


RESEARCH

Open Access



EEG emotion recognition based on differential entropy feature matrix through 2D-CNN-LSTM network

Teng Wang¹, Xiaoqiao Huang^{1,2*} , Zenan Xiao¹, Wude Cai^{1*} and Yonghang Tai^{1,2}

*Correspondence:
hxq@ynnu.edu.cn;
caiwude@sohu.com

¹ School of Physics and Electronic Information, Yunnan Normal University, Kunming 650500, Yunnan, China

² Yunnan Key Lab of Optic-Electronic Information Technology, Kunming 650500, Yunnan, China

Abstract

Emotion recognition research has attracted great interest in various research fields, and electroencephalography (EEG) is considered a promising tool for extracting emotion-related information. However, traditional EEG-based emotion recognition methods ignore the spatial correlation between electrodes. To address this problem, this paper proposes an EEG-based emotion recognition method combining differential entropy feature matrix (DEFM) and 2D-CNN-LSTM. In this work, first, the one-dimensional EEG vector sequence is converted into a two-dimensional grid matrix sequence, which corresponds to the distribution of brain regions of the EEG electrode positions, and can better characterize the spatial correlation between the EEG signals of multiple adjacent electrodes. Then, the EEG signal is divided into equal time windows, and the differential entropy (DE) of each electrode in this time window is calculated, it is combined with a two-dimensional grid matrix and differential entropy to obtain a new data representation that can capture the spatiotemporal correlation of the EEG signal, which is called DEFM. Secondly, we use 2D-CNN-LSTM to accurately identify the emotional categories contained in the EEG signals and finally classify them through the fully connected layer. Experiments are conducted on the widely used DEAP dataset. Experimental results show that the method achieves an average classification accuracy of 91.92% and 92.31% for valence and arousal, respectively. The method performs outstandingly in emotion recognition. This method effectively combines the temporal and spatial correlation of EEG signals, improves the accuracy and robustness of EEG emotion recognition, and has broad application prospects in the field of emotion classification and recognition based on EEG signals.

Keywords: EEG signal, Differential entropy feature matrix, Deep learning, Emotion recognition

1 Introduction

Human emotion recognition plays an important role in human–computer interaction and has become an important research field in cognitive science, computer science, psychology, and other fields [1]. It is also considered a hot topic in neuroscience and artificial intelligence research because emotions are affective states that accompany cognition and awareness and have a crucial role in human social interaction.

Observing the external and internal reactions of humans can infer their emotional state, as different emotional states elicit different responses. At present, emotion recognition research methods include non-physiological signals and physiological signals, and non-physiological signals include facial expressions [2], speech [3], and body movements [4]. Physiological signals include electrocardiogram (ECG) signals [5], electromyogram (EMG) [5] signals, electrooculogram (EOG) [6] signals, and electroencephalogram (EEG) [7] signals. Compared with non-physiological signals [4], physiological [4] signals are not easily affected by external factors and subjective intentions, thus increasing the reliability and objectivity of the experiment. In recent years, with the progress of sensor technology, it has become possible to monitor, record, and analyze multi-channel neuro-physiological signals. EEG, as a non-invasive brain electrophysiological technology, only needs to place electrodes on the scalp, which is relatively safe and has a wide range of applications. At the same time, EEG signals are real-time and can monitor changes in the brain's electrical activity in real-time, which makes EEG the focus of many researchers. Therefore, the potential application scenarios of emotion recognition have become a hot topic in the research field, and more and more people pay attention to them [8–11].

In the field of emotion recognition, the analysis of EEG signals is widely used to understand an individual's emotional state. Emotions are complex and multi-dimensional experiences, often manifested in dynamic changes in time and spatial features. Time-domain features include frequency analysis, amplitude, and waveform shape, which can reflect the activity state of the brain at different time points. Alazrai et al. [12] introduced a novel emotion recognition method based on EEG, which uses an innovative time–frequency feature extraction technique. Specifically, the study uses a quadratic time–frequency distribution (QTFD) to establish a high-resolution representation of the time–frequency characteristics of EEG signals to effectively capture the spectral changes of EEG signals on the time axis, and experiments show that the average classification accuracy of their proposed method is between 73.8% and 86.2%. Li et al. [13] proposed an innovative approach a multi-domain adaptive graph convolutional network (MD-AGCN) that utilizes differential entropy (DE) as a feature extraction method. It cleverly integrates knowledge from both the frequency domain and the time domain to fully explore complementary information within EEG signals. Extensive experiments demonstrate that the introduced method consistently achieves excellent results across various experimental settings. Time-domain analysis is a key part of EEG research, and spatial information also provides valuable information for research. EEG signals are recorded on the scalp through an electrode array, forming a spatial topological structure. Each electrode corresponds to a specific region of the brain, so spatial features provide important information about the distribution of emotions in the brain. For example, Li et al. [14] fully considered the spatial information of EEG. The method of hierarchical neural network is used to classify emotions, and the classification results are good. Song et al. [15] proposed a novel dynamic graph convolutional neural network (DGCNN) to mine the spatial relationship of multi-channel EEG data. Tao et al. [16] proposed an attention-based convolutional recurrent neural network (ACRNN), which assigns different weights to different channels to make full use of channel information and improve the accuracy of emotion recognition. Combining time-domain information can help to capture the changing trend of emotion in time while combining spatial information can

model the difference of emotion expression in different parts so that the model can better adapt to the dynamic change of emotion expression. Zhang et al. [17] proposed a new deep learning framework called spatiotemporal recursive neural network (STRN). It captures remote context clues by traversing regions of space in different directions along each time slice. Subsequently, RNN layer learning is used to represent the time-dependent discriminant features of the generated sequences. Experimental results on datasets show that the proposed method achieves high classification performance. Rudakov et al. [18] proposed an innovative emotion recognition model, the multitask convolutional neural network (MT-CNN), which takes brain maps generated from EEG as input and outputs emotion states related to arousal and valence. Experimental results demonstrate that the proposed approach achieves high classification performance.

In the emotion recognition problem, most of the existing methods are based on machine learning, commonly used are support vector machine (SVM) [19] and k-nearest neighbors (KNN) [20]. With the increasing penetration of deep learning algorithms into various fields, deep learning has become a popular method for studying emotion recognition due to its superior performance and remarkable achievements. In recent years, several outstanding algorithms have been applied to emotion recognition, such as deep belief networks (DBN) [21], convolutional neural networks (CNN) [22], graph convolutional neural networks (GCNNs) [23], and capsule networks (CapsNet) [24]. Hwang et al. [25] compared with traditional LSTM networks, using information from the past and future biological signals to more effectively assign weights for emotion recognition under the current LSTM cell state, and integrating ant colony optimization (ACO) to find the optimal combination of features among many, thereby enhancing performance. Alhagry et al. [26] applied LSTM algorithms, extracting features from EEG signals, and finally performing classification through fully connected layers. Their method achieved average accuracies of 85.65% and 85.45% for arousal and valence classification, respectively, on the DEAP dataset. Tripathi et al. [27] ingeniously combined modern techniques such as dropout and linear units with CNN and classified pre-processed EEG data. Through extensive experiments on the DEAP dataset, the results indicated classification accuracies of 81.41% for emotion and 73.35% for arousal. Additionally, Song et al. [28] leveraged the significant advantages of CNN in graphics to classify multi-channel EEG signals for emotion recognition. By training and classifying using publicly available datasets, they achieved accuracies of 86.23% for valence, 85.54% for arousal, and 85.02% for dominant emotion classification.

From the literature, it is evident that applying deep learning for emotion recognition outperforms traditional machine learning methods. However, deep learning offers numerous advantages, two challenges need to be addressed. First, the common method of EEG classification processing is to extract features in the time domain, time-frequency domain, and spatial domain, and then use machine learning or deep learning to classify. Applying CNN to time-domain data often reveals features related to the frequency domain [29]. However, this method does not take into account the information characteristics of different frequency bands and the interrelationship of spatial information between different electrode channels. Second, applying CNN to the temporal dimension for extracting temporal features allows simultaneous extraction of spatiotemporal features. However, long time-series data, containing a wealth of information, may

pose challenges for traditional CNN structures, as they are prone to issues such as vanishing or exploding gradients. There is limited research that effectively combines both aspects.

To better integrate spatiotemporal features, a method of EEG signal characterization based on differential entropy feature matrix (DEFM) is proposed, and deep learning models will be used, especially the hybrid model combining 2D-CNN and LSTM. Two-dimensional-CNN is used for feature extraction in space to capture the relationship between different electrodes, while LSTM can effectively prevent the problem of gradient disappearance or gradient explosion. By combining these two structures, the model can better understand the overall context and more accurately identify patterns and laws in spatiotemporal sequences, reducing the number of parameters in the overall model and reducing the computational burden.

The main contributions of this paper are as follows:

- A new feature extraction method called differential entropy feature matrix (DEFM), based on differential entropy and spatial feature matrix, has been proposed. According to the relative positions of 32 electrodes in brain space, we construct a 9×9 feature matrix, which helps analyze the influence of electrode position on emotion. At the same time, we divided the 60-s EEG of each subject into 120 times windows of equal length of 0.5 s and calculated the DE of 32 electrodes in each time window. In this way, 2D images of each time window could be obtained, and the spatial and spectral information of the EEG signal could be captured by this method.
- We propose a 2D-CNN-LSTM network model for emotion classification. Two-dimensional-CNN can automatically extract features from the above 2D images through convolution continuously, and finally input them into LSTM through the connection layer, and make use of LSTM's advantages in learning time series for continuous training. Finally, emotion classification is carried out by the connection layer.
- To verify the effect of the proposed method on emotion classification, we conducted a large number of experiments on the DEAP dataset. The experimental results show that the average accuracy of valence and arousal is 91.92% and 92.31%, respectively. Therefore, our proposed method has a high classification effect in emotion classification.

The rest of this paper is organized as follows. In Sect. 2, we introduce the datasets and proposed method in detail. In Sect. 3, we report experiments and results.

2 Materials and methods

2.1 The overall framework of the proposed methodology

The general framework of the proposed method is shown in Fig. 1 and is divided into three steps in total:

Step 1 Preprocessing of EEG signals. Identification and processing of outliers and noise in the data used.

Step 2 Feature extraction. According to the relative positions of the electrodes in the brain distribution, the one-dimensional EEG vector sequence is converted into a two-dimensional network matrix sequence, to better represent the spatial correlation

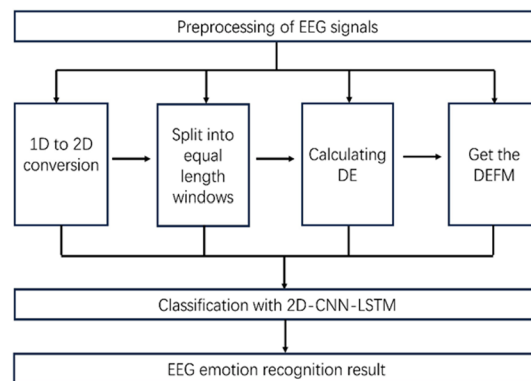


Fig. 1 The overall framework of the proposed methodology

Table 1 The details of the DEAP dataset

Attribute	Description
Subject	32 (16 males and females)
Stimulant	40 one-minute videos
Trials of each subject	40 times
Length of each trial	63 s (3-s pre-trial and 60 trial)
Recorded EEG signals	32 electrode positions
Rating scales	Arousal, valence, dominance, and liking
EEG data	Array shape $40 \times 32 \times 8064$ (trial \times channel \times data)
Labels	Array shape 40×4 (trial \times label)

between the electrodes. Then, a whole EEG signal is divided into several equal time windows using a sliding window, and the DE of each electrode in the time period is calculated, and the DEFM is obtained by combining the DE and the time window.

Step 3 Classification with 2D-CNN-LSTM. Two-dimensional-CNN-LSTM combines the advantages of CNN automatic feature extraction and LSTM which can better handle time series to achieve better classification results.

2.2 Dataset and preprocessing

This paper verifies the effectiveness of the proposed method based on the DEAP dataset [30]. The DEAP dataset is a large-scale open-source dataset containing physiological signals such as electroencephalography developed by a research team at Queen Mary University of London. The details of the DEAP are shown in Table 1. The dataset consisted of 32 brain electrical channels and eight channels that recorded other physiological signals caused by music videos of different emotional tendencies. In particular, 32 subjects watched 40 stimulus videos, recorded EEG signals at a sampling frequency of 512 Hz, and then down-sampled them to 128 Hz. After the viewing, 1–9 consecutive values were used to evaluate arousal, efficacy, preference, dominance, and familiarity. Forty of the stimulation videos were composed of three seconds of resting time and 60 s of video. In this paper, only EEG signals are used, so 32 channels of EEG data are selected to record.

To better identify emotions, arousal and valence are selected. We choose a threshold of 5, according to the evaluation value of these two indicators, if the evaluation value is greater than or equal to 5, it is marked as high arousal (HA) and high valence (HV), if it is less than 5, it is marked as low arousal (LA) and low valence (LV).

First, the EEG data are down-sampled, reducing the sampling rate to 128 Hz. To further filter out noise and eliminate artifacts, EEG data are bandpass filtered and restricted to a frequency range of 4–45 Hz.

2.3 Feature extraction

Differential entropy (DE) [31] is a concept in information theory used to measure the uncertainty of a random variable. In EEG research, DE is used to analyze the complexity and randomness of EEG signals, which has some advantages. At the same time, the one-dimensional EEG vector sequence is transformed into a two-dimensional network matrix sequence. Then, the whole EEG signal is divided into multiple time windows by sliding window. The DE of each electrode in the time window is calculated.

2.3.1 Differential entropy

DE is a method of measuring the uncertainty of random variables that can be used to describe the random nature of probability density functions. Similar to discrete entropy, differential entropy is also a non-negative real number, but it can be infinite, which is related to the infinity of continuous variables. Differential entropy has a wide range of applications in information theory, statistics, machine learning, and other fields, such as density estimation, source coding, channel coding, probability density estimation in machine learning, and other issues. For a continuous random variable x , its probability density function is $p(x)$, then its differential entropy calculated as shown in Eq. (1):

$$DE = - \int_a^b p(x) \log(p(x)) dx \quad (1)$$

where $p(x)$ represents the probability density function of the continuous signal $[a, b]$, it represents the interval of information value. For a signal of a specific length, the differential entropy calculation formula of an EEG with an approximate Gaussian distribution $N(\sigma^2)$ is shown in Eq. (2):

$$\begin{aligned} DE &= - \int_a^b \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \log \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \right) dx \\ &= \frac{1}{2} \log(2\pi e\sigma_i^2) \end{aligned} \quad (2)$$

Since DEAP includes baseline data of 3 s, which do not record any information, the data of 3 s are removed to avoid the impact on the EEG signal. In this work, we denoised the 60-s EEG signals of 32 subjects, respectively, and then divided the EEG signals into 120 equal small time windows, each time period of 0.5 s. In our experiment, this 0.5-s time window was mainly studied. Then, the DE of the 0.5-s time window is calculated according to Eq. (2), and the DE calculated in each time window is taken as the feature to form a feature vector.

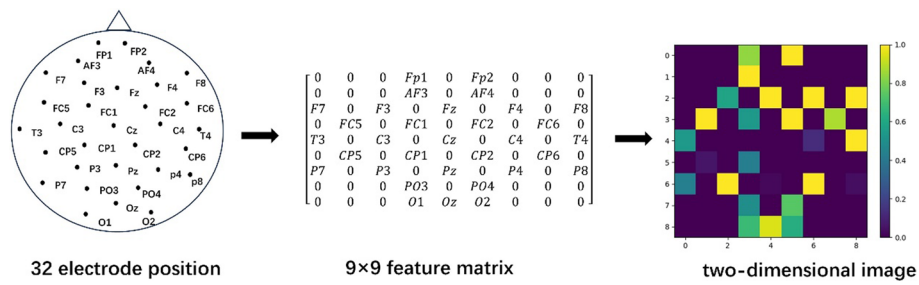


Fig. 2 Two-dimensional image flowchart

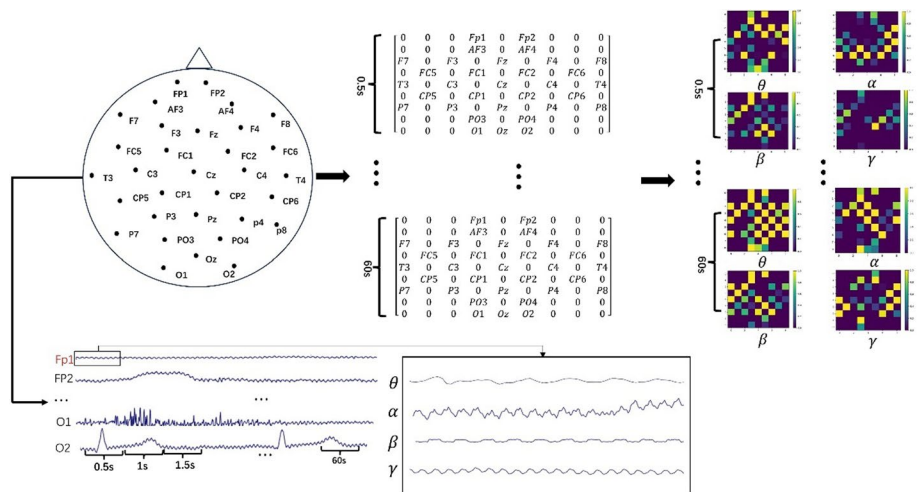


Fig. 3 Data processing flowchart

2.3.2 Two-dimensional EEG mesh feature conversion method based on DE

In order to better integrate the time-domain and spatial information of EEG signals, we will extract DE features from 32 channels to form a feature matrix, as shown in Fig. 2. Specifically, based on the relative distribution of 32 electrodes in the brain, we obtain a 9×9 feature matrix of the electrode distribution on a two-dimensional plane, where the positions of no electrodes are set to 0, and these 0's do not play any role in our experiment. Then, we perform normalization calculation on the DE value and get the right-most two-dimensional color image, on the right side of the color image is a color bar, you can see the relationship between the values of DE and color, different DEs have different colors.

To better describe the proposed method, we chose one of the subjects as an example, and the whole process is shown in Fig. 3. Taking the Fp1 channel as an example, the 60-s EEG signal is divided into 120 time windows, and each time window is 0.5 s. Four frequency bands $\theta(4 \leq \theta < 8 \text{ Hz})$, $\alpha(8 \leq \alpha < 15 \text{ Hz})$, $\beta(15 \leq \beta < 32 \text{ Hz})$, and $\gamma(32 \leq \gamma < 45 \text{ Hz})$ were, respectively, extracted from 120 time windows. At the same time, 9×9 color graphs of four frequency bands were obtained according to the method shown in Fig. 2. Fp1 could obtain 120×4 sample numbers. Therefore, a sample feature matrix of dimension $32 \times 32 \times 120 \times 4$ is generated in this experiment. Where

the first 32 represents the number of subjects, the last 32 represents the number of electrodes, 120 represents the number of time windows, and 4 represents the number of frequency bands.

Since the calculation of DE can lead to the presence of outliers, which might affect the performance of the model, it is necessary to normalize the feature matrices for each participant. These matrices should be scaled to a range between 0 and 1. The normalization is performed using Eq. (3) [32].

$$F' = \frac{F_{\max} - F}{F_{\max} - F_{\min}} \tag{3}$$

During this process, we first normalize the feature values using Eq. (3), where F represents the original feature value, F_{\max} and F_{\min} represent the maximum and minimum feature values, respectively, and F' represents the normalized feature value. After normalization, we gather the features of 32 channels in the same frequency band for each sample and construct a submatrix following the mapping rule illustrated in Fig. 1. The submatrix contains the average DE values of each corresponding channel, while the elements that correspond to channels without corresponding electrodes are set to 0 by default.

The feature extraction method adopted in this paper combines time–frequency and spatial features to provide richer EEG change information, which can be used to classify emotional states. Time-domain information provides insights about the dynamic changes in emotions, while spatial information provides insights about how emotions are distributed in the brain. Therefore, the method can provide more comprehensive information on changes in EEG signals, which can better classify different emotional states.

2.4 Fusion model of 2D-CNN and LSTM for emotion recognition

The overall architecture of CNNs and LSTMs for emotion recognition is shown in Fig. 4, it contains the CNN layer, LSTM layer, and dense layer. We employ a 2D-CNN to capture spatial features from each two-dimensional matrix of EEG data. Subsequently,

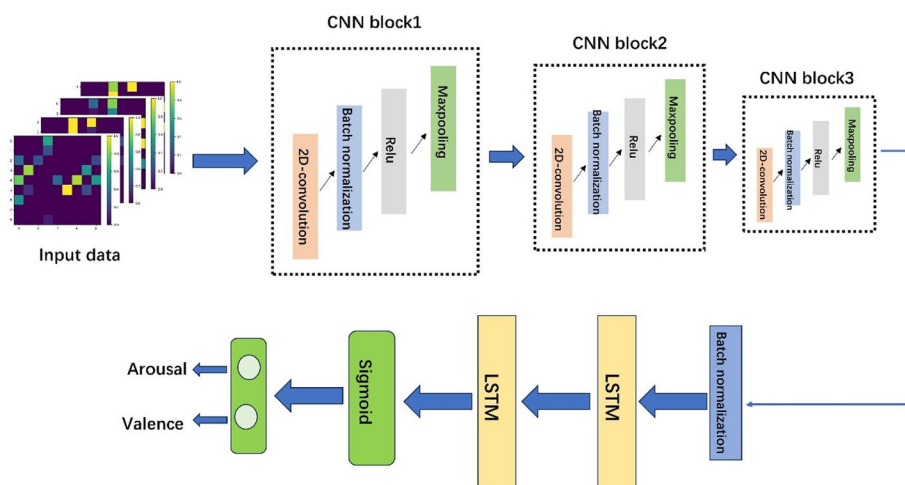


Fig. 4 Two-dimensional-CNN-LSTM combined model

these extracted spatial feature sequences are fed into an LSTM to further capture the temporal features of the EEG data. Following this, we utilize a fully connected layer to receive the output of the LSTM network at the last time step, thereby forming a feature vector. Lastly, this feature vector is passed through a dense connection layer (Dense) for the final emotion classification.

2.4.1 Two-dimensional-CNN

Two-dimensional-CNN refers to a 2D convolutional neural network. It is a deep learning neural network structure, which is widely used in computer vision tasks, such as image recognition, object detection, semantic segmentation, etc.

In a two-dimensional convolutional neural network, the input data are usually a two-dimensional image or video, and each input datum is represented as a matrix or tensor. The neural network processes the input data through the structures of the convolutional layer, pooling layer, and full connection layer to learn and extract image features.

The convolutional layer is the core component of the two-dimensional convolutional neural network. It uses a set of learnable convolution checks to carry out convolution operations on input data, to extract different features. The calculation formula is shown in Eq. (4). The pooling layer is used to down-sample the feature graphs output by the convolutional layer to reduce the dimension and computation amount of feature graphs while preserving important features. The fully connected layer is used to flatten the feature map output by the pooling layer and match it with the label to get the final prediction result.

$$N = (W - F + 2P)/S + 1 \quad (4)$$

where N is the output size, W is the input size, F is the convolution kernel size, P is the filling value size, and S is the step size.

The 2D-CNN structure has the advantages of hierarchical, automatic feature extraction, and multi-level feature learning, so it is widely used in computer vision tasks and has achieved excellent results in many application fields.

It should be noted that the calculation of these two stages is usually limited by the accuracy requirements, so we need to make certain adjustments and optimizations to improve the accuracy and efficiency of the algorithm.

2.4.2 LSTM

Inspired by the human brain, the LSTM uses selective input and selective forgetting mechanisms, introducing three “gate” structures (forgetting gate, input gate, and output gate) to control the flow of information in the form of filters. Through this mechanism, LSTMs can selectively retain and update past information while also remembering current information to better capture long-term dependencies in sequence data. The LSTM structure is shown in Fig. 5

The LSTM takes in three components as its input: the current moment input X_t , the previous moment's output value $h_{(t-1)}$ of the LSTM, and the state of the unit $C_{(t-1)}$. It provides two types of outputs: the current moment's LSTM output value H_t and the cell state C_t . The input gate calculation as shown in Eq. (5), which controls the amount of input information, the forget gate controls the amount of historical information retained,

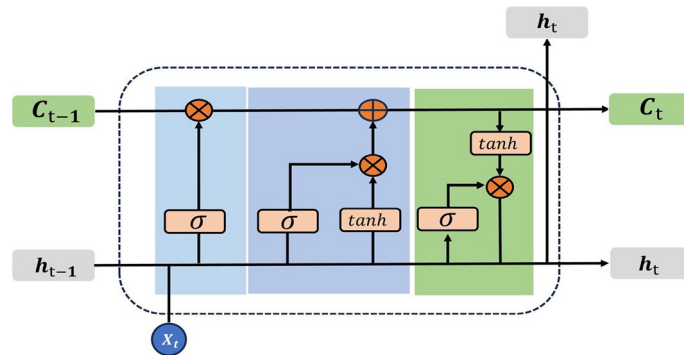


Fig. 5 LSTM unit

and the output gate calculation as shown in Eq. (6), which controls information from the current unit state is to be output to the current hidden state. The gates can be adjusted adaptively according to the network's needs to achieve better results.

input door:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

output door:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

Among them, $\sigma(\cdot)$ is a sigmoid function that outputs a value between 0 and 1. If the value of f is close to 0, the information will be forgotten, and the information close to 1 will be retained. When the LSTM network forgets part of the previous state information, it needs to absorb new memory from the current memory to fill the blank, and this process is realized by the input gate. At this time, the input gate will filter the input information, select some current information to enter the current cell state with a certain probability, and together with the forgetting gate, it selectively updates the current cell state with a certain probability for the current information and the information at the previous time.

The LSTM network replaces the neuron in the ordinary recurrent neural network with the above gating structure and effectively saves the historical information to help the current decision-making. The emergence of the LSTM network effectively overcomes the problems of gradient disappearance, gradient explosion, and other problems in the learning process of neural networks. The LSTM neural network is a logic unit with a "gate" structure added to each neuron, so that the error direction propagation can directly pass through the "gate," thus avoiding the gradient disappearance and explosion in the error reverse propagation so that the gradient of the LSTM network in the transmission process remains relatively stable and will not disappear completely.

2.4.3 Two-dimensional-CNN-LSTM

The 2D-CNN-LSTM model structure was proposed in this paper, as shown in Fig. 4. To better extract EEG signal features, we input the DEFM features in the four bands of θ , α , β , and γ extracted in Sect. 2.3 into the network 2D-CNN-LSTM in the form of

$N \times 9 \times 9 \times 4$, where N refers to the number of samples sent to the model each time, in our experiment, N selected the number of 10 samples, 9×9 refers to the size of the two-dimensional matrix, and 4 refers to the four frequency bands under study. These inputs contain temporal and spatial information of EEG data. Based on 2D-CNN feature extraction (time and space), LSTM is used to further extract time-series features. Finally, the output of the last time point of the LSTM network is received through the fully connected layer, and the feature vector is generated, and then, the feature vector is fed to the SoftMax layer for the final emotion classification. Combined with the temporal and spatial characteristics of EEG signals, this method improves the effect of emotion recognition.

Specifically, the 2D-CNN in the hybrid model is mainly composed of three convolution layers and three pooled layers, where each convolution layer has a convolution kernel of 32, 64, and 128, respectively, and is optimized using the ReLu activation function and the Adam optimizer, with a learning rate of 0.0005. In the hybrid model, the LSTM model contains two hidden layers with 64 and 128 neurons, respectively, and finally, it has 0.1 dropouts to prevent overfitting, and finally a fully connected layer with 258 neurons. The LSTM network is used to further calculate the relevant characteristics of EEG fragments in the time domain, making the features extracted by the model more objective and accurate.

2.5 Evaluation indices

To demonstrate the performance of the proposed method, there are several metrics commonly used to evaluate the quality of algorithms. Below are a few common and important evaluation metrics:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Accuracy is the most commonly used and intuitive metric, calculated as shown in Eq. (7) [33]. Here, TP represents the number of samples correctly identified as low arousal/negative valence emotions by the classification model (referred to as positives); TN represents the number of samples correctly identified as high arousal/positive valence emotions by the classification model (referred to as negatives); FP represents the number of samples where negative valence emotions are incorrectly classified as positive valence emotions; and FN represents the number of samples where positive valence emotions are incorrectly predicted as negative valence emotions.

The precision is calculated as shown in Eq. (8) [33]:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

The recall is calculated as shown in Eq. (9) [33]:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

The F-score, also known as the F1-score, is a harmonic average of precision and recall used to comprehensively evaluate the performance of a classification model, especially in unbalanced datasets. The F-score is calculated as shown in Eq. (10) [34]:

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

Precision represents accuracy, while recall indicates the recall rate. As seen from the calculation formula, the F1-score ranges between 0 and 1. The closer it is to 1, the better the model's performance, showcasing superior classification effectiveness. Furthermore, when precision and recall simultaneously achieve higher values, the corresponding F1-score also increases. The role of the F1-score is to strikingly balance precision and recall. This is particularly crucial when dealing with uneven sample distributions, where the impact of the F1-score becomes more pronounced.

The Kappa coefficient primarily measures the effectiveness of a classifier using statistical methods. Its characteristic lies in its thorough consideration of the model's randomness and continuous enhancement of the accuracy of random classification. It can be employed to assess the consistency of classification tasks. The calculation method for Kappa is shown in Eq. (11) [35]:

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (11)$$

P_o refers to the observed accuracy, which can be obtained by summing the diagonal elements of the confusion matrix. P_e refers to the accuracy of random classification in a completely random state. It can be calculated by summing the product of the true label frequency and the predicted label frequency for each category in the confusion matrix. From the calculation formula, it is evident that when the Kappa coefficient is closer to 1, the model's classification performance is better.

3 Results and discussion

3.1 Experiments results

In this study, each subject had 120×40 samples, 120 being the time window and 40 being the stimulus video, so there are a total of 4800 samples per subject. The network model is cross-verified by tenfold.

The 9th subject was selected to adjust the network parameters because subject 9 had a more uniform distribution of labels. We used the 2D-CNN-LSTM network to investigate the effect of the number of 2D-CNN convolutional layer and LSTM hidden layer cells on emotion classification. Through experiments, we find that the number of convolutional kernels has the greatest influence on the network model compared with the number of hidden layer cells. When the number of hidden layer cells is 2 and the number of convolutional kernels is 3, the accuracy of the network model is the highest. The network model parameter of 2D-CNN-LSTM is shown in Table 2.

To better assess the model's performance, tenfold cross-validation [36] is chosen for evaluation. In tenfold cross-validation, the dataset is divided into ten subsets, and the model is trained and evaluated ten times. In each iteration, nine subsets are used for training, and the remaining subset is used for validation. This process repeats to ensure that each subset serves as the validation set exactly once. Ultimately, the model's performance evaluation is the average of these ten validation results. This approach is effective in better assessing the model's generalization ability, reducing the risk of overfitting or underfitting, and identifying optimal hyperparameter

Table 2 Parameter settings

2D-CNN-LSTM model parameters	Values
The number of LSTM hidden layer cells	2
The number of 2D-CNN convolution layer	3
Model error threshold ϵ	0.8
Model learning rate γ	0.0005
2D-CNN activation function type	Sigmoid
LSTM activation function type	Sigmoid
Cross verification	Tenfold

settings. According to the above parameter settings, 32 subjects were tested, respectively, and the results are shown in Table 3.

As shown from Table 3, the classification result of our proposed method is relatively high. The average accuracy of 32 subjects in the arousal classification was 92.31%; the average accuracy of 32 subjects in the valence classification was 91.92%; the average F-score of 32 subjects in the arousal classification was 90.75%; and the average F-score of 32 subjects in the valence classification was 92.31%;

The average Kappa of 32 subjects in the arousal classification was 91.76%; the average Kappa of 32 subjects in the valence classification was 92.36%.

Accuracy is the most intuitive index to measure the performance of the model, and it is also the most important index. As shown in Fig. 6, in the 2D-CNN-LSTM model proposed in this paper, as the number of iterations increases, the training accuracy of arousal and valence classification accuracy to 97.2% and 96.8%, respectively. Meanwhile, the test accuracy also keeps improving, and finally, the test accuracy of arousal and valence classification accuracy to 92.31% and 91.92%, respectively. The results show that the proposed method has good flexibility and is effective in the field of emotion recognition.

The confusion matrix is an indicator to evaluate the model results and is part of the model evaluation. In addition, a confusion matrix is often used for judgment classifiers and is suitable for data models of different types. Figure 7 shows the confusion matrix of arousal and valence. From the results of the confusion matrix, about 7.6% of the valence was wrongly classified, and about 7.2% of the arousals were wrongly classified. The overall classification effect was good.

The ROC curve, also known as the receiver operating characteristic curve, is a popular visual metric used to evaluate the performance of binary classification models. It typically plots the true-positive rate (TPR) on the y-axis and the false-positive rate (FPR) on the x-axis. The TPR represents the proportion of actual positive cases that the model correctly identifies, while the FPR represents the proportion of actual negative cases that are incorrectly classified as positive. The shape of the ROC curve reflects the overall performance of the model, with curves closer to the upper-left corner indicating better performance. The classification accuracy of the model proposed in this paper is relatively high, as shown from Fig. 8.

Table 3 Two-dimensional-CNN-LSTM classification result

Subject	Binary classification of valence			Binary classification of arousal		
	Accuracy (%)	F-score (%)	Kappa (%)	Accuracy (%)	F-score (%)	Kappa (%)
01	92.16	89.73	88.89	90.23	90.36	91.33
02	90.14	90.56	91.02	90.58	91.22	91.27
03	91.33	89.77	90.59	91.36	90.88	92.47
04	90.65	90.25	91.76	91.27	91.88	88.87
05	91.58	90.27	91.01	89.77	90.28	89.67
06	89.89	90.21	91.02	92.11	91.29	92.02
07	91.35	91.89	92.03	90.23	91.22	90.78
08	91.54	92.32	92.01	86.54	89.67	88.73
09	91.39	92.86	90.89	91.36	92.24	91.69
10	91.11	92.25	92.42	91.50	93.51	92.61
11	91.58	92.19	92.54	91.54	90.32	91.29
12	90.51	91.40	91.12	90.38	90.88	90.71
13	91.15	92.89	91.25	90.33	90.58	91.58
14	89.91	92.43	90.78	92.38	90.05	91.11
15	92.63	91.97	92.18	92.56	91.78	92.82
16	90.91	91.88	91.61	92.24	91.89	92.99
17	91.58	91.54	92.56	92.69	91.38	91.72
18	91.56	92.69	92.65	92.58	93.11	92.89
19	91.67	91.91	92.68	90.78	91.32	91.96
20	90.94	91.37	91.82	91.84	92.77	92.65
21	93.01	91.56	92.22	92.49	91.88	91.84
22	89.68	90.28	92.28	91.64	91.56	91.77
23	91.37	91.36	91.28	91.21	91.85	91.29
24	89.89	91.76	92.88	91.59	90.68	91.21
25	90.16	91.58	90.36	90.71	91.36	92.21
26	90.83	92.49	91.34	90.88	90.21	91.06
27	91.56	90.99	91.56	89.78	89.88	90.09
28	90.81	91.21	92.02	90.36	91.26	91.28
29	92.02	91.25	92.45	91.03	91.45	91.17
30	91.06	90.85	91.49	90.88	90.87	92.85
31	91.33	91.54	92.55	91.02	90.65	88.68
32	90.88	91.09	92.35	91.36	90.88	89.75
Average	91.92	91.85	92.36	92.31	90.75	91.76

3.2 Recognition performance of different time windows

Since the length of the EEG signal determines the different emotional information it contains, this section will focus on the influence of different time windows on emotion recognition performance. We select four different time window periods, namely, $N \in [0.5, 0.2, 0.8, 1.0]$. The table shows the classification results of emotion recognition for valence and arousal under four different time windows. Table 4 shows that the recognition performance of the CNN-LSTM model is optimal when the time window is 0.5 s. The average recognition rates of valence and arousal classification were 91.9% and 92.3%, respectively. Compared with the other three times windows, when N is 0.5 s, the classification accuracy is increased by 4.19%, 5.99%, and 3.31% compared with 0.2 s, 0.4 s, and 1.0 s, respectively.

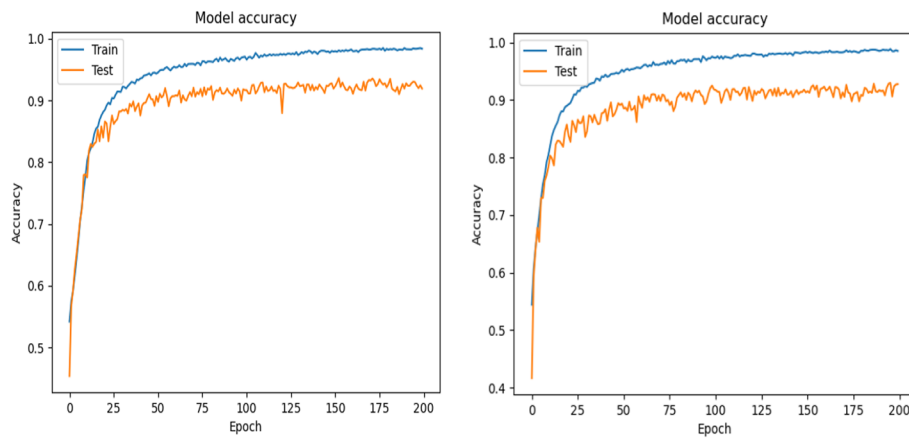


Fig. 6 Arousal classification accuracy (left) and valence classification accuracy (right)

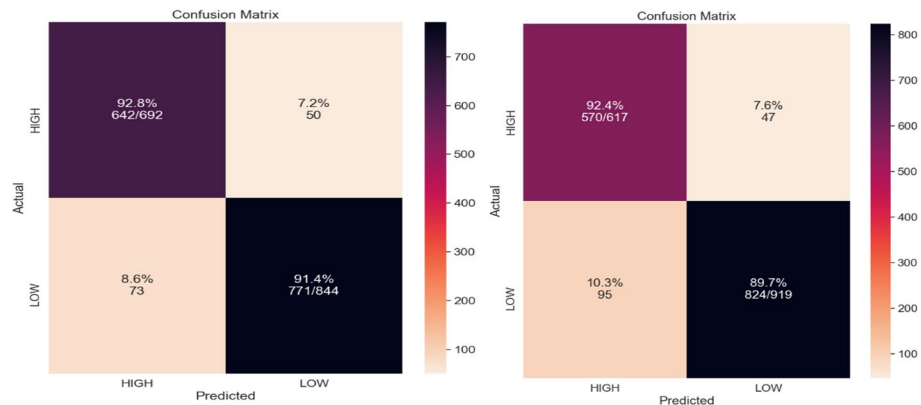


Fig. 7 Arousal confusion matrix (left) and valence confusion matrix (right)

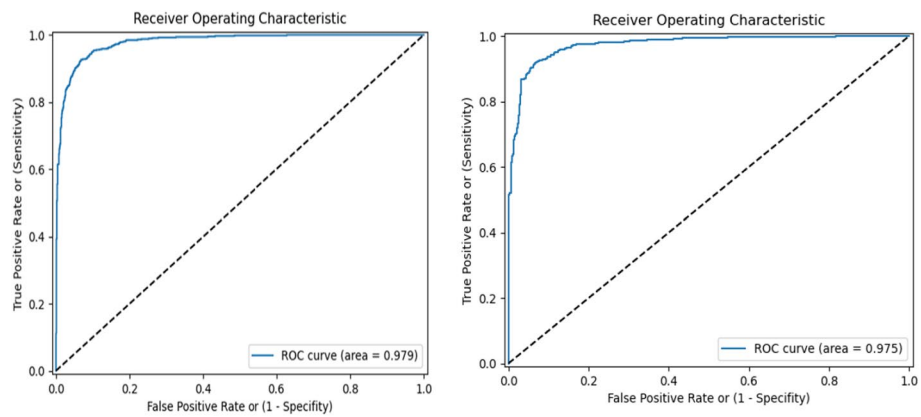


Fig. 8 Arousal ROC curve (left) and valence ROC curve (right)

3.3 Compared with other methods

To better show the superior performance of our proposed method, Table 5 lists the comparison between the 2D-CNN-LSTM network model proposed in this paper and other

Table 4 Two-dimensional-CNN-LSTM performance in different time windows (%)

N	Valence	Arousal
0.2	87.23 ± 1.52	85.11 ± 2.7
0.8	85.43 ± 2.36	86.32 ± 4.3
1.0	88.11 ± 1.2	87.97 ± 2.3
0.5	91.9 ± 0.43	92.3 ± 0.6

Table 5 Emotion recognition performance of different models

Reference index	Feature extraction methods	Classifier used	Database	Accuracy(%)
Yoon et al. [37]	FFT-based spectral power features extracted from EEG rhythms	Bayesian	DEAP	Arousal: 0.709 Valence: 0.701
Arnau-González et al. [38]	Spectral power, energy, and connectivity features	SVM	DREAMER	Arousal: 0.862 Valence: 0.854
Gupta et al. [39]	Information potential feature extracted in the FAWT domain of EEG signal	Random forest	DEAP	Arousal: 0.714 Valence: 0.799
Gupta et al. [40]	Graph-theoretic-based EEG features	RVM	DEAP	Arousal: 0.67 Valence: 0.69
Cheng et al. [41]	3D cube evaluated from EEG segment	CNN	DEAP	Arousal: 0.894 Valence: 0.904
Soleymani et al. [42]	EEG power	SVM	MAHNOB-HCI	Arousal: 0.52 Valence: 0.57
S. Katsigiannis, et al. [43]	Power spectral density-based features extracted from EEG signal	SVM	DREAMER	Arousal: 0.624 Valence: 0.621
Zhang et al. [44]	Temporal slices obtained from each channel EEG signal	Recurrent attention model	DREAMER/DDEAP	Arousal: 0.855 Valence: 0.836
Yin et al. [45]	EEG's differential entropy	ERDL	DEAP	Arousal: 0.848 Valence: 0.852
Topic et al. [46]	TOPO-FM	CNN + SVM	DEAP	Arousal: 0.806 Valence: 0.857
Liu et al. [47]	DE, statistical features	DCAA	DEAP	Arousal: 0.843 Valence: 0.856
Yang et al.[48]	Row signals	PCRNN	DEAP	Arousal: 0.913 Valence: 0.908
Gao et al.[49]	Time domain and frequency domain	SVM	DEAP	Arousal: 0.752 Valence: 0.805
Our method	EEG's differential entropy	2D-CNN-LSTM	DEAP	Arousal: 0.919 Valence: 0.923

network models. It can be seen from the experimental results that the proposed method has higher classification accuracy than other models. This is due to the comprehensive consideration of the spatial positioning of electrodes and its effect on emotion recognition; at the same time, time, space, and frequency information are extracted from EEG signals, where CNN learns the spatial characteristics of the two-dimensional grid data at each sampling point. LSTM further captures the global temporal dynamics between the

continuous sampling points in EEG samples, thus realizing the potential value of feature exploration and achieving higher recognition accuracy in binary classification. Therefore, the proposed feature fusion method has a strong spatiotemporal representation. The temporal and spatial characteristics of fusion significantly improved the accuracy of emotion recognition in EEG, and the accuracy of arousal and valence was more than 91%.

4 Conclusion

In this paper, we propose a method of EEG emotion recognition based on DEFM and 2D-CNN-LSTM. DEFM is a DE feature vector method for EEG signal characterization, which considers the time, space, and frequency of the EEG signal. The method converts the original one-dimensional chain channel information into two-dimensional grid spatial information, corresponding to the brain region distribution of EEG electrode positions, and effectively characterizes the spatial correlation between multiple adjacent electrodes in the physics of EEG signal. A time window is used to segment the two-dimensional grid sequence into equal-length time segments, which is a new data representation integrating the spatiotemporal correlation of EEG. In addition, an end-to-end, trainable hybrid deep neural network model for EEG emotion recognition is proposed, which combines 2D-CNN and LSTM networks to capture the spatial correlation of data between physically adjacent electrodes and the temporal dependence of EEG data streams. The model was evaluated for potency and arousal using 32 subjects in a large-scale DEAP dataset to evaluate the performance of the EEG spatiotemporal feature representation and the proposed hybrid deep learning model. The experimental results show that the average accuracy in valence and arousal is 91.92% and 92.31%, respectively, which is significantly better than the most advanced methods. Although our proposed method effectively combines the spatiotemporal correlation of EEG and improves the accuracy and robustness of EEG emotion recognition, it also has certain limitations, such as significant differences between EEG signals and other spatiotemporal information between different individuals. Future research can further carry out cross-paradigm, cross-device, and cross-population research on EEG emotion recognition.

Acknowledgements

We appreciate the editors and reviewers who processed and reviewed our manuscript to provide detailed professional comments on the technical contributions, logical structure, and content presentation of this paper.

Author contributions

TW was involved in conceptualization, data curation, validation, and writing—original draft preparation. ZNX and WDC were responsible for methodology, software, visualization, and writing—original draft preparation. HXQ and YHT contributed to funding acquisition, resources, supervision, and writing—reviewing and editing.

Funding

This work is supported by the National Natural Science Foundation of China (62266053, 62062070, 62365017, 62062069, and 62005235), the Natural Science Foundation of Yunnan Province (202101AT070100), and Yunnan Expert Workstation (No. 202305AF150012).

Availability of data and materials

Parts of the models, data, and codes that support the study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors approved the final manuscript and the submission to this journal.

Competing interests

The authors declare that they have no competing interests.

Received: 29 September 2023 Accepted: 2 April 2024

Published online: 08 April 2024

References

1. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor, Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* **18**(1), 32–80 (2001)
2. R. Adolphs et al., Recognition of facial emotion in nine subjects with bilateral amygdala damage. *Neuropsychologia* **37**, 1111–1117 (1999)
3. M. Chatterjee, D.J. Zion, M.L. Deroche, B.A. Burianek, C.J. Limb, A.P. Goren, A.M. Kulkarni, J.A. Christensen, Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Res.* **2015**(322), 151–162 (2015). <https://doi.org/10.1016/j.heares.2014.10.003>
4. P.D. Ross, L. Polson, M.H. Grosbras, Developmental changes in emotion recognition from full-light and point-light displays of body movement. *PLoS ONE* (2012). <https://doi.org/10.1371/journal.pone.0044815>
5. H. Chao, H.Z. Zhi, L. Dong, Y.L. Liu, Recognition of Emotions Using Multichannel EEG Data and DBN-GC-Based Ensemble Deep Learning Framework. *Comput. Intel. Neurosc.* (2018). <https://doi.org/10.1155/2018/9750904>
6. Y. Li, J. Huang, H. Zhou, H.Y. Zhou, N. Zhong, Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Appl. Sci.* (2017). <https://doi.org/10.3390/app7101060>
7. W.L. Zheng, B.N. Dong, B.L. Lu, Multimodal emotion recognition using EEG and eye tracking data, in *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Chicago, 2014). <https://doi.org/10.1109/EMBC.2014.6944757>.
8. W.L. Zheng, J.Y. Zhu, Y. Peng, B.L. Lu, EEG-based emotion classification using deep belief networks, *Proc. - IEEE Int. Conf. Multimed. Expo.* (2014). <https://doi.org/10.1109/ICME.2014.6890166>.
9. M. Bilalpur, S.M. Kia, M. Chawla, T.S. Chua, R. Subramanian, Gender and emotion recognition with implicit user signals, *ICMI 2017 - Proc. 19th ACM Int. Conf. Multimodal Interact.* **2017**, 379–387. (2017). <https://doi.org/10.1145/3136755.3136790>.
10. W. Liu, W.L. Zheng, B.L. Lu, Emotion recognition using multimodal deep learning. *Lect. Notes Comput. Sci.* **9948**, 521–529 (2016). https://doi.org/10.1007/978-3-319-46672-9_58
11. W. Liu, W.L. Zheng, B.L. Lu, Multimodal emotion recognition using multimodal deep learning. Available online: <https://arxiv.org/abs/1602.08225> (Accessed on 30 September 2016)
12. R. Alazrai, R. Homoud, H. Alwanni, M.I. Daoud, EEG-based emotion recognition using quadratic time-frequency distribution. *Sensors* **18**(8), 2739 (2018)
13. R. Li, Y. Wang, B.L. Lu, A multi-domain adaptive graph convolutional network for EEG-based emotion recognition, in *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 5565–5573). (2021)
14. J. Li, Z. Zhang, H. He, Hierarchical convolutional neural networks for eeg-based emotion recognition. *Cogn. Comput.* **10**, 368–380 (2018)
15. T. Song, W. Zheng, P. Song, Z. Cui, Eeg emotion recognition using dynamical graph convolutional neural networks, *IEEE Trans. Affect. Comput.* (2018)
16. W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, X. Chen, Eeg-based emotion recognition via channel-wise attention and self attention, *IEEE Trans. Affect. Comput.* (2020)
17. T. Zhang, W. Zheng, Z. Cui, Y. Zong, Y. Li, Spatial-temporal recurrent neural network for emotion recognition. *IEEE Trans. Cyber.* **49**(3), 839–847 (2019). <https://doi.org/10.1109/TCYB.2017.2788081>
18. E. Rudakov, L. Laurent, V. Cousin, A. Roshdi, R. Fournier, A. Nait-ali, S. Al Kork, Multi-task CNN model for emotion recognition from EEG Brain maps, in *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)* (pp. 1–4). IEEE. (2021)
19. V. Rozgić, S. Ananthakrishnan, S. Saleem, R. Kumar, R. Prasad, Ensemble of svm trees for multimodal emotion recognition, in *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference* pp. 1–4. IEEE. (2012)
20. M. Li, H. Xu, X. Liu, S. Lu, Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* **26**(S1), 509–519 (2018)
21. K. Murphy, Y. Weiss. The factored frontier algorithm for approximate inference in DBNs. *arXiv preprint arXiv:1301.2296*, (2013).
22. Y. Wei et al., CNN: Single-label to Multi-label, 6(1), 1–14. (2014). <https://doi.org/10.1109/TPAMI.2015.2491929>
23. S. Verma, Z.L. Zhang, Graph capsule convolutional neural networks, 2018. Available: <http://arxiv.org/abs/1805.08090>
24. R. Mukhometzianov, J. Carrillo, CapsNet comparative performance evaluation for image classification. *arXiv:1805.11195*, *arXiv.org*, pp. 1–14, 2018, [Online]. Available: <https://arxiv.org/ftp/arxiv/papers/1805/1805.11195.pdf>
25. W.H. Hwang, D.H. Kang, D.H. Kim, Brain lateralisation feature extraction and ant colony optimisation-bidirectional LSTM network model for emotion recognition. *IET Signal Proc.* **16**(1), 45–61 (2022)
26. S. Alhagry, A. Aly, R. A., Emotion recognition based on EEG using LSTM recurrent neural network, *Int. J. Adv. Comput. Sci. Appl.* (2017). <https://doi.org/10.14569/ijacsa.2017.081046>.
27. S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, S. Bhattacharya, Using deep and convolutional neural networks for accurate emotion classification on deap dataset

28. T. Song, W. Zheng, P. Song, Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* **11**(3), 532–541 (2020). <https://doi.org/10.1109/TAFFC.2018.2817622>
29. Schirmeister, R. Tibor, et al., Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Map* **38**(11), 5391–5420. (2017)
30. S. Koelstra et al., DEAP: A database for emotion analysis; Using physiological signals. *IEEE Trans. Affect. Comput.* **3**(1), 18–31 (2012). <https://doi.org/10.1109/T-AFFC.2011.15>
31. T.M. Cover, J.A. Thomas, Differential entropy. *Elements of Inf. Theory*, 224–238. (1991)
32. J. Fdez, N. Guttenberg, O. Witkowski, A. Pasquali, Cross-subject EEG-based emotion recognition through neural networks with stratified normalization. *Front. Neurosci.* **15**, 626277 (2021)
33. A. Baratloo, M. Hosseini, A. Negida, G. El Ashal, Part 1: simple definition and calculation of accuracy, sensitivity and specificity. (2015)
34. N. Salankar, P. Mishra, L. Garg, Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot. *Biomed. Signal Process. Control* **65**, 102389 (2021)
35. T.A.N.G. Wan, H.U. Jun, H. Zhang, W.U. Pan, H.E. Hua, Kappa coefficient: a popular measure of rater agreement. *Shanghai Arch. Psychiatry* **27**(1), 62 (2015)
36. X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, P. Marttinen, EEG based emotion recognition: A tutorial and review. *ACM Comput. Surveys* **55**(4), 1–57 (2022)
37. H.J. Yoon, S.Y. Chung, EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Comput. Biol. Med.* **43**(12), 2230–2237 (2013). <https://doi.org/10.1016/j.compbio.2013.10.017>
38. P. Arnau-Gonzalez, M. Arevalillo-Herraez, S. Katsigiannis, N. Ramzan, On the influence of affect in EEG-based subject identification. *IEEE Trans. Affect. Comput.* **12**(2), 391–401 (2021). <https://doi.org/10.1109/TAFFC.2018.2877986>
39. V. Gupta, M.D. Chopda, R.B. Pachori, Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals. *IEEE Sens. J.* **19**(6), 2266–2274 (2019). <https://doi.org/10.1109/JSEN.2018.2883497>
40. R. Gupta, K. ur Rehman Laghari, T. H. Falk, Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization, *Neurocomputing*, **174**, 875–884, (2016). <https://doi.org/10.1016/j.neucom.2015.09.085>
41. J. Cheng et al., Emotion recognition from multi-channel EEG via deep forest. *IEEE J. Biomed. Heal. Inf.* **25**(2), 453–464 (2021). <https://doi.org/10.1109/JBHI.2020.2995767>
42. S. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **3**(1), 42–55 (2012). <https://doi.org/10.1109/T-AFFC.2011.25>
43. P. Arnau-González, S. Katsigiannis, N. Ramzan, D. Tolson, and M. Arevalillo-Herráez, ES1D: A deep network for EEG-based subject identification, *Proc. - 2017 IEEE 17th Int. Conf. Bioinforma. Bioeng. BIBE 2017*, **2018**, 81–85, (2017). <https://doi.org/10.1109/BIBE.2017.00-74>.
44. D. Zhang, L. Yao, K. Chen, J. Monaghan, A convolutional recurrent attention model for subject-independent EEG signal analysis. *IEEE Signal Process. Lett.* **26**(5), 715–719 (2019). <https://doi.org/10.1109/LSP.2019.2906824>
45. Y. Yin, X. Zheng, B. Hu, Y. Zhang, X. Cui, EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput.* **100**, 106954 (2021). <https://doi.org/10.1016/j.asoc.2020.106954>
46. A. Topic, M. Russo, Emotion recognition based on EEG feature maps through deep learning network. *Eng. Sci. Technol. an Int. J.* **24**(6), 1442–1454 (2021). <https://doi.org/10.1016/j.jestech.2021.03.012>
47. W. Liu, J.L. Qiu, W.L. Zheng, B.L. Lu, Multimodal emotion recognition using deep canonical correlation analysis, (2019). arXiv preprint arXiv. 1908. 05349.
48. Y. Yang, Q. Wu, M. Qiu, Y. Wang, X. Chen, Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network, in *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). IEEE.
49. Q. Gao, Y. Yang, Q. Kang, Z. Tian, Y. Song, EEG-based emotion recognition with feature fusion networks. *Int. J. Mach. Learn. Cybern.* **13**(2), 421–429 (2022)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.