Depression Assessment Method: An EEG Emotion Recognition Framework Based on Spatiotemporal Neural Network

Hongli Chang[1,2†], Yuan Zong[1†], Wenming Zheng[1*], Chuangao Tang[1], Jie Zhu[1,2] and Xuejun Li[1]

[1] Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University, Nanjing, China,
[2] School of Information Science and Engineering, Southeast University, Nanjing, China

The main characteristic of depression is emotional dysfunction, manifested by increased levels of negative emotions and decreased levels of positive emotions. Therefore, accurate emotion recognition is an effective way to assess depression. Among the various signals used for emotion recognition, electroencephalogram (EEG) signal has attracted widespread attention due to its multiple advantages, such as rich spatiotemporal information in multi-channel EEG signals. First, we use filtering and Euclidean alignment for data preprocessing. In the feature extraction, we use short-time Fourier transform and Hilbert–Huang transform to extract time-frequency features, and convolutional neural networks to extract spatial features. Finally, bi-directional long short-term memory explored the timing relationship. Before performing the convolution operation, according to the unique topology of the EEG channel, the EEG features are converted into 3D tensors. This study has achieved good results on two emotion databases: SEED and Emotional BCI of 2020 WORLD ROBOT COMPETITION. We applied this method to the recognition of depression based on EEG and achieved a recognition rate of more than 70% under the five-fold cross-validation. In addition, the subject-independent protocol on SEED data has achieved a state-of-the-art recognition rate, which exceeds the existing research methods. We propose a novel EEG emotion recognition framework for depression detection, which provides a robust algorithm for real-time clinical depression detection based on EEG.

Keywords: depression, emotion recognition, electroencephalogram (EEG), convolutional neural network (CNN), long-short term memory network (LSTM)

1. INTRODUCTION

The recognition of emotion is a major research direction of affective computing, which had been widely used to detect depression (1, 2). Emotion is crucial to the quality and scope of human daily experience (3). With the development of the brain–computer interface (BCI) and the advancement of artificial intelligence, the recognition of emotions based on EEG signals has become an active research topic of emotion recognition. EEG signals contain a large amount of information related to emotions and have the characteristics of high time resolution, and are not effortless to disguise (4–6), which

shows tremendous advantages in the field of real-time emotion recognition. Accurate and real-time judgment of human emotional state through some technical means has great application value in many areas, for example, driving fatigue detection (7), depression monitoring (8), and real-time monitoring of critically ill patients (9).

The relationship between EEG and emotion has been reported in past studies. Brain regions implicated in emotional experience include the orbitofrontal cortex, insular cortex, and anterior and posterior cingulate cortices. The amygdala is involved in linking perception with automatic emotional responses and memory (3). The activation of the amygdala seemed to be more related to negative emotions, and the relative activation of the right frontal lobe correlated with negative emotions (such as fear or disgust) (10). Precisely, fear corresponds to the amygdala (11), anger is related to the orbitofrontal cortex and anterior cingulate cortex (12), sadness occurs in the amygdala and right temporal pole (13), and disgust is produced in the anterior insula and anterior cingulate cortex (14). In addition, the power of the alpha band and the asymmetry between the cerebral hemispheres relates to emotions (15–17), the changes in the gamma band connects with happiness and sadness, and the reduction of alpha waves on different sides of the temporal lobe correlates with joy and sorrow (left side is sad, happy on the right) (18, 19).

Extracting emotion-related features to make larger the distance between classes and smaller the distance within classes is helpful to solving cross-database problems. Emotion-related EEG signal feature extraction methods include time domain [such as Hjorth extraction activity, mobility and complexity of EEG signals (20), higher-order crossover features used to describe the oscillation mode of a time series (21) and magnitude squared coherence estimate (22)], frequency domain [such as power spectral density features (23, 24)], time-frequency domain [such as time-frequency spectrum [TFS] features (25)], auto-regressive (26), asymmetric spatial pattern (27), entropy [such as differential entropy (7), sample entropy (28) and approximate entropy (29)], maximum relevance minimum redundancy method (30), common spatial patterns (31), filter bank common spatial pattern (32), higuchi fractal dimension (33), and so on. Regarding EEG feature types, all frequency bands or some frequency bands of delta, theta, alpha, beta, and gamma are mainly utilized (34). These features characterize the signal from different aspects, so a variety of effective features extracted from the signal can be better classified.

To train an excellent model, the user usually needs to collect enough marker data for calibration. This calibration process is typically time-consuming and laborious, which is a significant problem of practical use in emotional brain computer interface. Therefore, reducing or even eliminating the calibration process and realizing Plug-and-Play is a critical challenge for the brain–computer interface from the laboratory to real life. Transfer learning is a crucial technology that can solve this problem by using annotation data from other auxiliary users to help new users build models (35). However, due to individual differences, i.e., different users have different neural responses to the same event, such that need first to perform data distribution adaptation to alleviate the individual differences of EEG features. (36). To

this end, in this paper we propose an unsupervised distributed adaptation method to align data between different users, that is, Euclidean alignment (EA) (37).

To improve EEG emotion recognition performance, performing deep neural networks to learn higher-level features would be useful to achieve good results, such as deep belief networks (7), recurrent neural networks (38), graph convolutional neural networks (39), transfer learning (40), and adversarial neural networks (41). Nevertheless, the recognition performance is limited to subject-dependent and cross-subject experiments under the same database, which is still far from realizing a practical emotional brain–computer interface. For this reason, we investigate an interesting and challenging problem in EEG emotion recognition, where training samples and test samples come from different emotional EEG databases. The preliminary research on EEG emotion recognition across data sets have demonstrated the significant drop of the recognition performance because of the inconsistency of feature distribution between the original training samples and test samples (42). Consequently, in this paper we will take advantages of the powerful high-level feature learning ability of deep learning technique to deal with the cross-database EEG emotion recognition problem.

The major contributions of this paper are summarized as follows:

(1) This paper proposes a novel recognition framework on the emotional EEG database, from raw data to recognition results, including preprocessing, feature engineering, classification recognition, and cross-database evaluation protocol.
(2) In feature engineering, we designed a time-frequency-spatial feature extraction method, combining forms of TFS, CNN, and bidirectional long and short memory network (BiLSTM) to extract multi-dimensional effectual features.
(3) Employing an unsupervised data alignment method to project data from different databases into the same space. While considering the inherent topological structure of the EEG electrodes, the preliminary TFS features are converted into three-dimensional tensors, which takes into account the information relationship between the electrodes.

This paper is organized as follows. Section 2 introduces emotion database, data processing methods, and experimental settings. Section 3 specifies the test results on the emotion database and the test results applied to the recognition of depression. Section 4 discusses the methods and results of this research. We conclude the paper in Section 5.

2. MATERIALS AND METHODS

As shown in **Figure 1**, this section mainly introduces emotion database and the algorithms of preprocessing engineering and feature engineering, including filtering, downsampling, EA, short-time Fourier transform, Hilbert–Huang transform (HHT), conversion of 1D sequence to 3D tensor, and the spatiotemporal
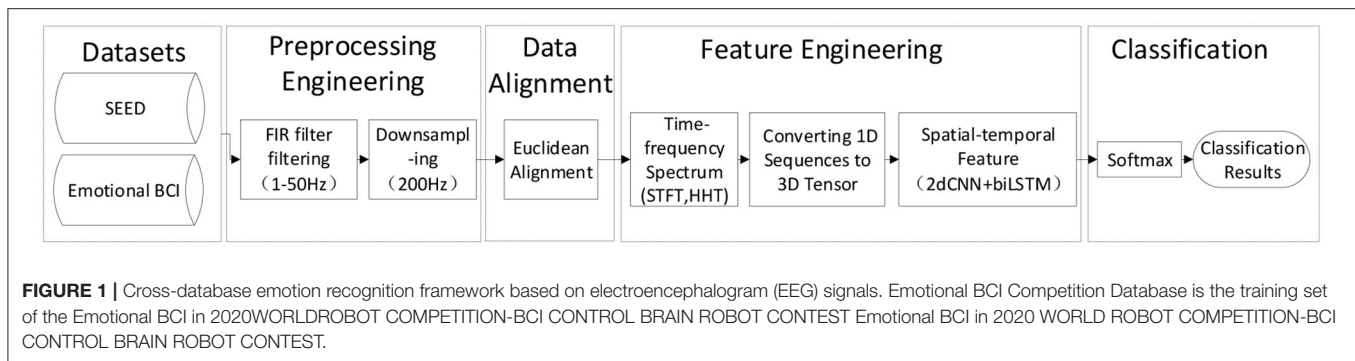
FIGURE 1 | Cross-database emotion recognition framework based on electroencephalogram (EEG) signals. Emotional BCI Competition Database is the training set of the Emotional BCI in 2020WORLDROBOT COMPETITION-BCI CONTROL BRAIN ROBOT CONTEST Emotional BCI in 2020 WORLD ROBOT COMPETITION-BCI CONTROL BRAIN ROBOT CONTEST.

TABLE 1 | Details of the two experimental databases.

	SEED	Emotional BCI Competition Database
Category	Positive, neutral, negative	happy, sad, neutral
Channel	62	62
Subject	15	23
Session	3	2
Positive	49,680	49,110
Neutral	52,650	50,722
Negative	50,400	56,694
Sum	152,730	156,526

feature extraction model combined with convolutional neural network (CNN) and BiLSTM.

2.1. Emotion Database

One of the databases used in this study is SEED (43). This database includes 15 subjects, three sessions for each subject and three emotion categories under video stimulation (i.e., positive, neutral and negative). The data were downsampled to 200 Hz. A bandpass frequency filter from 0 to 75 Hz was applied. The data are cut into one sample per second, with a total of 152,730. The other database comes from the training set of the Emotional BCI in 2020 WORLD ROBOT COMPETITION—BCI CONTROL BRAIN ROBOT CONTEST Emotional BCI in 2020 WORLD ROBOT COMPETITION— BCI CONTROL BRAIN ROBOT CONTEST (Emotional BCI Competition Database), which includes 23 subjects, two sessions (from A and B, respectively), and three emotion categories under video stimulation (i.e., happy, sad, and neutral). The data samples rate of the Emotional BCI Competition Database is 100 Hz. The EEG signals are segmented in seconds and hence results in a total of 156,520 samples. Details of the two databases are shown in **Table 1**. It can be seen from the table that the two databases have differences in categories, subjects, sessions, and the number of samples. In the subsequent processing, the three categories of happy, sad, and neutral in the emotional BCI database correspond to the positive, negative, and neutral emotion, respectively.

2.2. Preprocessing Engineering

EEG recordings measured by the scalp often contain noise and artifacts, such as blinking or movement, and cannot accurately represent signals from the brain. Therefore, it is necessary to preprocess the recorded EEG data. The preprocessing steps include converting or organizing the recorded EEG data, removing insufficient data, and segmenting the continuous original signal without changing the clean data. Appropriate band-pass filtering can effectively reduce the superimposed artifacts of various sources embedded in the EEG recording. Generally, the finite impulse response (FIR) filters are a good choice because they do not distort wave phases (44). EA maps each user's EEG signal to a new space so that the difference in the second-order statistics of the average covariance matrix of the mapped users is minimized, thereby implicitly reducing the difference in the original distribution. EA implements the above mapping for each user (auxiliary user and new user). Since different users have the same average covariance matrix after mapping, they tend to be more consistent in data distribution, meaning models trained on auxiliary users can be better applied to new users.

2.3. Data Alignment

EA is easy to perform and completely unsupervised, in which the basic idea of aligning EEG from different subjects (domains) is as follows (35): for all subjects, EA first calculates the arithmetic mean of all spatial covariance matrices.

$$\overline{R} = \frac{1}{N} \sum_{n=1}^{N} X_n (X_n)^T \qquad (1)$$

then performs the alignment by

$$\widetilde{X}_n = \overline{R}^{-\frac{1}{2}} X_n \qquad (2)$$

where $X_n \in \mathbb{R}^{c \times t}$ is the nth EEG trial, in which c is the number of EEG channels and t is the number of samples. The aligned EEG trials are whitened, and the average spatial covariance matrix of each subject is the identity matrix (45), so the EEG test distribution of different subjects is more consistent, which is meaningful for subsequent cross-database recognition.

2.4. Time-Fequency Spectrum

The EEG signal is non-linear and non-stationary, so its statistical properties (for example, spectral density) will change greatly over time. Spectrum estimation cannot identify its time-varying spectral components and cannot perform time-frequency positioning simultaneously. Time-frequency analysis technology is capable of revealing the time-varying frequency spectrum of non-stationary EEG signals and can provide a joint time-frequency distribution (TFD) of signal power (46). This paper adopts two methods of short-time Fourier transform (STFT) (47–49) and HHT (50, 51) for time-frequency spectrum (TFS) analysis. The method of calculating TFS using STFT and HHT comes from Song et al. (25).

STFT spectrum is calculated by

$$TFS_{STFT}(t,f) = \left| \int_{-\infty}^{+\infty} w(\tau - t)x(\tau)e^{-j2\pi f\tau}\, d\tau \right|^2 \quad (3)$$

where $x(t)$ is the time series and $w(\tau - t)$ is the short-time analysis window.

The Hilbert–Huang spectrum is calculated based on HHT. HHT processing non-stationary signals include three basic processes. First, the empirical mode decomposition (EMD) method is used to decompose a given signal into a number of intrinsic mode functions (IMF),

$$x(t) = \sum_{i=1}^{K} IMF_i(t) + r_K(t) \quad (4)$$

where $r_K(t)$ represents the residual of a constant or monotonic signal. These IMFs are components that meet certain conditions. Then, perform Hilbert transform on each IMF to obtain the corresponding Hilbert spectrum, that is, represent each IMF in the joint time-frequency domain. An analytic signal reconstructed by a conjugate pair (IMF and IMF_k^*) can be formulated as

$$Z_k = IMF_k(t) + jIMF_k^* = A_k(t)e^{j\theta_k(t)} \quad (5)$$
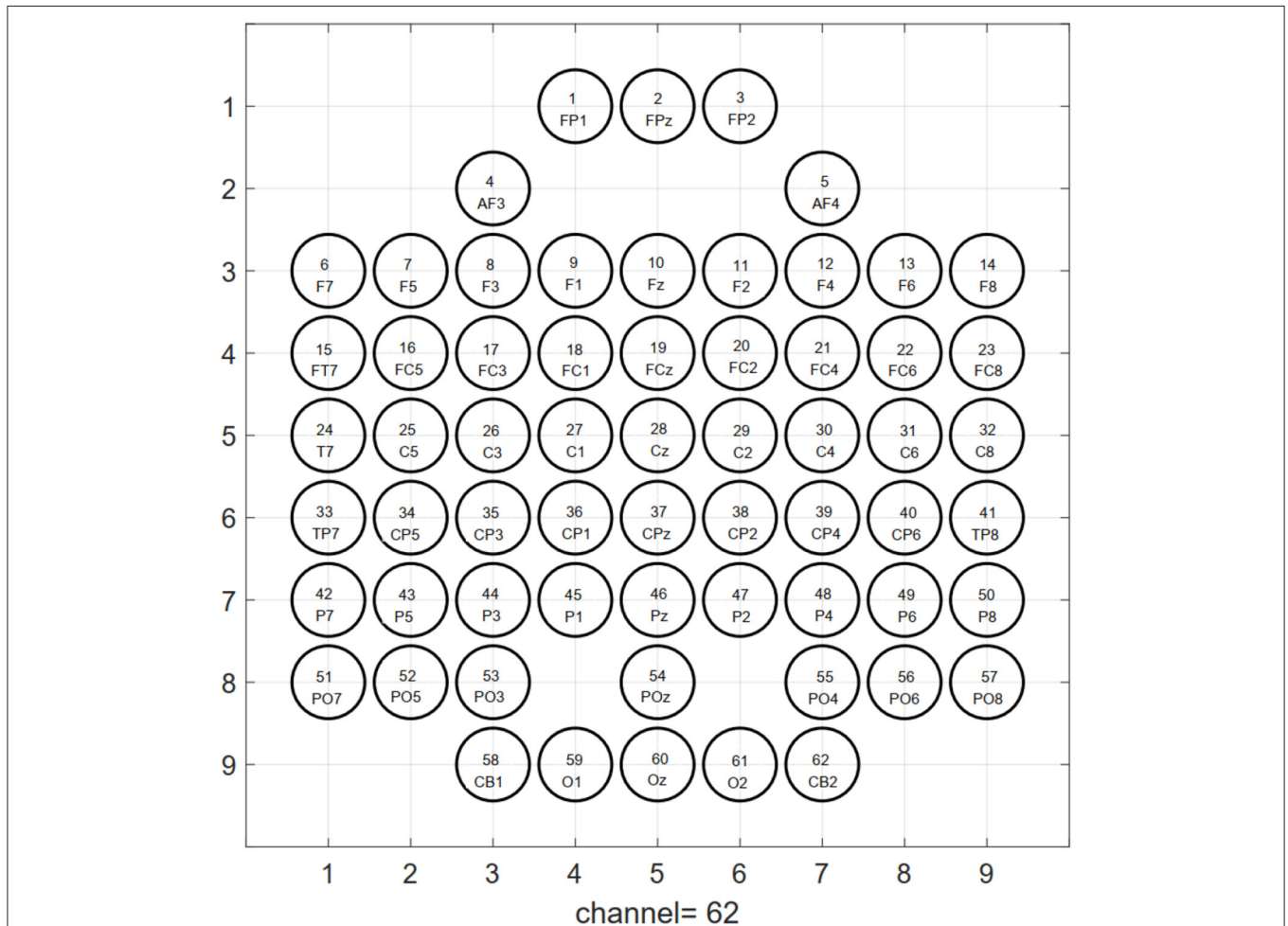

FIGURE 2 | Topological structure map of 64-electrode channels mapped to a two-dimensional picture. The circle represents the electrode, and the label inside is the serial number and name of the electrode. The left and right mastoid electrodes (M1, M2) of the 64-lead electrodes are reference electrodes when collecting signals, so they are not used as signal input for emotion recognition.

where $A_k(t)$ represents the instantaneous amplitude of $Z_k(t)$ and $\theta_k(t)$ denotes the instantaneous phase of IMFk(t). Finally, summarizing all Hilbert spectra of IMF will get the Hilbert spectra of the original signal. The original time series x(t) can be obtained by

$$x(t) = \sum_{i=1}^{K} A_k(t) e^{j2\pi \int f_k(t)\, dt} \tag{6}$$

and the instantaneous frequency can be evaluated by

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i}{dt} \tag{7}$$

where the squared amplitude $A_k^2(t)$ and instantaneous frequency $f_k(t)$ form the time-frequency spectrum.

2.5. Convert 1D Feature Sequence to 3D Tensor

Due to a large amount of noise in the EEG signal and the difficulty in capturing the unobvious relationship between the EEG signal and certain brain activities, the practical interpretation of the EEG signal is still challenging. Most of the existing studies only treat EEG as a chain sequence, ignoring the complex dependence between adjacent signals or the need to convert EEG, such as converting EEG waves into images (52).

According to the inherent topological structure of the EEG channel, as illustrated in **Figure 2**, the one-dimensional sequence

data $S_t = [s_f^1, ..., s_f^c, ..., s_f^C]$ (where s_f^c is the TFS feature of the cth electrode channel at frequency f) after extracting the TFS feature is mapped into a three-dimensional tensor $T_n \in R^{H \times W \times F}$, where the first dimension H is height, the second dimension W is width, and the third dimension F is channel (i.e., the number of features extracted per channel) of the nth EEG trial. The conversion function of 1D feature sequence to 3D tensor $T_n(H, W, f)$ is,

$$T_n(H, W, f) = \begin{bmatrix} 0 & 0 & 0 & s_f^1 & s_f^2 & s_f^3 & 0 & 0 & 0 \\ 0 & 0 & s_f^4 & 0 & 0 & 0 & s_f^5 & 0 & 0 \\ s_f^6 & s_f^7 & s_f^8 & s_f^9 & s_f^{10} & s_f^{11} & s_f^{12} & s_f^{13} & s_f^{14} \\ s_f^{15} & s_f^{16} & s_f^{17} & s_f^{18} & s_f^{19} & s_f^{20} & s_f^{21} & s_f^{22} & s_f^{23} \\ s_f^{24} & s_f^{25} & s_f^{26} & s_f^{27} & s_f^{28} & s_f^{29} & s_f^{30} & s_f^{31} & s_f^{32} \\ s_f^{33} & s_f^{34} & s_f^{35} & s_f^{36} & s_f^{37} & s_f^{38} & s_f^{39} & s_f^{40} & s_f^{41} \\ s_f^{42} & s_f^{43} & s_f^{44} & s_f^{45} & s_f^{46} & s_f^{47} & s_f^{48} & s_f^{49} & s_f^{50} \\ s_f^{51} & s_f^{52} & s_f^{53} & 0 & s_f^{54} & 0 & s_f^{55} & s_f^{56} & s_f^{57} \\ 0 & 0 & s_f^{58} & s_f^{59} & s_f^{60} & s_f^{61} & s_f^{62} & 0 & 0 \end{bmatrix} \tag{8}$$

which is the fth channel features. Among them, the positions without electrodes were filled with zeros. Each generated data grid contains spatial information of brain activity.

FIGURE 3 | Cascade 2dCNN+BiLSTM architecture.
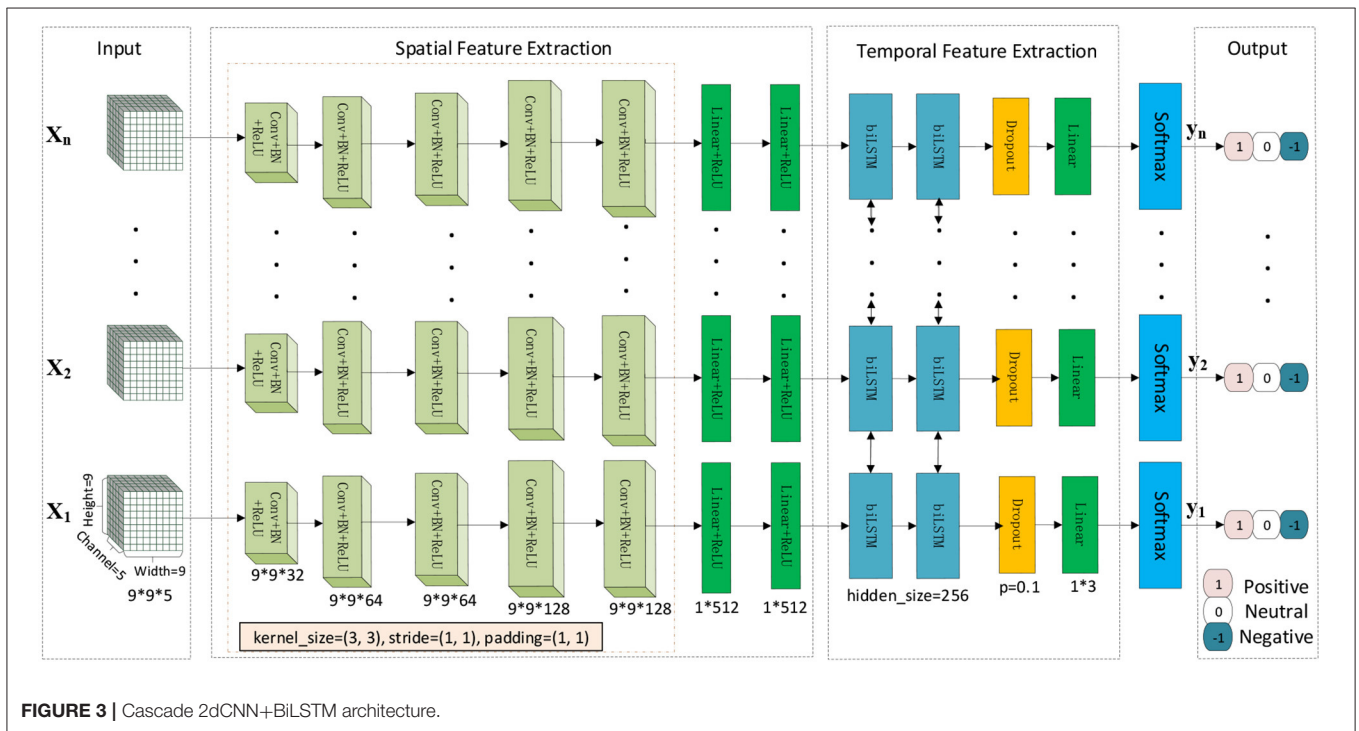
2.6. 2dCNN+BiLSTM

We designed a cascaded deep convolutional recurrent neural network framework, as shown in **Figure 3**, to capture the spatiotemporal features of EEG. The model's input is the converted 3D tensor T_n that a 3D data structure containing space and time information. First, 2D CNN extracts the spatial features of each data, BiLSTM extracts temporal features, a fully connected layer receives the output of the last step of BiLSTM, and then uses the softmax layer for final emotion prediction.

This study constructed a 2dCNN+BiLSTM model to learn a good spatiotemporal representation for multi-channel EEG. The diagram for this deep spatiotemporal network is illustrated in **Figure 3**. Since each EEG segment with the duration of 1 s is treated as one sample, we conduct time-frequency spectrum feature (ST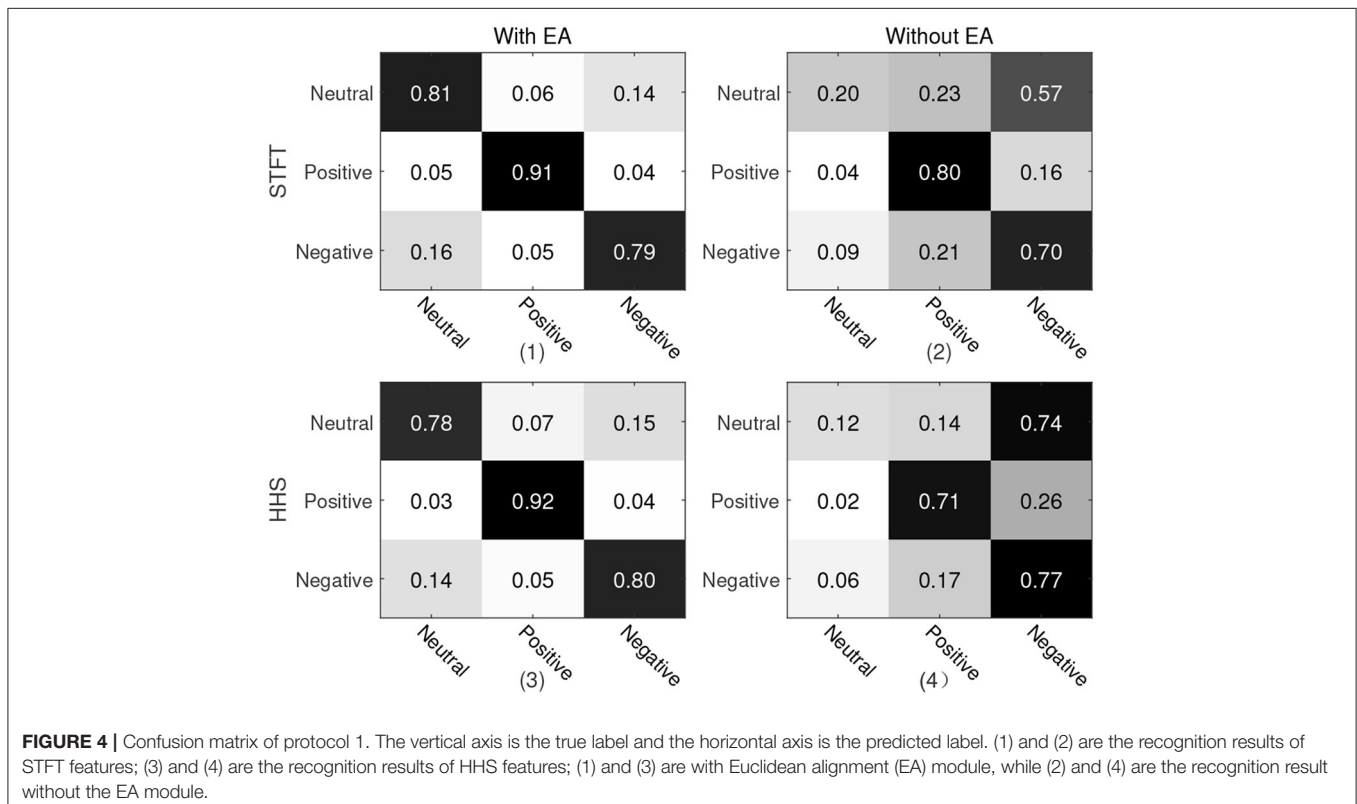FT and HHT) extraction for each sample, which was fed into the deep network for deep feature extraction. Each 1-s sample is denoted by $X_i (i = 1, 2, \ldots, n)$ and treated as a spatial image with five channels. Five convolutional layers were followed by ReLU to learn non-linear local spatial features, in which a 3 x 3 convolutional kernel was used. Following the convolutional layers, the fully connected layers were utilized to learn global spatial features. Existing studies showed that spatial features for a temporal signal are insufficient for discriminant information representation. We also employed the BiLSTM to learn temporal representation.

2.7. Experimental Settings

First, we utilize the FIR filter to perform 50-order 1–50 Hz band-pass filtering on the EEG original signal, downsampling on the Emotional BCI data to 200 Hz to be consistent with the SEED data, and then perform EA. Then extract the relative energy of the five frequency bands for each electrode channel [i.e. delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma (30–50 Hz)] using STFT and HHS, respectively. The number of features extracted from each sample is $5 * 62 = 310$, then converted into a 3D tensor of $9 * 9 * 5$. Then feed the 3D tensor to spatiotemporal network for training, the batch size is 32, the frame length is 12 (i.e., 12 s), the epoch set to 100, the cross-entropy used as loss function, the optimizer selects SGD, the learning rate initialized to 0.005. The update calculation is $lr = init\_lr * (0.95^{epoch//10})$, where $init\_lr$ is the initial learning rate.

TABLE 2 | Recognition results on emotional database.

No.	Protocol	STFT		HHS	
	Training set → Test set	Acc(%)	F1	Acc(%)	F1
1	Emotional BCI Competition Database → SEED	83.56	0.84	83.60	0.84
2	SEED → Emotional BCI Competition Database	74.33	0.72	70.26	0.70
3	Leave-One-Subject-Out	81.58	0.80	79.29	0.77

FIGURE 4 | Confusion matrix of protocol 1. The vertical axis is the true label and the horizontal axis is the predicted label. (1) and (2) are the recognition results of STFT features; (3) and (4) are the recognition results of HHS features; (1) and (3) are with Euclidean alignment (EA) module, while (2) and (4) are the recognition result without the EA module.

3. RESULTS

3.1. Emotion Recognition Results

In order to test the performance of the emotion recognition framework system built in various aspects, three protocols are proposed. In the three protocols proposed in this study, the training set and test set data are completely non-overlapping, and the test set data and labels are not used in the training process. The training set and test set of the first two protocols are from different databases. The third protocol is the leave-one-subject-out method. Considering the imbalance of the category, in addition to calculating the accuracy, the F1 score is also calculated. We applied this model to depression recognition and performed five-fold cross-validation.
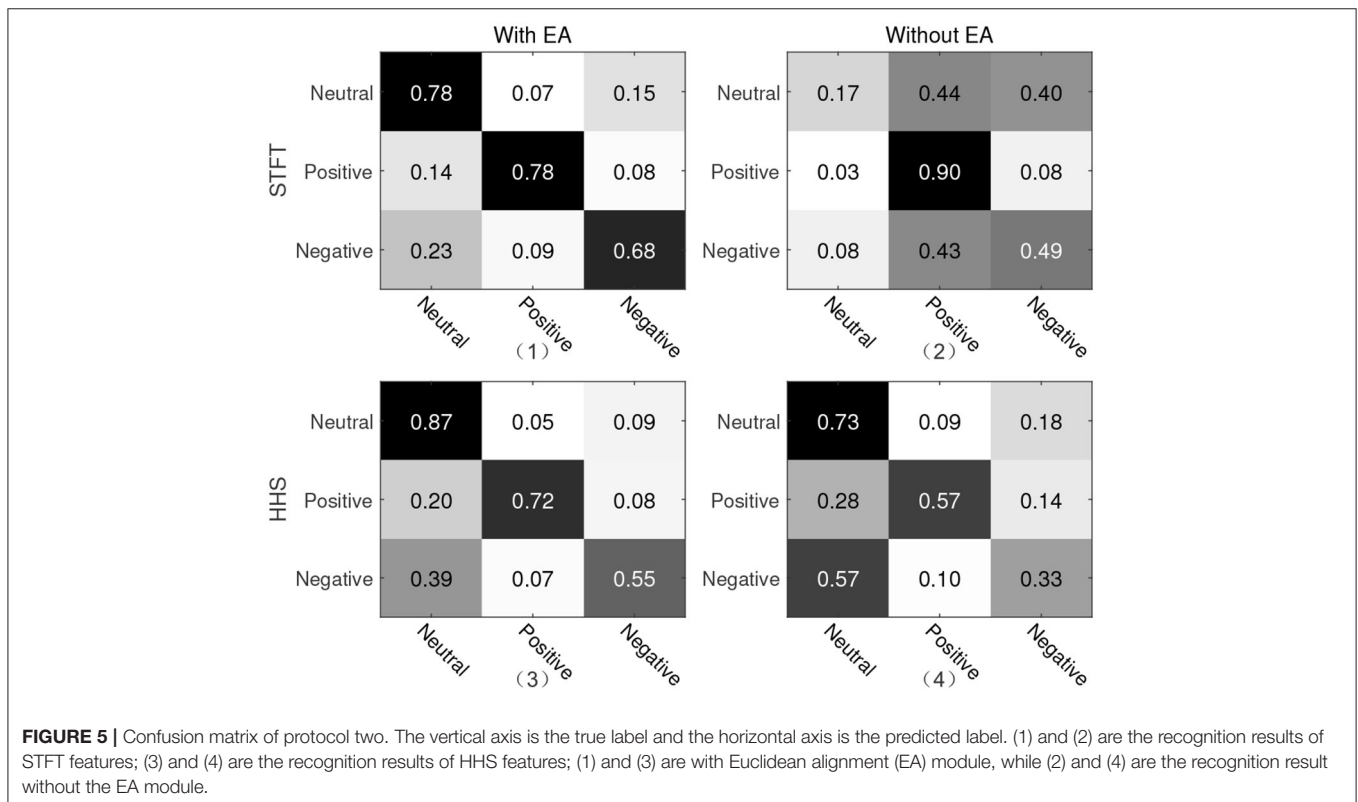
For the first protocol, all data of Emotional BCI competition database are used as the training set, and all data of SEED are used as the test set. For two different manual features, the recognition accuracy and weighted average F1-score are shown in **Table 2**. It can be seen from the table that the manual feature recognition effect extracted by the STFT method is better. In order to show the true prediction of each category, the confusion matrix of the classification accuracy is analyzed. we present a confusion matrix exploiting the features of STFT and HHS shown in **Figure 4**, from which we can see that neutral emotion has the highest recognition rate among the three types of emotions, whether it is STFT or HHS features. The recognition rate of the three types of emotions under the STFT feature is higher than that of the HHS feature.

For the second protocol, all data of SEED are used as the training set, and all data of Emotional BCI Competition Database are used as the test set. For two different manual features, the recognition accuracy and weighted average F1-score are shown in **Table 2**. The recognition rate under the STFT feature is 4.07% higher than that of the HHS feature, but it is about 9% lower than the protocol 1. Similarly, we present a confusion matrix using the features of STFT and HHS shown in **Figure 5**. The recognition rate

TABLE 3 | EA ablation experiment results.

Training set → Test set	TFS	EA	Acc(%)	F1
Emotional BCI Competition Database → SEED	STFT	With EA	**83.56**	**0.84**
		Without EA	57.29	0.54
	HHS	With EA	**83.60**	**0.84**
		Without EA	53.84	0.49
SEED → Emotional BCI Competition Database	STFT	With EA	**74.33**	**0.72**
		Without EA	52.40	0.48
	HHS	With EA	**70.26**	**0.70**
		Without EA	53.46	0.53

Bold value indicate the same experimental conditions, the maximum index with or without EA comparison.

FIGURE 5 | Confusion matrix of protocol two. The vertical axis is the true label and the horizontal axis is the predicted label. (1) and (2) are the recognition results of STFT features; (3) and (4) are the recognition results of HHS features; (1) and (3) are with Euclidean alignment (EA) module, while (2) and (4) are the recognition result without the EA module.

of the three categories under the STFT feature is relatively balanced, while the recognition rate of positive emotion

FIGURE 6 | The impact of time length selection on recognition rate. The length of time is in seconds.

under the HHS feature is significantly higher than the other two categories.

For the third protocol, the recognition results of two databases are shown in **Table 2** including the accuracy and weighted average F1-score where the recognition results are sorted according to the database (i.e., the recognition results of the subjects in each database are averaged) and the average recognition rate of all subjects is calculated. It can be seen from the table that the recognition result under the STFT feature is slightly higher than HHS.

In order to explore the influence of EA on experiment, an ablation experiment was performed on this module. As shown in **Figures 4**, **5** and **Table 3**, the difference between the recognition results of the EA module and the absence of the EA module is very obvious, whether it is protocol 1 or 2. At the same time, in order to explore the timing relationship between EEG emotional frames, the frame length is selected from 8 to 32, and the step size is 4 during training. Experiments were carried out on protocols one and two, and the experimental results are shown in **Figure 6**.
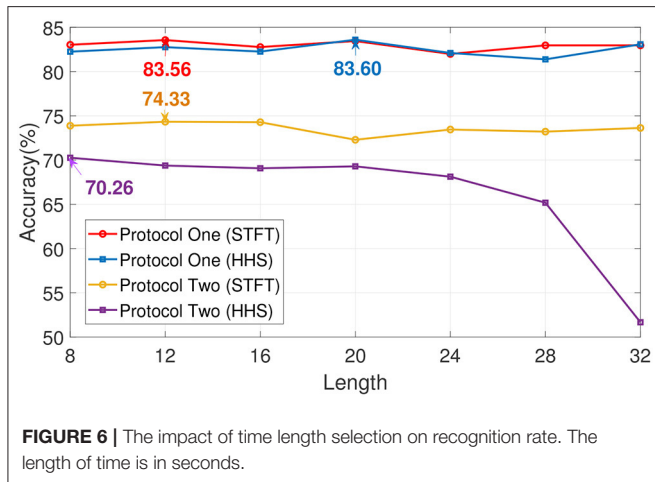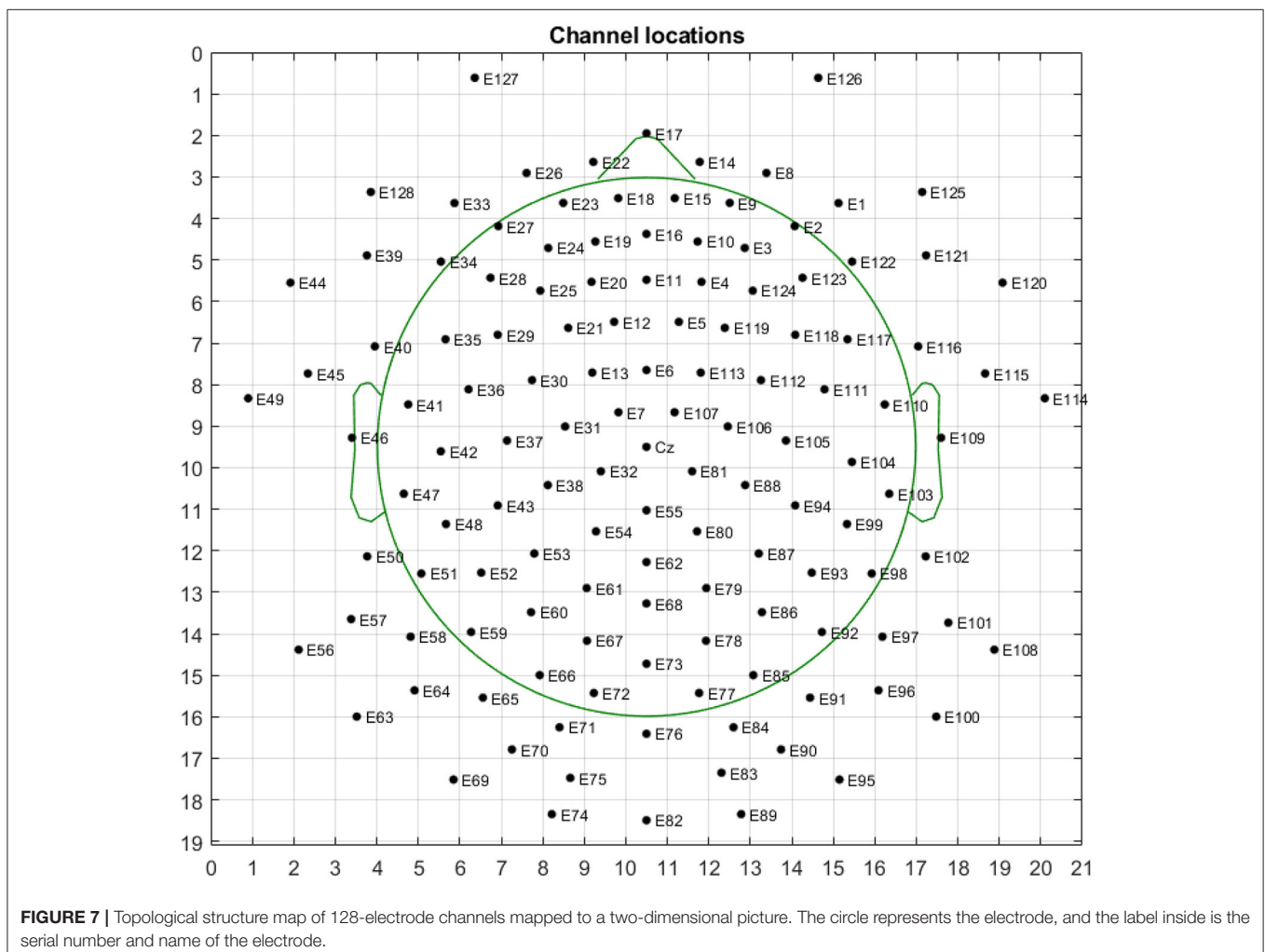

FIGURE 7 | Topological structure map of 128-electrode channels mapped to a two-dimensional picture. The circle represents the electrode, and the label inside is the serial number and name of the electrode.

3.2. Depression Recognition Results

We chose a multi-modal open dataset for depression recognition, i.e., the MODMA dataset. The dataset includes 128-channel event-related potential recordings, of which 24 major depressive disorder subjects and 29 healthy controls, the age range is 16–52 years old (53–55). Since the number of electrodes in the database is 128 and the topology is shown in **Figure 7**, the size of the three-dimensional matrix mapped to it is 21*19*5. Note that 53 subjects, including 24 outpatients and 29 healthy controls, were divided into fivefold. Due to uneven data, the first three folds made up of 5 depressed and 6 normal subjects of each one, and the last fold included 4 depressed and 5 normal subjects. The recognition result of each fold is shown in **Figure 8**. The ERP experiment is a dot-probe task, and its cue stimuli include three kinds of emotional-neutral face pairs, namely Happy-Neutral ("hcue"), Fear-Neutral ("fcue"), and Sad-Neutral ("scue"). Therefore, we not only tested all the experiments but also identified depressed patients and normal subjects on
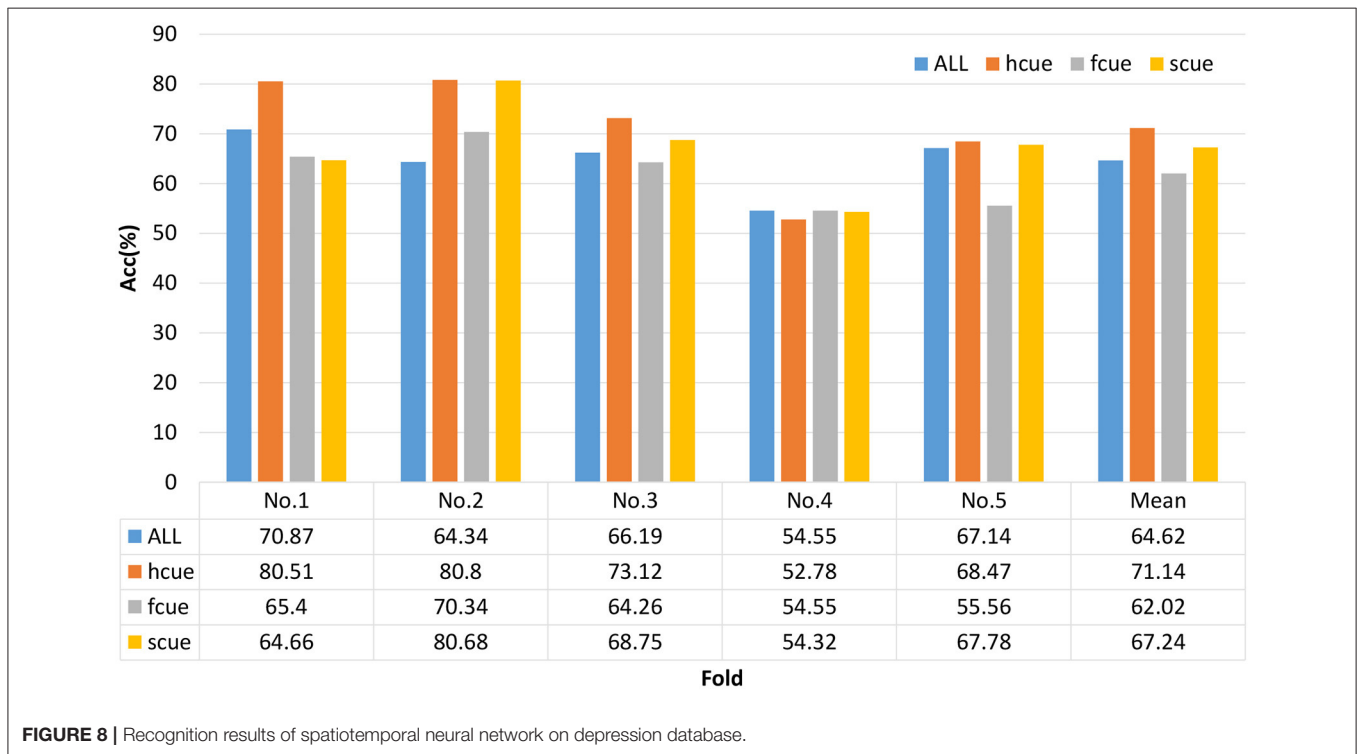

	No.1	No.2	No.3	No.4	No.5	Mean
ALL	70.87	64.34	66.19	54.55	67.14	64.62
hcue	80.51	80.8	73.12	52.78	68.47	71.14
fcue	65.4	70.34	64.26	54.55	55.56	62.02
scue	64.66	80.68	68.75	54.32	67.78	67.24

FIGURE 8 | Recognition results of spatiotemporal neural network on depression database.
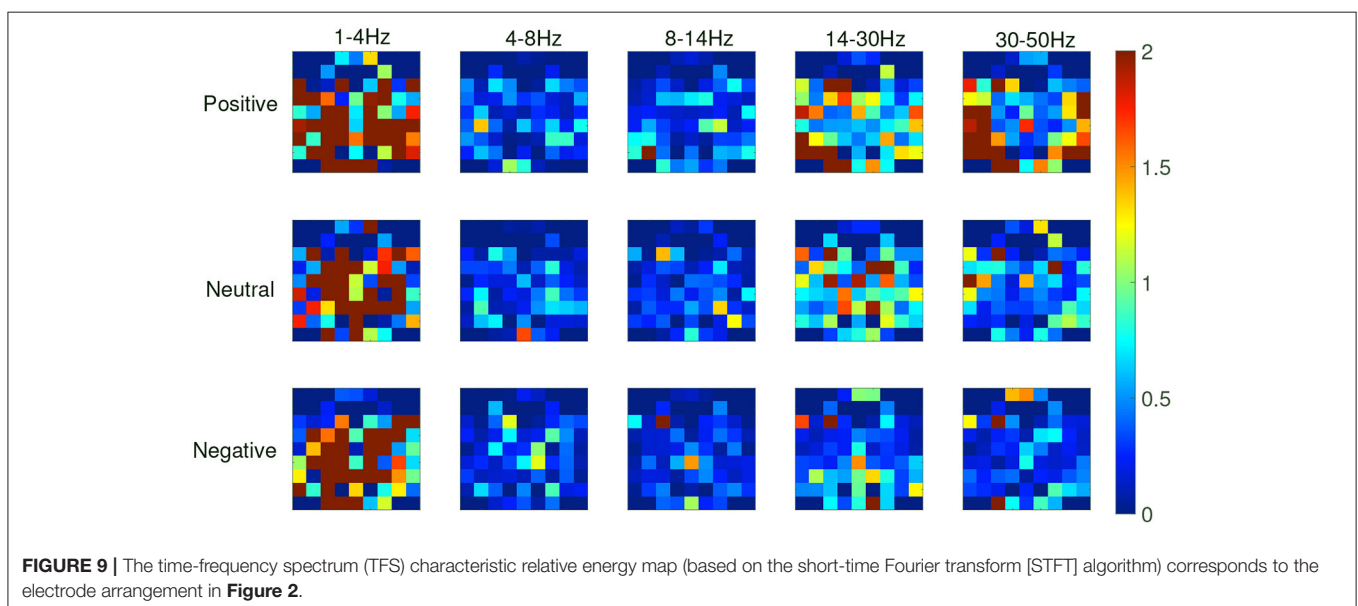

FIGURE 9 | The time-frequency spectrum (TFS) characteristic relative energy map (based on the short-time Fourier transform [STFT] algorithm) corresponds to the electrode arrangement in **Figure 2**.

different stimuli. Among them, the overall recognition rate on "hcue" is the highest, reaching 71.14%.

4. DISCUSSION

This paper proposes a complete pipeline from preprocessing to EEG-based emotion recognition, with a recognition rate of over 80%. The preprocessing part follows with the unsupervised EA method to map the data of different databases to the same space, where STFT, CNN, and BiLSTM are combined to extract multi-domain features in the time-frequency space. Before the CNN operation, according to the spatial arrangement of the EEG electrodes, the one-dimensional time series feature is converted into a three-dimensional tensor, such that the correlation between EEG electrodes can be fully considered. Moreover, we use 2D CNN to extract spatial features, and BiLSTM to capture the timing relationship of features.

It can be seen from the confusion matrix of protocol 1 that the recognition rates of the three categories under the two methods of STFT and HHS are relatively balanced, and the positive emotion recognition rate is the highest. The neutral emotion recognition rate under the HHS method of protocol 2 is the highest, and the negative emotion recognition rate is the lowest. There is a 9% difference between the accuracy of protocols 1 and 2. Since the three categories of data in the Emotional BCI database are more diversified (the first 15 people and the bottom eight people in the three categories of the Emotional BCI database in **Table 1** are different).
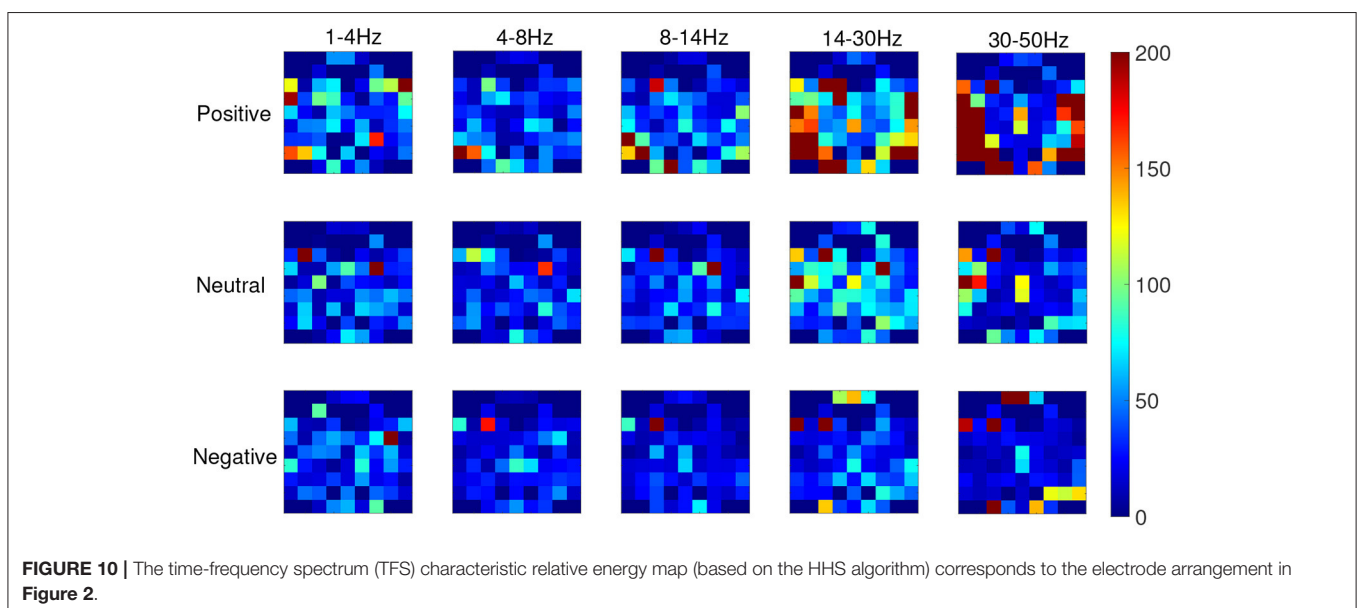
From the recognition results of all protocols, the accuracy and F1 score of the TFS features extracted by STFT are higher than those of the HHS method. **Figure 9** shows the STFT method and **Figure 10** shows the HHS method. The three categories are displayed in five frequency bands, and each spectrum is shown per the electrode arrangement in **Figure 2**. The features extracted by the two methods are pretty different in the high-frequency range. The relative energy of the two frequency bands, beta (14–30 Hz) and gamma (30–50 Hz), under the STFT method, is

relatively high, and the three categories have apparent differences. In contrast, the HHS method has relatively high positive and neutral relative energies in these two frequency bands. Negative emotions have always been low energy in the entire frequency band. Hence, the recognition rate of the HHS method is lower than that of STFT, and it performs well in positive and negative emotions.

To further validate the proposed method, we compared our model with the start-of-the-art methods. **Table 4** presents a summary of the current subject-independent recognition algorithms on the SEED database, including linear support vector machine (SVM) (56), kernel principal component analysis (KPCA) (57), transfer component analysis (TCA) (58), transductive parameter transfer (TPT) (59), domain adversarial neural network (DANN) (60), dynamical graph convolutional neural network (DGCNN) (39), bi-hemispheres domain adversarial neural network (BiDANN) (61), BiDANN-S (41), hierarchical spatial-temporal neural network (R2G-STNN) (62), and instance-adaptive graph (IAG) (63). It can be seen

TABLE 4 | The mean accuracies (Acc) and standard deviations (Std) on SEED dataset for subject-independent EEG emotion recognition experiment.

Method	Acc/Std(%)
SVM (56)	56.73/16.29
KPCA (57)	61.28/14.62
TCA (58)	63.64/14.88
TPT (59)	76.31/15.89
DANN (60)	75.08/11.18
DGCNN (39)	79.95/09.02
BiDANN (61)	83.28/09.60
BiDANN-S (41)	84.14/06.87
R2G-STNN (62)	84.16/07.63
IAG (63)	86.30/06.91
ours	**86.42/05.26**

FIGURE 10 | The time-frequency spectrum (TFS) characteristic relative energy map (based on the HHS algorithm) corresponds to the electrode arrangement in **Figure 2**.

from the table that our method has achieved the highest accuracy and the smallest standard deviation. Unlike these methods, our training set adds the Emotional BCI database. The training set's increase makes the training model's generalization better, proving that the proposed method can effectively extract spatiotemporal multi-view features and classify emotions well across databases or subjects.

5. CONCLUSIONS

This study designed a complete pipeline from preprocessing to the classification of emotion recognition based on EEG, which achieved a correct rate of more than 80%. It is significant that we apply this model to the recognition of depression based on EEG signals. The preprocessing combined with the unsupervised EA method maps the data of different databases to the same space. The three methods of STFT, CNN, and BiLSTM are combined to extract the time-frequency-space multi-domain features. Before the CNN operation, the one-dimensional time series feature was converted into a three-dimensional tensor according to the spatial arrangement of the EEG electrodes. In the future, we will study end-to-end real-time emotional brain–computer interfaces for depression recognition.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://bcmi.sjtu.edu.cn/~seed/index.html; http://modma.lzu.edu.cn/data/application/#data_1; https://oneuro.cn/n/index.html?code=d349d58c825b4041a0e53ea55b5157ae/state=#/chinabci, Emotional BCI.

AUTHOR CONTRIBUTIONS

HC and YZ: conceptualization. HC: methodology and writing and original draft preparation. WZ: formal analysis and funding acquisition. CT: investigation. JZ: resources. XL: data curation. CT and JZ: review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

REFERENCES

1. Daros AR, Haefner SA, Asadi S, Kazi S, Rodak T, Quilty LC. A meta-analysis of emotional regulation outcomes in psychological interventions for youth with depression and anxiety. *Nat Hum Behav*. (2021) 5:1443–57. doi: 10.1038/s41562-021-01191-9

2. Vanderlind WM, Millgram Y, Baskin-Sommers AR, Clark MS, Joormann J. Understanding positive emotion deficits in depression: from emotion preferences to emotion regulation. *Clin Psychol Rev*. (2020) 76:101826. doi: 10.1016/j.cpr.2020.101826

3. Dolan RJ. Emotion, cognition, and behavior. *Science*. (2002) 298:1191–4. doi: 10.1126/science.1076358

4. Balconi M, Lucchiari C. EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. *Neurosci Lett*. (2006) 392:118–23. doi: 10.1016/j.neulet.2005.09.004

5. Bekkedal MY, Rossi III J, Panksepp J. Human brain EEG indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neurosci Biobehav Rev*. (2011) 35:1959–70. doi: 10.1016/j.neubiorev.2011.05.001

6. Davidson PR, Jones RD, Peiris MT. EEG-based lapse detection with high temporal resolution. *IEEE Trans Biomed Eng*. (2007) 54:832–9. doi: 10.1109/TBME.2007.893452

7. Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans Auton Ment Dev*. (2015) 7:162–75. doi: 10.1109/TAMD.2015.2431497

8. Gross JJ, Mu noz RF. Emotion regulation and mental health. *Clin Psychol (New York)*. (1995) 2:151–64. doi: 10.1111/j.1468-2850.1995.tb00036.x

9. Jordan KG. Continuous EEG and evoked potential monitoring in the neuroscience intensive care unit. *J Clin Neurophysiol*. (1993) 10:445–75. doi: 10.1097/00004691-199310000-00006

10. Chanel G, Kierkels JJ, Soleymani M, Pun T. Short-term emotion assessment in a recall paradigm. *Int J Hum Comput Stud*. (2009) 67:607–27. doi: 10.1016/j.ijhcs.2009.03.005

11. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. *Nature*. (2005) 433:68–72. doi: 10.1038/nature03086

12. Blair R, Morris JS, Frith CD, Perrett DI, Dolan RJ. Dissociable neural responses to facial expressions of sadness and anger. *Brain*. (1999) 122:883–93. doi: 10.1093/brain/122.5.883

13. Adolphs R, Tranel D. Impaired judgments of sadness but not happiness following bilateral amygdala damage. *J Cogn Neurosci*. (2004) 16:453–62. doi: 10.1162/089892904322926782

14. Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron*. (2003) 40:655–64. doi: 10.1016/S0896-6273(03)00679-2

15. Balconi M, Mazza G. Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues.: ERS/ERD and coherence measures of alpha band. *Int J Psychophysiol*. (2009) 74:158–65. doi: 10.1016/j.ijpsycho.2009.08.006

16. Liu Y, Sourina O, Nguyen MK. Real-time EEG-based emotion recognition and its applications. In: *Transactions on Computational Science XII*. Berlin; Heidelberg: Springer; (2011). p. 256–77.

17. Jatupaiboon N, Pan-ngum S, Israsena P. Emotion classification using minimal EEG channels and frequency bands. In: *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. Khon Kaen: IEEE (2013). p. 21–4.

18. Li M, Lu BL. Emotion classification based on gamma-band EEG. In: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Minneapolis, MN: IEEE (2009). p. 1223–6.

19. Park KS, Choi H, Lee KJ, Lee JY, An KO, Kim EJ. Emotion recognition based on the asymmetric left and right activation. *Int J Medical Sci*. (2011) 3:201–209. doi: 10.1016/j.neuroimage.2016.05.059

20. Hjorth B. EEG analysis based on time domain properties. *Electroencephalogr Clin Neurophysiol*. (1970) 29:306–10. doi: 10.1016/0013-4694(70)90143-4

21. Petrantonakis PC, Hadjileontiadis LJ. Emotion recognition from EEG using higher order crossings. *IEEE Trans Learn Technol*. (2009) 14:186–97. doi: 10.1109/TITB.2009.2034649

22. Khosrowabadi R, Quek HC, Wahab A, Ang KK. EEG-based emotion recognition using self-organizing map for boundary detection. In: *2010 20th International Conference on Pattern Recognition*. Istanbul: IEEE (2010) p. 4242–5.

23. Bos DO. EEG-based emotion recognition. *Influence Visual Auditory Stimuli*. (2006) 56:1–17.

24. Musha T, Terasaki Y, Haque HA, Ivamitsky GA. Feature extraction from EEGs associated with emotions. *Artif Life Rob*. (1997) 1:15–9. doi: 10.1007/BF02471106

25. Song T, Zheng W, Lu C, Zong Y, Zhang X, Cui Z. MPED: a multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*. (2019) 7:12177–91. doi: 10.1109/ACCESS.2019.2891579

26. Hatamikia S, Maghooli K, Nasrabadi AM. The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. *J Med Signals Sens*. (2014) 4:194. doi: 10.4103/2228-7477.137777

27. Huang D, Guan C, Ang KK, Zhang H, Pan Y. Asymmetric spatial pattern for EEG-based emotion detection. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. Brisbane, QLD: IEEE (2012). p. 1–7.

28. Jie X, Cao R, Li L. Emotion recognition based on the sample entropy of EEG. *Biomed Mater Eng*. (2014) 24:1185–92. doi: 10.3233/BME-13 0919

29. Hosseini SA, Naghibi-Sistani MB. Emotion recognition method using entropy analysis of EEG signals. *Int J Image Graph Signal Process*. (2011) 3:30. doi: 10.5815/ijigsp.2011.05.05

30. Wang XW, Nie D, Lu BL. EEG-based emotion recognition using frequency domain features and support vector machines. In: *International Conference on Neural Information Processing*. Berlin; Heidelberg: Springer (2011). p. 734–43.

31. Pan J, Li Y, Wang J. An EEG-based brain-computer interface for emotion recognition. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. Vancouver, BC: IEEE. (2016). p. 2063–7.

32. Koelstra S, Patras I. Fusion of facial expressions and EEG for implicit affective tagging. *Image Vis Comput*. (2013) 31:164–74. doi: 10.1016/j.imavis.2012.10.002

33. Thammasan N, Fukui Ki, Numao M. Application of deep belief networks in eeg-based dynamic music-emotion recognition. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. Vancouver, BC: IEEE (2016). p. 881–8.

34. Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. *IEEE Trans Affect Comput*. (2017) 10:374–93. doi: 10.1109/TAFFC.2017.2714671

35. Wu D, Peng R, Huang J, Zeng Z. Transfer learning for brain-computer interfaces: a complete pipeline. *arXiv preprint* arXiv:200703746. (2020).

36. Wu D, Xu Y, Lu BL. Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016. *IEEE Trans Cogn Dev Syst*. (2020) 1. doi: 10.1109/TCDS.2020.3007453

37. He H, Wu D. Transfer learning for Brain-Computer interfaces: A Euclidean space data alignment approach. *IEEE Trans Biomed Eng*. (2019) 67:399–410. doi: 10.1109/TBME.2019.2913914

38. Zhang T, Zheng W, Cui Z, Zong Y, Li Y. Spatial-temporal recurrent neural network for emotion recognition. *IEEE Trans Cybern*. (2018) 49:839–47. doi: 10.1109/TCYB.2017.2788081

39. Song T, Zheng W, Song P, Cui Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans Affect Comput*. (2018) 11: 532–41. doi: 10.1109/TAFFC.2018.281 7622

40. Li J, Qiu S, Du C, Wang Y, He H. Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Trans Cogn Develop Syst*. (2019) 12:344–353. doi: 10.1109/TCDS.2019.29 49306

41. Li Y, Zheng W, Zong Y, Cui Z, Zhang T, Zhou X. A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. *IEEE Trans Affect Comput*. (2018) 12: 494–504. doi: 10.1109/TAFFC.2018.288 5474

42. Lan Z, Sourina O, Wang L, Scherer R, Müller-Putz GR. Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Trans Cogn Develop Syst*. (2018) 11:85–94. doi: 10.1109/TCDS.2018.2826840

43. Duan RN, Zhu JY, Lu BL. Differential entropy feature for EEG-based emotion classification. In: *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. San Diego, CA: IEEE (2013). p. 81–4.

44. Teplan M. Fundamentals of EEG measurement. *Meas Sci Rev*. (2002) 2:1–11. Available online at: http://www.edumed.org.br/cursos/neurociencia/MethodsEEGMeasurement.pdf

45. Zhang W, Wu D. Manifold Embedded Knowledge Transfer for Brain-Computer Interfaces. *IEEE Trans Neural Syst Rehabilitation Eng*. (2020) 28:1117–27. doi: 10.1109/TNSRE.2020.298 5996

46. Boashash B. *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*. Cambridge, MA: Academic Press (2015).

47. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. (2004) 134:9–21. doi: 10.1016/j.jneumeth.2003. 10.009

48. Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, et al. EEG-based emotion recognition in music listening. *IEEE Trans Biomed Eng*. (2010) 57:1798–806. doi: 10.1109/TBME.2010.204 8568

49. Sammler D, Grigutsch M, Fritz T, Koelsch S. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*. (2007) 44:293–304. doi: 10.1111/j.1469-8986.2007.00 497.x

50. Huang NE. *Hilbert-Huang transform and Its Applications, Vol. 16*. Singapore: World Scientific (2014).

51. Huang NE, Wu Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Rev Geophys*. (2008) 46:1–23. doi: 10.1029/2007RG000228

52. Zhang D, Yao L, Zhang X, Wang S, Chen W, Boots R, et al. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. In: *AAAI*, New Orleans. (2018). p. 1703–10.

53. Cai H, Gao Y, Sun S, Li N, Tian F, Xiao H, et al. MODMA dataset: a multi-modal open dataset for mental-disorder analysis. *arXiv preprint* arXiv:200209283. (2020).

54. Hu B, Rao J, Li X, Cao T, Li J, Majoe D, et al. Emotion regulating attentional control abnormalities in major depressive disorder: an event-related potential study. *Sci Rep*. (2017) 7:1–21. doi: 10.1038/s41598-017-13626-3

55. Li X, Li J, Hu B, Zhu J, Zhang X, Wei L, et al. Attentional bias in MDD: ERP components analysis and classification using a dot-probe task. *Comput Methods Programs Biomed*. (2018) 164:169–79. doi: 10.1016/j.cmpb.2018.07.003

56. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. (1999) 9:293–300. doi: 10.1023/A:1018628609742

57. Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput*. (1998) 10:1299–319. doi: 10.1162/089976698300017467

58. Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. *IEEE Trans Neural Netw*. (2010) 22:199–210. doi: 10.1109/TNN.2010.2091281

59. Sangineto E, Zen G, Ricci E, Sebe N. We are not all equal: personalizing models for facial expression analysis with transductive parameter transfer. In: *Proceedings of the 22nd ACM International Conference on Multimedia*, Lisboa. (2014). p. 357–66. doi: 10.1145/2647868.2654916

60. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. *J Mach Learn Res*. (2016) 17:1–35. Available online at: https://www.jmlr.org/papers/volume17/15-239/15-239.pdf

61. Li Y, Zheng W, Cui Z, Zhang T, Zong Y. A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm. (2018). p. 1561–7.

62. Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Trans Affect Comput*. (2019) 1. doi: 10.1109/TAFFC.2019.2922912

63. Song T, Liu S, Zheng W, Zong Y, Cui Z. Instance-adaptive graph for EEG emotion recognition. In: *AAAI*. (2020). p. 2701–8.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.