Emotional Lateralization-inspired Spatiotemporal Neural Networks for EEG-based Emotion Recognition

Dongmin Huang[†a], Dazhi Jiang[†a], Sijin Zhou[a], Liang Chen[a], Jiali Lin[b] and Erik Cambria[c,*]

[a]*Department of Computer Science, Shantou University, China*
[b]*Business School, Shantou University, China*
[c]*College of Computing and Data Science, Nanyang Technological University, Singapore*

ARTICLE INFO

Keywords:
Emotion Recognition
Channel Selection
Emotional Lateralization
CNN
LSTM

ABSTRACT

Due to the volume conduction effects in the brain, multi-channel electroencephalogram (EEG) contain redundant information. Thus it is necessary to employ the channel selection method to ascertain vital channel in the EEG emotion recognition task. Existing studies select channels by the correlation analysis or the supervised learning. However, such methods are impacted by the inherent limitations of the datasets, such as the limited numbers and unbalanced distribution, which tend to overfit high-frequency emotions and underfit low-frequency ones. As a consequence, these selected channels are highly related to certain emotions, but low related to the overall emotion. Hence, it is an essential problem in current research to select an effective channel subset using unsupervised learning to fully improve the universal adaptability in multi-label classification tasks. In this paper, we propose a channel selection method named emotional lateralization-inspired spatiotemporal neural networks (ELSTNN) for EEG emotion recognition. In ELSTNN, the neuroscience finding of emotional lateralization is applied to guide channel selection with unsupervised learning, which could remove redundant channels. Then, ELSTNN with three convolution layers and one long short-term memory layer are applied to mine and integrate the deep spatiotemporal feature from the selected EEG signals. To evaluate the effectiveness of ELSTNN, extensive experiments are carried out on two public datasets, DEAP and DREAMER. Our proposed method achieves better performance compared with various state-of-the-art methods, which obtains the mean accuracy 95.67% in valence and 94.97% in arousal on DEAP, and 93.72% in valence and 93.57% in arousal on DREAMER.

1. Introduction

Electroencephalogram (EEG) is the bioelectrical signal that is produced by the autonomic nervous system of humans, which directly reflects the neural activity of the brain cortex [1, 2]. As EEG has some natural advantages in emotion recognition, e.g. high temporal resolution, reliability, and objectivity [3], EEG emotion recognition has received a growing amount of attention in affective brain-computer interface (BCI) [4, 5, 6]. In BCI, EEG acquisition is the first step for EEG emotion recognition. Here the channel is defined as the sensor used to record EEG signal from the specific location in the cerebral cortex [7]. Recently, the performance of EEG emotion methods has been improved, which partly profited from that more and more neuronal activity is uncovered using an ever-increasing number of channels [8]. However, continuously increasing channels not only increase the redundant information caused by inevitable noise, but also lead to the exponential growth of EEG representation dimension [9]. Due to the small sample size of EEG dataset [10], the high-dimensional EEG representations will easily raise the overfitting in the problem solving. In addition, the increasing density of channels reduces the convenience and comfort of BCI, which may be the obstacle to popularization and daily application [11]. Therefore, these factors directly induce the urgency of finding a minimum-number but task-related channels subset.

To this end, various channel selection methods were proposed [12], such as information gain [13] and maximum relevance minimum redundance [14]. According to the way it is combined with the classifier, these channel selection methods could be subdivided into three kinds: filter-based method, iterative-based method, and learning-based method [15]. The filter-based methods use the sorting technique as the principal criterion of selection, and the correlation between these features and corresponding emotional labels will be ranked for the channel selection. Since this method is separated with the learning process of classifier, however, it may lead to the selected channel not being the optimal channel subset [16]. The iterative-based method solves the disadvantage of filter-based methods to some extent, which adopts a two-step strategy [17]. The first step is used to determine the subset of EEG channels by an algorithm, and the second is employed to evaluate the performance of subset using a specific classifier. The iterative-based method method will repeat the above steps to find the optimal subset [18]. However, the main weakness of the iterative-based method is the computational cost, which is higher than most of the filter-based methods [17]. The learning-based method attempts to use the regularization method to overcome the shortcoming of the above two methods, in which Xu et al. [11] preserve the discriminant information in the orthogonal subspace during the automatically channel selection. Nevertheless, the defect of it is that many highly related EEG features are preserved and depend on the label information.

[†]These authors contributed equally to this work.
[*]Corresponding author
ORCID(s):

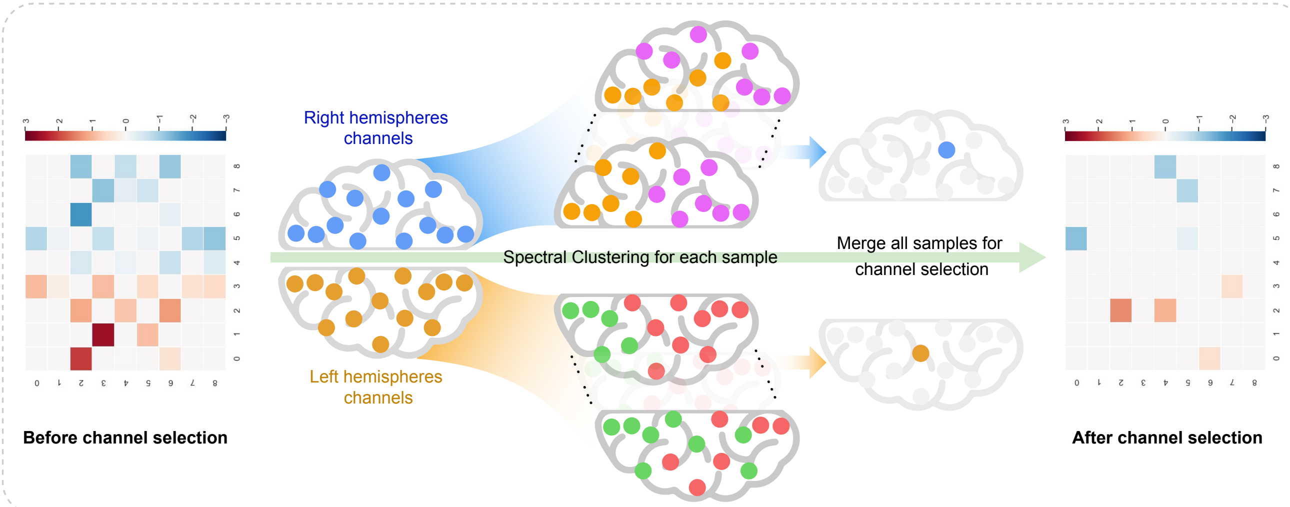

Figure 1: The main idea of emotional lateralization-based channel cluster selection. According to the 10-20 system, the EEG 2D representation before and after channel selection is visualized. Firstly, spectral clustering is used for the left and right hemispheres of each sample. Then, merging all samples is applied to choose the channel that most frequently appeared in different groups.

From the point of neuroscience, one-channel EEG signal is the linear combination of several underlying source signals caused by the volume conduction effects of the brain [19]. This fact reveals that the EEG feature extracted from these physically adjacent channels may be highly related but redundant [20, 21]. Furthermore, it is proved that the redundant EEG features provide less or even no useful extra information for the task [15, 22], so only one channel of these channels needs to be reserved. Since most channel selection methods use the relationship between emotional label and these channels to rank channels, the ranking of highly related channels tends to be ordered equally [17]. As a result, the channel subset obtained by these methods still has the redundant information. More importantly, as the number of channel increases, the dimension of EEG representation increases correspondingly, which further exacerbates the undesirable consequence mentioned above [23]. Therefore, it is necessary to further investigate how to remove the highly related but redundant channels to improve the convenience and performance of BCI. In addition, the above channel selection methods depend greatly on abundant emotional labels. Specifically, the correlation between channel features and specific emotional labels is used to rank and choose the key channels [15], or specific emotional labels is employed to filter the channels in the regularization method during the supervised learning [11]. Due to the unbalanced distribution of emotion and the small number of samples in EEG datasets [10], those methods tend to overfit high-frequency emotions and underfit low-frequency ones. Consequently, the richness of these selected channels is limited by particularly emotional state, and lacks universal adaptability to overall emotional states. Hence, it is desired to investigate channel selection method based on unsupervised learning, which could enhance the universal adaptability of channel subset in multi-label classification task.

To deal with the above-mentioned issues, we propose a novel emotional lateralization-inspired spatiotemporal neural networks (ELSTNN) for channel selection method. The basic idea of model is to apply the phenomenon of emotional lateralization (the left and right hemispheres respond differently to emotion stimuli [24, 25]) to guide the removal of highly related but redundant channels, and then to explore the deep spatiotemporal EEG feature for emotion recognition. To remove the highly related but redundant channels, spectral clustering is introduced to cluster channels located the left and right hemisphere of every samples as two groups. This way, the channels belonging to the same group are considered to be highly related with each other inherently. Then for the hemispheres, the clustering results of all samples are merged to choose the channel that would appear frequently in different groups. The main idea of this process is shown in Fig. 1.

To get rid of the limitations of emotional labels, ELSTNN separates the channel selection process from the classifier and adopts unsupervised learning for the channel selection. This way, the redundant channels are filtered out without the emotional label information. To improve the performance of classifier, the 2D EEG representation is converted to 3D representation with spatiotemporal information. After that, ELSTNN is used for EEG emotion recognition with three convolutional layers and one long short-term memory (LSTM) layer. To capture the emotional specificity and contribution at different times, three convolutional layers are applied to mine the temporal and spatial features of emotion per second. Then, one LSTM layer is used to integrate these temporal and spatial features for exploiting the continuity of emotion in time. Extensive experiments are carried out on DEAP [26] and DREAMER [27], where the label is set as high/low valence (HV/LV) and high/low arousal (HA/LA).

The experimental results show that ELSTNN achieve the mean accuracy 95.67% and 94.97% in valence and arousal on DEAP, while 93.72% and 93.57% in valence and arousal on DREAMER. Compared with various state-of-the-art methods, our proposed method is effective and superior. To sum up, the contribution of this paper can be summarized as follows.

1. We design ELSTNN for EEG emotion recognition, where the continuity and specificity of emotion in time are considered. ELSTNN uses three convolutional layers to explore the deep spatiotemporal information of one second with the goal of learning the temporal specificity. And one LSTM layer of ELSTNN is used to integrate all deep feature to mine the continuity information of emotion in time.
2. In the process of experiment, the effect of time on channel stability is considered, where various of different-length time windows are applied to conduct experiments to evaluate the time stability of selected channels. The channels selected by ELSTNN have good time stability, which may provide a reference for the design of portable BCI for long-term emotional monitoring.
3. In this paper, the emotional lateralization finding is used to achieve the channel selection combining with the unsupervised learning. Though visualized analysis, the 3D EEG representation formed by the selected channel can preserve the information about emotional lateralization. To our best knowledge, we are the first to take advantage of the combination of neuroscience phenomena and unsupervised learning for channel selection method in EEG emotion recognition task.

The remainder of this paper is structured as follows: Section 2 is the related work; the detailed operation of ELSTNN is presented in Section 3; in Section 4, the detailed description about dataset, model implementations, experimental results, and visual analysis are given; finally, the conclusion of this paper is shown in Section 5.

2. Related Work

2.1. Deep learning for EEG Emotion recognition

The goal of BCI is to render computers the power of recognizing, understanding, and representing human emotion for the concordant human-machine interaction [28]. As the primary link of this goal, emotion recognition can be realized via the objective methods (physiological signals) and the subjective methods (face, voice and body action) [26, 29, 30]. Comparing with other medium, EEG is receiving increasing attention in BCI due to the advantages of objectivity, reliability and uncontrollability by human will [3, 31]. Most EEG emotion recognition researches follow such paradigm: from EEG acquisition and preprocessing, feature extraction, to classifier design. In the early stage, various handcraft features constructed by extracting emotion-related information from multi-channel EEG signals were the focus of research [32]. Petrantonakis et al. [33] proposed a higher order crossings method based on the time domain to capture the oscillatory pattern of EEG. Duan et al. [34] proposed a differential entropy feature to explore energy distribution in different frequency bands, which is still widely used at present [35]. Liu et al. [36] applied the empirical mode decomposition and multiple feature extraction methods to explore the EEG representation.

However, the construction process of these handcraft features is always uncertain to some extent and dependent on human experience, which leads to the loss of emotional information of EEG [37]. With the success of deep learning in other fields [38, 39, 40, 41, 42, 43], automatic feature extraction and classification using deep learning has gradually become the mainstream method of EEG emotion recognition. Zhang et al. [44] use LSTM to explore spatiotemporal information with the sequence of different channel traversal. Song et al. [45] proposed a method based on graph convolutional network (GNN) to learn the functional connection of all channels. Smith et al. [46] used CNN with different kernels to automatically extract time-frequency feature.

Liang et al. [47] proposed an EEG code method based on autoencoder to learn the generalized representation. Among various deep learning-based methods for EEG emotion recognition, it has become the major trend to design the model according to the characteristics of EEG signals to explore the temporal and spatial feature.

2.2. EEG Channel selection

Due to the volume conduction effects of the brain [19], each signal of multi-channel EEG is the result of linear combination of several underlying source signals [48]. As a result, the EEG features extracted from these physically adjacent channels are highly related but redundant [22]. It has been proven that these redundant features can't provide useful information in classification task [15]. And the EEG dimension increases with the number of channels, which can easily lead to the overfitting problem and poor performance of classifier. In addition, the daily application of BCI depends on the convenience and comfort of wearable devices, but the dense channels hinder its popularization [15].

Hence, researchers pay increasing attention on exploring the key channel related with task to implement the goal of BCI in the view of theoretical study and practical application [15]. According to the way combining with classifier, the channel selection methods could be summarized into three kinds: filter-based method, iterative-based method, and learning-based method [49]. For the filter-based method, the contribution of channel in emotion recognition would be ranked according to certain criteria, the channel is selected according to the ranking. Chen et al. [13] apply the gender-related correlations between EEG channel features and specific emotional states using information gain method for channel selection. Zabihi et al. [50] use the conditional mutual information maximization method and multi-domain features to choose channel.

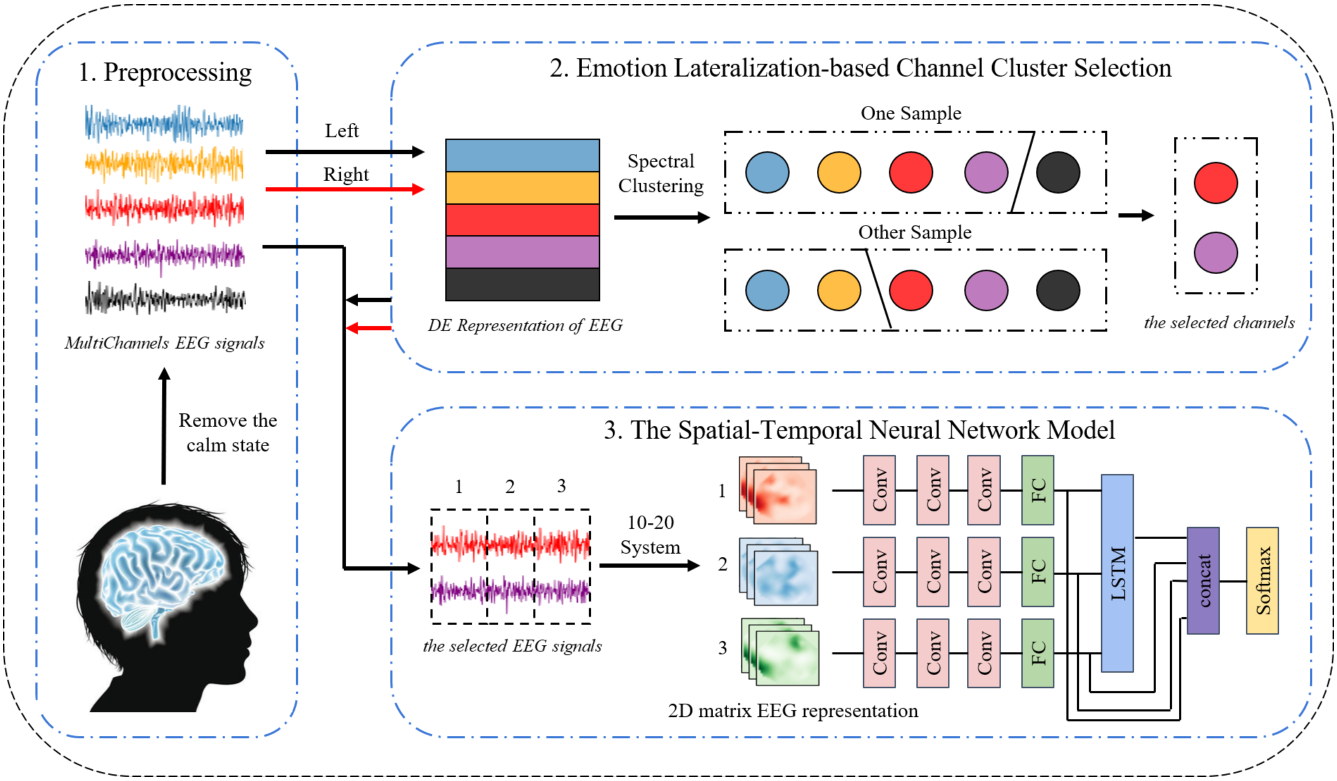

Figure 2: The whole framework of ELSTNN. The first step is the data preprocessing, where the effect of calm state is removed to ensure the pure of EEG. The second step is the channel selection using ELCCS to reserve the internal-external differences between hemispheres. The last step is to build an emotion recognition model, in which the shallow and deep information of EEG signals is used for emotion recognition.

However, these methods ignore the learning process of classifier, which may make it difficult to achieve optimal channel subset. In order to overcome the above shortcoming, iterative-based methods use an iterative way to choose channel. Channel subsets are firstly generated randomly, and then the subset performance is evaluated by classifier [17]. Handiru et al. [51] search the most relevant channels using an iterative method by exploiting the functional correlation of channels. Nakisa et al. [52] automatically select the channels by the evolutionary computation. Nevertheless, the calculation cost of these methods is always more expensive than filter-based method [17].

As an alternative to solving the shortcoming of aforementioned methods, learning-based methods attempt to select channels during model optimization [15]. Tao et al. [53] introduced the attention mechanisms to automatically select the critical channels. Xu et al. [11] preserved more discriminative information in channel subset using orthogonal regression. However, the specific emotion information is applied to channel selection among these methods, which may limit the comprehensiveness and the possibility of EEG representation. Therefore, it should be further investigated to adopt unsupervised learning method for exploring the EEG channel subset with powerful generalization ability.

3. Method

In this section, ELSTNN is described in detail from EEG preprocessing to channel selection, and finally to model design. The whole framework is shown in Fig. 2. In the first part, the emotional calm state is removed to ensure that EEG is pure under the emotional stimulation. The key point of second part is how spectral clustering is combined with channel selection to model emotional lateralization. For the last part, the model is designed by integrating the spatiotemporal emotional information for emotion recognition.

3.1. Data Preprocessing

During the process of EEG data acquisition, the baseline signal B is collected as EEG signals collected without any form of stimulation, which represents the calm state of a subject. For the sake of description, the EEG signal collected during emotional stimulation is marked as O. It has been proven that removing EEG signals from the calm state is necessary to improve the emotion recognition performance [54]. The operation of removing the effect of calm state is shown in Fig. 3 and can be expressed by the following mathematical formula.

$$E_i = O_i - m(B) \tag{1}$$

where $m(\cdot)$ denotes the mean operation to obtain the mean value of baseline signal, O_i denotes the stimulated EEG segment equal in length to the mean baseline signal, and E_i

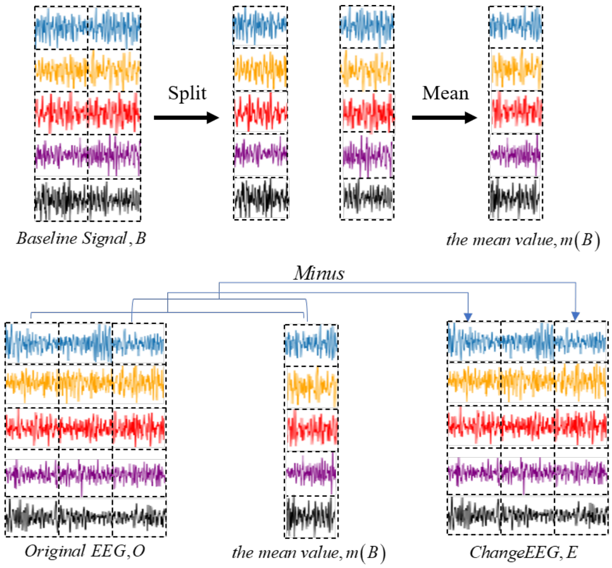

Figure 3: The operation of EEG preprocessing. Firstly, the mean of baseline signal is calculated to represent the calm state. Then, the EEG signals recorded under stimulation minus the mean of baseline signal.

denotes the EEG segment removing the calm state. Therefore, the EEG change value under emotional stimulation is employed for the channel selection and the input of neural network.

3.2. Emotional Lateralization-based Channel Cluster Selection

Due to the brain volume conduction effects [19], there is a large amount of highly related but redundant information in channel features extracted from physically similar EEG channels [22, 55]. However, the redundant information of EEG feature not only does not have the ability to improve the emotion recognition performance [56], but also tends to cause overfitting of classifiers due to the small EEG dataset and the high dimensional EEG vector [10]. Hence, the redundant information of EEG feature should be removed. In this paper, we propose a method called emotional lateralization-based channel cluster selection (ELCCS), and its goal is to remove the redundant channel using spectral clustering to model the emotional lateralization phenomenon. In brief, all channels are first roughly divided into left and right subsets according to the channel distribution of the International 10–20 System. Then the spectral clustering is introduced for every subset to cluster channels into different groups, in which the channels of one group are highly related. In the end, summary all training samples and select channels which appear in different groups. The main idea of ELCCS can be represented using Fig. 4, and the specific operation of ELCCS is described as follows.

Suppose that one EEG sample of subject in dataset is marked as $S_i = [X_i, Y_i]$, where X_i and Y_i denote the i-th preprocessed EEG signal and the corresponding emotional label. Furthermore, X_i contains multi-channel EEG signals, which is marked as $X_i = [c_1, c_2, \cdots, c_n] \in \mathbb{R}^{n \times p}$. n and p denote the number of channels and the sampling points. Due to the frequency-domain analysis reflecting activity differences among multiple brain regions, the spectral power analysis is firstly carried out using the feature extracted from X_i in frequency domain. Among a large number of spectral power analysis methods, it has been proven the effectiveness and excellence of differential entropy (DE) [34] in the emotion recognition task. Therefore, the handcraft feature is accomplished by applying DE for X_i, which can be implemented by the following equation.

$$\begin{aligned} DE &= -\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(t-\mu)^2}{2\sigma^2}\right)\right) \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right. \\ &\quad \left.\times \exp\left(\frac{(t-\mu)^2}{2\sigma^2}\right)\right) dt \\ &= \frac{1}{2}\log 2\pi e \sigma^2 \end{aligned} \tag{2}$$

where t denotes the EEG signal. Through the above formula, the EEG X_i is translated into the handcraft feature $H_i = [c_1^h, c_2^h, \cdots, c_n^h] \in \mathbb{R}^{c \times b}$, where n and b denote the number of channels and bands. And c_i^h denotes the DE feature of the i-th channel. The frequency band is generally divided into five sub-bands: δ (1–3 Hz), θ (4–7 Hz), α (8–14 Hz), β (14–30 Hz), and γ (31–45 Hz). These sub-bands play different roles in the affective system of brain and contribute to emotion in different degrees [31]. However, the δ band is more related with the unconscious of deep sleep [31, 36]. Hence, the frequency information of θ, α, β and γ bands are chosen in our method, which means b is equal to 4. After that, channels are roughly divided into subsets of channels in the left and right hemispheres according to the relative physical location of channels of international 10-20 system, which are marked as H_i^l and H_i^r.

After the above operation, the spectral clustering is applied for every frequency band of each subset to bring together highly related channels using the handcraft feature of left and right hemispheres. The goal of spectral clustering is to split a complete spectrum G into different separate subgraphs $\{A_1, A_2, \cdots, A_k\}$ with two conditions.

The first condition is to achieve the lowest possible weighted sum of the edges between different subgraphs, and the second is to achieve the highest possible weighted sum of the edges within the subgraphs. In our task, H_i^l and H_i^r are considered as complete spectrum G. Firstly, the construction of similar graphs is accomplished by the fully connected method based on Gaussian-based similarity function, which is a common approach in different spectral clustering applications. The Gaussian-based similarity function, measuring the similarity between channels is defined as follows.

$$s(c_i^h, c_j^h) = e^{-\frac{\left\|c_i^h - c_j^h\right\|^2}{2\sigma^2}} \tag{3}$$

where $s(c_i^h, c_j^h)$ means the similarity between the i-th and j-th channel. The similarity between two pairs of all channels is calculated using (3), and the similarity matrix is constructed. Then, the Laplace matrix of the graph is calculated as follows.

$$L = D - W \tag{4}$$

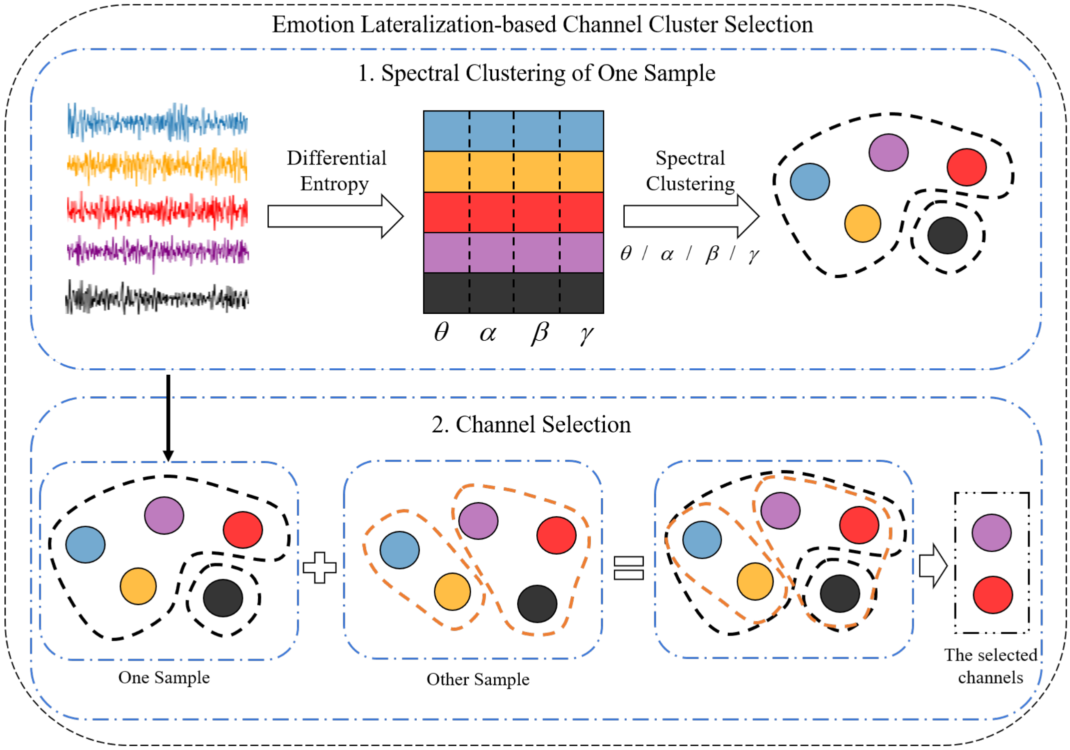

Figure 4: The main idea of emotional lateralization-based channel cluster selection. In this method, a one-second long window is used to slice the EEG of whole trial. Then the differential entropy (DE) of these segments is calculated in four bands. Next, the spectral clustering is introduced for each band of each segment to obtain the clustering of channels in the current segment. The above operations correspond to the top half of the figure. Last, by comparing the channel clustering of all the fragments, the channels appearing in different clusters are selected, which corresponds to the lower part of the figure

where W denotes the similarity matrix, and D denotes the degree matrix of G, in which $\{d_1, d_2, \cdots, d_n\}$is the elements of diagonal matrix D. These elements are referenced as follows.

$$d_i = \sum_{j=1}^{n} s_{ij} \tag{5}$$

where d_i denotes the sum of similarity between the i-th channel and other channels. In the end, the normalized cut (Ncut) is used to cut G into different separate subgraphs, which is proposed by Shi et al. [57]. The specific operation of Ncut can be described as an optimization problem, which is shown as follows.

$$\min Ncut(A_1, A_2, \cdots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \overline{A_i})}{vol(A_i)} \tag{6}$$
$$s.t. \overline{A_i} = \{c_r | c_r \in V and\ c_r \notin A_i\}$$

where $vol(A_i)$ denotes the sum of similarity of all channels in A_i, and $W(A_i, \overline{A_i})$ denotes the tangent weight between the subset and the remaining subset, which is shown as follows.

$$W(A_i, \overline{A_i}) = \sum_{i\in A_i, j \in \overline{A_i}} s_{ij} \tag{7}$$

Since the optimization formula (6) is difficult to calculate, therefore, the indicator vector $h_j \in \{h_1, h_2, \cdots, h_n\}, j = 1, 2, \cdots, n$ is introduced to simplify the optimization problem using the Laplace matrix of graph L. Every h_j has n dimensions that are equal to the number of channels in every channel subset. The definition of h_j is as follows.

$$h_{ij} = \begin{cases} 0 & , c_i \notin A_j \\ \frac{1}{\sqrt{vol(A_j)}} & , c_i \in A_j \end{cases} \tag{8}$$

Then for $h_i^T L h_i$, the following identity transformation exists:

$$\begin{aligned} h_i^T L h_i &= \frac{1}{2}\sum_{m=1}\sum_{n=1} s_{mn}(h_{im} - h_{in})^2 \\ &= \frac{1}{2}(\sum_{m\in A_i, n\notin A_i} s_{mn}(\frac{1}{\sqrt{vol(A_j)}} - 0)^2 + \\ &\sum_{m\notin A_i, n\in A_i} s_{mn}(0 - \frac{1}{\sqrt{vol(A_j)}})^2) \\ &= \frac{1}{2}(\sum_{m\in A_i, n\notin A_i} s_{mn}\frac{1}{vol\left(A_i\right)} + \sum_{m\notin A_i, n\in A_i} s_{mn}\frac{1}{vol\left(A_i\right)}) \\ &= \frac{1}{2}(\sum_{m\in A_i, n\notin A_i} \frac{W(A_i, \overline{A_i})}{vol\left(A_i\right)} + \sum_{m\notin A_i, n\in A_i} \frac{W(\overline{A_i}, A_i)}{vol\left(A_i\right)}) \\ &= \sum_{i=1}^{k} \frac{W(A_i, \overline{A_i})}{vol(A_i)} \end{aligned} \tag{9}$$

Therefore, the optimization problem is transformed as following.

$$
\begin{aligned}
&\min Ncut(A_1, A_2, \cdots, A_k) \\
&= \sum_{i=1}^{n} h_i^T L h_i = \sum_{i=1}^{n} (H^T L H)_{ii} = tr(H^T L H) \\
&s.t. H^T D H = I
\end{aligned}
\tag{10}
$$

Now, the indicator matrix is transformed as $H = D^{-1/2}F$. Due to $H^T L H = F^T D^{-1/2} L D^{-1/2} F, H^T D H = F^T F = I$. The optimization problem will be changed as follows.

$$
\begin{aligned}
&\min Ncut(A_1, A_2, \cdots, A_k) = tr(H^T L H) \\
&= tr(F^T D^{-1/2} L D^{-1/2} F) \\
&s.t. F^T F = I
\end{aligned}
\tag{11}
$$

After the above operation, K-means is applied to feature matrix F to obtain the final clustering result of left and right channel subsets of one sample. The selected channel is referenced as the channels that appear in different clusters in various emotional samples. In this method, spectral clustering is applied to select the important channels in each frequency band from the set of the left and right hemispheric channels. During the processing, the highly related channels of hemispheres are filtered down to just one. Furthermore, the internal and external differences between the left and right hemispheres are modelled using these channels to construct the discriminative EEG features with small dimensions.

3.3. Spatiotemporal Neural Network(STNN)

In ELSTNN, we propose a deep learning-based model called spatiotemporal neural network (STNN) to mine the temporal stability of the selected channel, in which the temporal and spatial information of the emotional lateralization is extracted by STNN for EEG emotion recognition. The framework of STNN is shown in Fig. 2. Specifically, the CNN layers are used to extract the spatiotemporal information of emotional lateralization within one second. Then the LSTM layer is employed to mine the deep spatiotemporal information of emotional lateralization.

In the end, all shallow and deep spatiotemporal information is applied to EEG emotion recognition. The specific operation is described below. In previous studies of EEG emotion recognition, the LSTM and CNN layers are applied to structure the neural network to extract the spatiotemporal information within one second, respectively. Li et al. [58] use different channel traversal methods combined with LSTM to ensure the completeness of spatial information among channels for emotion recognition. He et al. [59] construct a 2D EEG matrix based on the relative physical position of the channel of the EEG acquisition device and use CNN for emotion recognition. However, RNN always requires that the EEG channels of samples to have a chain-like structure to extract the deep feature. In the International 10-20 system

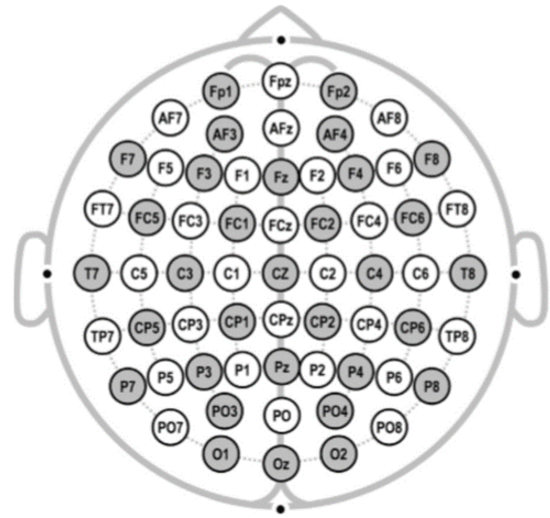

Figure 5: The international 10-20 system. It represents the electrical channel layout of the brain, and is used to capture multi-channels EEG signals at specific locations in the brain.

that is shown in Fig. 5, there are several physically adjacent channels around each channel. As a result, the chain-like structure of channels is relatively difficult to represent the spatial information among channels. Therefore, the CNN layers are chosen to extract the spatiotemporal information within one second from EEG. One second of chain-like EEG $X \in \mathbb{R}^{c^s \times p}$ is required to transform into the 2D EEG matrix $X \in \mathbb{R}^{p \times r \times c}$ according to the channel layout of international 10-20 system, where c^s denotes the number of the selected channel using ELCCS. r and c denote the length and width of the channel layout, which corresponding to the maximum number of channels in the vertical and horizontal directions in the acquisition device. The transformation matrix of the above operation relating to 32-channels 10-20 system in DEAP dataset can be represented by the following mathematical formula.

$$
f_t = \begin{bmatrix}
0 & 0 & 0 & Fp1 & 0 & Fp2 & 0 & 0 & 0 \\
0 & 0 & 0 & AF3 & 0 & AF4 & 0 & 0 & 0 \\
F7 & 0 & F3 & 0 & Fz & 0 & F4 & 0 & F8 \\
0 & FC5 & 0 & FC1 & 0 & FC2 & 0 & FC6 & 0 \\
T7 & 0 & C3 & 0 & CZ & 0 & C4 & 0 & T8 \\
0 & CP5 & 0 & CP1 & 0 & CP2 & 0 & CP6 & 0 \\
P7 & 0 & P3 & 0 & Pz & 0 & P4 & 0 & P8 \\
0 & 0 & 0 & PO3 & 0 & PO4 & 0 & 0 & 0 \\
0 & 0 & 0 & O1 & Oz & O2 & 0 & 0 & 0
\end{bmatrix}
\tag{12}
$$

where the zero elements denote the blank position of collecting EEG devices where the EEG signals could not be collected. The non-zero elements denote the channels placed on the human scalp. After ELCCS, the value of unselected channels is also set to 0 to ensure the validity of channel selection. After that, Z-score is used to remove the effect of the values by normalizing each non-zero elements in the matrix, which is shown as follows.

$$z^i = \frac{f^i - \overline{f}}{\sigma_f} \tag{13}$$

where $\overline{f}$ and σ_f denote the mean value and standard deviation of all selected elements, respectively.

After obtaining the 2D EEG representation of one second, the EEG sample of one subject $X \in \mathbb{R}^{s \times l \times p \times r \times c}$ is formed with the observation period $[t, t+1, \cdots, t+l]$, where s, l and p denote the total number of samples, the duration of observation period, and the sampling points. Then, the spatiotemporal information of each second during the observation period is extracted using three consecutive CNN layers, as shown in Fig. 2. These layers have the same setup, except that the first one has 32 filters, the second and third have 64 and 128 filters. And in computer vision, the 3×3 kernel is often chosen to extract information. However, the 4×4 kernel is chosen for three CNN because it can learn more correlation among more channels than the 3×3 kernel. And the set of padding models is the same to reserve the edge information of each transformation matrix. The activation function is relu. In addition, the batch normalization is used after each CNN to prevent overfitting and speed up the training process. After then, the full connection layer is used to integrate and compress the spatiotemporal information of every second.

Let $X_i = [x_1, x_2, \cdots, x_l] \in \mathbb{R}^{l \times p \times r \times c}, x_j \in \mathbb{R}^{p \times r \times c}, i = 1, 2, \cdots, s, j = 1, 2, \cdots, l$ denote one of EEG samples. The spatiotemporal information of every second after the above operation can be formulated as follows.

$$h_i = C(x_i) = W_2 \cdot (Relu(Conv2D(x_i, W_1))) + b_1, i = 1, 2, \cdots, l \tag{14}$$

where $C(\cdot)$ denotes the operation of using three CNN and one FC to obtain the information. w_1, w_2, and b_1 denote the parameters that need to be learned through back propagation. And the representation of one EEG sample can be represented as follows.

$$H_i = C(X_i) = [h_1, h_2, \cdots, h_i] \in \mathbb{R}^{l \times s}, i = 1, 2, \cdots, l \tag{15}$$

where s denotes the number of hidden units of FC, which is set as 128. After that, LSTM is introduced to extract the deep spatiotemporal information from H_i related to emotional accumulation with the advantage of LSTM in extracting sequence data. Here, the output of above operation is set as the last hidden units of LSTM, which is often considered as the emotional state at the current moment. Then, the final EEG emotional representation vector is formed by concatenating the shallow and deep spatiotemporal information, H_i and $\mathcal{L}(H_i)$, where $\mathcal{L}(\cdot)$ denotes the LSTM operation. Last, FC with softmax activation function is employed to output the emotion label, which can be described as follows.

$$the final representation : R = \left[h_1, h_2, \cdots, h_i, \mathcal{L}(H_i)\right] \tag{16}$$

Table 1
The detail of emotion recognition datasets. Here, V, A, D, L and F denote valence, arousal, dominance, liking, and familiarity.

Attribute	DEAP	DREAMER
Audio-visual stimuli		
Videos	40	18
Duration	60 s	65 - 393s
Experiment information		
Subjects	16 m, 16 f	14 m, 11 f
Age	19 – 37	22 – 33
Rating scales	V, A, D, L, K	V, A, D
Rating value	1 - 9	1 - 5
EEG format for each subject		
Baseline	40 * 32 * 384	18 * 14 * 7808
Data	40 * 32 * 7680	23 * 18 * 25472
Label	40 * 5 (V, A, D, L, K)	40 * 3 (V, A, D)

$$emotionlabel : p = softmax(W_3 \cdot R + b_2) \tag{17}$$

where R denotes the final EEG representation vector, and p denotes the probability of predicted label. During training, the cross entropy is set as the loss function and is defined as follows.

$$L = \frac{1}{s} \sum_{i=1}^{s} -[y_i \cdot log(p_i) + (1 - y_i) log(1 - p_i)] \tag{18}$$

where y_i denotes the ground truth label. All weighting parameters are learned by backpropagation. In this method, ELSTNN is to mine spatiotemporal features among the selected channel within one second from the 2D EEG feature matrix, and then use LSTM to extract emotional information accumulated over time. These shallow and deep information extracted by ELSTNN is integrated for emotion recognition.

4. Experiments

In this section, the experiments will be described from four aspects: the dataset of EEG emotion recognition, the implementation of ELSTNN, the experimental results, and the analysis of ELSTNN.

4.1. Datasets Introduction

4.1.1. DEAP Dataset

In EEG emotion recognition task, DEAP is widely used as a benchmark dataset [26]. There are 32 healthy subjects joining the experiment, half of whom are males. Every subject is required to watch 40 music videos, while a variety of physiological data of their are recorded, including 32-channel EEG signals. Before the video is played, the first three seconds of data are collected as the physiological state in the calm state. When the video is over, their emotional state is evaluated by themselves in five dimensions on a scale of 1 to 9. There is a short break to restore the calm emotional state. For 32-channel EEG signals, a sampling rate

Table 2
The subject-dependent experimental results of DEAP using ELSTNN and the strategy of 5-fold cross-validation. Here, the channels containing F3, C3, P7, PO3, Fp2, T8, CP6 and CP2 are used to construct 2D matrix EEG representation.

	Accuracy of Valence (%)			Accuracy of Arousal (%)		
subject	3 second	4 second	5 second	3 second	4 second	5 second
1	98.62	98.00	98.95	97.75	97.66	96.66
2	94.50	94.33	93.12	92.49	93.99	92.70
3	97.99	98.16	98.12	96.62	96.66	97.08
4	93.12	91.99	90.95	93.12	93.49	92.91
5	96.25	96.16	96.87	97.00	98.33	97.91
6	89.62	87.99	90.91	95.00	94.33	94.58
7	93.25	93.83	92.91	98.12	95.16	94.37
8	96.12	97.16	96.25	98.87	98.66	97.70
9	96.49	95.99	95.00	95.75	93.83	94.70
10	98.04	98.50	98.12	97.75	97.83	97.08
11	89.87	89.50	89.79	90.25	88.33	89.16
12	95.24	95.49	94.58	89.62	88.69	89.72
13	90.49	91.33	90.83	86.87	85.49	87.33
14	92.50	93.16	91.87	93.99	92.83	93.00
15	97.12	98.16	96.25	98.75	98.83	98.33
16	97.74	98.33	97.49	98.75	98.33	98.75
17	91.49	91.16	90.83	88.37	89.50	88.54
18	96.49	94.66	93.33	94.49	96.33	95.83
19	97.99	96.83	97.91	94.75	94.99	94.33
20	98.49	98.16	98.12	96.87	94.16	91.45
21	97.62	96.16	97.08	90.62	91.82	91.25
22	97.74	98.83	98.33	97.62	97.99	98.12
23	96.25	96.66	97.08	97.25	96.66	98.54
24	99.12	97.16	98.12	94.87	95.33	94.62
25	94.00	95.99	93.33	90.74	94.16	91.25
26	96.37	95.83	95.20	94.74	95.33	95.08
27	97.99	96.66	96.24	97.12	97.83	97.91
28	94.49	93.66	92.50	93.87	93.99	93.54
29	98.99	98.99	98.75	98.05	97.66	98.12
30	98.37	96.33	96.04	98.37	98.83	98.12
31	93.12	93.99	94.79	95.25	95.49	94.58
32	95.99	97.33	96.45	95.37	95.49	92.91
Mean	95.67	95.52	95.19	94.97	94.94	94.57

of 512 Hz is first used to collect these signals, followed by the sampling rate of 128Hz. To ensure the purity of EEG and the maximum correlation with the emotion, the EOG is removed and a bandpass frequency filter from 4.0–45.0 bands is used, because the neuroscience has discovered that the information of this band is related to the emotional activity. And more details of DEAP can be found in Table 1. However, in this paper, we refer to the emotional state as Russel's emotion definition [60], which is a two dimensions model assessing pleasure and activation. One is from unpleasure to pleasure called valence, while the other is from mild to intense called arousal. Hence, the five is set as the threshold for these two dimensions [61, 62], the low/high valence/arousal label (LV/HV, LA/HA) is defined as emotional label.

4.1.2. DREAMER Dataset

In DREAMER dataset [27], 14 males and 9 females are invited to participate in the emotion experiments. Each of them is asked to watch 18 film clips. As they watch each of clip, their ECG and 14-channel EEG are collected for emotion recognition. Before and after the film clip is played, the subject is asked to watch the same neutral film clip to return the neutral emotional state. During the above period, the multi-channel EEG signals are collected to represent the calm state. At the end of each clips, they evaluate their emotional state in three dimensions (arousal, valence and dominance) on a scale from 1 to 5. And EEG signals are recorded at the 128Hz sampling rate. And ECG is recorded at 256Hz sampling rate. The more detail of DREAMER is also shown in Table. 1. However, we use emotional definition proposed by Russel [60] as the final emotion definition (LV/HV, LA/HA). The threshold for these dimensions is set at 2.5 [63, 62].

4.2. Implementation

In process of data pre-processing, the EEG signal representing the calm emotional state is obtained by calculating

Table 3
In DEAP, the comparison results between ELSTNN and other works.

Method	Valence (%)	Arousal (%)
Luo et al [64]	78.17	73.89
Kim et al. [65]	78.72	79.03
Zheng et al. [66]	84.75	82.16
Liu et al. [36]	86.46	84.90
Piho et al. [67]	89.61	89.84
Yin et al [68]	90.45	90.60
Yang et al [54]	90.80	91.03
Ma et al. [69]	92.30	92.87
Tao et al. [53]	93.72	93.38
Huang et al. [37]	94.38	94.72
Xu et al. [70]	94.56	94.81
ELSTNN	**95.67**	**94.97**

the mean EEG signals in s seconds with a 1 second long non-overlapping sliding window. For DEAP, s is equal to the corresponding with the length of baseline signals. Then the EEG signals under stimulation minus the EEG signal representing the calm state using the sliding windows with no overlap. The above operation has been proven to improve emotional recognition performance [54]. Then, for the left and right hemispheric channel sets, ELSTNN is applied to channel selection by using the DE representation of processed EEG signal of each band. In ELSTNN, the redundant channels are eliminated to only one in each band by spectral clustering and all of the subject's train samples, while the representation of the selected channel is more emotionally discriminating. Since four emotion-related frequency bands are chosen and there are left and right channel sets, 8 channels are finally selected: F3, C3, P7, PO3, Fp2, T8, CP6, and CP2. Next, the 1D EEG signals are transformed into a 2D matrix representation to better capture the spatial relationship among EEG channels. Here, the effect of time length is considered on the temporal stability of the selected channels. Hence, the EEG signals are sliced using a non-overlapping sliding window of length s. Hence, the final EEG samples of one subject in DEAP can be marked as $X \in \mathbb{R}^{n \times s \times p \times r \times c}$, where r and c denote the length and width of a 2D representation, respectively. s and p denote the length of time and the sampling points of one second. And n denotes the number of samples. In DEAP, r, c and p are equal to 9, 9, and 128. And in the experiment, s is set to different values to mine the time stability of the channel, 3, 4, and 5. Hence, the value of n is changed with s, which is equal to 40 (trail) $\times$ 60 (length of one trail) / s.

For DREAMER, the same pre-processing operation is carried out, in which s is the last 3 seconds of the neutral film clip to ensure subject was in the calm state before stimulation. For a fair and effective comparison with other methods, the stimulus EEG signals of last 60 seconds of film clip are used for emotional analysis. Then ELSTNN is applied for channel selection using DE extracted from the channel of each hemisphere. Since four frequency bands are chosen, 8 channels are F3, FC5, T7, P7, T8, FC6, F4, and F8. Next, the 1D EEG signal is translated into a 2D matrix representation. Hence, the final EEG samples of one subject in DREAMER could be marked as $X \in \mathbb{R}^{n \times s \times p \times r \times c}$, in which r, c, and p are equal to 7, 7, and 128. And the value of s is set as 3, 4, and 5 to keep it the consistent with DEAP. Hence, n is equal to 18 (trail) $\times$ 60 (length of one trail) / s.

4.3. The Experimental Results

In this paper, we focus on the subject-dependent experiments, where the data coming from one subject is applied to train and test model. All the experiments use the 5-fold cross-validation as the experimental result to evaluate the performance of ELSTNN in LV/LH or LA/HA task, where the Adawn optimizer was applied to train all models with 100 epoch, 64 batch size, and 1e-4 learning rate.

The experimental results carried out in DEAP are shown in Table 2, which shows that for two different emotional states, our proposed method achieves good performance in different time windows. It also indicates the effectiveness of ELSTNN in long-time emotion recognition, and the selected channel has the time stability. Furthermore, to verify the superiority of ELSTNN in DEAP, various state-of-the-art methods [65, 64, 36, 54, 69, 68, 67, 53, 70, 66] are applied to compare with our proposed method. Luo et al. [64] use GAN-based data enhancement methods to augment EEG samples in the form of DE representation. Kim et al. [65] mine the complementation between electrocardiogram and EEG by deep learning model for emotion recognition. Liu et al. [36] use empirical mode decomposition to mine the EEG fragments with the most emotional information. Piho et al. [67] find out the EEG fragment containing the highest emotional information by the mutual information method for emotion recognition. Yin et al. [68] apply the graph convolutional neural networks to extract the deep channel correlation. Yang et al. [54] use the spatiotemporal information extracted by LSTM. Ma et al. [69] proposed a residual-based LSTM to learn the emotional information hidden in multimodal data. Tao et al. [53] use a deep model based on the attentional mechanism to mine the relationships among channels. Huang et al. [37] use the neuroscience discovery of emotional lateralization to build the new EEG representation. The detailed comparison result is shown in Table. 3. It shows that the performance of ELSTNN is better than other methods, which illustrates the effectiveness and superiority of ELSTNN.

For DREAMER, the experimental result is shown in Table 4. We find that ELSTNN obtains the excellent performance in the valence and arousal classification task. Furthermore, to verify the superiority of ELSTNN in DREAMER, our proposed method is compared with a large number of various state-of-the-art methods [73, 74, 45, 71, 72, 75, 66, 76]. Zhang et al. [75] propose an attention-based model to learn temporal dynamics of EEG. Yang et al. [72] use the multi-layer perceptron and decision tree to explore the emotional boundaries. Song et al. [45] use GCNN to learn the functional correlation between channels. Cheng et al. [73] mine the deep information from EEG signals by using the

Table 4
The subject-dependent experimental results of DREAMER using ELSTNN and the strategy of 5-fold cross-validation. Here, the channels containing F3, FC5, T7, P7, T8, FC6, F4, and F8 are used to construct 2D matrix EEG representation.

	Accuracy of Valence (%)			Accuracy of Arousal (%)		
subject	3 second	4 second	5 second	3 second	4 second	5 second
1	91.88	92.81	91.65	92.88	91.44	91.67
2	94.16	97.40	94.89	91.16	93.14	92.82
3	93.88	92.22	93.76	91.88	91.59	92.23
4	92.16	92.59	93.29	93.16	92.96	93.76
5	95.49	95.18	95.82	93.49	94.18	94.89
6	90.83	92.59	88.43	90.83	92.96	91.95
7	96.38	96.29	95.83	96.38	96.29	95.47
8	91.50	90.00	92.61	92.50	89.99	93.53
9	93.33	93.70	92.12	93.33	94.81	91.65
10	88.44	90.11	88.32	89.44	90.00	87.95
11	97.49	97.77	97.70	97.49	97.40	96.78
12	93.61	94.81	94.01	91.61	92.81	92.46
13	94.49	95.92	95.82	97.49	96.29	96.28
14	95.55	94.07	94.43	95.55	94.07	90.73
15	88.17	89.62	92.11	87.67	88.88	91.67
16	97.22	96.66	97.23	97.22	96.29	94.90
17	96.22	97.03	95.36	97.22	97.03	97.67
18	96.66	96.66	95.38	94.66	96.14	95.67
19	95.55	93.33	93.67	93.55	94.07	92.16
20	94.66	94.92	94.06	96.66	96.66	95.38
21	92.61	92.70	92.96	91.61	92.33	90.26
22	93.05	92.22	92.24	91.05	92.18	92.46
23	92.27	92.18	92.83	95.27	94.07	95.83
Mean	93.72	93.95	93.67	93.57	93.72	93.40

Table 5
In DREAMER, the comparison results between ELSTNN and other works.

Method	Valence (%)	Arousal (%)
Zheng et al. [71]	56.65	70.30
Zheng et al. [66]	81.55	80.23
Yang et al. [72]	84.54	84.84
Song et al. [45]	89.59	88.93
Cheng et al. [73]	89.03	90.41
Liu et al. [74]	90.57	88.99
Zhang et al. [75]	92.27	93.03
ELSTNN	**93.72**	**93.57**

deep forest. Liu et al. [74] use the deep canonical correlation analysis for emotion recognition. Zheng et al. [71] automatically select channels by canonical correlation analysis during the supervised learning. The detailed comparison result is shown in Table. 4.2, which shows that the superiority of our proposed method in emotion recognition task. Through the above experiments, we can find that ELSTNN could achieve the excellent performance in the valence and arousal task, and it has more outstanding performance.

4.4. Model Analysis

In this part, ELSTNN is analyzed comprehensively to understand how the method models the phenomenon of emotional lateralization and applies it to emotion recognition.

First, the 2D matrix EEG representations of DEAP before and after channel selection by ELCCS are visualized, in which the first, the middle, and the last frames of the EEG representation of different emotional states are chosen to understand the channel changes in times. The visualization result of subject 1 of DEAP is shown in Fig. 6. Here, the 8 channels containing F3, C3, P7, PO3, Fp2, T8, CP6, and CP2 are selected by ELSTNN, while the 32 channels denote all channels used to collect EEG signals in DEAP. It is easy to observe the phenomenon that for the voltage change of channels, most of the channels located on the left side are higher than those on the right side in the state of LV and LA, while most of the channels situated on the left side are more passive than those on the right side in the state of HV and HA. And this phenomenon has continuity in time for participants receiving the same emotional stimulus. The important thing is that this phenomenon can be well represented by our proposed 2D matrix representation of 32 channels, which explains the reasons about the increasing number of channels can improve the EEG recognition performance to a certain extent. However, the selected channel chosen by ELSTNN could still preserve the information of the left and right differences. Furthermore, it should be noted that most of these selected channels are located at the boundary of the left and right difference and record EEG signals from different brain regions, which means that the emotional activities are associated with the inter-connect of

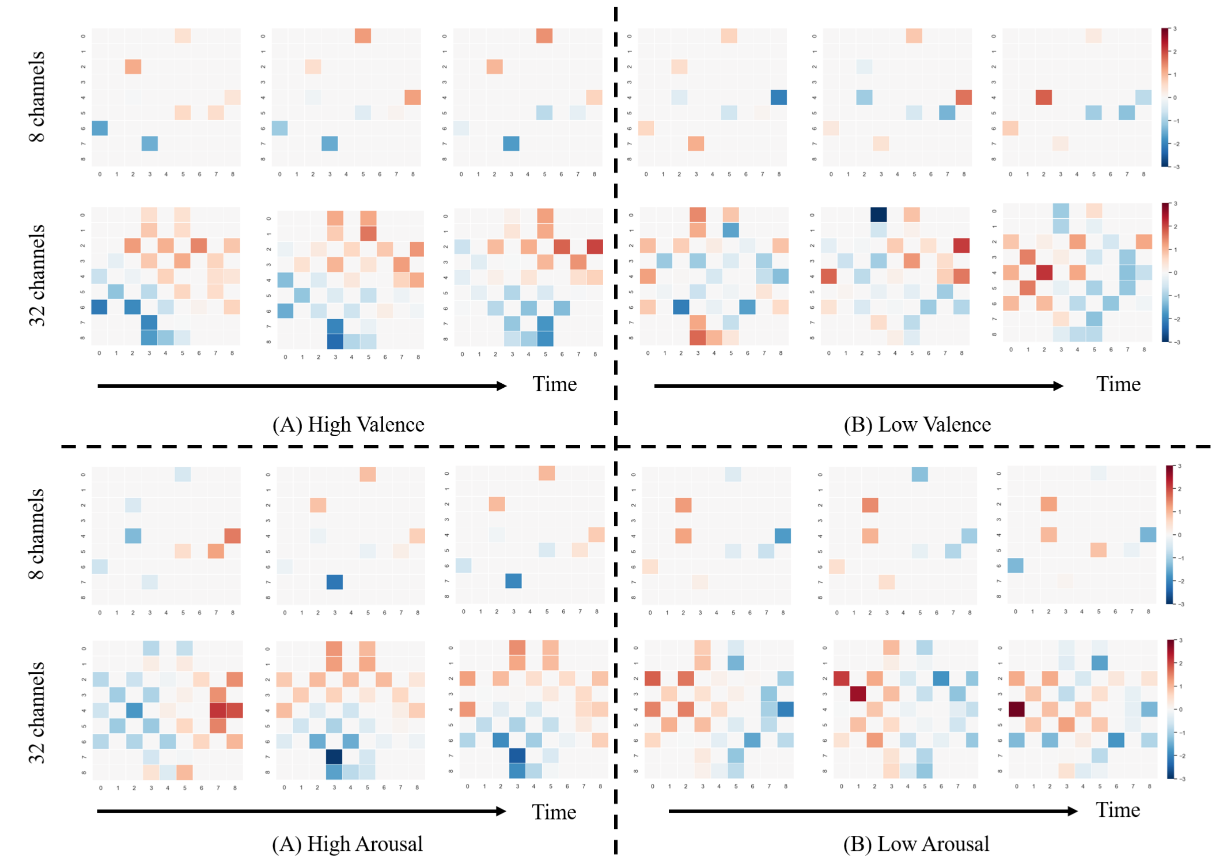

Figure 6: The visualization of the input matrix representing multi-channel EEG signals. Each subgraph is the 2D matrix representation consisting of 8, 32 channels. And the first, the middle, and the last frame of the sample is shown to indicate the effect of stimulus duration on the channel.

boundaries between brain regions. Hence, it is necessary to explore the functional connections between brain regions to further improve the performance, which is also the basis of recent research [77, 78]. In addition, the visualization is consistent with the neuroscientific findings [79, 80] that the left hemisphere is more dynamic in HV state and the right hemisphere is higher in LV state. In summary, the 2D matrix EEG representation with the channels selected by ELSTNN could preserve the differential information between the left and right hemispheres and eliminate redundant and highly related EEG channels. As a result, it could be applied for channel selection to provide more discriminative and low-dimension features to recognize various emotional states under the guidance of neuroscientific findings.

Furthermore, to explore the mutual effect between the different number of channels and our proposed model, an additional experiment is carried out on DEAP. In this experiment, 32 channels (all EEG channels in DEAP), 16 channels, 12 channels, and 8 channels are used to convert multi-channel 1D EEG signals into the 2D representation, respectively, where all channel subsets of DEAP are structured by ELSTNN. 16 channels are composed of two channels selected from the clustering effect of each of four bands of each hemisphere, in which these channels are in different clusters of different emotion samples. And according to the previous studies, high-frequency channel information is more related to emotional activity, especially β band, and γ band. Hence, the formation process of 12 channels is as follows: one channel with the above effect is selected in the clustering of each of the two low-frequency bands of each hemisphere, and two channels are selected in each of the two high-frequency bands of each hemisphere. After that, the four-channel combinations are combined with STNN, respectively, to mine the effect of different channel numbers on the model performance by visualizing the relationship between the loss and the accuracy of train and test. And the window size is set as 3 seconds. As shown in in Fig. 7, we can find that the performance of STNN could converge to a relatively stable range with different numbers, which means that a large number of repetitive or even redundant EEG channel features cannot improve the emotion recognition performance. Furthermore, our proposed method could eliminate redundant EEG information and improve the performance of emotion recognition using less information of channel when facing different emotional states.

Through all experiments and the above analysis, the conclusions are summarized that our proposed method could screen out various redundant EEG channels while preserving the differential boundary between the two hemispheres to form emotionally discriminative representation for emotion recognition. Furthermore, by comparing with other outstanding methods, the ELSTNN is effective and superior

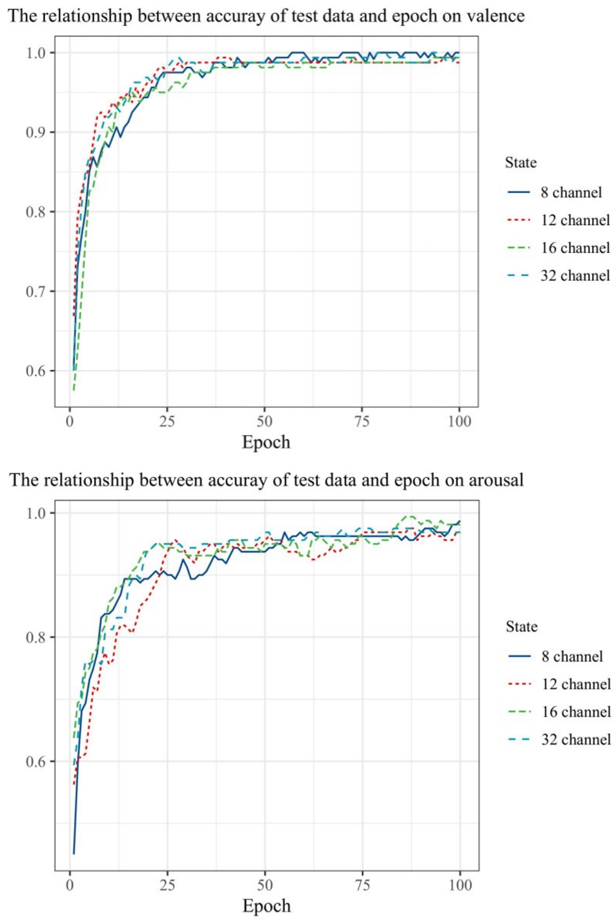

Figure 7: The relationship between the accuracy of test data and epoch under the emotional state of valence and arousal. 32 channel denotes all channels recording EEG signals in DEAP. The rest is the subset of 32 channels using ELSTNN to constitute.

for emotion recognition. The validity of ELSTNN can be summarized as follows: (1) with the guidance of neuroscience finding about emotional lateralization, ELSTNN can remove the highly related channels while preserving the lateralization information. (2) the 3D spatiotemporal representation is constructed with characteristics of sparsity and emotional discrimination. (3) ELSTNN explores the deep spatiotemporal feature from 3D representation for EEG emotion recognition.

However, there are some limitations in our work that need to be addressed in future research. Firstly, our study primarily focuses on binary emotion classification, failing to consider the complexity of multi-dimensional emotional states. Emotions are not simply high or low; they encompass a range of categories, including happiness, sadness, anger, and more. Future research should explore multi-class emotion recognition to capture the diversity of human emotions more comprehensively. Secondly, our experiment adopts a subject-dependent design aimed at exploring the emotional response characteristics of the selected channels, laying the foundation for the development of personalized models while also providing a priori feasibility analysis for subsequent cross-participant validation. In the future, it will be essential to expand our research to a subject-independent setting to further validate the robustness of our findings.

5. Conclusion

In this paper, we propose a channel selection method named ELSTNN for EEG emotion recognition, where the highly related but redundant channels are removed in an unsupervised learning method, and the neural phenomenon of emotional lateralization is modelled to improve EEG emotion recognition performance. The extensive experiments are driven on two public dataset DEAP and DREAMER. Our proposed method obtains a mean accuracy of 95.67% and 94.97% in valence and arousal on DEAP, while achieving 93.72% and 93.57% in valence and arousal on DREAMER. The experimental results demonstrate that the performance of ELSTNN has achieved a better performance than most state-of-the-art methods, and the selected channels have the time stability. With the visualization process and the comprehensive analysis, the good performance of our proposed method can be attributed to three reasons: (1) Neuroscientific phenomena is applied in channel selection process to remove the highly related and redundant channels while retaining discriminative feature. To our best knowledge, this is the first time that the combination of neuroscience phenomena and unsupervised learning has been used for EEG channel selection. (2) The sparse 3D EEG representation is formed with more spatiotemporal information by converting view of EEG according to the 10-20 system. (3) With the respective advantages of CNN and LSTM, ELSTNN is employed to explore and integrate the deep spatiotemporal information for emotion recognition. This work is especially useful for the channel design of BCI, as it reduces the number of required EEG channels while maintaining high recognition accuracy, thereby enhancing the practicality and accessibility of EEG devices in real-world scenarios.

Acknowledgments

The authors would like to respect and thank all reviewers for their constructive and helpful review. This research is funded by the National Natural Science Foundation of China (62372283, 62206163), Science and Technology Major Project of Guangdong Province(STKJ2021005, STKJ202209002, STKJ2023076), Natural Science Foundation of Guangdong Province (2024A1515010239), and Key Project of Colleges and Universities in Guangdong Province (Natural Science) (2024ZDZX1019).

References

[1] Zhang, H., Zhou, Q.Q., Chen, H., Hu, X.Q., Li, W.G., Bai, Y., Han, J.X., Wang, Y., Liang, Z.H., Chen, D., et al., 2023. The applied principles of eeg analysis methods in neuroscience and clinical neurology. Military Medical Research 10, 67.

[2] Yan, S., Hu, Y., Zhang, R., Qi, D., Hu, Y., Yao, D., Shi, L., Zhang, L., 2024. Multilayer network-based channel selection for motor imagery brain–computer interface. Journal of Neural Engineering 21, 016029.

[3] García-Martínez, B., Martinez-Rodrigo, A., Alcaraz, R., Fernández-Caballero, A., 2018. A review on nonlinear methods using electroencephalographic recordings for emotion recognition. IEEE Transactions on Affective Computing , 1–1doi:10.1109/TAFFC.2018.2890636.

[4] Li, Z., Zhang, G., Wang, L., Wei, J., Dang, J., 2023. Emotion recognition using spatial-temporal eeg features through convolutional graph attention network. Journal of Neural Engineering 20, 016046.

[5] Samal, P., Hashmi, M.F., 2024. Role of machine learning and deep learning techniques in eeg-based bci emotion recognition system: a review. Artificial Intelligence Review 57, 1–66.

[6] Wang, T., Mao, R., Liu, S., Cambria, E., Ming, D., 2025. Explainable multi-frequency and multi-region fusion model for affective brain-computer interfaces. Information Fusion 118, 102971.

[7] Homan, R.W., Herman, J., Purdy, P., 1987. Cerebral location of international 10–20 system electrode placement. Electroencephalography and Clinical Neurophysiology 66, 376–382.

[8] Becker, H., Fleureau, J., Guillotel, P., Wendling, F., Merlet, I., Albera, L., 2020. Emotion recognition based on high-resolution eeg recordings and reconstructed brain sources. IEEE Transactions on Affective Computing 11, 244–257. doi:10.1109/TAFFC.2017.2768030.

[9] Wang, H., Wu, X., Yao, L., 2020. Identifying cortical brain directed connectivity networks from high-density eeg for emotion recognition. IEEE Transactions on Affective Computing , 1–1doi:10.1109/TAFFC.2020.3006847.

[10] Zhan, C., She, D., Zhao, S., Cheng, M.M., Yang, J., 2019. Zero-shot emotion recognition via affective structural embedding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[11] Xu, X., Wei, F., Zhu, Z., Liu, J., Wu, X., 2020. EEG feature selection using orthogonal regression: Application to emotion recognition, in: ICASSP, pp. 1239–1243.

[12] Qu, G., Wang, F., Bi, J., An, Z., Hu, F., Li, H., Han, S., 2024. A hybrid critical channel selection framework for eeg emotion recognition. IEEE Sensors Journal 24, 14881–14893.

[13] Chen, J., Hu, B., Moore, P., Zhang, X., Ma, X., 2015. Electroencephalogram-based emotion assessment system using ontology and data mining techniques. Applied Soft Computing 30, 663–674. doi:https://doi.org/10.1016/j.asoc.2015.01.007.

[14] Wang, X.W., Nie, D., Lu, B.L., 2011. Eeg-based emotion recognition using frequency domain features and support vector machines, in: Lu, B.L., Zhang, L., Kwok, J. (Eds.), Neural Information Processing, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 734–743.

[15] Xu, X., Jia, T., Li, Q., Wei, F., Ye, L., Wu, X., 2021. Eeg feature selection via global redundancy minimization for emotion recognition. IEEE Transactions on Affective Computing , 1–1doi:10.1109/TAFFC.2021.3068496.

[16] Zhang, R., Nie, F., Li, X., Wei, X., 2019. Feature selection with multi-view data: A survey. Information Fusion 50, 158–167. doi:https://doi.org/10.1016/j.inffus.2018.11.019.

[17] Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517.

[18] Estevez, P.A., Tesmer, M., Perez, C.A., Zurada, J.M., 2009. Normalized mutual information feature selection. IEEE Transactions on Neural Networks 20, 189–201. doi:10.1109/TNN.2008.2005601.

[19] van den Broek, S., Reinders, F., Donderwinkel, M., Peters, M., 1998. Volume conduction effects in eeg and meg. Electroencephalography and Clinical Neurophysiology 106, 522–534. doi:https://doi.org/10.1016/S0013-4694(97)00147-8.

[20] Lin, X., Chen, J., Ma, W., Tang, W., Wang, Y., 2023. Eeg emotion recognition using improved graph neural network with channel selection. Computer Methods and Programs in Biomedicine 231, 107380.

[21] Cai, H., Liu, X., Ni, R., Song, S., Cangelosi, A., 2023. Emotion recognition through combining eeg and eog over relevant channels with optimal windowing. IEEE Transactions on Human-Machine Systems .

[22] Hejrati, B., Fathi, A., Abdali-Mohammadi, F., 2017. Efficient lossless multi-channel eeg compression based on channel clustering. Biomedical Signal Processing and Control 31, 295–300. doi:https://doi.org/10.1016/j.bspc.2016.08.024.

[23] Odabaee, M., Tokariev, A., Layeghy, S., Mesbah, M., Colditz, P.B., Ramon, C., Vanhatalo, S., 2014. Neonatal eeg at scalp is focal and implies high skull conductivity in realistic neonatal head models. NeuroImage 96, 73–80. doi:https://doi.org/10.1016/j.neuroimage.2014.04.007.

[24] Dimond, S.J., Farrington, L., Johnson, P., 1976. Differing emotional response from right and left hemispheres. Nature 261, 690–692.

[25] Zatorre, R.J., Jones-Gotman, M., Evans, A.C., Meyer, E., 1992. Functional localization and lateralization of human olfactory cortex. Nature 360, 339–340.

[26] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. Deap: A database for emotion analysis ;using physiological signals. IEEE Transactions on Affective Computing 3, 18–31. doi:10.1109/T-AFFC.2011.15.

[27] Katsigiannis, S., Ramzan, N., 2017. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. IEEE journal of biomedical and health informatics 22, 98–107.

[28] Picard, R.W., 2000. Affective computing. MIT press.

[29] Zhou, S., Huang, D., Liu, C., Jiang, D., 2022. Objectivity meets subjectivity: A subjective and objective feature fused neural network for emotion recognition. Applied Soft Computing 122, 108889.

[30] Liu, H., Lou, T., Zhang, Y., Wu, Y., Xiao, Y., Jensen, C.S., Zhang, D., 2024. Eeg-based multimodal emotion recognition: A machine learning perspective. IEEE Transactions on Instrumentation and Measurement .

[31] Alarcão, S.M., Fonseca, M.J., 2019. Emotions recognition using eeg signals: A survey. IEEE Transactions on Affective Computing 10, 374–393. doi:10.1109/TAFFC.2017.2714671.

[32] Adolphs, R., Anderson, D.J., 2018. The neuroscience of emotion: A new synthesis. Princeton University Press.

[33] Petrantonakis, P.C., Hadjileontiadis, L.J., 2009. Emotion recognition from eeg using higher order crossings. IEEE Transactions on information Technology in Biomedicine 14, 186–197.

[34] Duan, R.N., Zhu, J.Y., Lu, B.L., 2013. Differential entropy feature for eeg-based emotion classification, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE. pp. 81–84.

[35] Zhang, G., Yu, M., Liu, Y.J., Zhao, G., Zhang, D., Zheng, W., 2021. Sparsedgcnn: Recognizing emotion from multichannel eeg signals. IEEE Transactions on Affective Computing .

[36] Liu, Z.T., Xie, Q., Wu, M., Cao, W.H., Li, D.Y., Li, S.H., 2018. Electroencephalogram emotion recognition based on empirical mode decomposition and optimal feature selection. IEEE Transactions on Cognitive and Developmental Systems 11, 517–526.

[37] Huang, D., Chen, S., Liu, C., Zheng, L., Tian, Z., Jiang, D., 2021. Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for eeg emotion recognition. Neurocomputing 448, 140–151.

[38] Hu, G., Xin, Y., Lyu, W., Huang, H., Sun, C., Zhu, Z., Gui, L., Cai, R., Cambria, E., Seifil, H., 2025. Recent trends of multimodal affective computing: A survey from NLP perspective. arXiv preprint arXiv:2409.07388 .

[39] Ma, Y., Zhang, D., Li, S., Cambria, E., Zhou, G., 2025. UECO: Unified editing chain for efficient appearance transfer with multimodality-guided diffusion. Expert Systems with Applications 270, 126510.

[40] Wang, R., Guo, C., Cambria, E., Rida, I., Yuan, H., Piran, M.J., Feng, Y., Zhu, X., de Compiegne, M., 2025. CIME: Contextual interaction-based multimodal emotion analysis with enhanced semantic information. The Journal of Supercomputing .

[41] Xiao, L., Mao, R., Zhang, X., He, L., Cambria, E., 2024. Vanessa: Visual connotation and aesthetic attributes understanding network for multimodal aspect-based sentiment analysis, in: EMNLP, pp. 11486–11500.

[42] Liang, B., Gui, L., He, Y., Cambria, E., Xu, R., 2024. Fusion and discrimination: A multimodal graph contrastive learning framework for multimodal sarcasm detection. IEEE Transactions on Affective

Computing 15, 1874–1888.
[43] Zhang, H., Meng, Z., Luo, M., Han, H., Liao, L., Cambria, E., Fei, H., 2025. Towards multimodal empathetic response generation: A rich text-speech-vision avatar-based benchmark, in: Proceedings of WWW.
[44] Zhang, T., Zheng, W., Cui, Z., Zong, Y., Li, Y., 2018. Spatial–temporal recurrent neural network for emotion recognition. IEEE transactions on cybernetics 49, 839–847.
[45] Song, T., Zheng, W., Song, P., Cui, Z., 2020. Eeg emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing 11, 532–541. doi:10.1109/TAFFC.2018.2817622.
[46] Khare, S.K., Bajaj, V., 2020. Time-frequency representation and convolutional neural network-based emotion recognition. IEEE transactions on neural networks and learning systems .
[47] Liang, Z., Zhou, R., Zhang, L., Li, L., Huang, G., Zhang, Z., Ishii, S., 2021. Eegfusenet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional eeg with an application to emotion recognition. arXiv preprint arXiv:2102.03777 .
[48] Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., Pfurtscheller, G., 2007. A fully automated correction method of eog artifacts in eeg recordings. Clinical Neurophysiology 118, 98–104. doi:https://doi.org/10.1016/j.clinph.2006.09.003.
[49] Wu, X., Xu, X., Liu, J., Wang, H., Hu, B., Nie, F., 2021. Supervised feature selection with orthogonal regression and feature weighting. IEEE Transactions on Neural Networks and Learning Systems 32, 1831–1838. doi:10.1109/TNNLS.2020.2991336.
[50] Zabihi, M., Kiranyaz, S., Ince, T., Gabbouj, M., 2013. Patient-specific epileptic seizure detection in long-term eeg recording in paediatric patients with intractable seizures. IET Conference Proceedings , 7.06–7.06(1)URL: https://digital-library.theiet.org/content/conferences/10.1049/cp.2013.2060.
[51] Handiru, V.S., Prasad, V.A., 2016. Optimized bi-objective eeg channel selection and cross-subject generalization with brain–computer interfaces. IEEE Transactions on Human-Machine Systems 46, 777–786. doi:10.1109/THMS.2016.2573827.
[52] Nakisa, B., Rastgoo, M.N., Tjondronegoro, D., Chandran, V., 2018. Evolutionary computation algorithms for feature selection of eeg-based emotion recognition using mobile sensors. Expert Systems with Applications 93, 143–155. doi:https://doi.org/10.1016/j.eswa.2017.09.062.
[53] Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., Chen, X., 2020. Eeg-based emotion recognition via channel-wise attention and self attention. IEEE Transactions on Affective Computing .
[54] Yang, Y., Wu, Q., Qiu, M., Wang, Y., Chen, X., 2018. Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–7.
[55] Xu, X., Chen, X., Zhang, Y., 2018. Removal of muscle artefacts from few-channel eeg recordings based on multivariate empirical mode decomposition and independent vector analysis. Electronics Letters 54, 866–868.
[56] Zhang, R., Nie, F., Li, X., 2018. Feature selection under regularized orthogonal least square regression with optimal scaling. Neurocomputing 273, 547–553. doi:https://doi.org/10.1016/j.neucom.2017.07.064.
[57] Hagen, L., Kahng, A.B., 1992. New spectral methods for ratio cut partitioning and clustering. IEEE transactions on computer-aided design of integrated circuits and systems 11, 1074–1085.
[58] Li, Y., Wang, L., Zheng, W., Zong, Y., Qi, L., Cui, Z., Zhang, T., Song, T., 2021. A novel bi-hemispheric discrepancy model for eeg emotion recognition. IEEE Transactions on Cognitive and Developmental Systems 13, 354–367. doi:10.1109/TCDS.2020.2999337.
[59] Li, J., Zhang, Z., He, H., 2018. Hierarchical convolutional neural networks for eeg-based emotion recognition. Cognitive Computation 10, 368–380.
[60] Russell, J.A., 1980. A circumplex model of affect. Journal of personality and social psychology 39, 1161.
[61] Dhara, T., Singh, P.K., Mahmud, M., 2024. A fuzzy ensemble-based deep learning model for eeg-based emotion recognition. Cognitive Computation 16, 1364–1378.
[62] Singh, M.K., Singh, M., et al., 2023. A deep learning approach for subject-dependent & subject-independent emotion recognition using brain signals with dimensional emotion model. Biomedical Signal Processing and Control 84, 104928.
[63] Yuvaraj, R., Thagavel, P., Thomas, J., Fogarty, J., Ali, F., 2023. Comprehensive analysis of feature extraction methods for emotion recognition from multichannel eeg recordings. Sensors 23, 915.
[64] Luo, Y., Lu, B.L., 2018. Eeg data augmentation for emotion recognition using a conditional wasserstein gan, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE. pp. 2535–2538.
[65] Kim, B.H., Jo, S., 2020. Deep physiological affect network for the recognition of human emotions. IEEE Transactions on Affective Computing 11, 230–243. doi:10.1109/TAFFC.2018.2790939.
[66] Zheng, W., Pan, B., 2024. A spatiotemporal symmetrical transformer structure for eeg emotion recognition. Biomedical Signal Processing and Control 87, 105487.
[67] Piho, L., Tjahjadi, T., 2018. A mutual information based adaptive windowing of informative eeg for emotion recognition. IEEE Transactions on Affective Computing 11, 722–735.
[68] Yin, Y., Zheng, X., Hu, B., Zhang, Y., Cui, X., 2021. Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm. Applied Soft Computing 100, 106954.
[69] Ma, J., Tang, H., Zheng, W.L., Lu, B.L., 2019. Emotion recognition using multimodal residual lstm network, in: Proceedings of the 27th ACM international conference on multimedia, pp. 176–183.
[70] Xu, F., Pan, D., Zheng, H., Ouyang, Y., Jia, Z., Zeng, H., 2024. Eescn: A novel spiking neural network method for eeg-based emotion recognition. Computer Methods and Programs in Biomedicine 243, 107927.
[71] Zheng, W., 2017. Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis. IEEE Transactions on Cognitive and Developmental Systems 9, 281–290. doi:10.1109/TCDS.2016.2587290.
[72] Yang, Y., Wu, Q., Fu, Y., Chen, X., 2018. Continuous convolutional neural network with 3d input for eeg-based emotion recognition, in: International Conference on Neural Information Processing, Springer. pp. 433–443.
[73] Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A., Chen, X., 2020. Emotion recognition from multi-channel eeg via deep forest. IEEE Journal of Biomedical and Health Informatics 25, 453–464.
[74] Liu, W., Qiu, J.L., Zheng, W.L., Lu, B.L., 2019. Multimodal emotion recognition using deep canonical correlation analysis. arXiv preprint arXiv:1908.05349 .
[75] Zhang, D., Yao, L., Chen, K., Monaghan, J., 2019. A convolutional recurrent attention model for subject-independent eeg signal analysis. IEEE signal processing letters 26, 715–719.
[76] Cambria, E., Howard, N., Hsu, J., Hussain, A., 2013. Sentic blending: Scalable multimodal fusion for continuous interpretation of semantics and sentics, in: IEEE SSCI, pp. 108–117.
[77] Li, P., Liu, H., Si, Y., Li, C., Li, F., Zhu, X., Huang, X., Zeng, Y., Yao, D., Zhang, Y., et al., 2019. Eeg based emotion recognition by combining functional connectivity network and local activations. IEEE Transactions on Biomedical Engineering 66, 2869–2881.
[78] Wu, X., Zheng, W.L., Lu, B.L., 2020. Investigating eeg-based functional connectivity patterns for multimodal emotion recognition. arXiv preprint arXiv:2004.01973 .
[79] Sonoh, S., Aou, S., Horio, K., Tamukoh, H., Koga, T., Yamakawa, T., 2010. A human robot interaction by a model of the emotional learning in the brain. Journal of Automation Mobile Robotics and Intelligent Systems 4, 48–54.
[80] Harmon-Jones, E., Gable, P.A., Peterson, C.K., 2010. The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. Biological psychology 84, 451–462.