,2	94.48±2.60	93.53±2.83	93.99±2.18	94.57±3.32	94.39±2.60
W/O FR	95.05±3.16	96.59±3.02	94.41±3.69	92.56±5.40	97.14±2.52
W/O TCN	95.23±3.37	95.76±4.02	94.68±3.83	94.02±6.22	96.26±3.88
W/O FR+TCN	94.31±2.84	95.32±3.78	93.70±3.08	92.54±5.03	95.80±4.07
BFTCNet	95.98±3.47	97.04±2.92	95.47±4.04	94.21±5.96	97.50±2.59

At the same time, because the model BFTCNet proposed in this paper is an improvement based on the model EEGNet-8, 2<sup>[9]</sup>, the ablation experiment section sets EEGNet-8, 2 as a single-scale network model.

Firstly, all evaluation indicators will decline after removing any of three modules, indicating that all modules contribute to improving the classification ability of the network.

Secondly, after the removal of FR+TCN module from BFTCNet, the classification performance decreases the most with the classification accuracy decreases by 1.67% and *F1* by 1.77%, which proves that the combination of FR module and TCN module could improve the classification ability of the network

Thirdly, it can be seen that after the removal of FR module and TCN module from BFTCNet respectively, the classification ability of both two sub-networks decreases.

On one hand, the FR module helps the network to obtain more context information through the context-aware mechanism, and the squeeze operation of the FR module can obtain the global information of the input data with a faster convergence speed, which improves recognition accuracy and sample quality by distinguishing the importance of different time-domain features on the depression EEG classification results, that is, enhancing useful features and suppressing poor performance features, while exploiting the complementary information between different features.

On the other hand, the TCN module are able to expand the size of their perceptual fields exponentially with a linear increase in the number of parameters, unlike traditional CNN where the size of the perceptual field only increases linearly. A series of causal convolutional layers with different expansion coefficients are applied to take advantage of a wider historical coverage and to extract rich temporal features at each time step.

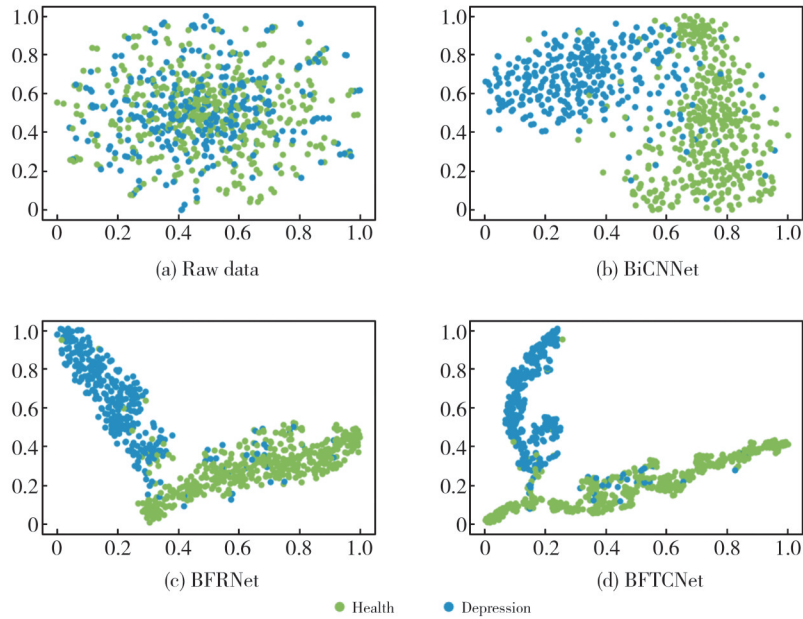
Fourthly, the sensitivity metric value of the proposed model BFTCNet is lower than the specific metric value, indicating that the model has better recognition accuracy for negative samples (health), and the recognition accuracy of positive samples (depression) is slightly

lower than that of positive samples.

### 3.2 t-SNE visualization features

The t-distributed stochastic neighbor embedding (t-SNE) is used to carry out qualitative evaluation and classification tasks on MODMA data sets<sup>[23]</sup>. t-SNE is an effective visualization method for visualizing embedded features, and its main idea is to convert the Euclidean distance between high-dimensional features into conditional probability distribution representing similarity, so that similar features have a higher probability of being selected. The experiment compared the discriminatory capabilities of different network structures in extracting general features from the original EEG data. The vectors before the classifier were used as representative features, and the success of the model in extracting features from the original data was verified through visual clustering. The proposed model BFTCNet is analyzed step by step, and three model variants are specifically derived as follows. 1) BiCNNNet: network model consists of only Bi\_CNN, corresponding to W/O FR+TCN. 2) BFRNet: network model composes of Bi\_CNN with the introduction of FR, corresponding to W/O TCN. 3) BFTCNet: network model composed of BFRNet fused with TCN.

Fig. 5 shows the feature representation of the visualized network using the t-SNE algorithm in the test phase with the addition of the BiCNN module, FR module, and TCN network in sequence. The t-SNE algorithm can be used for dimensionality reduction visualization, where the *x* and *y* coordinate positions represent the similarity of the data sample. In general, similar data points will be clustered together, and data points of different categories will form different clusters. In order to facilitate visual display, the data are normalized and unified into the same coordinate interval. Fig. 5(a) shows the inseparability in the original EEG data. Fig. 5(b)–5(d) show the two-dimensional feature maps of BiCNNNet, BFRNet, and BFTCNet, respectively. This further shows the advantages of BFTCNet in extracting EEG separable features and also proves the effectiveness of the proposed model.



**Fig. 5 Visual feature representation of method proposed method at different network stages**

### 3.3 Comparison of different deep learning models

The comparison experiment is conducted on the same dataset (MODMA), and structure of the comparison model is based on the parameters determined in the original paper, while data processing, training, and evaluation follow the same procedures defined in this study. The BFTCNet model is compared with the common classification models in EEG classification tasks, including EEGNet-4, 2<sup>[9]</sup>, EEGNet-8, 2<sup>[9]</sup>, EEG-TCNet<sup>[15]</sup>, and ShallowConvNet<sup>[24]</sup>, with classification accuracy, precision, *F1* score, sensitivity, and specificity as evaluation metrics.

EEGNet is a compact EEG analysis neural network

**Table 3 Comparison of experimental results of different deep learning models**

Model	<i>Acc</i> /%	<i>Pre</i> /%	<i>F1</i> /%	<i>Sen</i> /%	<i>Spe</i> /%
EEGNet-4, 2	89.73±2.42	87.70±4.49	88.99±2.58	90.73±4.52	88.92±4.85
EEGNet-8, 2	94.48±2.60	93.53±2.83	93.99±2.18	94.57±3.32	94.39±2.60
EEG-TCNet	91.43±1.20	90.77±1.22	90.62±1.35	90.51±2.04	92.21±1.68
ShallowConvNet	95.62±2.06	96.38±2.74	95.13±2.44	94.16±4.76	96.86±2.49
<b>BFTCNet (ours)</b>	<b>95.98±3.47</b>	<b>97.04±2.92</b>	<b>95.47±4.04</b>	<b>94.21±5.96</b>	<b>97.50±2.59</b>

### 3.4 Comparison with related studies

This section compares the classification ability of the proposed method with other EEG classification methods based on MODMA data sets.

Sun et al.<sup>[20]</sup> extracted the functional connectivity features of EEG and obtained its classification results using Phase Lag Index (PLI) and support vector machine (SVM). Zhang et al.<sup>[25]</sup> integrated gender and age factors into one-dimensional CNN (1D CNN)

model designed for conventional EEG signal recognition tasks. In this paper, both of its two structures (EEGNet4\_2 and EEGNet8\_2) are compared experimentally.

EEG-TCNet is an extension of EEGNet. TCN structure is connected after the feature extraction layer of EEGNet. ShallowConvNet is often used to distinguish the network structure of normal EEG signals from pathological EEG signals, in order to decode different frequency band features of EEG signals

The comparison results are shown in Table 3, where the bold data are the optimal values among the comparing models. It can be seen that the proposed model (BFTCNet) outperforms the other comparison models except for the sensitivity index values.

through attention mechanisms, which can prompt 1D CNN to explore the complex association between EEG signals and demographic factors. Wang et al.<sup>[14]</sup> used time-frequency analysis to extract the instantaneous amplitude and instantaneous frequency of EEG, and then obtained classification results using a modified temporal convolution network (L-TCN). Liu et al.<sup>[11]</sup> obtained classification results using the classical end-to-end EEG decoding model EEGNet. Liu et al.<sup>[12]</sup> converted EEG signal data into brain spectrograms with

frequency domain and spatial features, and then used CNN and GRU to extract sequence features to classify depression. Deng *et al.*<sup>[26]</sup> proposed a new model SparNet with attention mechanism to extract spatial-frequency domain features of EEG in local and whole brain regions. The model BFTCNet proposed in this paper mainly uses the raw EEG signal to realize the end-to-end automatic

classification of depression, and requires less domain knowledge of features.

Table 4 shows the comparison results between the proposed method and other EEG classification methods for depression based on MODMA data set, where the bolded data is the optimal value of the corresponding index.

**Table 4 Comparison of experimental results with other related studies**

References	Channel	Classifier	<i>Acc</i> /%	<i>F1</i> /%	<i>Sen</i> /%	<i>Spe</i> /%
Sun <sup>[20]</sup>	16	SVM	82.31			
Zhang <sup>[25]</sup>	3	1DCNN	75.29	71.60	66.20	83.00
Wang <sup>[14]</sup>	128	L-TCN	86.87	90.51	90.15	83.83
Liu <sup>[11]</sup>	128	EEGNet	90.98	90.83	90.59	
Liu <sup>[12]</sup>	128	CNN+GRU	90.62	88.79	87.81	87.48
Deng <sup>[26]</sup>	73	SparNet	94.37	94.40	95.07	93.66
<b>Ours</b>	<b>16</b>	<b>BFTCNet</b>	<b>95.98</b>	<b>95.47</b>	<b>94.21</b>	<b>97.50</b>

Compared with the existing EEG classification methods that use the same MODMA dataset, the classification accuracy and *F1* value of the proposed BFTCNet model in this paper have reached 95.98% and 95.47%, respectively. This indicates that the model effectively captures the EEG semantic information related to the task and fully utilizes the time-frequency domain features of the EEG signal, verifying the superiority of the model proposed in this paper. In addition, compared with other methods<sup>[11,12,14,26]</sup>, fewer EEG electrode channels were used in this study, and the best classification results were achieved in comparison with the above depressed EEG classification models that took the MODMA as their experimental data sets, which could provide a better solution for future real-time applications.

### 3.5 Discussion of time window length

Setting different window lengths will result in different fragment lengths and quantities. According to Table 5, the differences reported for the indicators corresponding to 6 s, 8 s, and 10 s fragment lengths are smaller than those between 4 s and 2 s fragment lengths. However, the variances of the evaluation indicators corresponding to the 6 s, 7 s, and 10 s fragments are significantly greater than those of the 4 s and 2 s fragment lengths. This indicates that as the fragment length increases, the number of available samples decreases, the sample dimension increases, and the classification performance of the model does not continue to improve with the increase in the number of sampling points in the samples, and the model tends to become unstable. This phenomenon suggests that if the window length is set too short, the EEG signal fragments captured are too short,

and the number of captured fragments will increase, resulting in a larger total sample size. The small amount of EEG information contained in different frequency bands in each fragment may not be sufficient to reflect the differences in brain rhythm activities between normal people and patients with depression. If the window length is set too long, the number of EEG signal fragments captured will decrease, resulting in a smaller total sample size, and the model is prone to overfitting, and each EEG signal fragment corresponds to a relatively longer time domain, and an excessively high data dimension will also reduce the data processing efficiency of the subsequent network<sup>[27]</sup>. In conclusion, this paper finally selects a 4 s window to divide the 300 s EEG data.

**Table 5 Classification accuracy and *F1* scores for different time windows**

Time/s	<i>Acc</i> /%	<i>F1</i> /%
2	92.60±2.62	91.66±3.10
4	95.98±3.47	95.47±4.04
6	97.38±3.92	97.07±4.41
8	97.86±4.15	97.51±4.89
10	97.92±4.61	97.53±5.51

## 4 Conclusions

An end-to-end hybrid model BFTCNet based on multi-lead scalp resting state EEG signals was proposed and applied to the classification of EEG signals in depression. Firstly, BFTCNet used the feature extraction module (Bi\_CNN) to integrate the frequency domain features and spatial features learned from different channels by tensor splicing. Secondly, through FR module, the most effective information in time-frequency domain features was highlighted, and the less important features were suppressed. Finally, the

temporal features and multi-channel enhanced features were fused by TCN to achieve depression EEG classification.

The ablation experiments, feature classification ability visualization experiments, and comparison experiments verified the validity of the proposed model and its sub models from different aspects, and also verified the feasibility and effectiveness of temporal features of enhanced dual-channel features.

At the present stage, the study only focuses the classification ability of dual-channel fusion features and temporal features. In the next step, the research will focus on the multi-scale model, and further improve the structure and classification ability of the network from the perspective of neuroscience, so as to further deepen the research of this study.

## Acknowledgement

This work was supported by Natural Science Foundation of Gansu Province (No. 21JR11RA062); University Innovation Fund of Gansu Province (No.2022A-047).

## Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

## References

- [ 1 ] SMITH K. Mental health: a world of depression. *Nature*, 2014, 515(7526): 181.
- [ 2 ] ETTMAN C K, ABDALLA S M, COHEN G H, et al. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. *JAMA Network Open*, 2020, 3(9): e2019686.
- [ 3 ] SAEIDIM, KARWOWSKI W, FARAHANI F V, et al. Neural decoding of EEG signals with machine learning: a systematic review. *Brain Sciences*, 2021, 11(11): 1525.
- [ 4 ] YAHYA N, MUSA H, ONG Z Y, et al. Classification of motor functions from electroencephalogram (EEG) signals based on an integrated method comprised of common spatial pattern and wavelet transform framework. *Sensors*, 2019, 19(22): 4878.
- [ 5 ] GRIN-YATSENKO V A, BAAS I, PONOMAREV V A, et al. Independent component approach to the analysis of EEG recordings at early stages of depressive disorders. *Clinical Neurophysiology*, 2010, 121(3): 281-289.
- [ 6 ] RAHMAN M A, HOSSAIN M A, KABIR M R, et al. Optimization of sleep stage classification using single-channel EEG signals//2019 4th International Conference on Electrical Information and Communication Technology, December 20-22, 2019, Khulna, Bangladesh. New York: IEEE, 2020: 1-6.
- [ 7 ] ACHARYA U R, OH S L, HAGIWARA Y, et al. Automated EEG-based screening of depression using deep convolutional neural network. *Computer Methods and Programs in Biomedicine*, 2018, 161: 103-113.
- [ 8 ] UYULAN Ç, ERGÜZEL T, UNUBOL H, et al. Major depressive disorder classification based on different convolutional neural network models: deep learning approach. *Clinical EEG and Neuroscience*, 2020, 52: 38-51.
- [ 9 ] LAWHERN V J, SOLON A J, WAYTOWICH N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 2018, 15(5): 056013.
- [10] YAN D D, ZHAO L L, SONG X W, et al. Automated detection of clinical depression based on convolution neural network model. *Biomedizinische Technik Biomedical Engineering*, 2022, 67(2): 131-142.
- [11] LIU B, CHANG H L, PENG K, et al. An end-to-end depression recognition method based on EEGNet. *Frontiers in Psychiatry*, 2022, 13: 864393.
- [12] LIU W, JIA K B, WANG Z Z, et al. A depression prediction algorithm based on spatiotemporal feature of EEG signal. *Brain Sciences*, 2022, 12(5): 630.
- [13] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018: arXiv: 1803.01271. <https://arxiv.org/abs/1803.01271>
- [14] WANG Y X, LIU F R, YANG L J. EEG-based depression recognition using intrinsic time-scale decomposition and temporal convolution network. *The Fifth International Conference on Biological Information and Biomedical Engineering*, New York: ACM, 2021: 1-6.
- [15] INGOLFSSON T M, HERSCHE M, WANG X Y, et al. EEG-TCNet: an accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces//2020 IEEE International Conference on Systems, Man, and Cybernetics, October 11-14, 2020. Toronto, ON, Canada. New York: IEEE, 2020: 2958-2965.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7132-7141.
- [17] CAI H S, YUAN Z Q, GAO Y W, et al. A multi-modal open dataset for mental-disorder analysis. *Scientific Data*, 2022, 9: 178.
- [18] SEGAL D L. Diagnostic and statistical manual of mental disorders (DSM-IV-TR). *The Corsini Encyclopedia of Psychology*, 2010: 1-3.
- [19] GRAMFORT A, LUSSINI M, LARSON E, et al. MNE software for processing MEG and EEG data. *NeuroImage*, 2014, 86: 446-460.
- [20] SUN S T, LI J X, CHEN H Y, et al. A study of resting-state EEG biomarkers for depression recognition. 2020: arXiv: 2002.11039. <https://arxiv.org/abs/2002.11039>.
- [21] SEAL A, BAJPAI R, AGNIHOTRI J, et al. DeprNet: a

- deep convolution neural network framework for detecting depression using EEG. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 2505413.
- [22] JAS M, ENGEMANN D A, BEKHTI Y, et al. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 2017, 159: 417-429.
- [23] MAATEN L, HINTON G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008(9): 2579-2605.
- [24] SCHIRMEISTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 2017, 38(11): 5391-5420.
- [25] ZHANG X W, LI J L, HOU K C, et al. EEG-based depression detection using convolutional neural network with demographic attention mechanism//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, July 20-24, 2020. Montreal, QC, Canada. New York: IEEE, 2020: 128-133.
- [26] DENG X, FAN X F, LÜ X W, et al. SparNet: a convolutional neural network for EEG space-frequency feature learning and depression discrimination. *Frontiers in Neuroinformatics*, 2022, 16: 914823.
- [27] MAO W D. Deep learning based research on recognition of depression patients using eeg data. Lanzhou: Lanzhou University, 2019.

## 融合时间卷积和特征校正的抑郁脑电信号分类

孙芳林, 翟凤文\*, 金 静

兰州交通大学, 电子与信息工程学院, 甘肃 兰州 730070

**摘要:** 针对单尺度神经网络模型存在特征提取不充分和卷积神经网络在抑郁症脑电信号分类中无法处理时序任务的问题, 提出了一种基于特征重新校正(FR)的双分支卷积神经网络(Bi\_CNN)和时间卷积网络(TCN)的混合模型(BFTCNet), 来进行抑郁症患者和健康对照组的脑电信号分类。该模型使用Bi\_CNN模块提取不同频带间和不同通道间的混合特征, FR模块加强表示Bi\_CNN提取的重要特征, 用带膨胀因果卷积(Dilated causal convolution)的时间卷积网络TCN进行时间序列学习, 捕捉特征之间的时间依赖性。本文使用公共数据集MODMA的128通道静息态(闭眼)脑电图作为实验数据, 其中包括29名健康对照组和24名抑郁症患者, 通过十折交叉验证(10-fold cross validation)方法评估模型的性能。实验结果表明, BFTCNet的分类准确率达到95.98%,  $F1$ 分数值为95.47%, 敏感性和特异性分别为94.21%和97.50%。相比单尺度网络模型EEGNet-8,2, 分类准确率和 $F1$ 值分别提高了1.5%和1.48%。同时, 消融实验证明每个子模块对模型分类能力的高都有贡献。

**关键词:** 多导联脑电信号; 双分支卷积神经网络; 特征重新校正; 时间卷积网络

**引用格式:** SUN Fanglin, ZHAI Fengwen, JIN Jing. Classification of EEG signals in depression by fusing temporal convolution and feature recalibration. *Journal of Measurement Science and Instrumentation*, 2025, 16(4): 547-557. DOI: 10.62756/jmsi.1674-8042.2025053