EEG emotion recognition based on Efficient-Capsule Network with Convolutional Attention

Abstract

EEG-based emotion recognition, as a pivotal component in human-computer interaction, has garnered considerable scholarly interest. And finding EEG features with stronger time-space-frequency correlation as well as reducing the computational overhead while enhancing the model structure have been the focus of research in this field. Therefore, in this paper, a deep learning model, ECNCA, is optimized. Firstly, by mining the temporal, frequency and spatial features in the EEG data, four frequency bands, namely theta, alpha, beta and gamma, are spliced and fused to make full use of the information in the EEG data for emotion classification. Secondly, the input data was strengthened by CNN and attention mechanism, and Efficient-Capsule was used to complete the classification of emotions as a way to achieve the goal of high accuracy with low overhead. Finally, we conducted various experiments on the SEED dataset and DEAP dataset, and the highest accuracy of the emotion triple classification task and quadruple classification were 94.96% and 91.34%, respectively. In addition, the experiments demonstrate that ECNCA also has an advantage over CNN and CapsNet in terms of computational overhead. This study can provide some reference for emotional experience and emotional computing.

Keywords : Deep learning, Electroencephalogram (EEG), Emotion recognition, Efficient-Capsule network, Attention mechanisms.

1. Introduction

In recent years, emotion recognition technology has increasingly been applied across a range of fields including psychological research, personalized medicine, intelligent education, driving safety, and human-computer interaction[1-3]. Emotion computing primarily utilizes the analysis of physiological signals to ascertain individuals' emotional states, with common signals including Electroencephalogram (EEG), Electromyogram (EMG), and Electrocardiogram (ECG). EEG signals, in particular, offer a direct reflection of brain activity with high temporal resolution without involving radiation or the need for extensive equipment. Compared to other sources, EEG signals are prized for their reliability, safety, and portability, making them a significant source of data for emotion recognition[4-6].

Non-invasive EEG signals capture spontaneous bioelectric potentials transmitted from the brain cortex to the scalp over a period, through continuous contact between electrodes and the scalp[7]. From a neuroscience perspective, there are specific regions within the brain cortex closely associated with emotions, such as the ventromedial prefrontal cortex, orbitofrontal cortex, frontal lobes, and amygdala[8]. EEG signals are typically divided into five frequency bands: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), and gamma (>25 Hz). Each frequency band reflects different brain activities and is closely linked to current emotional states[9]. By extracting information from these bands, it is possible to ascertain human emotional states, including boredom, disgust, excitement, joy, and satisfaction.

EEG-based emotion recognition relies on EEG features with sufficient resolution. Initially, features were primarily categorized into time, frequency, and time-frequency domains, which were all manually extracted[10-12]. However, these features alone could not provide robust differentiation for emotion recognition, serving only as identifiers for specific tasks. Subsequently, researchers began using indirect measures such as entropy or power from EEG signals as features for emotion classification[13-14]. Using these indirect measures enhanced the dimensionality of EEG features, yielding better classification results and becoming a mainstream approach in subsequent studies.

Currently, feature fusion techniques, which combine two or more features for a single task, are predominantly used in EEG emotion recognition[15]. But most studies only utilize a portion of the features available in raw EEG signals, not fully exploiting all available information. Inadequate feature extraction can lead to the loss or oversight of crucial information within the EEG data, affecting the model's understanding of emotional states. Therefore, how to thoroughly extract features from EEG data remains a primary challenge.

Additionally, enhancing network structures is a key focus for researchers. The rapid rise of deep learning in the 21st century has offered new perspectives for EEG emotion recognition. Having achieved significant success in fields like computer vision and natural language processing, many scholars have begun to apply deep learning to emotion recognition. Deep learning algorithms like CNN enable models to learn higher-dimensional and more abstract features from data[16], proving advantageous in emotion classification tasks. Meanwhile, CNN face limitations, such as the lack of large annotated datasets required for EEG data and the inability of CNN kernels to fully address the hierarchical and spatial relationships between EEG features. Considering these issues, Hinton[17] proposed Capsule to replace convolutional layers. Capsules, as components of the network, help capture relationships between objects more effectively, and their structure allows information to be passed through the network without being discarded. While networks containing Capsule generally involve an iterative process, and the problems of high computational overhead and low computational efficiency limit their further development[18-19]. Based on this, Mazzia et al. [20] introduced deep separable convolution and non-iterative, highly parallel routing algorithms, significantly reducing the model's computational burden. Whereas, the effectiveness of its proposed efficient Capsule in the field of emotion recognition is yet to be further verified. Thus, enhancing network structures to accurately model emotional states from EEG data is our second crucial consideration.

Addressing the tow aforementioned challenges, this study first constructs multi-dimensional EEG data incorporating temporal, spatial, and frequency band features as input, thoroughly extracting features across all dimensions of the original EEG data. Furthermore, we propose a deep learning framework named ECNCA (Efficient Capsule Network with Convolutional Attention). ECNCA utilizes a convolutional module to extract higher-level features and an attention module to suppress irrelevant feature inputs, ensuring model performance is not compromised by feature redundancy. Additionally, we have designed a channel selection module to further reduce computational costs by discarding channels that contribute minimally to emotion recognition. The contributions of this paper are as follows:

1. We have extensively mined multi-dimensional EEG features related to time, space, frequency, and frequency bands, merging the theta, alpha, beta, and gamma bands. A feature extraction module tailored for EEG emotion recognition tasks, the Convolutional Attention (CA) module, has been designed, endowing the network model with robust representational learning capabilities.
2. We introduce the ECNCA model for EEG emotion recognition, which enhances input data through CNN and Attention mechanisms and employs efficient Capsules for emotion classification, achieving high model performance with reduced computational expenditure.
3. The model has achieved accuracies of 94.96% and 91.34% on emotion triple and quadruple classification tasks, respectively, in two public EEG datasets, SEED and DEAP, demonstrating its efficacy in the domain of emotion recognition.

The structure of this paper is as follows: Section 2 reviews related work, Section 3 details the data processing and the proposed model, Section 4 applies our model to the public datasets DEAP and SEED for validation, and finally, Section 5 concludes the paper.

2. Related work

2.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of feedforward neural network particularly suited for processing data with a grid-like topology, such as images and audio. CNN is principally composed of convolutional layers, pooling layers, activation functions, and fully connected layers. Their convolutional kernels can respond to units within their receptive field, and the advantage of weight sharing significantly enhances training efficiency. In recent years, CNN has been widely used in the field of emotion recognition. Yang et al.[21] proposed a multi-column CNN model for EEG-based emotion recognition, employing multiple decision-making modules for weighted decision-making, which improved accuracy compared to single decision-making modules. However, EEG data typically comprises multiple channel data, and this approach did not account for spatial features between channels. Consequently, Zhang et al.[22] leveraged CNN's potent capability to process spatial features by mapping the three-dimensional positions of EEG electrodes onto a two-dimensional plane to simulate their spatial structure, thereby enhancing model performance. Hwang et al.[23] introduced differential entropy features that preserve topological characteristics into CNN for emotion recognition, achieving exceptional performance on the SEED dataset. Given the temporal and spatial characteristics of EEG signals, Khademi et al.[24] combined CNN with Recurrent Neural Networks (RNN) to construct a hybrid deep learning model, integrating contextual information while exploring inter-channel correlations. Shen et al.[25], aiming to amalgamate more useful EEG features, integrated the frequency, spatial, and temporal information of multi-channel EEG signals, validating the effectiveness of multi-dimensional feature input. Three-dimensional convolution supports a higher volume of data than two-dimensional convolution, making it more suitable for analyzing high-dimensional EEG signals. Hence, Wang et al.[26] introduced Emotion-Net, a three-dimensional CNN tailored for EEG-based emotion recognition, achieving state-of-the-art results at the time.

Despite certain advantages of three-dimensional convolution, two-dimensional convolution is more suited to the data used in this study and offers computational cost benefits. Therefore, this paper will employ two-dimensional convolution for feature extraction from EEG data.

2.2 Attention mechanism

The attention mechanism, inspired by human visual and cognitive systems, enables neural networks to focus on relevant parts of input data, thereby autonomously learning and selectively attending to important information within the inputs. Since its introduction, the attention mechanism has achieved tremendous success in fields such as Natural Language Processing (NLP) and Computer Vision (CV), proving to enhance overall model performance. Tao et al.[27] proposed an Attention-based Convolutional Recurrent Neural Network (ACRNN), which demonstrated superior recognition accuracy compared to the most advanced methods at the time. Given that EEG data is stored across multiple channels, often containing a significant amount of noise, Huang et al.[28] employed attention mechanisms to balance the weight values of different electrode channels, so as to avoiding redundant information. Moreover, attention modules can focus on various features, and Liu et al.[29] utilized the three-dimensional structure of inputs, inserting attention modules, spectral attention, and temporal attention into convolutional layers, enabling the model to focus on parts of the data that significantly influence emotion recognition. Beyond its successful implementation in CNN, the attention mechanism has also found widespread application in Graph Neural Networks. Lin et al.[30] combined the advantages of one-dimensional convolution and graph convolution by incorporating attention modules in their model, enabling automatic selection of effective channels and adapting the model to EEG emotion classification tasks with varying cost and accuracy requirements.

Considering the multidimensional features of the data post-feature extraction, and in order to avoid redundancy, this study employs a channel attention mechanism to assign weights to input features. Additionally, in practical applications, some channels may not be particularly important for emotion recognition. Hence, we utilized a spatial attention mechanism to make reasoned selections of the original EEG channels.

2.3 Capsule Network

Capsule, introduced by the pioneer of deep learning, Hinton, aim to address the issue of CNN neglecting the relationships between features and the loss of information due to pooling operations. Capsule is a vector entity composed of multiple scalar neurons, capable of identifying visual entities and outputting attributes such as the entity's orientation, size, and relative positions between objects[17]. The introduction of capsules garnered significant attention from scholars. Chao et al.[31] were the first to apply CapsNet to multi-channel EEG data, affirming its effectiveness in the field of emotion recognition. Liu et al.[32] proposed a multi-level feature-guided capsule network, addressing the issue of CNN failing to adequately represent the intrinsic relationships between different EEG channels. Furthermore, the integration of attention mechanisms with capsules has led to the development of many excellent models. Wei et al.[33] applied the Transformer module to CapsNet, capturing both global and local features of EEG signals and achieving state-of-the-art results at the time. Although CapsNet-based models have high classification accuracy, they still cannot avoid the problem of high computational overhead. Efficient-Capsule offers several advantages over traditional capsules. Wang et al.[34] applied CapsNet and Efficient-CapsNet to facial expression recognition, comparing the performance of both models across three public datasets and highlighting Efficient-Capsule's ability to significantly enhance training efficiency while maintaining accuracy and requiring fewer parameters.

Though Efficient-Capsule has been applied in other domains[35-37], its effectiveness in EEG-based emotion recognition remains to be validated, posing a challenge for our research. Therefore, we adjusted the Efficient-Capsule to suit emotion classification tasks and conducted experiments to further explore its performance, aiming to maximize its capabilities.

3. Proposed model

This section first introduces our data processing and feature extraction methods, followed by a detailed description of each module within our model, as outlined in Figure 1.

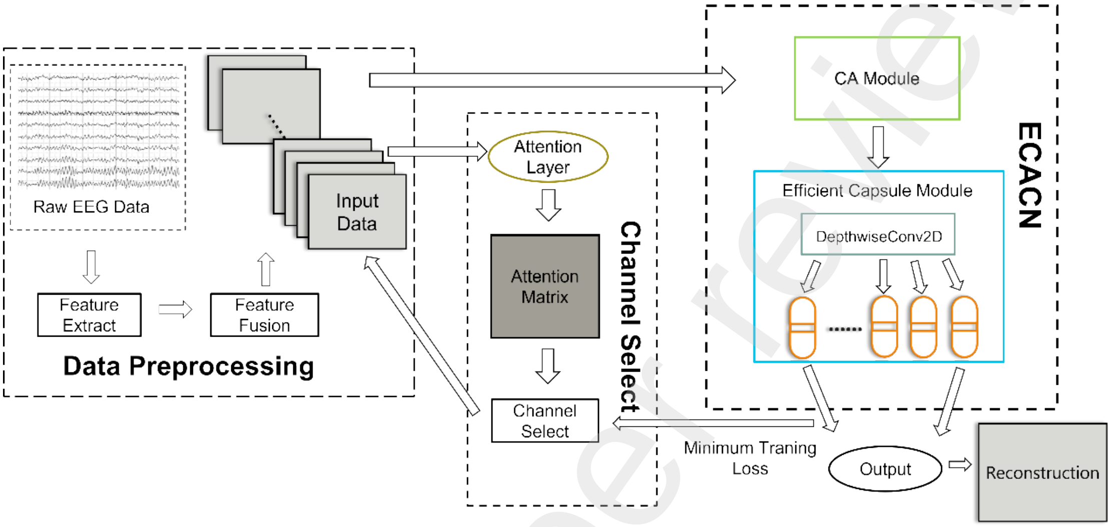


FIG. 1 The overall process of emotion recognition

3.1 Feature extraction

EEG offers extremely high temporal resolution, capturing millisecond-level brain activities, making it highly useful for studying and understanding the brain's rapid dynamic processes. To fully mine the temporal, frequency, and spatial dimensions of EEG data, we constructed a feature structure encompassing three dimensions.

Starting with the temporal and frequency dimensions, the optimal time segment for representing emotions is 1s-4s[38]. Considering the overall data volume, we used a 1.5s time window for non-overlapping segmentation, as shown in Equation (1), where n represents the number of windows, and x_i represents individual data segments, with each segment assigned the original data's label.

$$X \rightarrow \{x_1, x_2, x_3, x_4, \ldots\ldots, x_n\} \tag{1}$$

Then, each data segment underwent denoising and frequency band extraction using a Butterworth filter, selecting specific bands for subsequent fusion. To capture the dynamic features of EEG signals, differential entropy (DE) was extracted from each band. DE, as one of the most reliable features for emotion recognition[39], is commonly used in time-series signal analysis and is defined as shown in Equation (2).

$$DE = -\int_a^b p(x)\log(p\ (x))\,dx \tag{2}$$

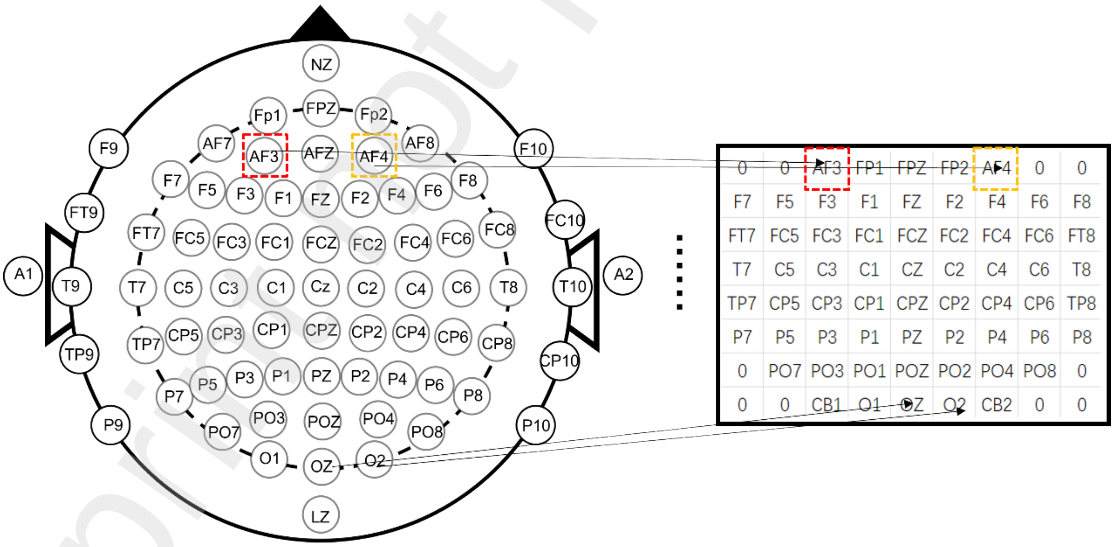
where $p\ (x)$ denotes the probability density of continuous information and [a, b] denotes the value space of the information. For a particular length of approximately Gaussian distributed EEG data $N(\mu,\sigma_i^2)$, DE can be represented as shown in (3):

$$DE = \frac{1}{2}\log(2\pi\sigma_i^2) \tag{3}$$

Normalized data can accelerate model convergence and improve generalizability, enhancing training outcomes. This paper employs Z-score normalization, defined in Equation (4) for samples $x_1, x_2, x_3, x_4, \ldots\ldots, x_n$.

$$y_i = (x_i - \bar{x})/s \tag{4}$$

For the spatial dimension of EEG data, it is essential to preserve the structural information of EEG electrodes in space. Based on the international 10/20 electrode placement system, all electrodes were mapped to a two-dimensional matrix according to their spatial relationships. where unused electrodes as well as unprojected locations are filled with a value of 0, thus allowing the model to learn the spatial characteristics of the EEG data. As shown in Figure 2, the created two-dimensional matrix is more suitable for CNN to extract spatial features[40].

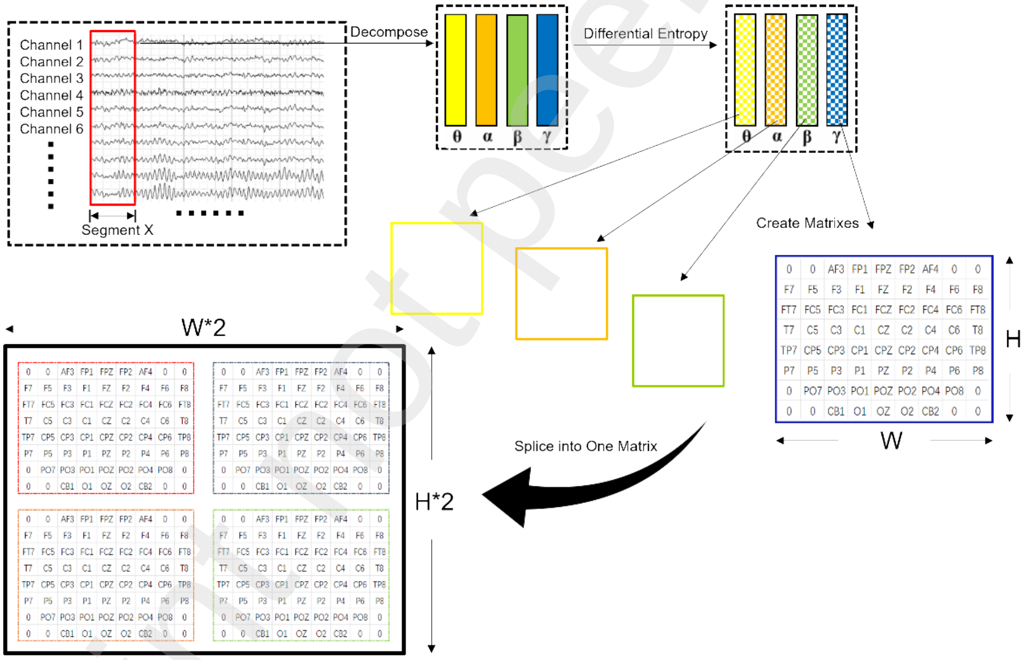


0	0	AF3	FP1	FPZ	FP2	AF4	0	0
F7	F5	F3	F1	FZ	F2	F4	F6	F8
FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6	FT8
T7	C5	C3	C1	CZ	C2	C4	C6	T8
TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6	TP8
P7	P5	P3	P1	PZ	P2	P4	P6	P8
0	PO7	PO3	PO1	POZ	PO2	PO4	PO8	0
0	0	CB1	O1	OZ	O2	CB2	0	0

FIG. 2 The image on the left is the international 10/20 electrode placement system, the right is mapped with an example of the electrodes used in the SEED dataset, and the arrows represent the positions projected from the electrodes into the matrix.

3.2 Feature fusion

The fusion of EEG data from different frequency bands, which contain varied emotional information, offers a richer informational context than individual bands. Delta waves (0.5-4Hz), prominent during deep sleep stages and often associated with physical recovery and relaxation, are not directly correlated with emotions and are therefore excluded from this study. Instead, we focus on the theta (4 - 8 Hz), alpha (8 - 12 Hz), beta (12 - 25 Hz), and gamma (> 25 Hz) frequency bands for experimentation. The feature fusion operation employed concatenating the extracted feature matrices from each frequency band in a 2x2 manner, resulting in a final matrix of dimensions 16x18. This matrix, compared to traditional and single-band features, effectively integrates the frequency, band, spatial, and temporal characteristics of multiple EEG channels, offering a comprehensive representation of EEG signal variations. The entire data processing workflow is depicted in Figure 3.



0	0	AF3	FP1	FPZ	FP2	AF4	0	0
F7	F5	F3	F1	FZ	F2	F4	F6	F8
FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6	FT8
T7	C5	C3	C1	CZ	C2	C4	C6	T8
TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6	TP8
P7	P5	P3	P1	PZ	P2	P4	P6	P8
0	PO7	PO3	PO1	POZ	PO2	PO4	PO8	0
0	0	CB1	O1	OZ	O2	CB2	0	0

FIG. 3 Taking SEED data set as an example, the data of the four frequency bands are spliced horizontally and vertically to form a new matrix.

3.3 ECNCA

In the ECNCA model, the first component is the Convolutional Attention (CA) module, comprised of multiple two-dimensional convolutional layers and channel attention layers. This module is designed to enhance the model's ability to understand and process data, thereby improving classification performance. As shown in Figure 4, by stacking two-dimensional convolutional layers, each convolution operation captures deeper features, enriching the feature representation for the classification task. Additionally, the batch_norm layer is added after each convolution operation to stabilize the learning process. The channel attention layer assigns different weights to each feature, emphasizing features more relevant to the classification task and suppressing redundant inputs. This dynamic focus on valuable features further strengthens the model's discriminative capabilities. The attention mechanism is formulated as shown in Equation (5). Following these operations, a series of two-dimensional feature maps are produced and fed into the Efficient-Capsule module (Mazzia et al. 2021) for final classification.

$$F_C = sigmoid[MLP(Avg\ Pool(F)) + MLP(MaxPool(F))] \tag{5}$$

The Efficient-Capsule module constructs capsules using depthwise separable convolutions, significantly reducing the number of parameters. Unlike traditional neural networks, which rely on single neurons, it employs multiple capsules with vector outputs, enabling the capture of a richer array of information. Capsules are activated via an activation function, limiting the length of the output capsule between 0 and 1, therefore enhancing the model's expressive capacity. Information flow between capsules is facilitated by a self-attention routing algorithm, which routes the prediction vectors of the next layer to the current layer's capsules by computing weighted sums. This process achieves affine transformations between capsule layers and adjusts the coupling coefficients between capsules through a self-attention mechanism, optimizing the information transmission pathway.

The output of this module is no longer traditional scalars but vectors, which represent not only the probability of categories but also contain attribute information. The Efficient-Capsule module employs a linear combination of Margin Loss and Reconstruction Loss as the loss function, aiming to improve classification accuracy and enhance the model's generalization ability. By adjusting the weights of the two types of losses, the model achieves a balance between classification performance and data reconstruction capability. Overall, the structural and algorithmic innovations of the

Efficient-Capsule, compared to traditional capsule networks, endow it with a more efficient and powerful model representation capability.

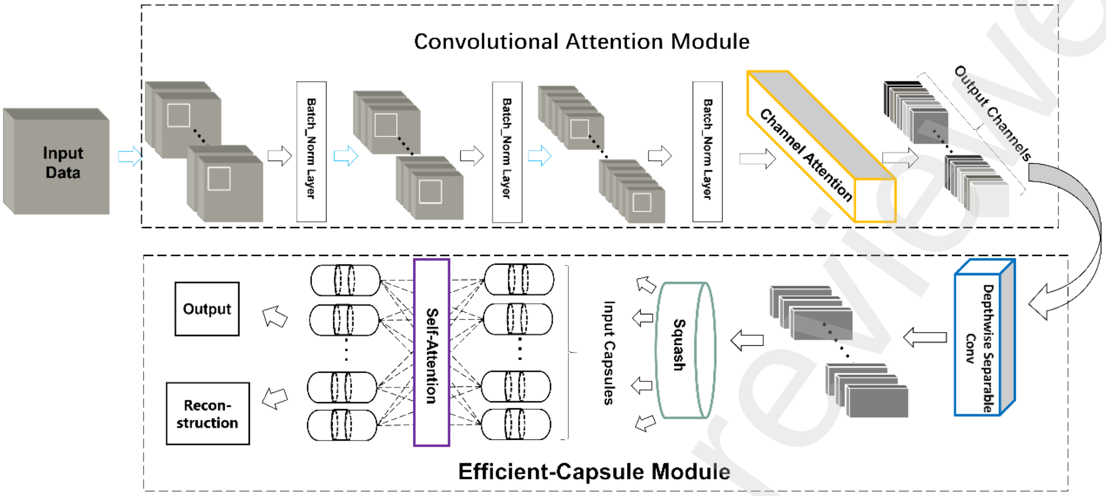


FIG. 4 The full operation of ECNCN, and the blue arrow represents the convolution operation.

3.4 Channel selection

EEG data typically encompasses multiple original channels, with each channel capturing distinct information. In some applications, due to equipment limitations or specific requirements, not all EEG channels are necessary or available. Research[41] has shown that selecting EEG channels can reduce the computational load and complexity of models, as well as lower the costs associated with EEG data acquisition. The main objective of this module is to optimize the EEG data processing to minimize computational expenses, while also assessing the model's practicality when the number of channels is restricted. Prior to inputting data into ECNCA, spatial attention layers are added to calculate the attention weights for each channel, as defined by the following equation:

$$F_S = Sigmoid(Conv([AvgPool(F);MaxPool(F)])) \tag{6}$$

Given the variance in the number of EEG channels across datasets, we standardize the selection process by setting thresholds to retain a specific percentage of channels, including 100%, 80%, 60%, 40%, and 20%. The detailed flow of the algorithm is as follows:

Algorithm 1 Code for the channel selection module

```
Algorithm 1    EEG channel selection
Input data: number of current training rounds I; discard channel ratio R; EEG data E; attention weight matrix A.
E obtains A through the attention layer
while I<=num_epoch；
    if train_loss < min_loss then
            train_loss = min_loss;
            New_A = A;
Return New_A;
num_drop_channel = num_E_channel * R;
Sort New_A in ascending order and give the index of each value as New_A_index;
drop_index . append(New_A [ 0 : num_drop_channel ] );
for J in all_channel_index:
    if J == drop_index then
            E [ J ] = 0;
```

4. Experiments

4.1 Dataset configuration

This section evaluates the model's performance using two widely applied public datasets in the EEG emotion recognition field: DEAP and SEED. Below is a detailed description of these datasets.

4.1.1 SEED

The SEED dataset[42] involved 15 subjects (7 males and 8 females) who watched 15 different themed videos of approximately 4 minutes each, with EEG signals collected using the international standard 10-20 electrode placement system. Each video was carefully edited to evoke coherent emotions, interpreted without explanation. Each video should elicit a desired target emotion. Subjects rated the emotions evoked after viewing each video, pre-labeled as negative, neutral, or positive. The ratio of videos inducing positive, neutral, and negative emotions was 1:1:1. The raw EEG signals were downsampled to 200Hz and processed with a 1-75Hz band-pass filter.

4.1.2 DEAP

The DEAP dataset[43] collected EEG data from 32 subjects (16 males and 16 females), all in good physical and mental health, using the international standard 10-20 electrode placement system. Subjects watched 40 one-minute video clips and rated the evoked emotions on a scale from 1 to 9 after viewing. The original signal frequency was downsampled to 128Hz and processed with a 4-45Hz band-pass filter. Each sample was 63s long, which included a 3s baseline time for recording data in the experimenter's calm state. In the data processing of this paper, the subtraction of the baseline has been done for the smoothness of the data. The labels were divided into four categories based on valence and arousal: high valence-high arousal, high valence-low arousal, low valence-low arousal, and low valence-high arousal.

4.2 Model performance

Table 1 SEED is used as an experimental dataset to compare the effects of different convolutional layer settings and number of capsules on model accuracy and standard deviation.

Model	Convolutional Layer Setup	Number of capsules	Accuracy (%)
A	256@3*3 // 512@3*3 // 64@3*3	8	93.42±1.24
B	256@3*3 // 512@3*3 // 128@3*3	16	93.75±0.96
C	256@3*3 // 512@3*3 // 256@3*3	32	93.93±1.08
D	256@3*3 // 512@3*3 // 256@3*3 // 64@3*3	8	94.74±0.84
E	256@3*3 // 512@3*3 // 256@3*3 // 128@3*3	16	94.87±0.69
F	256@3*3 // 512@3*3 // 256@3*3 // 256@3*3	32	94.96±0.65

As shown in Table 1, we first explore the effect on the models in different convolutional layer settings and number of capsules, and six models are designed for the study in this paper. The SEED dataset is used as input to achieve three emotion classifications: negative, neutral, and positive. Data segmentation was conducted by subject-dependent, following the data processing procedures outlined in Section 3. Optimal parameters were identified through ten-fold cross-validation, dividing the data into ten equal parts, using one part as the testing set and the rest for training. This cycle was repeated ten times with different data as the testing set each time, initializing model parameters before each cycle. This approach ensured that all data were tested, yielding more reliable experimental results. During training, a learning rate scheduler was employed to automatically adjust the learning rate, starting at 0.0002 and reducing to a minimum of 9e-5, with a batch_size of 40 for 70 epochs training. The model achieved the classification of negative, positive, and neutral emotions on the SEED dataset. Possibly due to the fact that deeper convolutional layers can respond to more complex and abstract features, and that increasing the number of capsules leads to stronger representation and generalization capabilities. So the results in Table 1 show that increasing the number of convolutional layers and capsules can play a role in model accuracy improvement, and the maximum difference in the accuracy of the six models reaches 1.54%. It is noteworthy that the accuracy of each of these models reaches more than 93%. The standard deviation of the ten-fold cross-validation is kept at a low level, which proves that the proposed model has strong learning ability and stability.

After establishing the best model parameters, experiments were conducted on data from individual frequency bands to validate the effectiveness of frequency band fusion. As shown in Figure 5, two experimental designs were proposed. The first maintained the same input data size, using four identical frequency band data for concatenation instead of combining data from four different frequency bands. The second approach directly used data from a single frequency band, with the size being one-fourth of the fused frequency band data size. Four categories of emotions were classified on the DEAP dataset: high valence-high arousal, high valence-low arousal, low valence-low arousal, and low valence-high arousal.

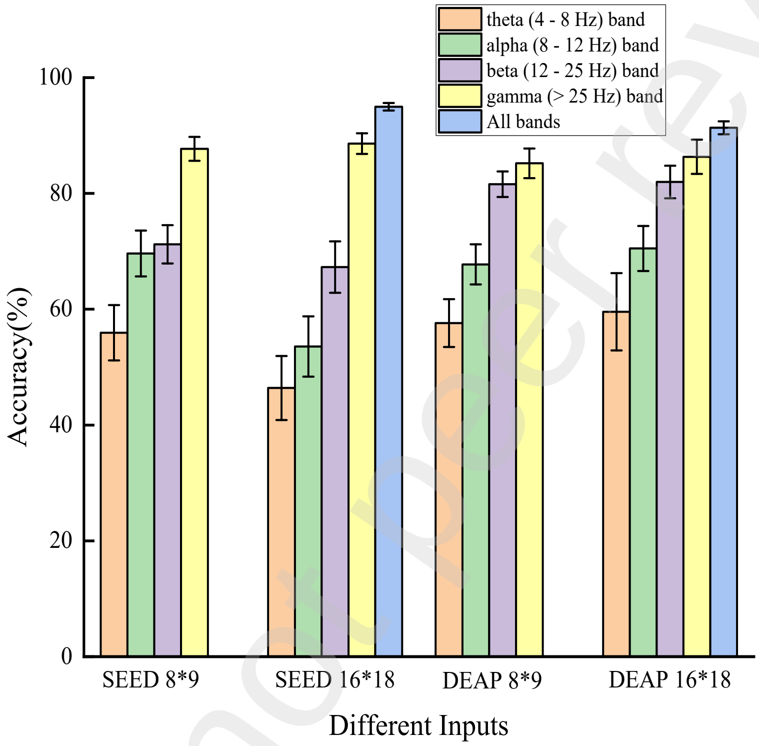


FIG. 5 Compare the accuracies of SEED and DEAP under different frequency bands.

The results indicated that the gamma frequency band contained more features relevant to emotion recognition, while lower frequency bands performed poorly in the model. The gamma band is typically associated with advanced cognitive functions and complex brain activities, which shows higher accuracy in the experiments[44]. Moreover, as each frequency band is linked to different brain activities, using all bands could provide a more comprehensive view of brain activities, leading to the highest accuracy. In addition, each frequency band is associated with different functional activities of the brain, so using all the frequency bands may provide more comprehensive information about brain activities compared to using a single band for classification. The frequency band fusion method can effectively improve the model classification accuracy, and the experimental results show that the accuracy on SEED and DEAP is improved by 6.35% and 5.02%, respectively.

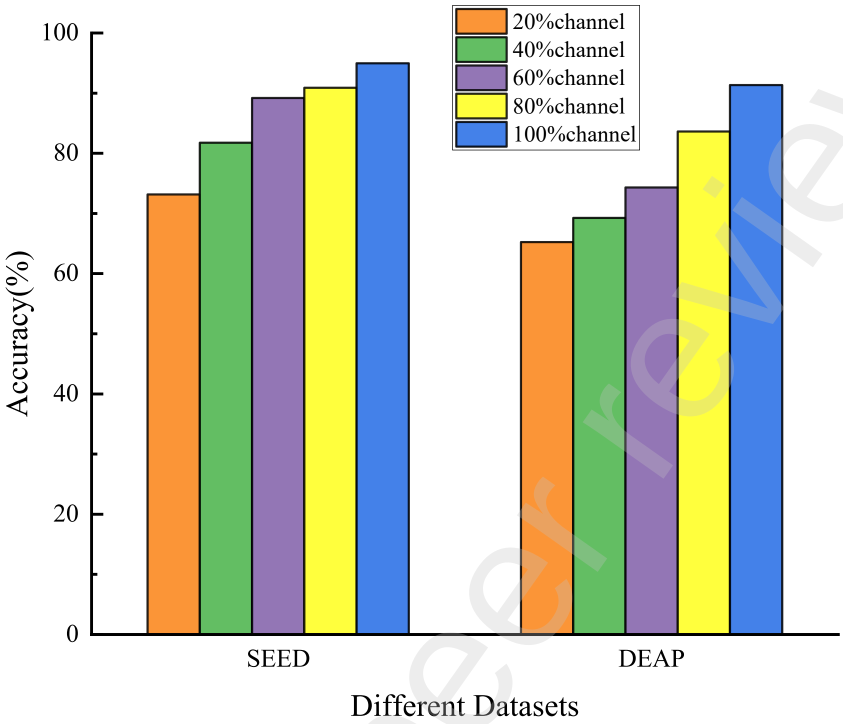


FIG. 6 Compare the accuracies of the SEED dataset and the DEAP dataset under different channel numbers.

The final study focused on the impact of using different numbers of channels on the model, and the results are shown in Figure 6. Through the results, we found that experiments on both datasets revealed a gradual decrease in average model accuracy with fewer channels used, likely due to the loss of diverse emotional information each channel contains. Discarding channels reduced the number of features available to the model, and using zero values to fill in discarded channels might have made the model more sensitive, leading to significant accuracy drops. This effect was particularly evident in the DEAP dataset, which uses fewer original channels. Although accuracies dropped to some extent on both datasets, it was observed that maintaining 80% and 60% of the channels in the SEED dataset still resulted in accuracies close to using 100% of the channels, with accuracies above 80% for using 40% or more of the channels. This suggests that the channel selection process likely removed channels that were less beneficial to the model, achieving the experimental objective.

4.3 Model comparison

With the proliferation of mobile, IoT, and embedded systems, there's a growing need for deep learning models to run in resource-constrained environments. The most prominent advantage of the model proposed in this paper over CapsNet is that it can use less computational resources to get better performance, which is in line with the development trend of deep learning models. We compare ECNCA with CNN, CapsNet, and ECN. The results are shown in Figure 7. Through experiments, our proposed feature extraction module CA is shown to be effective in improving model accuracy. We can see that on SEED and DEAP, ECNCA has about 4% and 6% improvement in accuracy compared to Efficient-Capsule Network. In addition, it beats the conventional CNN in terms of accuracy. This may be because our model achieves a maximally preserved outcome in information transfer. Notably, ECNCA requires significantly fewer parameters than CapsNet, using less than one-fifth.

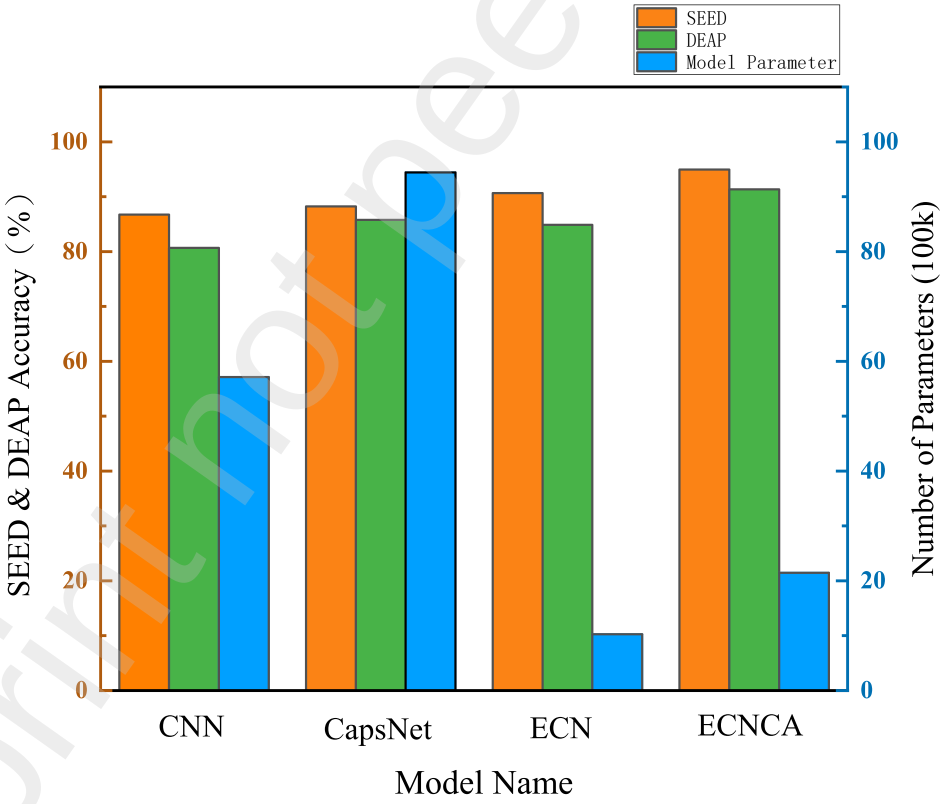


FIG. 7 Comparison of ECNCA with other models

Table 2 provides a comparison between our model and previous emotion classification studies. We selected recent articles that used CapsNet as a baseline model for comparison. Most of these studies are based on binary classification experiments on multiple emotional dimensions in a subject-dependent setting, achieving accuracies exceeding 98%. In our experiments, the SEED and DEAP datasets achieved optimal performances of 94.96% and 91.34% on ECNCA, approaching state-of-the-art accuracies. It's noteworthy that human emotions are diverse, thus classifying multiple emotions on datasets bears more practical significance.

Table 2 Comparison of the model proposed in this paper with other different CapsNet models in recent years, *represents subject-independent experiment.

Author	Classificationtarget	Dataset	Model	Classificationaccuracy(%)
Chao et al.[31]	High or low-valence/arousal/dominance	DEAP	CapsNet	Arousal68.28;valence66.73;dominance67.25
Liu et al.[45]	DEAP＆DREAMER:High or low-valence/arousal SEED:Negative or neutral or .positive	DEAP,SEED,DREAMER	DA-CapsNet	DEAP:valence73.62;arousal75.25SEED:84.63 DREAMER:valence81.63;arousal81.09
Liu et al.[32]	High or low-valence/arousal/dominance	DEAP,DREAMER	MLF-CapsNet	DEAP:valence97.97;arousal98.31;dominance98.32 DREAMER:valence94.59;arousal95.26;dominance95.13
Li et al.[46]	High or low-valence/arousal/dominance	DEAP,DREAMER	MTCA-CapsNet	DEAP:valence97.25;arousal97.41;dominance98.35 DREAMER:valence94.96;arousal95.54;dominance95.52
Liu et al.[47]*	High or low-valence/arousal/dominance	DEAP	AP-CapsNet	DEAP:valence93.89;arousal95.04;dominance95.08 Subject-independent:62.71;63.51;64.00
Wei et al.[33]	High or low-valence/arousal/dominance	DEAP,DREAMER	TC-Net	DEAP:valence98.76;arousal98.81;dominance98.82 DREAMER:valence98.59;arousal98.61;dominance98.67
Chen et al.[48]	High or low-valence/arousal/dominance	DEAP,DREAMER	Caps-EEGNet	DEAP:valence96.67;arousal96.75;dominance96.64 DREAMER:valence91.12;arousal92.60;dominance93.74
Ours	DEAP:High-valence&high-arousal or high-valence& low-arousal or low-valence& low-arousal or low-valence&high-arousal SEED:Negative or neutral or positive	DEAP,SEED	ECNCA	SEED:94.96 DEAP:91.34

We also compared our model with other advanced deep learning models, including graph neural networks and recurrent neural networks. The results, as shown in Table 3, indicate that our model maintains a high standard of accuracy compared to these other deep learning models.

Table 3 Comparison of the model proposed in this paper with other deep learning models, *represents a subject-independent experiment.

Author	Classification target	Dataset	Model	Classification Accuracy(%)
Lin et al.[30]	DEAP:High-valence & high- arousal or high-valence & low- arousal or low-valence & low- arousal or low-valence & high- arousal SEED:Negative or neutral or positive	DEAP,SEED	CSGNN	DEAP:91 SSED:90.22
Song et al.[49]	High or low- valence/arousal	DEAP	ECA-CRNN	Arousal 95.70; valence 95.33
Li et al.[50]	High or low- valence/arousal	DEAP	STC-CNN	Arousal 96.79; valence 96.89
Li et al.[51]*	DEAP:High or low-valence/arousal SEED Negative or positive	DEAP,SEED	MSRN	DEAP:arousal 71.92; valence 71.29 SEED:87.05
Ours	DEAP:High-valence & high- arousal or high-valence & low- arousal or low-valence & low- arousal or low-valence & high- arousal SEED:Negative or neutral or positive	DEAP,SEED	ECNCA	SEED:94.96 DEAP:91.34

5. Discussion and Conclusion

It is demonstrated in the experiments in Section IV that increasing the number of capsules improves the model accuracy. This may be due to the fact that as the number of capsules increases enables more features to be input into the classification model, thus improving the overall model performance. Taking the SEED dataset as an example, we boosted the number of capsules to 64 for the experiment, as shown in Table 4. Although the model accuracy is still improved, the ensuing problem is that the computational overhead of the model increases and the complexity increases. Therefore, how to balance the model accuracy and computational overhead is an issue we need to consider. In addition, subject-independent experiments have been a topic of interest to researchers, and follow-up work will be conducted on subject-independent work.

Table 4 Comparison of accuracy and parameters for different number of capsules

Number of capsules	**Accuracy(%)**	**Model parameter(100k)**
8	SEED:94.74	18.7
64	SEED:95.02	25.27

This study introduces a multi-frequency EEG data-based emotion feature extraction method, effectively integrating the frequency, spatial, and temporal information of EEG data. To solve the problems in CNN and CapsNet, ECNCA is introduced in this paper for the multi-classification task of emotions. It achieved accuracies of 94.96% and 91.34% on SEED and DEAP, respectively. While not surpassing the state-of-the-art in EEG emotion recognition, ECNCA showcases efficiency and lightweight advantages. ECNCA is easier to deploy in a variety of devices, including mobile and wearable devices, and is more adaptable to resource-constrained environments.

References

[1] Torres E P, Torres E A, Hernández-Álvarez M, et al. EEG-based BCI emotion recognition: A survey[J]. Sensors, 2020, 20(18): 5083. https://doi.org/10.3390/s20185083.

[2] Sheykhivand S, Rezaii T, Mousavi Z, et al. Automatic detection of driver fatigue based on EEG signals using a developed deep neural network[J]. Electronics, 2022, 11(14): 2169. https://doi.org/10.3390/electronics11142169.

[3] Sourina O, Liu Y, Nguyen M K. Real-time EEG-based emotion recognition for music therapy[J]. Journal on Multimodal User Interfaces, 2012, 5(1-2): 27-35. https://doi.org/10.1007/s12193-011-0080-6.

[4] Li B, Cheng T, Guo Z. A review of EEG acquisition, processing and application[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1907(1): 012045. https://doi.org/10.1088/1742-6596/1907/1/012045.

[5] Yuvaraj R, Thagavel P, Thomas J, et al. Comprehensive analysis of feature extraction methods for emotion recognition from multichannel EEG recordings[J]. Sensors, 2023, 23(2): 915. https://doi.org/10.3390/s23020915.

[6] Chen J, Lin X, Ma W, et al. EEG-based emotion recognition for road accidents in a simulated driving environment[J]. Biomedical Signal Processing and Control, 2024, 87: 105411. https://doi.org/10.1016/j.bspc.2023.105411.

[7] Soufineyestani M, Dowling D, Khan A. Electroencephalography (EEG) technology applications and available devices[J]. Applied Sciences, 2020, 10(21): 7453. https://doi.org/10.3390/app10217453.

[8] Lotfi E, Akbarzadeh-T. M R. Practical emotional neural networks[J]. Neural Networks, 2014, 59: 61-72. https://doi.org/10.1016/j.neunet.2014.06.012.

[9] Zhang Y, Yan G, Chang W, et al. EEG-based multi-frequency band functional connectivity analysis and the application of spatio-temporal features in emotion recognition[J]. Biomedical Signal Processing and Control, 2023, 79: 104157. https://doi.org/10.1016/j.bspc.2022.104157.

[10] Jenke R, Peer A, Buss M. Feature extraction and selection for emotion recognition from EEG[J]. IEEE Transactions on Affective Computing, 2014, 5(3): 327-339. https://doi.org/10.1109/TAFFC.2014.2339834.

[11] Mehmood R M, Lee H J. EEG-based emotion recognition from human brain using Hjorth parameters and SVM[J]. International Journal of Bio-Science and Bio-Technology, 2015, 7(3): 23-32. https://doi.org/10.14257/ijbsbt.2015.7.3.03.

[12] Atkinson J, Campos D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers[J]. Expert Systems with Applications, 2016, 47: 35-41. https://doi.org/10.1016/j.eswa.2015.10.049.

[13] Murugappan M, Ramachandran N, Sazali Y. Classification of human emotion from EEG using discrete wavelet transform[J]. Journal of Biomedical Science and Engineering, 2010, 3(4): 390-396. https://doi.org/10.4236/jbise.2010.34054.

[14] Ende M, Louis A K, Maass P, et al. EEG signal analysis by continuous wavelet transform techniques[C]//Nonlinear analysis of physiological data. Springer Berlin Heidelberg, 1998: 213-219.

[15] Kamble K, Sengupta J. A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals[J]. Multimedia Tools and Applications, 2023, 82(18): 27269-27304. https://doi.org/10.1007/s11042-023-14489-9.

[16] Yin Z, Zhao M, Wang Y, et al. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model[J]. Computer Methods and Programs in Biomedicine, 2017, 140: 93-110. https://doi.org/10.1016/j.cmpb.2016.12.005.

[17] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[J]. Advances in neural information processing systems, 2017, 30.. https://doi.org/10.48550/arXiv.1710.09829.

[18] Peer D, Stabinger S, Rodriguez-Sanchez A. Limitation of capsule networks[J]. Pattern Recognition Letters, 2021, 144: 68-74. https://doi.org/10.1016/j.patrec.2021.01.017.

[19] Haq M U, Sethi M A J, Rehman A U. Capsule network with its limitation, modification, and applications—A survey[J]. Machine Learning and Knowledge Extraction, 2023, 5(3): 891-921. https://doi.org/10.3390/make5030047.

[20] Mazzia V, Salvetti F, Chiaberge M. Efficient-CapsNet: capsule network with self-attention routing[J]. Scientific Reports, 2021, 11(1): 14634. https://doi.org/10.1038/s41598-021-93977-0.

[21] Yang H, Han J, Min K. A multi-column CNN model for emotion recognition from EEG signals[J]. Sensors, 2019, 19(21): 4736. https://doi.org/10.3390/s19214736.

[22] ZHANG D, YAO L, ZHANG X, et al., Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1)[2023-11-12]. https://doi.org/10.1609/aaai.v32i1.11496.

[23] Hwang S, Hong K, Son G, et al. Learning CNN features from DE features for EEG-based emotion recognition[J]. Pattern Analysis and Applications, 2020, 23(3): 1323-1335. https://doi.org/10.1007/s10044-019-00860-w.

[24] Khademi Z, Ebrahimi F, Kordy H M. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals[J]. Computers in Biology and Medicine, 2022, 143: 105288. https://doi.org/10.1016/j.compbiomed.2022.105288.

[25] Shen F, Dai G, Lin G, et al. EEG-based emotion recognition using 4D convolutional recurrent neural network[J]. Cognitive Neurodynamics, 2020, 14(6): 815-828. https://doi.org/10.1007/s11571-020-09634-1.

[26] Wang Y, Huang Z, McCane B, et al. EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition[C]//2018 International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro: IEEE, 2018: 1-7[2023-11-12]. 10.1109/IJCNN.2018.8489715.

[27] Tao W, Li C, Song R, et al. EEG-based emotion recognition via channel-wise attention and self attention[J]. IEEE Transactions on Affective Computing, 2023, 14(1): 382-393. https://doi.org/10.1109/TAFFC.2020.3025777.

[28] Huang Z, Ma Y, Wang R, et al. A model for EEG-based emotion recognition: CNN-Bi-LSTM with attention mechanism[J]. Electronics, 2023, 12(14): 3188. https://doi.org/10.3390/electronics12143188.

[29] Liu J, Zhao Y, Wu H, et al. Positional-spectral-temporal attention in 3D convolutional neural networks for EEG emotion recognition[C]//2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2021: 305-312.. http://arxiv.org/abs/2110.09955.

[30] Lin X, Chen J, Ma W, et al. EEG emotion recognition using improved graph neural network with channel selection[J]. Computer Methods and Programs in Biomedicine, 2023, 231: 107380. https://doi.org/10.1016/j.cmpb.2023.107380.

[31] Chao H, Dong L, Liu Y, et al. Emotion recognition from multiband EEG signals using CapsNet[J]. Sensors, 2019, 19(9): 2212. https://doi.org/10.3390/s19092212.

[32] Liu Y, Ding Y, Li C, et al. Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network[J]. Computers in Biology and Medicine, 2020, 123: 103927. https://doi.org/10.1016/j.compbiomed.2020.103927.

[33] Wei Y, Liu Y, Li C, et al. TC-Net: A transformer capsule network for EEG-based emotion recognition[J]. Computers in Biology and Medicine, 2023, 152: 106463. https://doi.org/10.1016/j.compbiomed.2022.106463.

[34] Wang K, He R, Wang S, et al. The Efficient-CapsNet model for facial expression recognition[J]. Applied Intelligence, 2023, 53(13): 16367-16380. https://doi.org/10.1007/s10489-022-04349-8.

[35] Wang H, Zhang T, Cheung K M C, et al. Application of deep learning upon spinal radiographs to predict progression in adolescent idiopathic scoliosis at first clinic visit[J]. eClinicalMedicine, 2021, 42: 101220. https://doi.org/10.1016/j.eclinm.2021.101220.

[36] Zou B, Cao C, Wang L, et al. FACILE: A capsule network with fewer capsules and richer hierarchical information for malware image classification[J]. Computers & Security, 2024, 137: 103606. https://doi.org/10.1016/j.cose.2023.103606.

[37] Maitre J, Bouchard K, Gaboury S. Data filtering and deep learning for enhanced human activity recognition from UWB radars[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(6): 7845-7856. https://doi.org/10.1007/s12652-023-04596-8.

[38] Kristianto W, Candra H. EEG–based emotion classification using convolutional neural networks[C]//2019 2nd International Conference on Applied Engineering (ICAE). Batam, Indonesia: IEEE, 2019: 1-4[2023-07-12]. https://ieeexplore.ieee.org/document/9221673/.

[39] Duan R N, Zhu J Y, Lu B L. Differential entropy feature for EEG-based emotion classification[C]//2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). San Diego, CA, USA: IEEE, 2013: 81-84[2023-11-12]. http://ieeexplore.ieee.org/document/6695876/.

[40] Cui F, Wang R, Ding W, et al. A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition[J]. Mathematics, 2022, 10(4): 582. https://doi.org/10.3390/math10040582.

[41] Wang Z M, Hu S Y, Song H. Channel selection method for EEG emotion recognition using normalized mutual information[J]. IEEE Access, 2019, 7: 143303-143311. https://doi.org/10.1109/ACCESS.2019.2944273.

[42] Wei-Long Zheng, Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks[J]. IEEE Transactions on Autonomous Mental Development, 2015, 7(3): 162-175. https://doi.org/10.1109/TAMD.2015.2431497.

[43] Koelstra S, Muhl C, Soleymani M, et al. DEAP: A database for emotion analysis using physiological signals[J]. IEEE Transactions on Affective Computing, 2012, 3(1): 18-31. https://doi.org/10.1109/T-AFFC.2011.15.

[44] Pan C, Shi C, Mu H, et al. EEG-based emotion recognition using logistic regression with Gaussian kernel and Laplacian prior and investigation of critical frequency bands[J]. Applied Sciences, 2020, 10(5): 1619. https://doi.org/10.3390/app10051619.

[45] Liu S, Wang Z, An Y, et al. DA-CapsNet: A multi-branch capsule network based on adversarial domain adaption for cross-subject EEG emotion recognition[J]. Knowledge-BasedSystems, 2024, 283: 111137. https://doi.org/10.1016/j.knosys.2023.111137.

[46] Li C, Wang B, Zhang S, et al. Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism[J]. Computers in Biology and Medicine, 2022, 143: 105303. https://doi.org/10.1016/j.compbiomed.2022.105303.

[47] Liu S, Wang Z, An Y, et al. EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network[J]. Knowledge-Based Systems, 2023, 265: 110372. https://doi.org/10.1016/j.knosys.2023.110372.

[48] Chen K, Jing H, Liu Q, et al. A novel caps-EEGNet combined with channel selection for EEG-based emotion recognition[J]. Biomedical Signal Processing and Control, 2023, 86: 105312. https://doi.org/10.1016/j.bspc.2023.105312.

[49] Song Y, Yin Y, Xu P. A customized ECA-CRNN model for emotion recognition based on EEG signals[J]. Electronics, 2023, 12(13): 2900. https://doi.org/10.3390/electronics12132900.

[50] Li T, Fu B, Wu Z, et al. EEG-based emotion recognition using spatial-temporal-connective features via multi-scale CNN[J]. IEEE Access, 2023, 11: 41859-41867. https://doi.org/10.1109/ACCESS.2023.3270317.

[51] Li J, Hua H, Xu Z, et al. Cross-subject EEG emotion recognition combined with

connectivity features and meta-transfer learning[J]. Computers in Biology and Medicine, 2022, 145: 105519. https://doi.org/10.1016/j.compbiomed.2022.105519.