Research article

BiTCAN: An emotion recognition network based on saliency in brain cognition

Yanling An[1], Shaohai Hu[1,*], Shuaiqi Liu[2,3,4,*] and Bing Li[4]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[2] College of Electronic and Information Engineering, Hebei University, Baoding 071000, China
[3] Machine Vision Technology Innovation Center of Hebei Province, Baoding 071000, China
[4] The State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

* **Correspondence:** Email: shhu@bjtu.edu.cn, shdkj-1918@163.com.

Abstract: In recent years, with the continuous development of artificial intelligence and brain-computer interfaces, emotion recognition based on electroencephalogram (EEG) signals has become a prosperous research direction. Due to saliency in brain cognition, we construct a new spatio-temporal convolutional attention network for emotion recognition named BiTCAN. First, in the proposed method, the original EEG signals are de-baselined, and the two-dimensional mapping matrix sequence of EEG signals is constructed by combining the electrode position. Second, on the basis of the two-dimensional mapping matrix sequence, the features of saliency in brain cognition are extracted by using the Bi-hemisphere discrepancy module, and the spatio-temporal features of EEG signals are captured by using the 3-D convolution module. Finally, the saliency features and spatio-temporal features are fused into the attention module to further obtain the internal spatial relationships between brain regions, and which are input into the classifier for emotion recognition. Many experiments on DEAP and SEED (two public datasets) show that the accuracies of the proposed algorithm on both are higher than 97%, which is superior to most existing emotion recognition algorithms.

Keywords: EEG; emotion recognition; spatio-temporal features; Bi-hemispheric discrepancy; spatial attention; attention mechanism

1.   Introduction

Emotion refers to a spontaneous and uncontrollable reaction that occurs when a person is affected by the surrounding environment. It is physically manifested as sweating or rapid heartbeat, and psychologically manifested as sadness or pleasure. Emotions change rapidly with the volatility of the surrounding environment or their own physiological needs in a short period of time. The complexity of the brain's neural mechanisms causes individuals to produce stressful or continuous emotional responses [1]. There are roughly two kinds of existing emotion recognition algorithms: One kind is based on non-physiological signals, for example, facial expression [2], speech intonation [3] and morphological posture [4], etc; the other is based on physiological signals [5,6] (such as electrooculogram [7], electromyography [8] and electrocardiogram [9], etc.). Non-physiological signals have the advantages of easy extraction of data and easy deployment, but such signals are highly subjective, easy to disguise and hide [10], and cannot reflect the emotional state truly and effectively. Physiological signals can objectively and truly reflect the variation of the central nervous system, which is hard to hide and can more effectively express real emotions. However, most of the physiological signals are extremely weak in emotion-related components, and the impact of noise is very obvious, which often leads to low recognition accuracy of the emotion recognition algorithm. EEG can directly reflect the activity of the human brain with the change of emotion, which is objective, reliable and easy to identify. In addition, EEG is non-invasive, so it has great development potential in emotion recognition [11].

Recently, with the development of brain-computer interfaces, the new portable and wearable EEG devices have become a part of people's lives. As one hot research issue in the field of brain-computer interface, many scholars proposed effective emotion recognition methods [12,13]. To extract EEG signal features effectively, researchers have explored the feature extraction of EEG signals in time the domain, frequency domain and time-frequency domain. At the same time, researchers constructed many emotion recognition networks that are suitable for EEG signals. For example, in [14], An et al. extracted and fused the differential entropy features from different frequencies of EEG signals, and they also constructed the spatial-frequency domain features by combining them with the placement of electrodes. Finally, the spatial-frequency domain features were input into the convolutional autoencoder (CAE) for emotion classification, which achieved a good effect. In [15], Liu et al. presented a dynamic differential entropy features extraction algorithm from EEG signals by combining differential entropy (DE) and empirical mode decomposition (EMD). Then, CNN was used to classify the extracted dynamic differential entropy features to obtain a better emotion recognition effect. In [16], Liu et al. used convolutional neural network and deep neural network (DNN) to achieve good emotion recognition effect. In [17], Cheng et al. first made baseline correction of the original EEG signals, and then they constructed the two-dimensional frame structure of the EEG signal by combining the spatial structure between electrodes, and used deep forest to distinguish emotion based on multi-channel EEG signals. This algorithm solves the shortcomings of the deep neural network with too many hyperparameters and too large training data, and achieves high accuracy of emotion classification. In [18], Song et al. proposed a feature extraction algorithm from EEG signals based on the variational instance-adaptive graph (V-IAG). This algorithm can capture the individual dependence between different EEG electrodes simultaneously, so as to obtain more discriminative

features. Simultaneously, Song et al. also gives a multi-channel EEG signal features extraction algorithm aggregated by multi-stage multi-graph convolution operation. They also used sparse constraints to obtain more robust features to achieve better emotion recognition effects. In [19], Cui et al. constructed an end-to-end network named regional asymmetric convolutional neural network (RACNN) and applied it to emotion recognition. RACNN can obtain the characteristics of the differences between the right hemispheres of the brain and the left hemispheres of the brain stimulated by emotion.

Although the above emotion recognition algorithms have achieved good emotion recognition effect, these algorithms often fail to take into account the intrinsic characteristics of EEG signals, and seldom consider how to use fewer electrodes to achieve better recognition effect in practical application. EEG signal is a time-varying and continuous biological time series. Continuous time EEG contains dynamic features, which are of great significance to the classification of emotions. The placement of electrode caps is orderly, and is based on the spatial characteristics of EEG signals. Therefore, the orderly relative position of electrodes will be a great help to the classification of emotions [20]. In addition, neurological studies show that there is a very close relationship between the cerebral cortex and the production of emotion [21]. Due to the phenomenon of not completely symmetrical right and left hemispheres in the brain, in [22], Li et al. proposed a Bi-hemispheric discrepancy model (BiHDM) to extract the asymmetry feature of EEG for emotion recognition. In [23], Huang et al. proposed a bi-hemisphere discrepancy convolutional neural network model (BiDCNN) for EEG emotion recognition. To achieve more accurate emotion recognition, we construct a new spatio-temporal convolution spatial attention network to achieve EEG emotion recognition by combining saliency in brain cognition and the characteristics of EEG signals mentioned above.

Besides, with the development of brain-computer interfaces and portable EEG devices, the number of electrodes should be minimized without affecting the recognition effect. Therefore, we explore the importance of EEG triggered by different brain regions in emotion recognition. Experiments show that the asymmetry of the frontal and temporal lobes plays an important role in emotion recognition tasks, which also indicates that the frontal and temporal lobes are the main brain regions related to emotion in the brain. These findings are consistent with the findings of neuropsychology. This conclusion provides a theoretical basis for channel selection in EEG emotion recognition, in order to reduce the computational complexity of emotion recognition models with a small number of EEG channels, and for the development of convenient EEG devices. The following are the paper's main contributions:

(1) In this paper, a three-dimensional (3D) spatio-temporal matrix is constructed by using the original EEG signal. The three-dimensional spatio-temporal matrix combined with the spatio-temporal convolution module can extract the spatio-temporal features of EEG signals efficiently, while the 3D spatio-temporal matrix combined with the Bi-hemispheric discrepancy module can effectively extract saliency in brain cognition of emotion.

(2) In this paper, spatio-temporal features and saliency features are fused and sent into the attention module. The attention module can reduce feature redundancy and extract the internal spatial relations between multi-channel EEG signals globally, so as to obtain better emotion recognition results.

(3) In this paper, the contribution of brain asymmetry characteristics in different brain regions to emotion recognition provides a theoretical basis for how to select electrodes when developing portable EEG devices, and provides ideas for the development of brain-computer interfaces (BCIs).

The rest of the paper is structured as follows. The second section describes the EEG data preparation approach and the proposed network in depth. The third section introduces the data collection and experimental settings employed in this paper, as well as the specifics of model training, before comparing and analyzing the experimental outcomes. At the same time, exploratory experiments were designed to demonstrate the good performance of the network constructed in this paper in emotion recognition tasks. The fourth section summarizes the methods proposed in this paper.

2. Methodology

To fully use the spatial and temporal properties of EEG signals, as well as the asymmetry of EEG signals in the left and right hemispheres of the brain, we construct a deep network based on Bi-hemispheric discrepancy and spatio-temporal convolution attention, named BiTCAN, and apply it to emotion recognition. In this paper, we first preprocess the original EEG signal and intercept fixed time period to spatially code each channel based on information on the physical location of the electrodes. Finally, the spatial and temporal features of EEG signals with position information are obtained by combining the three-dimensional spatio-temporal convolution module. In this paper, we calculate the difference between the right and left hemispheres of EEG signals by using a bi-hemispheric discrepancy module to obtain the asymmetric characteristics of EEG signals. Finally, the attention model is used to combine spatio-temporal and asymmetry information into the softmax classifier for emotion recognition. The schematic diagram of the proposed BiTCAN for emotion recognition is shown in Figure 1. The proposed algorithm includes a three-dimensional spatio-temporal matrix construction module, a bi-hemispheric discrepancy module, a three-dimensional spatio-temporal convolution module, an attention module, and an emotion classification module. Each module of the proposed algorithm is described below.
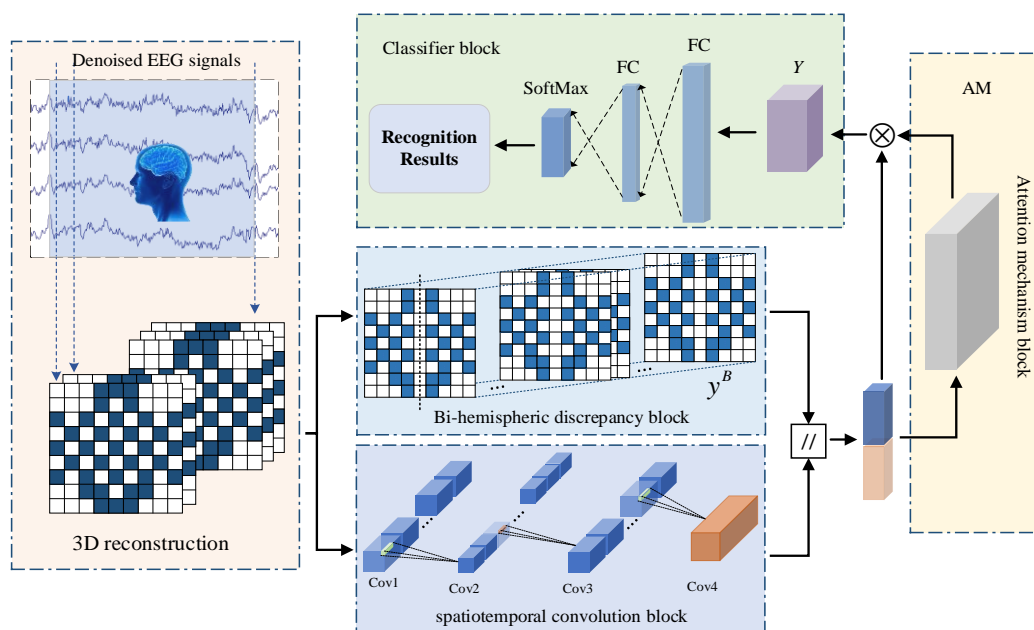
Figure 1. The BiTCAN model's framework for EEG emotion recognition.

2.1. 3D spatio-temporal matrix construction module

In the research of EEG emotion recognition, most algorithms only consider the experimental signal in the experiment, while ignoring the influence of the baseline signal (signal without emotion) on the emotion recognition algorithm. The human EEG signals are unstable because human EEG signals are vulnerable to small environmental changes. At the same time, the EEG signals generated by emotional stimulation are also affected by the emotional state before stimulation to a certain extent, so baseline removal can effectively inhibit the EEG signals unrelated to emotion and raise the effect of emotion recognition.

In the proposed model, $X^T = \left(v_1^E, v_2^E, \cdots, v_t^E \right) \in R^{E \times T}$ is defined as EEG signals containing T time points, where E is the number of electrodes, and $v_t^E = \left(s_t^1, s_t^2, \cdots, s_t^E \right) \in R^E \left(t \in \{1, 2, \cdots, T\} \right)$ is the EEG signals of all E electrodes collected at a certain time point t. The baseline signal and experimental signal in the original EEG signal are divided into K segment and I segment. The k-th baseline signal segment and the i experimental signal segment are represented by B_k and X_i, respectively. To acquire the EEG data after baseline removal, all of the baseline signals were averaged, and then the average value of the baseline signal was subtracted from the experimental signal. This step can be written as follows:

$$\overline{B} = \frac{\sum_{k=1}^{K} B_k}{K}, \tag{1}$$

$$X_i^T = X_i - \overline{B}, \tag{2}$$

where $\overline{B}$ represents the average of all baseline signal segments and the EEG data after paragraph i is removed from baseline is defined as X_i^T.

EEG signals are produced by sensors collecting different parts of the cerebral cortex, so the EEG signals generated by the same area of the cerebral cortex are related. In deep learning algorithms, if the network directly processes one-dimensional EEG signals, it often leads the network to find the neighborhood correlation of EEG signals and ignore the spatial correlation of signals, which will affect the final emotion recognition results. Therefore, inspired by reference [24], we map EEG signals from one-dimensional matrices into the two-dimensional matrix to maintain the spatial information between adjacent electrode channels in EEG signals and further excavate the spatial correlation of EEG signals.

In this paper, EEG electrodes are arranged into a two-dimensional mapping matrix (whose size is $M \times N$) by using the 10–20 international standard to preserve the spatial features between channels, where M is the number of electrodes used in the vertical direction and N is the number of electrodes used in the horizontal direction. Since the maximum number of channels tested in our paper is 62, the electrode channels of each data set can be arranged in a matrix of 9 × 9. Figure 2(a) is the international 10–20 standard electrode position topographic map. Figure 2(b) is the two-dimensional mapping matrix of electrode position. In order not to introduce other irrelevant noise, in this paper, we do not interpolate the position without an electrode but directly set the value of the position without an electrode to 0.
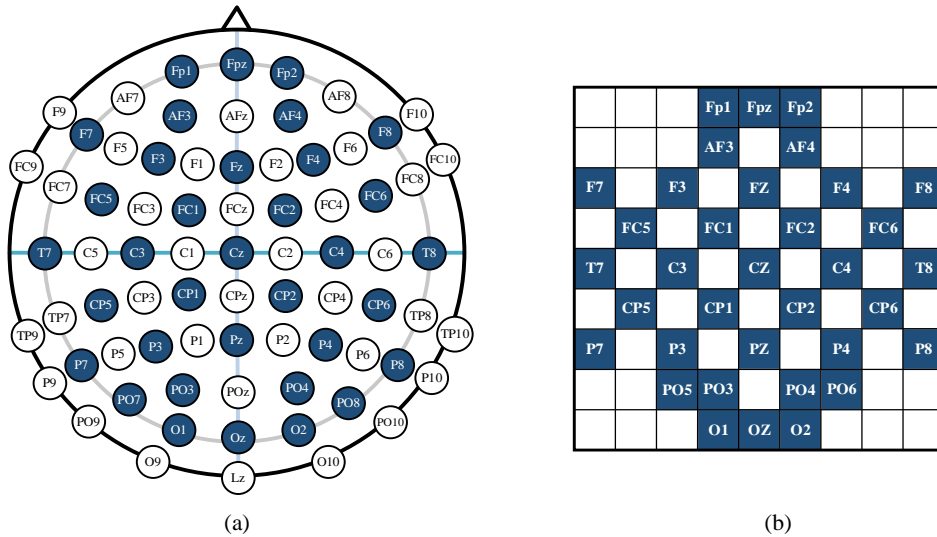
Figure 2. The electrode placement and the mapping matrix.

To preserve the spatial position information between each electrode and adjacent electrodes, each one-dimensional data vector $v_t^E = \left(s_t^1, s_t^2, \cdots, s_t^E \right)$ is converted to a two-dimensional matrix according to the mapping matrix shown in Figure 2. The two-dimensional matrix f_t at t time can be expressed as:

$$
f_t = \begin{bmatrix}
0 & 0 & 0 & s_t^1 & s_t^2 & s_t^3 & 0 & 0 & 0 \\
0 & 0 & 0 & s_t^4 & 0 & s_t^5 & 0 & 0 & 0 \\
s_t^6 & 0 & s_t^7 & 0 & s_t^8 & 0 & s_t^9 & 0 & s_t^{10} \\
0 & s_t^{11} & 0 & s_t^{12} & 0 & s_t^{13} & 0 & s_t^{14} & 0 \\
s_t^{15} & 0 & s_t^{16} & 0 & s_t^{17} & 0 & s_t^{18} & 0 & s_t^{19} \\
0 & s_t^{20} & 0 & s_t^{21} & 0 & s_t^{22} & 0 & s_t^{23} & 0 \\
s_t^{24} & 0 & s_t^{25} & 0 & s_t^{26} & 0 & s_t^{27} & 0 & s_t^{28} \\
0 & 0 & s_t^{29} & s_t^{30} & 0 & s_t^{31} & s_t^{32} & 0 & 0 \\
0 & 0 & 0 & s_t^{33} & s_t^{34} & s_t^{35} & 0 & 0 & 0
\end{bmatrix}. \tag{3}
$$

For each two-dimensional data vector, in order to reduce the degree of individual aggregation, in this paper, we use the Z-score normalization proposed in reference [25] to normalize all elements, that is:

$$
s_t^{e\prime} = \frac{s_t^e - \mu_{f_t}}{\sigma_{f_t}}, \tag{4}
$$

where $s_t^{e\prime}$ is the EEG signal value of a certain electrode e in time dimension t after normalization, and μ_{f_t} is the mean value of all channels' EEG signals at time t, σ_{f_t} is the standard deviation of these signal values, and the two-dimensional matrix at time t after normalization is expressed as $f_t^{\prime}$. The output EEG data after constructing the module with the three-dimensional spatio-temporal matrix are $X^{T\prime} = \left(f_1^{\prime}, f_2^{\prime}, \cdots, f_T^{\prime} \right) \in R^{M \times N \times T}$.

2.2. Bi-hemispheric discrepancy module

According to neuroscience research, in brain cognition, the right and left hemispheres of the brain have different stress signals to emotional stimuli, but simple convolution operations cannot obtain long-distance dependence messages of symmetry positions. In order to highlight saliency in brain cognition, we use the asymmetry difference between the two hemispheres of the brain caused by emotional stimuli and we conduct subtraction operation on the electrode pair at the symmetrical position of the brain and obtain more accurate emotion recognition accuracy. Based on the electrodes Fz, CZ, Pz and Oz on the central line, two electrodes with the same vertical distance were found as an electrode pair. For example, Fp1–Fp2, AF3–AF4, and C3–C4 were symmetrical electrode pairs.

Each electrode element $s_t^{e'}$ in the normalized two-dimensional matrix f_t' is operated as follows:

$$\tilde{s}_t^{(i,j)'} = s_t^{(i,j)'} - s_t^{(M+1-i,j)'},$$ (5)

where $s_t^{(i,j)'}$ denotes the electrode value at the position (i,j) at time t, $s_t^{(M+1-i,j)'}$ denotes the symmetric electrode of $s_t^{(i,j)'}$, where $i = 1, 2, \cdots m$, $j = 1, 2, \cdots n$.

After processing by the Bi-hemispheric discrepancy module, the two-dimensional matrix $\tilde{f}_t'$ can be defined as:

$$\tilde{f}_t'(i,j) = \alpha \tilde{s}_t^{(i,j)'}, \quad \alpha = \begin{cases} 1, & i \le \dfrac{M+1}{2} \\ -1, & i > \dfrac{M+1}{2} \end{cases}.$$ (6)

The Bi-hemispheric discrepancy module not only obtains the asymmetric difference characteristics of EEG signals in different hemispheres in the brain, but also retains the position structure relationship between EEG channels. Bi-hemispheric discrepancy module is used to extract the hemispheric asymmetry of the 3D spatio-temporal matrix $X^{T'} = \left(f_1', f_2', \cdots, f_T' \right) \in R^{M \times N \times T}$, which can be calculated by the following equation:

$$y^{B_i} = F_{B_i}\left(X^{T'} \right) \in R^{M \times N \times T}.$$ (7)

Among them, $y^{B_i} = \left(\tilde{f}_1', \tilde{f}_2', \cdots, \tilde{f}_T' \right) \in R^{M \times N \times T}$ is the asymmetry feature, and F_{B_i} is defined as the function of the Bi-hemispheric discrepancy module.

The extraction process of hemispheric asymmetric features is shown in Figure 3. Figure 3 shows how the three-dimensional spatio-temporal matrix of EEG signals is constructed, and the EEG signals of each electrode at each time point are taken as a pixel value in the matrix. Then the hemispheric asymmetry characteristics of EEG signals are extracted by using the Bi-hemispheric discrepancy module.
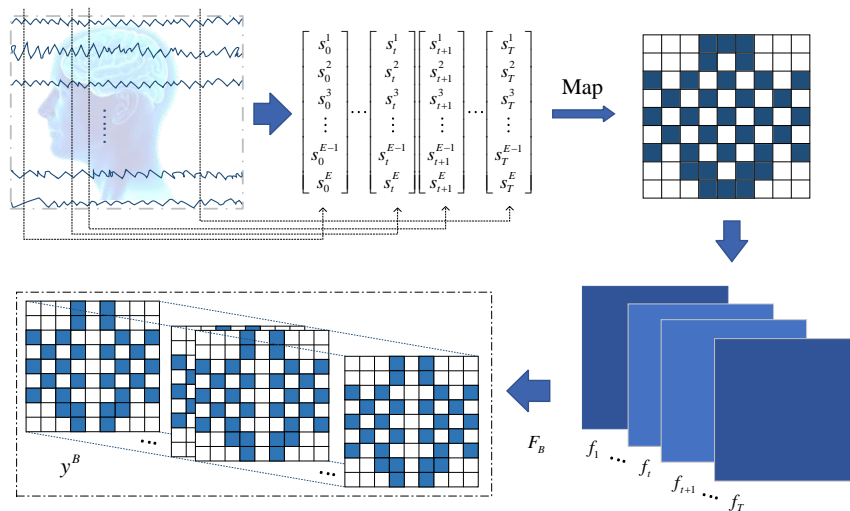
Figure 3. Feature extraction process of bi-hemispheric discrepancy module.

2.3. 3D spatio-temporal convolution module

In traditional two-dimensional convolution, the input layer, convolution layer, pooling layer, and output layer are normally present. The network's input and output layers are the input and output layers, respectively. The local features of the same place in the feature map acquired in the previous layer are extracted by each neuron in the convolution layer. The feature maps extracted from the convolution layer are down-sampled in the pooling layer, resulting in scale invariance in the convolution neural network, decreasing data calculation and keeping of important information. However, the two-dimensional convolution operation is only suitable for two-dimensional images and can only extract the image's spatial features. Three-dimensional convolution can combine temporal and spatial information from EEG signals, allowing it to learn more representative spatial and temporal aspects in signals than two-dimensional convolution.

In this paper, a three-dimensional spatio-temporal convolution module is constructed to extract the spatio-temporal information of EEG signals, and the local position information in the two-dimensional plane and the deep temporal dimension information of EEG signals are extracted by the three-dimensional convolution operation. Figure 4 shows the schematic diagram of 3D convolution.

Spatial Temporal Convolutional
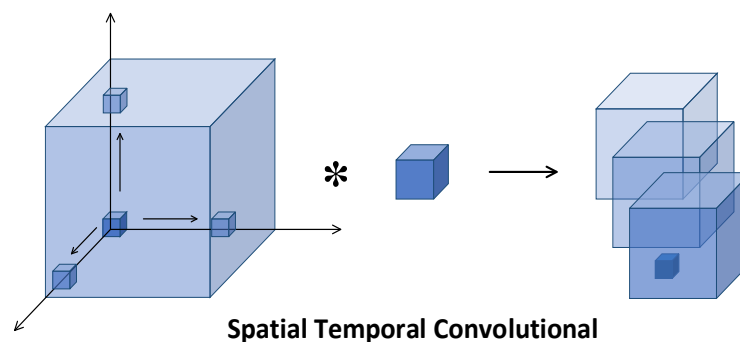
**Figure 4.** The schematic diagram of 3D convolution.

It can be seen from Figure 4 that the 3D convolution kernel can move in three directions (height, width and channel). Assuming that the element $s_{ij}^{x_0 y_0 z_0}(x_0 y_0 z_0)$ is an eigenvalue at (x_0, y_0, z_0) in the position of the j-th feature map in the i-th layer, the 3D convolution process can be expressed as:

$$s_{ij}^{x_0 y_0 z_0}(x_0 y_0 z_0) = \partial\left(b_{ij} + \sum_c \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijc}^{pqr} s_{(i-1)c}^{(x_0+p)(y_0+q)(z_0+r)}\right), \tag{8}$$

where the activation function is ∂, the size of the three-dimensional convolution kernel is R_i, and the value of the convolution kernel connecting the c feature map of the $i-1$ layer and the j feature map of the i layer at the (p, q, r) point is w_{ijc}^{pqr}. For the convolution layer, we employ the ReLu activation function, which is stated as:

$$ReLu(s) = \phi(s) = max(0, s). \tag{9}$$

The three-dimensional pooling layer can shrink the feature matrix and extract the signal's primary features, which is conducive to reducing the over-fitting risk of the network. After the pooling operation, the feature will have a certain loss. In this paper, the data after the three-dimensional spatio-temporal matrix construction module are $X^{T'} = \left(f_1', f_2', \cdots, f_T'\right) \in R^{M \times N \times T}$, where M and N are 9. The shape of the data is obviously small. If the pooling layer is still processed, too much information will be lost. Therefore, the 3D spatio-temporal feature extraction module in this paper does not need to add a pooling layer to compress the feature map. In our method, the structure of the 3D spatio-temporal feature extraction module is shown in Figure 5 (taking the SEED dataset as an example).
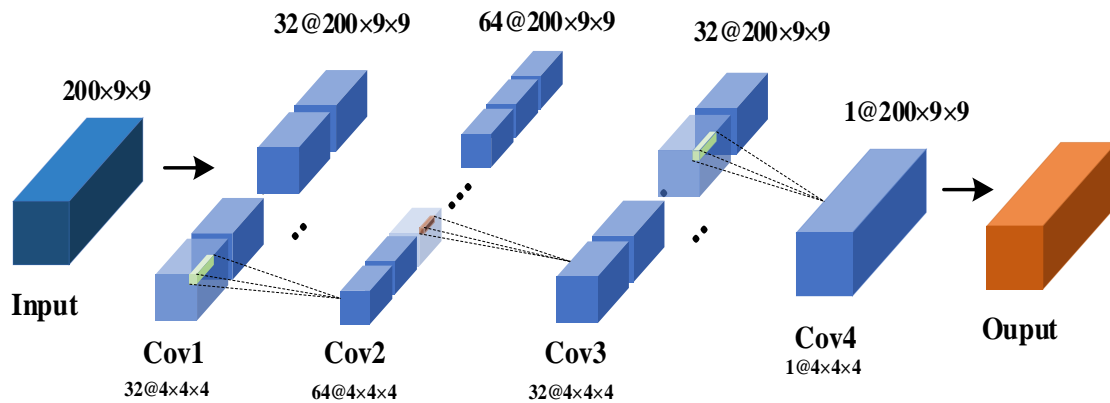
Figure 5. Module of spatio-temporal feature extraction.

Figure 5 shows that the spatio-temporal feature extraction module based on 3D CNN is composed of four continuous three-dimensional convolution layers, in which the convolution kernel size of each convolution layer is $4 \times 4 \times 4$. Continuous convolution can extract the local features between EEG signals more carefully and increase the nonlinear expression ability of the model. Compared with the common $3 \times 3 \times 3$ convolution kernel, the $4 \times 4 \times 4$ convolution kernel can better retain the spatial relationship between electrodes. The features captured by the spatio-temporal feature extraction module can be expressed as:

$$y^{3D} = F_{3D}\left(X^{T'}\right) \in R^{M \times N \times T}, \tag{10}$$

where y^{3D} denotes the spatio-temporal feature, and F_{3D} is the spatio-temporal feature extractor.

2.4. Space attention module

For different emotional states, the activation state of different brain regions is different. In the process of emotion recognition, the prefrontal and temporal lobe brain areas have differing weights. The effect of EEG emotion identification is affected by strengthening or suppressing a certain brain region. The attention mechanism can learn the weight map of the different EEG channels, and improve the properties of EEG signals that are conducive to emotion recognition, and suppress useless information that interferes with the recognition effect. In this paper, we use a spatial attention module to extract important information from asymmetric features and spatio-temporal features. The structure of the spatial attention module is shown in Figure 6. In Figure 6, we can find that the spatial attention module consists of an average pooling layer, maximum pooling layer, convolution layer and dense layer.
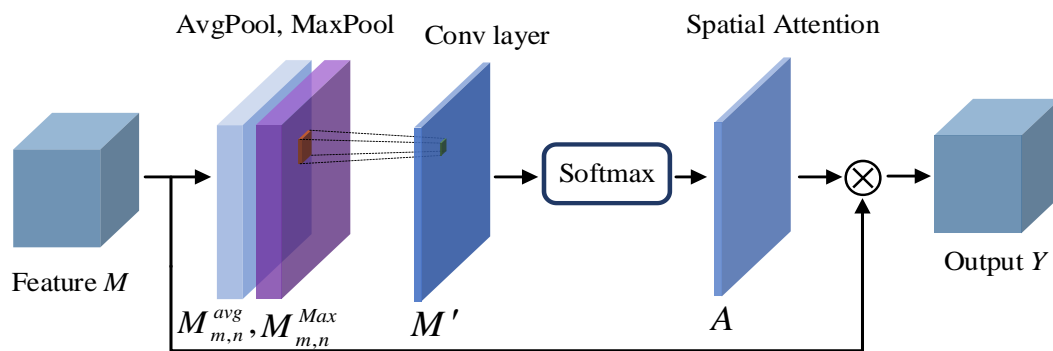
Figure 6. The structure of space attention module.

Firstly, the extracted asymmetric features and spatio-temporal features are spliced to obtain the feature M, namely:

$$v\,M = y^{B_i} \,\|\, y^{3D} \in R^{M \times N \times C}. \tag{11}$$

Then, to reduce the complexity of the calculation, we conduct global average pooling and maximum pooling of feature M, which are defined as:

$$M_{m,n}^{avg} = F_{GAP}\left(M_{m,n}\right) = \frac{1}{C} \sum_{c=1}^{C} M_{m,n}(c), \tag{12}$$

$$M_{m,n}^{Max} = F_{GMP}\left(M_{m,n}\right) = \underset{c=1}{\overset{C}{Max}}\, M_{m,n}(c), \tag{13}$$

where $M^{avg} \in R^{M \times N}$ and $M^{Max} \in R^{M \times N}$ are the average or maximum values of all channels C of the feature M, respectively; F_{GAP} and F_{GMP} represent the average pooling function and the maximum pooling function; $M_{m,n}$ is a vector, representing the eigenvalues of all channels of M at the position of (m,n).

In this paper, the results of average pooling and maximum pooling of features M are spliced into the feature matrix $M_{a,m}$, and then the feature map M' is obtained by convolution with the convolution of channel number 1. Then the spatial attention weight matrix is calculated by the dense layer with Softmax activation function, that is:

$$A = softmax\left(WM' + b\right), \tag{14}$$

where W and b are learnable network parameters. $A \in R^{M \times N \times 1}$ is the spatial attention weight matrix.

After the attention matrix is generated, the entire attention process can be defined as:

$$Y = A \otimes X, \tag{15}$$

where Y is a feature enhanced by attention and $\otimes$ represents a multiplication by element.

2.5. Classifier

In this paper, after feature enhancement by using a spatial attention module, the obtained features are input into the classifier module to achieve high-precision emotion classification. A flattened layer, two completely connected layers with the ReLu activation function, and a fully connected layer with the Softmax activation function make up the classifier module. Dropout is added to the FC1 and FC2 layers of the full connection layer to prevent over-fitting in network training. Dropout is an effective tool to prevent overfitting, which can not only reduce the computation of the network, but also enhance the generalization ability of the network.

Firstly, the feature Y is input into the Flatten layer for flattening, so that the three-dimensional feature Y is converted to one-dimensional y^1, which is convenient for the subsequent calculation of the full connection layer. The input of FC1 is y, and the output is:

$$y_1 = tanh(W_1 y^1 + b_1), \tag{16}$$

where W_1 is the weight of FC1 in the full connection layer, b_1 is the bias, and $tanh()$ is the activation function.

The input of FC2 is y_1, and the output is:

$$y_2 = tanh(W_2 y_1 + b_2), \tag{17}$$

where W_2 is the weight of FC2 in the full connection layer, b_2 is the bias, and $tanh()$ is the activation function.

The last layer of the classifier module adopts the full connection layer with Softmax activation for emotion recognition, that is:

$$P(c|x) = \frac{exp(y_c)}{\sum_1^C exp(y_i)}, \quad c = 1, 2, \cdots C, \quad (18)$$

$$\hat{e} = \arg\max_c P(c|x). \quad (19)$$

The classification cross-entropy is utilized as the loss function in this paper, and it is defined as follows:

$$L = -\sum_{c=1}^C y_{c,x} \log\left(P(c|x)\right), \quad (20)$$

where C denotes the total number of classes, $P(c|x)$ denotes the probability that x belongs to class c, $\hat{e}$ is the prediction value and $y_{c,x}$ denotes the binary indicator (0 or 1) in the class label.

3. Dataset and experimental setup

3.1. Dataset

To evaluate the effectiveness of the proposed algorithm, we employ the public database SJTU emotion EEG dataset (SEED) [26] and the multi-channel database of human emotional state, the database for emotion analysis using physiological signals (DEAP) [27, 28], for emotion recognition.

SEED dataset contains EEG signals triggered by the emotions from watching videos. In this dataset, there are 15 subjects, which contain 7 males and 8 females. The EEG signals are triggered by watching 15 Chinese film clips. Before each movie clip, a 5-second start prompt is given. Each film clip is around 4 minutes long. After watching each movie clip, each participant had 45 seconds to complete an evaluation questionnaire. In order to avoid fatigue in long-term experiments, a 15-second rest time is reserved for each film clip. The order of emotions presented in the selected segment is [1, 0, −1, 0, 1, −1, 0, 1, 0, 0, 1, 0, −1, 0, 1, −1], where 1 represents positive emotions, 0 represents neutral emotions, and -1 represents negative emotions. In three different periods, each participant carried out repeated tests, and there were 45 experiments in total. In data collection, the electrode cap was placed in accordance with the international 10–20 standard system [29]. Three types of emotions (neutral, positive, and negative) were recorded through 62-lead EEG acquisition equipment. The sampling frequency of the experiment was 1000 Hz, which was then down-sampled to 200 Hz during data preprocessing.

Koelstra et al. [30] created the DEAP dataset, which is a multimodal dataset. The DEAP dataset contains emotionally stimulated responses from 32 healthy subjects. In the DEAP dataset, EEG signals with 32 channels and a set of physiological signals of peripheral physiological signals with 8 channels are recorded when watching 40 1-minute emotion-related music videos. The experiment included a 2-second screen display of the test number, a 5-second baseline recording, a 1-minute music video presentation, and the remaining time for participants to self-evaluate (1 to 9) arousal, valence, liking, and dominance. After 20 experiments, participants took a brief break. The final EEG data for each

participant consists of 60 seconds of experimental data and 3 seconds of baseline signal. Electrode placement during EEG acquisition follows the international 10–20 system. The signal is down-sampled to 128 Hz after being recorded at 512 Hz sampling frequency.

In this work, as the raw EEG signal has a large amount of noisy information unrelated to emotion, it will interfere with emotion recognition and affect the accuracy of emotion recognition. In this paper, the EEGLAB toolkit was used to pre-process the raw signal and remove the physiological noise unrelated to the EEG signal, including eye movement, EMG, and industrial frequency interference. The EEGLAB toolkit also performed spherical interpolation on the bad electrodes with high noise impact.

In the experiment of this paper, the experiment with the SEED dataset is classified into three categories (positive, neutral, and negative). Each participant had 45 experiments (15×3). In this paper, we believed that the video was not enough to stimulate emotion at the beginning. Therefore, the first three seconds of the video were taken as the baseline signal, and the EEG signal at 31–90 s (total 60 s) was taken as the experimental signal. The size of the final processed EEG signal was $2700 \times 200 \times 9 \times 9$, where 2700 was the length of the time dimension of the data (45 experiments, 60 s of each experiment), 200 was the number of sampling points per second and 9×9 was the size of the two-dimensional matrix constructed according to the electrode position. For the DEAP dataset, only EEG signal channels (32 channels) are used to select the data from the valence and arousal dimensions. Two classification tasks were performed on each dimension, namely, high/low valence and high/low arousal (low: ≤5, high: >5). The first 3 s were selected as the baseline signal, and the last 60 s were selected as the experimental signal for baseline correction. The final size of the processed EEG signal was $2400 \times 128 \times 9 \times 9$, of which 2400 was the length of the data time dimension (40 trials, 60 s data per trial), 128 was the number of sampling points per second, and 9×9 was the size of the constructed two-dimensional matrix according to the electrode position.

3.2. Parameter setting

The method proposed in this paper was experimented on Python 3.7 based on the Tensorflow 2.2 platform using Keras 2.3.1. On a cluster server with an NVIDIA GeForce TITAN X GPU, training and testing are carried out. The network is optimized by using the adaptive moment estimation algorithm (Adam) optimizer, with a learning rate of 10–4. The experiment employs ten-fold cross-validation. The accuracy of a subject is determined by the average value of each fold, and the accuracy of the entire emotion recognition is determined by the average value of all subjects' accuracy. The experiment's test data will be randomly disturbed. The proportion of the training set and test set is divided according to the proportion of training set and test set in the experiment to investigate the influence of training set scale on network performance. In each iteration of the experiment, 10 or 20 times, batch_size was 64.

In order to verify the accuracy and robustness of our algorithm, three sets of experiments are conducted to verify. First, compared with the existing emotion recognition algorithms, the classifier's performance of the BiTCAN network is verified by a tenfold cross-validation method for a single subject. Second, we design ablation experiments to demonstrate the important role of each module in the proposed network and its impact on the accuracy of EEG-based emotion recognition tasks. Finally, we also designed experiments on the contribution of different brain regions to emotion recognition, and

selected brain regions that are more useful for EEG emotion recognition based on neurological mechanisms, which provides a theoretical basis for the development of portable EEG devices.

4. Experimental results and analysis

4.1. Emotional recognition results

To verify the performance of the proposed algorithm, emotion recognition experiments were conducted on the SEED dataset and the DEAP dataset, respectively. In the experiment, each participant used a ten-fold cross validation method. We evaluated the performance of the model by using multiple indicators to present the results more clearly. The experimental results of the SEED dataset are shown in Figure 7. Similarly, we conducted the same experiment on the DEAP dataset and calculated the statistical standard deviation for multiple indicators on both datasets. The smaller the standard deviation, the more stable the performance of the proposed network is. The final experimental results of the two datasets are shown in Table 1.
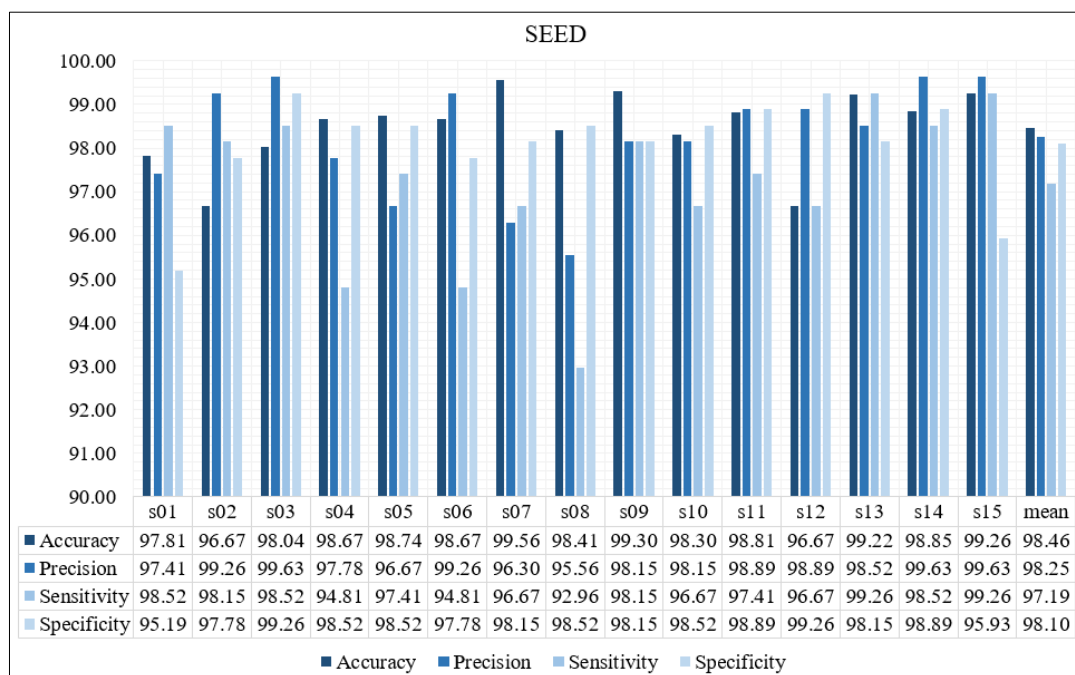
Figure 7. Recognition results of different subjects in the SEED dataset.

In Figure 7, the classification accuracy of the proposed model has reached over 97% among all subjects, with 11 subjects having a Precision of over 98%, 6 subjects having a Sensitivity of over 98%, and 12 subjects having a Specification of over 98%, demonstrating the superior performance of the proposed model.

Table 1. Evaluation results of different indicators for two datasets.

| Dataset / Index	SEED (%)	DEAP (%)	
		Valence	Arousal
Accuracy	98.46/0.84	97.65/1.20	97.73/1.19
Precision	98.25/1.24	99.35/1.05	98.57/0.98
Sensitivity	97.19/1.74	98.14/1.26	99.16/0.98
Specificity	98.10/1.09	98.33/1.48	97.91/1.04

From Table 1, it can be seen that our model has achieved very good results on both the SEED dataset and the DEAP dataset, with standard deviations below 2 on multiple evaluation indicators, proving that the proposed model is robust.

On the SEED and DEAP datasets, a comparison is done with 10 other usual approaches in order to better validate the advantages of the proposed algorithm (referred to as BiTCAN). These 10 algorithms include: (1) A dynamic graph convolutional neural networks-based emotion recognition algorithm was proposed in [31], which can be named DGCNN; (2) A channel fusion dense convolution network-based emotion recognition algorithm was proposed in [32], which can be named CDCN; (3) A novel hybrid model-based emotion recognition algorithm was proposed in [33], which can be named DGGN; (4) An improved temporal convolutional network-based emotion recognition algorithm was proposed in [34], which can be named SITCN; (5) An attention-based LSTM with domain discriminator model-based emotion recognition algorithm was proposed in [35], which can be named ATDD-LSTM; (6) A novel model-based emotion recognition algorithm was proposed in [36], which can be named DE-CNN-BiLSTM; (7) A fused CNN-LSTM deep learning model-based emotion recognition algorithm was proposed in [37], which can be named CNN-LSTM; (8) A temporal relative (TR) encoding mechanism and self-attention mechanism in the transformer-based emotion recognition algorithm was proposed in [38], which can be named TR&CA; (9) A simple EEG-based recognition network was proposed in [39], which can be named SEER-Net; (10) A dual-branch dynamic graph convolution-based adaptive transformer feature fusion network with adapter-finetuned transfer learning for EEG emotion recognition was proposed in [40], which can be named DBGC-ATFFNet. The comparison results are shown in Table 2.

It can be seen from Table 2 that BiTCAN obtains the best classification accuracy. In the SEED dataset, the accuracy of the three-classification task of BiTCAN reaches 98.46%, while the standard deviation is only 0.84%. The accuracy of BiTCAN is 1.15 percentage points more than the second-best algorithm, while the standard deviation is 0.63 percentage points lower than that of the next-best algorithm. In the DEAP dataset, the accuracy of BiTCAN in two dimensions is 97.65 and 97.73%, respectively, which is 0.26 and 0.32% higher than that of the next best algorithm, while BiTCAN has the lowest standard deviation among the existing algorithms. The reason why the BiTCAN algorithm has excellent emotion recognition performance is attributed to three important modules in BiTCAN network construction. The bi-hemispheric discrepancy module extracts the subtly different features of emotional response between the two hemispheres of the human brain. Three-dimensional spatio-temporal matrix and three-dimensional spatio-temporal convolution module extract more advanced spatio-temporal features related to emotion from continuous EEG signals. The attention module can enhance the important information related to emotion in EEG and suppress useless interference information.

Table2. Effect of each emotion recognition algorithm on the seed dataset and DEAP dataset.

Model	SEED (%)	DEAP (%)	
		Valence	Arousal
DGCNN	90.40/8.49	92.55/3.53	93.50/3.93
CDCN	90.63/4.34	92.24/–	92.92/–
DGGN	97.28/2.70	96.98/2.23	97.19/2.56
SITCN	88.84/–	95.02/–	95.29/–
ATDD-LSTM	91.08/6.43	90.91/12.95	90.87/11.32
DE-CNN-BiLSTM	94.82/–	94.02/–	94.86/–
CNN-LSTM	93.74/–	97.39/–	97.41/–
TR&CA	–	95.18/2.46	95.58/2.28
SEER-Net	90.73/3.38	–	–
DBGC-ATFFNet	97.31/1.47	–	–
BiTCAN	**98.46/0.84**	**97.65/1.20**	**97.73/1.19**

(a)   (b)   (c)
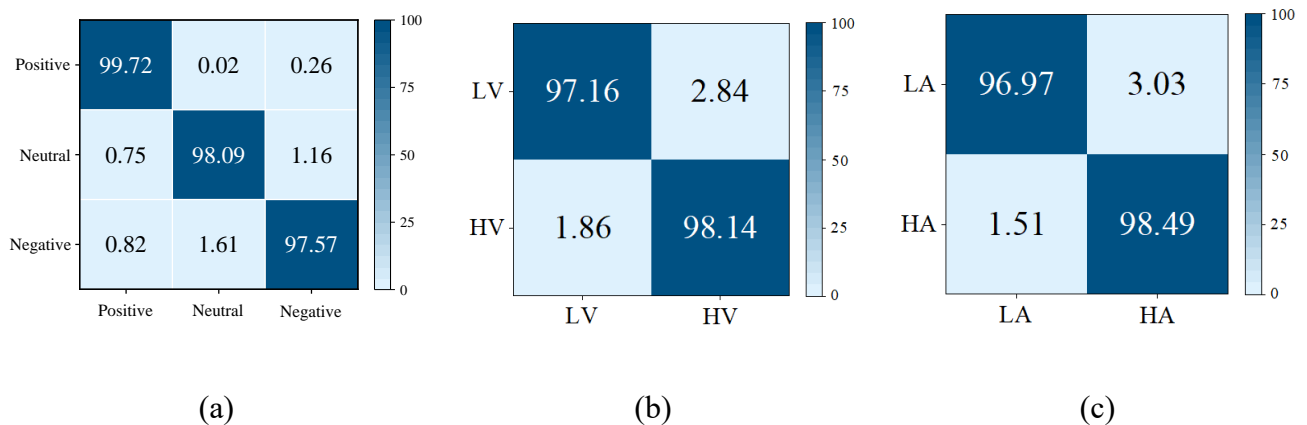
Figure 8. Confusion matrix of two databases. (a) The average confusion matrix of the SEED database; (b) The average confusion matrix of the DEAP database in valence; (c) The average confusion matrix of the DEAP database in arousal.

In order to observe the emotion recognition effect of the proposed algorithm more intuitively, Figure 8 shows the confusion matrix obtained by BiTCAN in the SEED dataset and DEAP dataset. Figure 8(a) shows that when tested on the SEED dataset, positive and negative emotions are easily distinguished, while neutral and negative emotions are more difficult to distinguish, which may be that positive emotional stimuli cause more resonance among participants [41]. It can be seen from Figure 8(b) and (c) that it is relatively difficult to distinguish between low valence and low arousal when testing on the DEAP dataset, which indicates that negative emotions are more easily confused.

4.2. The ablation experiments

In order to demonstrate the validity of the modules in the proposed algorithm, an ablation study was carried out. The network structure after the ablation of the three modules is given in Figure 9. Figure 9(a) represents the network structure after removing the bi-hemispheric discrepancy module,

named w/o Bi-h block. Figure 9(b) represents the network structure after removing the 3D spatio-temporal convolution module, named w/o 3D block. Figure 9(c) represents the network structure after removing the spatial attention module, named w/o AM block. The ablation experiments were performed by using a ten-fold cross-validation method for each subject, and the average accuracy of all subjects in the SEED or DEAP dataset was used as the final accuracy of the ablation model. The final emotion recognition results are shown in Figure 10. In the experimental results, we can find that the recognition accuracy of the model after removing a module is slightly worse than the BiTCAN results, but its accuracy is still higher than the baseline methods, thus proving the importance of the three modules proposed in this paper.
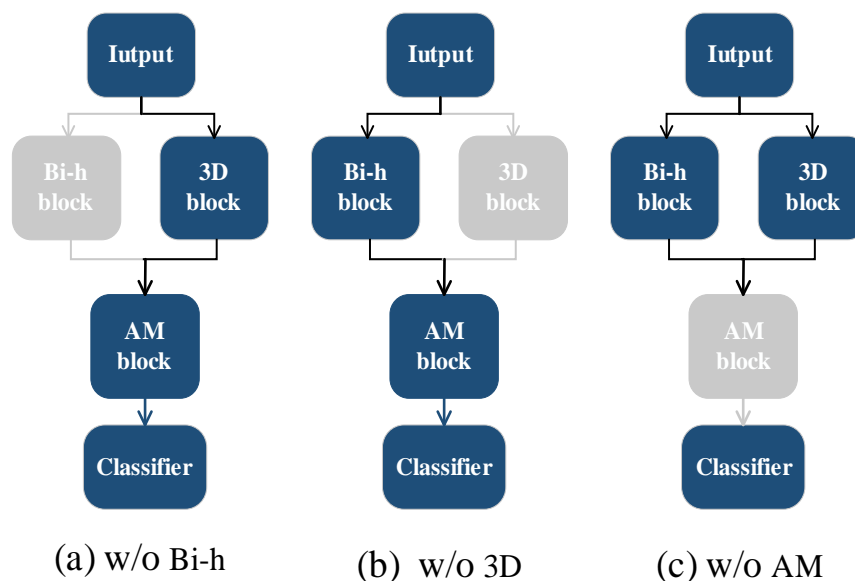
**Figure 9.** The diagram of the ablation experiment.

Figure 10 shows that the performance of the BiTCAN network is optimal whether it is a SEED dataset or DEAP dataset, and all the three modules in the BiTCAN network contribute to the final classification effect. Figure 9 shows that the accuracy and stability of the model in emotion recognition are greatly affected after removing the three-dimensional spatio-temporal convolution module, while the bi-hemispheric discrepancy module has a significant impact on the accuracy of classification and the spatial attention module has a great influence on the stability of the network. Therefore, we can conclude that the discrepancy information in the bi-hemispheric module is indeed helpful to the task of EEG emotion recognition. The 3D convolution module can not only extract spatial and temporal features, but also fuse the hemispheric discrepancy and spatial and temporal features. The focus of the attention module on different brain regions makes the high-correlation EEG channel reduce the feature redundancy, which is more conducive to emotion recognition.
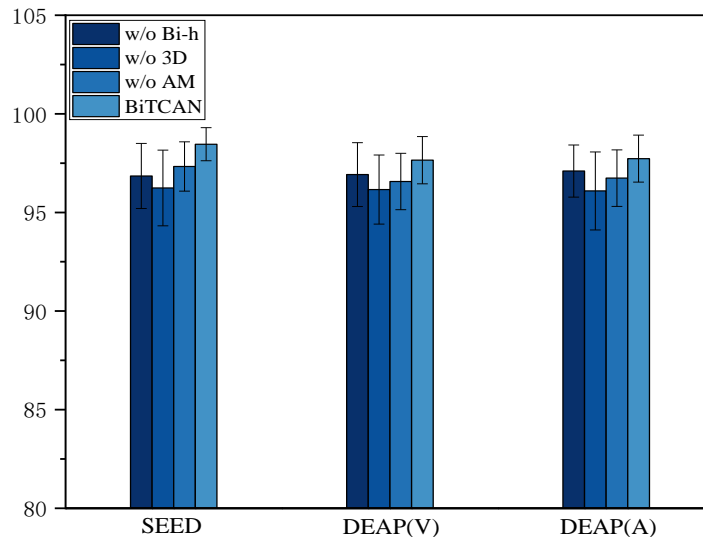
Figure 10. The results of the ablation experiment.

4.3. The contribution of brain regions to emotion recognition

Neurological studies show that the brain responds to different emotions with different activation values in different brain regions [42,43]. The attention mechanism added to the proposed model can effectively focus on the differences in activation degrees of different brain regions caused by emotions, which can increase emotion recognition accuracy. According to the neural mechanism, the brain is divided into four different regions [44,45], as indicated in Figure 11, the temporal lobe, the frontal lobe, the occipital lobe and the parietal lobe. Figure 11(a) is the brain partition map of the SEED dataset, and Figure 11(b) is the brain partition map of the DEAP dataset. The specific electrode settings are shown in Table 3, the frontal lobe is represented by F, the temporal lobe is represented by T, the parietal lobe is represented by P and the occipital lobe is represented by O.
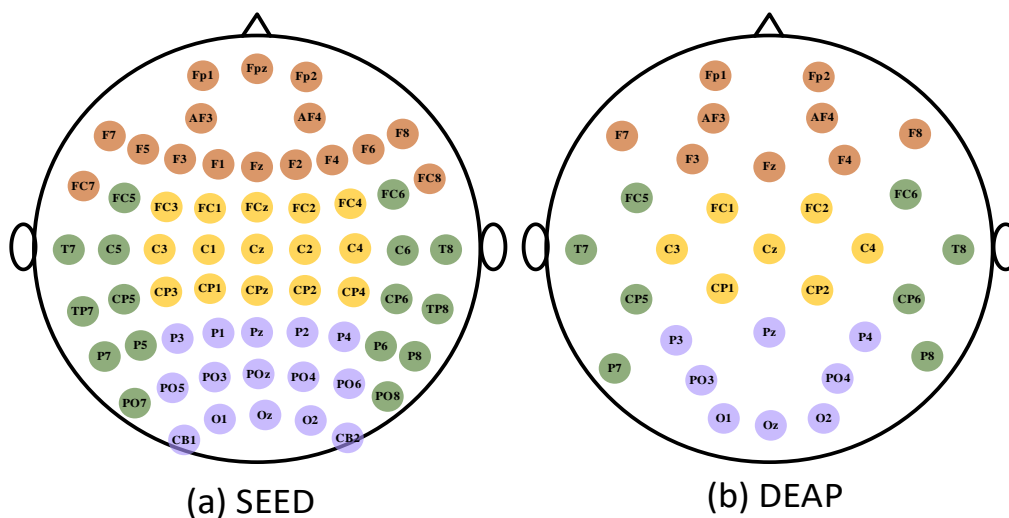
Figure 11. The brain region division on two datasets.

In emotion recognition, we explore the contribution of different brain regions, the electrodes in each brain region were selected to construct a three-dimensional spatio-temporal matrix in the experiment, while the electrodes in other useless positions were set to zero. Each subject is tested by tenfold cross-validation. The final accuracy of a specific brain area is the average accuracy of all subjects. The experimental results are shown in Table 4. It can be seen from Table 4 that the frontal and temporal lobes play the most important roles in emotion recognition, and emotion classification accuracy is the highest. In the emotion recognition task, the standard deviations of frontal and temporal lobes are lower and more stable than those of other brain regions, which is consistent with the research in [42,43]. The response of the occipital lobe and parietal lobe to emotion is basically the same, but the standard deviation of the occipital lobe is greater and the volatility is greater. Therefore, in the design of portable EEG emotion recognition equipment, the electrode selection should focus on the frontal lobe and temporal lobe.

Table 3. The electrode blocks in different brain regions on two datasets.

Brain region	SEED	DEAP
Frontal	Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FC7, FC8	Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8
Temporal	FC5, T7, C5, TP7, CP5, P7, P5, PO7, FC6, T8, C6, TP8, CP6, P8, P6, PO8	FC5, T7, CP5, P7, FC6, T8, CP6, P8
Parietal	FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4	FC1, FC2, C3, Cz, C4, CP1, CP2
Occipital	P3, P1, Pz, P2, P4, PO5, PO3, POz, PO4, PO6, CB1, O1, Oz, O2, CB2	P3, Pz, P4, PO3, PO4, O1, Oz, O2

The above experiments not only explored the contribution of each brain area to emotion recognition, but also proved that the proposed network in this paper reduced the number of electrodes to a quarter (for the SEED dataset, there are 16 electrodes in frontal and temporal lobes, 15 electrodes in parietal and occipital lobes; for the DEAP dataset, when there are nine electrodes in the frontal lobe, eight electrodes in the occipital lobe and the temporal lobe, and seven electrodes in the temporal lobe), and the accuracy of emotion classification is still almost the same as that of the whole brain electrode. This is mainly because the space-time matrix constructed in this paper can better preserve the space-time features. The nature of EEG signals determines that the data contain a large amount of time information but lack spatial information. The spatio-temporal matrix constructed in this paper increases spatial information. At the same time, the bi-hemispheric discrepancy module extracts the differences in emotional reactions between the left and right hemispheres of the brain. The three-dimensional convolution module and the attention module jointly extract the spatio-temporal features of more powerful and higher resolution depth features, so as to obtain better emotional recognition results.

**Table 4.** Effect of emotion recognition based on brain regions.

| | SEED | DEAP | |
		Valence	Arousal
F	96.82/2.27	96.42/1.93	96.74/1.65
T	96.77/2.04	96.28/2.03	97.28/1.65
P	93.19/3.32	95.96/1.97	95.91/1.70
O	94.94/4.20	95.60/2.78	96.04/2.18

F represents the electrode used only in the frontal lobe, T represents the electrode used only in the temporal lobe, P represents the electrode used only in the parietal lobe and O represents the electrode used only in the occipital lobe.

In previous studies, researchers in neuroscience have also explored the effects of asymmetry between different hemispheres in the brain on emotion [46–48]. It is pointed out in some literature that positive emotions can better stimulate the electrode activity of the left frontal lobe of the brain, while negative emotions can better activate the electrode activity of the right frontal lobe of the brain [49–51]. So, in this paper, we use the bi-hemispheric discrepancy module to extract the asymmetric information of the left and right hemispheres to obtain results more conducive to emotional classification and recognition features. In addition, in the proposed method, we also add the attention module to enhance the key information in EEG signals from the global perspective to strengthen the distinguishing feature.

In order to observe the electrode activity under different emotions more intuitively, we directly input the EEG features extracted by the bi-hemispheric discrepancy module into the attention module, and map the attention activation value to the corresponding electrode position, which can be seen in Figures 12 and 13. In Figure 12, we can find the electrode activity mapping of the SEED three-class dataset under positive, neutral and negative emotions. Figure 13 is the electrode activity mapping of the DEAP dataset under four emotions (high valence and high arousal (HVHA), high valence and low arousal (HVLA), low valence and high arousal (LVHA) and low valence and low arousal (LVLA)) in two emotional dimensions. The deeper the color is, the greater the contribution made under specific emotions.

(a) Positive   (b) Neutral   (c) Negative
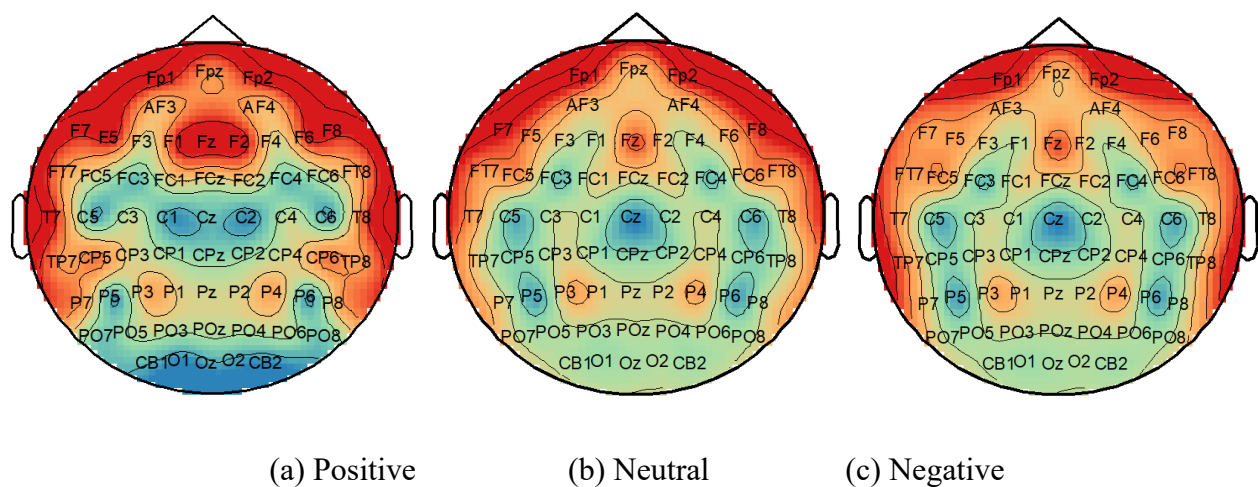
Figure 12. The electrode activity maps under three emotions in the SEED dataset.

(a) HVHA   (b) HVLA   (c) LVHA   (d) LVLA
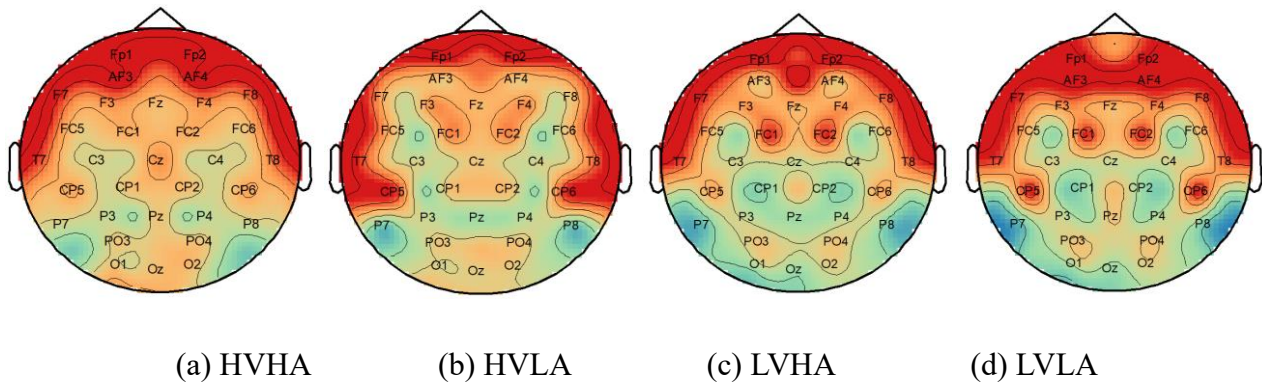
Figure 13. The electrode activity maps of four emotions in the DEAP dataset with two dimensions.

Figure 12 shows that the emotional induction is focused on the anterior frontal lobe, frontal lobe and temporal lobe, which is in line with the findings of the experiments in Table 4. Figures 12(a)–(c) show that the electrode activity map under positive emotions is more obvious than that under neutral emotions and negative emotions, the electrode activity of the frontal and occipital lobes are not obvious, and the difference in the parietal lobe is not significant.

Moreover, the electrode activity map under positive emotions is more different from that under the other two emotions, which is also the reason why positive emotions are easier to distinguish than the other two emotions in the emotion recognition experiment. The difference between neutral emotion and negative emotion is mainly that the electrode activity of the temporal lobe is slightly different, so the difference between these two emotions is small and difficult to distinguish, which may be related to the brain action mechanism of human beings under negative emotion and neutral emotion. It can be seen from Figure 13 that the four emotional categories in the two dimensions of the DEAP dataset are also basically concentrated in the anterior frontal lobe, frontal lobe, and temporal lobe. In the emotional state of high valence, the occipital lobe also made some contributions. Compared with the valence dimension, the emotions in high and low arousal dimensions are easier to distinguish in the electrode activity diagram (especially the electrodes of CP5 and CP6 temporal lobes), which also corresponds to the experimental results in Table 4. In both datasets, the asymmetry of frontal and temporal lobes plays an important role in emotion recognition tasks. These findings are consistent with neuropsychological findings that the frontal and temporal lobes are the main brain regions associated with emotion.

5. Conclusions

In this paper, we propose an EEG emotion recognition model based on BiTCAN, which improves the classification accuracy of EEG emotion recognition tasks. BiTCAN consists of a three-dimensional space-time matrix construction module, a Bi-hemispheric discrepancy module, a three-dimensional space-time convolution module, an attention module, and an emotion classification module. A 3D spatio-temporal matrix combined with a 3D spatio-temporal convolution module can effectively extract the spatio-temporal features of EEG signals, and a 3D spatio-temporal matrix combined with a Bi-hemispheric discrepancy module can effectively extract the asymmetric features of EEG signals.

Finally, the spatio-temporal features and asymmetric features are fused and sent to the attention module. The attention module can reduce the feature redundancy and extract the global internal spatial relationship between the multi-channel EEG signals, so as to obtain better emotional recognition results. The accuracy of our emotion recognition is above 97% in both of the two public datasets DEAP and SEED, which is better than the existing algorithms. In addition, we investigate the role of different brain regions in emotion recognition, concluding that the asymmetry of the frontal and temporal lobes is important in the emotion recognition task, which is consistent with neuropsychological findings. We also give some experiments to show how to select channels in EEG emotion recognition to reduce the computational complexity of emotion recognition models with a small number of EEG channels and provide a theoretical basis for the development of convenient EEG devices. Although the proposed model achieved satisfactory results in a single-subject task, it performed poorly in cross-subject experiments for a simple feature. In the future, we'll concentrate on the multimodality emotion recognition task and the cross-subject classification task based on EEG signals, which can provide more ideas for the development of BCI.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

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.

References

1. J. D. Dołżycka, J. Nikadon, P. P. Weis, C. Herbert, M. Formanowicz, Linguistic and emotional responses evoked by pseudoword presentation: An EEG and behavioral study, *Brain Cognition*, **168** (2023), 105973. https://doi.org/10.1016/j.bandc.2023.105973

2. M. Sajjad, F. U. M. Ullah, M. Ullah, G. Christodoulou, F. A. Cheikh, M. Hijji, et al., A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines, *Alexandria Eng. J.*, **68** (2023), 817–840. https://doi.org/10.1016/j.aej.2023.01.017

3. L. Trinh Van, T. Dao Thi Le, T. Le Xuan, E. J. S. Castelli, Emotional speech recognition using deep neural networks, *Sensors*, **22** (2022), 1414. https://doi.org/10.3390/s22041414

4. N. Ma, Z. Wu, Y. M. Cheung, Y. Guo, Y. Gao, J. Li, et al., A survey of human action recognition and posture prediction, *Tsinghua Sci. Technol.*, **27** (2022), 973–1001. https://doi.org/10.26599/TST.2021.9010068

5. V. Chaturvedi, A. B. Kaur, V. Varshney, A. Garg, G. S. Chhabra, M. Kumar, Music mood and human emotion recognition based on physiological signals: a systematic review, *Multimedia Syst.*, **28** (2022), 21–44. https://doi.org/10.1007/s00530-021-00786-6

6. X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, et al., Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine, *IEEE Trans. Cybern.*, **51** (2020), 4386–4399. https://doi.org/10.1109/tcyb.2020.2987575

7. H. Zhu, C. Fu, F. Shu, H. Yu, C. Chen, W. Chen, The effect of coupled electroencephalography signals in electrooculography signals on sleep staging based on deep learning methods, *Bioengineering*, **10** (2023), 573. https://doi.org/10.3390/bioengineering10050573

8. M. Xu, J. Cheng, C. Li, Y. Liu, X. Chen, Medicine, Spatio-temporal deep forest for emotion recognition based on facial electromyography signals, *Comput. Biol. Med.*, **156** (2023), 106689. https://doi.org/10.1016/j.compbiomed.2023.106689

9. J. A. Lee, K. C. Kwak, Personal identification using an ensemble approach of 1D-LSTM and 2D-CNN with electrocardiogram signals, *Appl. Sci.*, **12** (2022), 2692. https://doi.org/10.3390/app12052692

10. X. Chen, W. Liu. Research on Positive Emotion Recognition Based on EEG Signals, in *2023 5th International Conference on Communications, Information System and Computer Engineering (CISCE)*, (2023), 70–79. https://doi.org/10.1109/CISCE58541.2023.10142342

11. M. Wu, W. Teng, C. Fan, S. Pei, P. Li, Z. Lv, An investigation of olfactory-enhanced video on eeg-based emotion recognition, *IEEE Trans. Neural Syst. Reh. Eng.*, **31** (2023), 1602–1613. https://doi.org/10.1109/TNSRE.2023.3253866

12. Z. Tian, D. Huang, S. Zhou, Z. Zhao, D. Jiang, Personality first in emotion: a deep neural network based on electroencephalogram channel attention for cross-subject emotion recognition, *Roy. Soc. Open Sci.*, **8** (2021), 201976. https://doi.org/10.1098/rsos.201976

13. D. Huang, S. Zhou, D. Jiang, Generator-based domain adaptation method with knowledge free for cross-subject eeg emotion recognition, *Cogn. Comput.*, **14** (2022), 1316–1327. https://doi.org/10.1007/s12559-022-10016-4

14. Y. An, S. Hu, X. Duan, L. Zhao, C. Xie, Y. Zhao, Electroencephalogram emotion recognition based on 3D feature fusion and convolutional autoencoder, *Front. Comput. Neurosc.*, **15** (2021), 743426. https://doi.org/10.3389/fncom.2021.743426

15. S. Liu, X. Wang, L. Zhao, J. Zhao, Q. Xin, S. H. Wang, et al., Subject-independent emotion recognition of EEG signals based on dynamic empirical convolutional neural network, *IEEE/ACM Trans. Comput. Bi.*, **18** (2020), 1710–1721. https://doi.org/10.1109/tcbb.2020.3018137

16. J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, et al., EEG-based emotion classification using a deep neural network and sparse autoencoder, *Front. Syst. Neurosc.*, **14** (2020), 43. https://doi.org/10.3389/fnsys.2020.00043

17. J. Cheng, M. Chen, C. Li, Y. Liu, R. Song, A. Liu, et al., Emotion recognition from multi-channel EEG via deep forest, *IEEE J. Biomed. Health Inf.*, **25** (2020), 453–464. https://doi.org/10.1109/jbhi.2020.2995767

18. T. Song, S. Liu, W. Zheng, Y. Zong, Z. Cui, Y. Li, et al., Variational instance-adaptive graph for EEG emotion recognition, *IEEE Trans. Affect. Comput.*, **14** (2021), 343–356. https://doi.org/10.1109/taffc.2021.3064940

19. H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, X. Chen, EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network, *Knowl.-Based Syst.*, **205** (2020), 106243. https://doi.org/10.1016/j.knosys.2020.106243

20. J. Yang, X. Huang, H. Wu, X. Yang, EEG-based emotion classification based on bidirectional long short-term memory network, *Procedia Comput. Sci.*, **174** (2020), 491–504. https://doi.org/10.1016/j.procs.2020.06.117

21. J. Jungilligens, S. Paredes-Echeverri, S. Popkirov, L. F. Barrett, D. L. Perez, A new science of emotion: implications for functional neurological disorder, *Brain*, **145** (2022), 2648–2663. https://doi.org/10.1093/brain/awac204

22. Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, et al., A novel bi-hemispheric discrepancy model for EEG emotion recognition, *IEEE Trans. Cogn. Dev. Syst.*, **13** (2020), 354–367. https://doi.org/10.1109/TCDS.2020.2999337

23. D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, D. Jiang, Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition, *Neurocomputing*, **448** (2021), 140–151. https://doi.org/10.1016/j.neucom.2021.03.105

24. Y. Yang, Q. Wu, Y. Fu, X. Chen, Continuous convolutional neural network with 3D input for EEG-based emotion recognition, in *Neural Information Processing. ICONIP 2018. Lecture Notes in Computer Science*, Springer, (2018), 433–443.

25. H. Chen, S. Sun, J. Li, R. Yu, N. Li, X. Li, et al., Personal-zscore: Eliminating individual difference for eeg-based cross-subject emotion recognition, *IEEE Trans. Affect. Comput.*, **14** (2021), 2077–2088. https://doi.org/10.1109/taffc.2021.3137857

26. S. Liu, X. Wang, L. Zhao, B. Li, W. Hu, J. Yu, et al., 3DCANN: A spatio-temporal Convolution attention neural network for EEG emotion recognition, *IEEE J. Biomed. Health.*, **26** (2022), 5321–5331. https://doi.org/10.1109/JBHI.2021.3083525

27. C. Domingos, J. L. Marôco, M. Miranda, C. Silva, X. Melo, C. Borrego, Repeatability of brain activity as measured by a 32-channel EEG system during resistance exercise in healthy young adults, *Int. J. Environ. Res. Public Health*, **20** (2023), 1992. https://doi.org/10.3390/ijerph20031992

28. S. Liu, Z. Wang, Y. An, B. Li, X. Wang, Y. Zhang, DA-CapsNet: A multi-branch capsule network based on adversarial domain adaption for cross-subject EEG emotion recognition, *Knowl-based. Syst.*, **283** (2024), 111137. https://doi.org/10.1016/j.knosys.2023.111137.

29. J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, Y. Zhang, Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks, *IEEE Access*, **7** (2019), 44317–44328. https://doi.org/10.1109/access.2019.2908285

30. S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, et al., Deap: A database for emotion analysis; using physiological signals, *IEEE Trans. Affect. Comput.*, **3** (2011), 18–31. https://doi.org/10.1109/t-affc.2011.15

31. T. Song, W. Zheng, S. Liu, Y. Zong, Z. Cui, Y. Li, Graph-embedded convolutional neural network for image-based EEG emotion recognition, *IEEE Trans. Affect. Comput.*, **10** (2021), 1399–1413. https://doi.org/10.1109/tetc.2021.3087174

32. Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, G. Chen, A channel-fused dense convolutional network for EEG-based emotion recognition, *IEEE Trans. Cogn. Dev. Syst.*, **13** (2020), 945–954. https://doi.org/10.1109/tcds.2020.2976112

33. Y. Gu, X. Zhong, C. Qu, C. Liu, B. Chen, A domain generative graph network for EEG-based emotion recognition, *IEEE J. Biomed. Health Inf.*, **27** (2023), 2377–2386. https://doi.org/10.1109/JBHI.2023.3242090

34. L. Yang, Y. Wang, X. Yang, C. Zheng, Stochastic weight averaging enhanced temporal convolution network for EEG-based emotion recognition, *Biomed. Signal Proces.*, **83** (2023), 104661. https://doi.org/10.1016/j.bspc.2023.104661

35. X. Du, C. Ma, G. Zhang, J. Li, Y. K. Lai, G. Zhao, et al., An efficient LSTM network for emotion recognition from multichannel EEG signals, *IEEE Trans. Affect. Comput.*, **13** (2020), 1528–1540. https://doi.org/10.1109/TAFFC.2020.3013711

36. F. Cui, R. Wang, W. Ding, Y. Chen, L. Huang, A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition, *Mathematics*, **10** (2022), 582. https://doi.org/10.3390/math10040582

37. M. Ramzan, S. Dawn, Fused CNN-LSTM deep learning emotion recognition model using electroencephalography signals, *Int. J. Neurosci.*, **133** (2023), 587–597. https://doi.org/10.1080/00207454.2021.1941947

38. G. Peng, K. Zhao, H. Zhang, D. Xu, X. Kong, Temporal relative transformer encoding cooperating with channel attention for EEG emotion analysis, *Comput. Biol. Med.*, **154** (2023), 106537. https://doi.org/10.1016/j.compbiomed.2023.106537

39. D. Kuang, C. Michoski, SEER-net: Simple EEG-based Recognition network, *Biomed. Signal Proces.*, **83** (2023), 104620. https://doi.org/10.1016/j.bspc.2023.104620

40. M. Sun, W. Cui, S. Yu, H. Han, B. Hu, Y. Li, A dual-branch dynamic graph convolution based adaptive transformer feature fusion network for EEG emotion recognition, *IEEE Trans. Affect. Comput.*, **13** (2022), 2218–2228. https://doi.org/10.1109/TAFFC.2022.3199075

41. T. Song, W. Zheng, P. Song, Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks, *IEEE Trans. Emerg. Top. Comput.*, **11** (2018), 532–541. https://doi.org/10.1109/taffc.2018.2817622

42. J. A. Coan, J. J. B. Allen, Frontal EEG asymmetry as a moderator and mediator of emotion, *Biol. Psychol.*, **67** (2004), 7–50. https://doi.org/10.1016/j.biopsycho.2004.03.002

43. C. Wang, Y. Li, L. Wang, S. Liu, S. Yang, A study of EEG non-stationarity on inducing false memory in different emotional states, *Neurosci. Lett.*, **809** (2023), 137306. https://doi.org/10.1016/j.neulet.2023.137306

44. A. S. Reis, E. L. Brugnago, R. L. Viana, A. M. Batista, K. C. Iarosz, I. L. Caldas, Effects of feedback control in small-world neuronal networks interconnected according to a human connectivity map, *Neurocomputing*, **518** (2023), 321–331. https://doi.org/10.1016/j.neucom.2022.11.008

45. S. Halder, D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, et al., Neural mechanisms of brain–computer interface control, *Neuroimage*, **55** (2011), 1779–1790. https://doi.org/10.1016/j.neuroimage.2011.01.021

46. B. C. Gibson, A. Vakhtin, V. P. Clark, C. C. Abbott, D. K. Quinn, Revisiting hemispheric asymmetry in mood regulation: implications for rTMS for major depressive disorder, *Brain Sci.*, **12** (2022), 112. https://doi.org/10.3390/brainsci12010112

47. X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, et al., EEG based emotion recognition: A tutorial and review, *ACM Comput. Surv.*, **55** (2022), 1–57. https://doi.org/10.1145/3524499

48. J. D. Herrington, W. Heller, A. Mohanty, A. S. Engels, M. T. Banich, A. G. Webb, et al., Localization of asymmetric brain function in emotion and depression, *Psychophysiology*, **47** (2010), 442–454. https://doi.org/10.1111/j.1469-8986.2009.00958.x

49. B. Mishra, S. Tarai, V. Ratre, A. Bit, Medicine, Processing of attentional and emotional stimuli depends on retrospective response of foot pressure: Conceptualizing neuron-cognitive distribution in human brain, *Comput. Biol. Med.*, **164** (2023), 107186. https://doi.org/10.1016/j.compbiomed.2023.107186

50. Y. Liu, O. Sourina, M. K. Nguyen, Real-time EEG-based emotion recognition and its applications, in *Transactions on Computational Science XII*, Springer, (2011), 256–277. https://doi.org/10.1007/978-3-642-22336-5_13

51. N. Jatupaiboon, S. Pan-Ngum, P. Israsena. Emotion classification using minimal EEG channels and frequency bands, in *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, (2013), 21–24. https://doi.org/10.1109/JCSSE.2013.6567313