Physica A 603 (2022) 127700

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A Transformer based neural network for emotion recognition and visualizations of crucial EEG channels

Jia-Yi Guo [a], Qing Cai [a], Jian-Peng An [a], Pei-Yin Chen [a], Chao Ma [a,*], Jun-He Wan [b,*], Zhong-Ke Gao [a]

[a] School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
[b] Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong A Sciences), No. 2, Huiying street, Qingdao 266318, China

ARTICLE INFO

Article history:
Received 14 February 2022
Received in revised form 15 April 2022
Available online 18 June 2022

Keywords:
EEG
Emotion recognition
Transformer model
Deep learning
Time series analysis

ABSTRACT

With the rapid development of artificial intelligence and sensor technology, electroencephalogram-based (EEG) emotion recognition has attracted extensive attention. Various deep neural networks have been applied to it and achieved excellent results in classification accuracy. Except for classification accuracy, the interpretability of the feature extraction process is also considerable for model design for emotion recognition. In this study, we propose a novel neural network model (DCoT) with depthwise convolution and Transformer encoders for EEG-based emotion recognition by exploring the dependence of emotion recognition on each EEG channel and visualizing the captured features. Then we conduct subject-dependent and subject-independent experiments on a benchmark dataset, SEED, which contains EEG data of positive, neutral, and negative emotions. For subject-dependent experiments, the average accuracy of three classification tasks is 93.83%. For subject-independent experiments, the average accuracy of three classification tasks is 83.03%. Additionally, we assess the importance of each EEG channel in emotional activities by the DCoT model and visualize it as brain maps. Furthermore, satisfactory results are obtained by utilizing eight selected crucial EEG channels: FT7, T7, TP7, P3, FC6, FT8, T8, and F8, both in two classification tasks and three classification tasks. Using a small number of EEG channels for emotion recognition can reduce equipment costs and computing costs, which is suitable for practical applications.

1. Introduction

Human emotion is a psychological and physiological reaction accompanied by the process of human consciousness. Emotion plays a significant role in daily communication, which affects people's work efficiency and lifestyle. There is no doubt that accurate identification of people's emotions and emotion analysis will make people's life convenient and promote social development. With the rapid growth of artificial intelligence technology, for instance, cognitive science, computer science, artificial intelligence, and so on, emotion recognition [1] has become a popular investigation in many fields. Establishing exact as well as reliable computational systems is crucial for developing practical applications of emotion recognition, such as drivers' emotion detection, psychotherapy, and criminal investigation. We can recognize people's emotions by studying their physical appearance, for example, facial expressions [2–4], body language [5], etc.

* Corresponding authors.
E-mail addresses: chao.ma@tju.edu.cn (C. Ma), wan_junhe@qlu.edu.cn (J.-H. Wan).

https://doi.org/10.1016/j.physa.2022.127700
0378-4371/

Besides, physiological signal recognition is also a suitable method for recognizing people's emotions, as the physiological signals cannot be concealed compared with physical appearance. Currently, with the in-depth development of dry electrode techniques [6], wearable devices applications [7], and computer technologies, it is common to use EEG signals to reflect affective activities in various psychophysiology studies [8–12].

Recently, various studies have put efforts into exploring efficient methods to recognize emotions via nonlinear and complex EEG signals. Studies show that introducing handcraft feature extraction to classification models can improve emotion recognition performance to a certain extent. A variety of handcraft features have been employed with the excellent ability to reinforce the characteristics of different emotions. For example, higher-order crossing features [13] and Hjorth features [14], which belong to time-domain features, are capable of extracting the temporal information of signals. Power spectral density (PSD) [15], wavelet transform [16,17], discrete wavelet transform [18], and so on, with the ability to capture local features of the frequency domain, have been utilized to process EEG signals. Besides, establishing brain networks also have been employed as a feature extraction method by exploring the relationships between EEG channels. For instance, some works are based on the Pearson correlation coefficient [19], mutual information [20], and so on to construct brain networks, which are used for emotion recognition with models. Additionally, many works have adopted entropy measures to extract discriminative features by measuring the complexity of EEG signals, such as Shannon entropy (ShEn) [21], sample entropy (SampEn) [22], differential entropy (DE) [23], and so on. Differential entropy, with superior robustness and feature extraction ability, has been wildly used in analyzing EEG signals, especially in EEG-based emotion recognition [24,25]. We apply the DE feature in this work.

In the existing studies on exploring effective methods for EEG-based emotion recognition, machine learning models have achieved good results. Atkinson et al. [26] used a support vector machine (SVM) for two-category classification (Categories are divided by the arousal and valence) and have achieved a classification accuracy of 73.06%. Liu et al. [27] achieved accuracies of 69.9% and 71.2% by applying a random forest (RF) model and a KNN model in two classification tasks. Besides, Tuncer et al. [28] proposed a fractal pattern feature extraction approach and multi-machine learning method for emotion recognition. With the development of machine learning, it has been demonstrated that deep learning is outstanding in dealing with complex problems by virtue of its stronger learning ability and has been introduced into EEG-based emotion recognition. Zheng et al. [24] put forward a multi-frequency bands method to detect crucial EEG frequency bands of emotion recognition with deep belief networks (DBN). The finding suggests that the gamma band (30–45 Hz) outperforms other frequency bands for emotion detection, and it attains an average subject-dependent accuracy of 86.65% on the SEED dataset, an openly available dataset. Zhang et al. [29] introduced the spatial–temporal recurrent neural network (STRNN) model for EEG-based emotion detection. They also split the EEG signals into five frequency bands for extracting DE features. The subject-dependent experiments are based on the SEED dataset and have achieved an average maximum recognition result of 89.5%. Besides, Krishna et al. [30] proposed a unique mixture model, which is well-performed on EEG signals interfered by noises based on asymmetric distribution.

Except for the aforementioned works in subject-dependent emotion recognition, various studies have recently conducted numerous attempts to design models for subject-independent emotion recognition, which can extract common features to overcome individual differences. For instance, Song et al. [25] proposed a dynamical graph convolutional neural networks (DGCNN) model on multi-frequency bands EEG data. The work suggests that the DGCNN model can achieve the maximum average accuracies of 90.4% and 79.95% for subject-dependent and subject-independent experiments in the SEED dataset. Li et al. [31] introduced a combination model of a variational autoencoder and long short-term memory unit (LSTM) into EEG decoding for subject-independent emotion detection. They have achieved an average classification result of 85.81% for positive and negative emotions recognition. In addition, Komolovaitė et al. [32] developed the Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) to deal with limited EEG samples and has achieved a good performance on the EEG-based emotion recognition. Furthermore, a deep CNN model was designed by Maheshwari et al. [33] for real-time automated EEG-based affective computing, and the system has achieved good results on the DEAP dataset. Siddharth et al. [34] proposed a transfer learning-based multi-modal, which is capable of overcoming the inconsistencies between different datasets.

The attention mechanism has been a focus in deep learning fields since it was proposed by Bahdanau et al. [35]. The core thought of the attention mechanism is to help the neural network focus on the significant part of input data to improve learning efficiency, which makes its application attains excellently achievement in Natural Language Processing (NLP) fields, Computer Vision (CV) fields, and so on. Then, a well-known self-attention-based model, i.e., Transformer, was proposed by Vaswani et al. [36], and the model has caused a leap forward in capabilities for sequence modeling in NLP. Due to its great success, the Transformer has been introduced into translation, image generation, and other fields. The utilization of the attention mechanism in the Transformer helps capture the long-term dependences of data and also increases interpretability. Furthermore, Li et al. [37] introduced the attention mechanism into the deep learning models, capturing the interdependence of informative signals. Chen et al. [38] applied the attention mechanism to capture deep-level features from EEG signals in brain–computer interface fields.

In this work, inspired by the Vision Transformer model [39], we propose a novel model, DCoT, composed of depthwise convolution and Transformer encoders for emotion recognition. The DCoT model combines the advantage of convolution in capturing low-level features of different EEG frequency bands and the advantages of self-attention in associating long-term dependencies as well as exploring global features. Besides, the attention mechanism allows us to discover the crucial EEG channels in emotion recognition, and it makes the learning process of the DCoT more interpretable. This paper's main contributions are illustrated as follows:

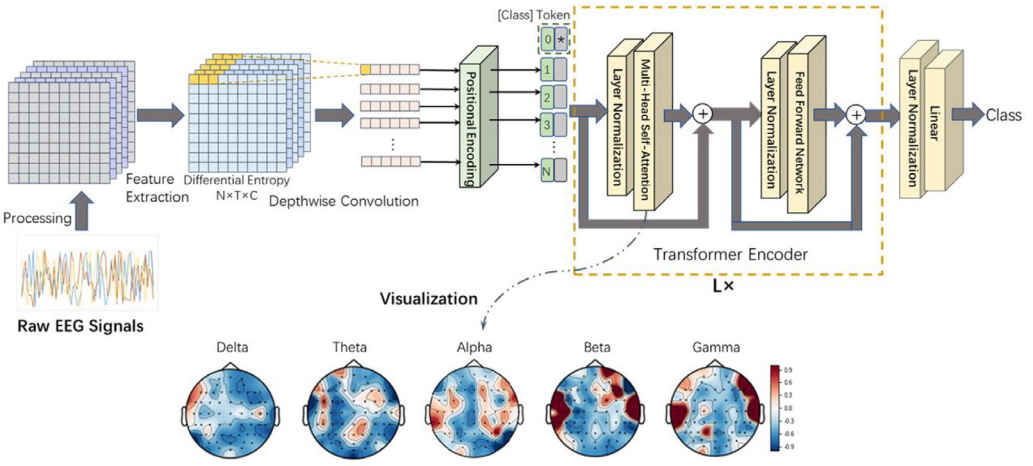

Fig. 1. The structure of the proposed DCoT model for emotion recognition.

(1) A novel model, DCoT, is proposed in this work for EEG-based emotion recognition tasks. The application of convolutional structure and Transformer encoders makes the model capable of extracting both local features and the global dependences of emotion recognition on different EEG channels. Experimental results display that our DCoT model outperforms the previous works on recognition accuracy considered in the comparison.

(2) The depthwise convolution is introduced to fuse multi-frequency-domain information of EEG signals. This enables the DCoT model to enhance feature extraction capability while maintaining independence between EEG channels. Experiments show that considering the independence between each EEG channel while designing models can improve the performance of EEG classification.

(3) The learning process of the model is more interpretable, as the learning process is the exploration of the importance of each EEG channel in emotion recognition. We can visualize the extracted features by the DCoT model as brain maps, and the active brain areas in emotional activities are displayed. This is an advantage over other existing methods.

2. Methodology

2.1. Handcraft feature extraction

Differential entropy (DE) as a nonlinear entropy measure has shown outstanding performance for EEG signal recognition, especially for EEG signals of emotions [23,24]. The stander differential entropy is defined as:

$$h(X) = -\int_x f(x)\log(f(x))dx, \tag{1}$$

where X is a random variable, and $f(x)$ represents the probability density function of X. As EEG signals have been proved to obey the Gaussian distribution $N\left(\mu, \sigma^2\right)$ [23], the differential entropy can be calculated as:

$$h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp\frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} exp\frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2}\log 2\pi e\sigma^2, \tag{2}$$

Meanwhile, for a fixed-length EEG segment, differential entropy is equivalent to the logarithm energy spectrum in a certain frequency band. We extract DE features from five EEG frequency bands (delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, and gamma: 31–50 Hz), respectively, utilizing 256-point Short-Time Fourier Transform with a nonoverlapped Hanning window of one second. Therefore, we get $N \times 1 \times C$ dimensions DE features per second, where N means the number of EEG channels and C is the number of EEG bands, and attain $N \times T \times C$ dimensions DE features for a T seconds length sample.

2.2. Model

Exploring the dependence of emotion recognition on each EEG channel and active brain areas under different emotions are valuable investigations for emotion recognition. In this paper, we propose the DCoT model to provide an efficient way to capture the interdependence hidden between the multiple EEG channels while recognizing emotions.

An overview of the DCoT model is illustrated in Fig. 1. The model input is represented as $X \in \mathbb{R}^{N\times T\times C}$, where N denotes the number of EEG Channels, T denotes the length of an EEG sample, C denotes the number of EEG frequency bands of

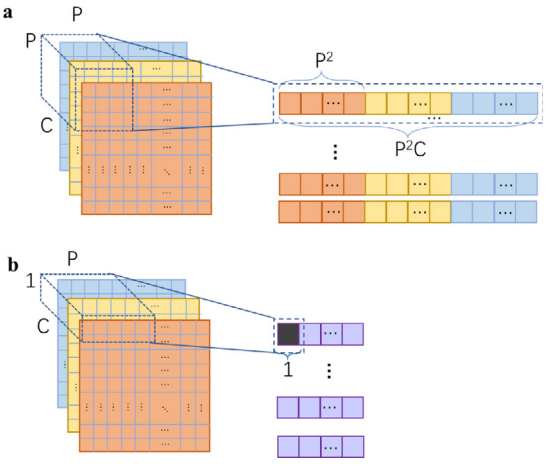

Fig. 2. (a) An overview of the canonical patch embedding process. (b) An overview of the depthwise convolution process in the DCoT model.

DE features. In model design, we based on the original Vision Transformer model [39] proposed in the computer vision (CV) fields, composed of multiple stacking with identical compositions. The essential components of the DCoT include depthwise convolution (DW-CONV) layer, position embeddings, learnable embeddings, Transformer encoders, and linear layers. Besides, the Transformer encoders consist of layer normalizations (LN), multi-head self-attention (MSA) layers, and feed-forward networks (FFN).

2.2.1. Depthwise convolution layer

The input DE features are fed into a depthwise convolution layer, and this layer plays a significant role in the model. The DCoT model is able to extract complete information for multi-frequency data with the depthwise convolution layer. For the canonical Transformer based models, for instance, the Vision Transformer model, the input data $y \in \mathbb{R}^{H\times W\times C}$ is segmented into n patches $\left\{x_p^1, x_p^2, \ldots, x_p^i, \ldots, x_p^n\right\}$, where $x_p^i \in \mathbb{R}^{P\times P\times C}$, then these patches are reshaped into n flattened vectors $\left\{x_L^1, x_L^2, \ldots, x_L^i, \ldots, x_L^n\right\}$ as shown in Fig. 2a, where $x_L^i \in \mathbb{R}^{P^2C}$. Nevertheless, it makes the DE feature sequences time-incoherent that simply concatenates DE features of different frequencies in the time dimension. Such operations cause the loss of temporal order characteristics as well as the interferences at the joint. Hence, the convolution layer is introduced to eliminate the negative influences. Implementing the convolutional layer further fuses the DE features of different frequencies and generates new features. As the convolution kernels move along the time dimension, we achieve a time-coherent feature matrix, which is shown in Fig. 2b. In this way, not only inherent relationships between different frequency bands at the same data point are considered, but also these unnecessary outside interferences are avoided. Second, with the excellent ability to capture local features, the utilization of the convolution layer also can capture frequency domain features from per EEG channel and decrease the computations of the following encoder blocks. To summarize, the depthwise convolution layer application provides more effective feature information, which can enhance the emotion recognition results.

Meanwhile, to ensure the independence of the features of each EEG channel, the canonical convolution is not applicable, so the depthwise convolution is employed. In the computer vision field, depthwise convolution is introduced to extract the image features of red (R), green (G), and blue (B) channels. Hence, we utilize the depthwise convolution layer to further extract features of the frequency-temporal DE features in different EEG channels, respectively. In this case, we regard the EEG channels as the image channels.

There are N convolutional kernels in the depthwise convolution layer with the size of $C\times f$ where C is in the frequency domain, and f is in the time domain, and the stride is s. Then, the output is a feature matrix $X_0 = \left(x_0^1, x_0^2, \ldots, x_0^N\right)^T$ with the size of $N \times D_f$, where $D_f = 1 + (T - f)/s$. We use N to present the number of EEG channels. To maintain the relative independence of each channel, our DCoT model splits the EEG feature matrix $X_0 \in \mathbb{R}^{N\times D_f}$ into a sequence of one-dimensional vectors according to EEG channels. Therefore, the following Transformer encoder obtains N inputs. The process is described more clearly in Fig. 1. The green boxes followed by the depthwise convolution layer in Fig. 1 are feature sequences of different EEG channels. The depthwise convolution layer will be called the DW-CONV layer for short in the following paper.

2.2.2. Position embedding and learnable embedding

Position embedding. To import the input orders of feature vectors of different channels, we add one-dimensional learnable position embeddings to the input sequences of the encoders to hold positional information. The gray boxes

Table 1
Details of the DCoT model hyperparameters.

Model	Convolution kernel size	Encoder layers	Self-attention heads	FFN dimension
DCoT	5 × 3	5	8	1024

in Fig. 1 represent the position embeddings. The formulations of the position embeddings are based on sine and cosine functions of different frequencies:

$$PE\,(pos, 2i) = \sin(pos/10000^{2i/d_{model}}),$$
$$PE\,(pos, 2i+1) = \cos(pos/10000^{2i/d_{model}}), \tag{3}$$

where d_{model} represents the number of input sequences, $pos \in \{0, 1, \ldots, N-1\}$ is the sequence order, and i indicates the dimension of each input sequence. Therefore, a sinusoid with wavelengths from 2π to $10000 \cdot 2\pi$ is assigned to each dimension of positional encoding, and the sequence relative positions can be recorded.

Learnable embedding. An extra [class] token is prepended to the input sequence called learnable embedding, shown as the box with '*' in Fig. 1. As known in NLP, BERT's [class] token means the whole sentences represent after encoding. Similarly, we prepend a learnable embedding x_{class}, which serves as the EEG feature representation. Besides, the output of the Transformer encoder Z_L^0 is utilized to represent x_{class}.

2.2.3. Transformer encoders

The intrinsic associations between EEG channels are mainly explored by a series of stacked Transformer encoders, and each of them consists of two major components: MSA and FFN. We apply residual connections around MSA and FFN, respectively, as shown in Fig. 1, followed by layer normalization. We utilize five layers of the encoder in this paper.

MSA. To understand the multi-head self-attention layer, we first describe how the self-attention (SA) layer operates. When the SA layer processes a feature vector of one position, the SA layer allows this sequence to associate with others. In other words, the SA layer integrates the sequences of all the relevant channels into the sequence processing. In order to implement this process, the input vector x_p is linearly transformed into QKV space, i.e., queries Q, keys K, and values V. These matrixes are linear projections of input sequences $X \in \mathbb{R}^{(N+1)\times D_f}$. Hence, we obtain $Q = W_q X$, $Q \in \mathbb{R}^{(N+1)\times d_k}$, $K = W_k X$, $K \in \mathbb{R}^{(N+1)\times d_k}$, and $V = W_v X$, $V \in \mathbb{R}^{(N+1)\times d_v}$, where $\mathrm{W}_q \in \mathbb{R}^{D_f \times d_k}$, $\mathrm{W}_k \in \mathbb{R}^{D_f \times d_k}$ and $\mathrm{W}_v \in \mathbb{R}^{D_f \times d_v}$ are trainable parameters. The computational process is defined as:

$$SA\,(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \tag{4}$$

where SA is the output of the SA layer. By extending the SA, we get the MSA, which applies several SA heads parallel to learning different interdependencies. For multi-head self-attention, h groups of matrixes Q, K, V are obtained by using h different groups of learnable parameters W_q, W_k, W_v (in the DCoT $h = 8$). Then the input embedding is mapped to h different representation subspaces, which means we attain h output from the SA layer. Then we concatenate these matrixes and multiply them by the learnable parameter W^0 to get the final MSA layer output:

$$MSA(X) = concat\,(SA_1, \ldots, SA_h)\, W^0, \tag{5}$$

FFN. Feed-forward neural networks are applied with powerful learning abilities. Per Feed-forward neural network consists of two fully connected layers, and each one is followed by a dropout layer. The dropout value is set as 0.2. Notably, the activation function is a GELU non-linearity. The FFN dimension is 1024 in this method.

To sum up, the sequence of input of the first Transformer encoder is marked as Eq. (6):

$$\mathrm{Z}_0 = \left[x_{class}; x_0^1, x_0^2, \ldots, x_0^N\right] + E_{pos}, \tag{6}$$

where $x_0^1, x_0^2, \ldots, x_0^N$ are the output sequences of the DW-CONV layer, X_{class} is the learnable embedding, and E_{pos} is the position embedding.

Consequently, the output of each encoder is :

$$Z_l' = MSA\,(LN\,(Z_{l-1})) + Z_{l-1}, \quad l = 1, 2, \ldots, L \tag{7}$$
$$Z_l = MLP\left(LN\left(Z_l'\right)\right) + Z_l', \quad l = 1, 2, \ldots, L \tag{8}$$

where Z_{l-1} means the input of the $l-1$ encoder, and Z_l represents the output of this encoder. Besides, $L = 5$ encoders are utilized in the DCoT. Table 1 describes the values of some major hyperparameters of the DCoT model. Furthermore, to achieve a minimum loss during model training, we combine the adaptive gradient descent optimizer with the learning rate which is based on the cosine annealing algorithm. The maximum learning rate is 0.00008.

3. Dataset

3.1. Dataset

We validate the effectiveness of the proposed approach based on the SJTU Emotion EEG Dataset [24] (SEED) in the following experiments. Fifteen subjects (7 males and 8 females, 23.27 $\pm$ 2.37 years) took part in emotional experiments to collect the EEG data, and each subject's high-quality EEG signals were captured by watching 15 emotionally stimulating film clips (each is about 4 min long). These films clips were selected to stimulate positive emotion, neutral emotions emotion and negative emotion, and each had five corresponding film clips. Immediately after watching each movie clip (trial), these participants were asked to record their emotional reactions to the movie clip by answering a questionnaire for feedback. Each subject conducted this experiment every two weeks for a total of three times. Therefore, there are 15 trials per session for each subject and a total of 3 sessions for each subject.

3.2. Processing

The EEG data as a neurophysiology signal collected from electrodes under various environments has characteristics of high dimensionality, noise, and redundancy. Hence, before the data analysis, we need to preprocess the raw EEG data. The default preprocessing technique for the SEED dataset is as follows: (1) The raw EEG signals (1000 Hz) are downsampled to 200 Hz; (2) Independent Component Correlation Algorithm (ICA) is applied to remove the EOG signal as well as blink artifacts interference. (3) A bandpass filter between 0–50 Hz is utilized to process the EEG data to eliminate noise. (4) Selecting a proper size for the sliding window in signal segmentation can provide sufficient samples for training and let segments have enough data points for feature extraction. The EEG signals corresponding to the duration of each movie clip (about 4 min length) are extracted, and the data of 15 trails (movie clips) per session are split into the same-length 10s samples without overlapping, respectively. Additionally, we use the data in the range of 1000–37000 epochs of each trial. Subsequently, five rhythms are extracted, which are termed delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, and gamma: 31–50 Hz.

4. Experiment

4.1. Subject-dependent experiments and results

We first conduct subject-dependent experiments to display the emotion recognition capability of our model. In this section, we compare the performances of the model under five different frequency bands and all five bands. Additionally, to demonstrate the DW-CONV layer effectivity, we conduct an ablation experiment in which we utilize a DCoT model without the DW-CONV layer. We utilize the samples of 12 trials per session of each subject as training data while using the other 3 trials per session as testing data. 10-fold cross-validation is applied to select the most appropriate hyperparameters. After exploring the most appropriate hyperparameters, we train the model with the training data of each subject and select the best-trained models for 15 subjects, respectively. Then, we measure the classification capabilities of these models on testing data of each subject. The average accuracies of 15 subjects represent the model performance under different classification tasks. Furthermore, we conduct 10 experiments with random initialization of the model parameters for each subject to demonstrate the reproducibility of the results. We set up 10 different random seeds of the training model and recorded the testing results. We adopt the average accuracies as the experiment results.

Fig. 3 presents the accuracies of identifying negative and positive emotions in different frequency bands. The bar chart shows that the recognition accuracies are higher when using the DW-CONV layer than without using the DW-CONV layer in all six cases. The bar charts directly display that the response capability of EEG signals to emotional activities varies from frequency band to frequency band, and using EEG features in all bands can contain more information to improve emotion recognition capability. In addition, the model with the DW-CONV layer has a better performance in every frequency band than without it, which demonstrates the efficiency of the introduction of the DW-CONV layer. In addition, we apply two other metrics to further demonstrate the performance of our method in the emotion recognition task: F1-score (F1) and area under the curve (AUC) value. Our method achieves an F1-score of 99.91% and an AUC value of 99.89% in two classification tasks.

The recognition results of positive, neutral, and negative emotions are described in the bar chart in Fig. 4. It is similar to the results above that our model performs best when using all bands of EEG data. Moreover, the gamma band has always been the best in classification accuracy in contrast to the other four bands, which displays that EEG signals in the gamma band can reflect more specific information in emotional activities. In addition, more comprehensive features are included while using EEG data in all frequency bands. Additionally, the average classification accuracy and standard deviation of DW-CONV is 93.83%/3.63% while using all frequency bands data, and without DW-CONV are 86.79%/3.94%. Obviously, the performances of applying the DW-CONV layer are more prominent in three emotion classifications than in two emotion classifications. Consequently, the gap between emotion recognition capabilities of using the DW-CONV layer and without using the DW-CONV layer is wider in three classification tasks. As using the DW-CONV layer in our model is an optimal

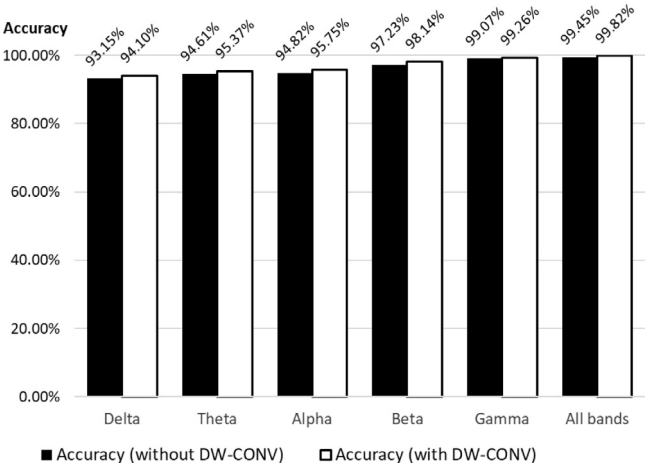

Fig. 3. The averaged recognition accuracies of positive and negative emotions under different EEG frequency bands. The dark bar charts represent the model performance without depthwise convolution, and the light bar charts represent the model performance with depthwise convolution.

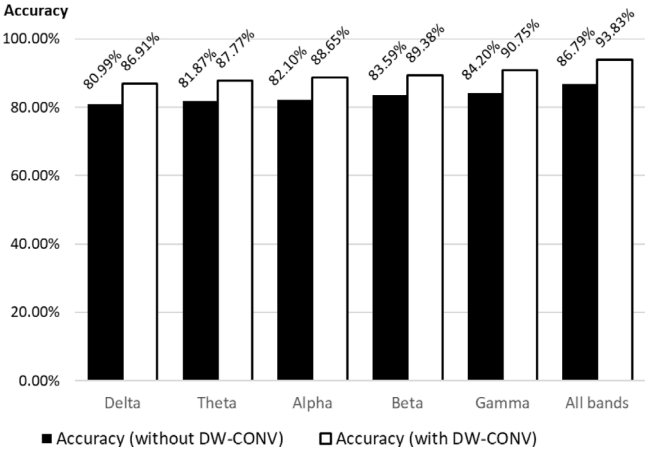

Fig. 4. The averaged recognition accuracies of positive, neutral, and negative emotions under different EEG frequency bands. The dark bar charts represent the model performance without depthwise convolution, and the light bar charts represent the model performance with depthwise convolution.

choice, all experiments are based on utilizing the DW-CONV layer in the following sections. Fig. 5(a) displays the confusion matrix of three emotions in subject-dependent experiments. The macro averaging F1-score (MF1) and macro-averaging AUC values are 93.78% and 97.46%, respectively. For natural, positive, and negative emotions, the AUC values are 93.91%, 99.35%, and 98.73%, respectively, which suggest that the proposed method has a high discrimination ability on the positive emotion, which is consistent with other works.

4.2. Subject-independent experiments and results

In this section, in order to measure the capability of the DCoT model in cross-subjects emotion recognition tasks, we adopt a leave-one subject-out (LOSO) cross-validation strategy. Specifically, in the LOSO cross-validation, the EEG data of 14 subjects are used as the training dataset, and the EEG data of the remaining subject is used as the testing dataset. Hence, for the SEED dataset, this experiment is conducted in 15 runs. The average classification accuracies based on DE features in all frequency bands are calculated as the model performance.

Additionally, we also conduct 10 experiments with random initialization of the model parameters for each subject. For the 10 random initialization experiments, the average accuracy of two classification tasks is distributed between 87% and 90%. The average accuracy of three classification tasks is distributed between 81% and 85% for random initialization experiments. We adopt the average accuracies as the experiment results. Table 2 summarizes the cross-subjects recognition accuracies in two classification tasks and three classification tasks. In addition, the proposed model achieves an average accuracy of 88.37% with a standard deviation of 5.26% in two classification tasks and an average

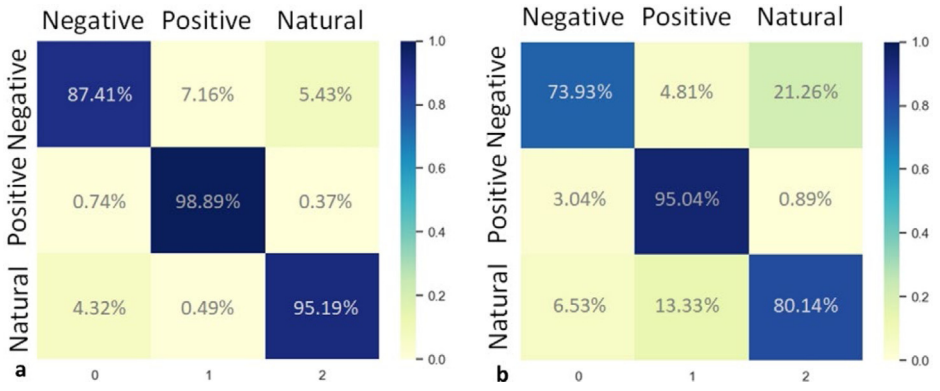

Fig. 5. The confusion matrixes of three emotions of subject-dependent experiments (a) and subject-independent experiments (b).

Table 2
The cross-subjects emotion recognition performance with the DCoT model.

Class	Accuracy							
	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8
Pos-Neg	88.38%	83.95%	84.51%	87.98%	83.67%	84.76%	90.92%	88.49%
	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Mean
	86.17%	79.58%	93.47%	94.91%	83.79%	97.64%	97.37%	88.37%
	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8
Pos-Neu-Neg	82.96%	77.67%	79.63%	84.83%	77.97%	80.91%	89.09%	84.16%
	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Mean
	82.95%	70.48%	86.65%	89.84%	76.76%	90.07%	91.43%	83.03%

Table 3
Performance comparison between this work and the state-of-art literature.

Work	Model	ACC (Pos-Neg)	ACC (Pos-Neu-Neg)
–	SVM	–	56.73%
Collobert et al. [40]	T-SVM	–	72.52%
Li et al. [41]	Automatic Feature Selection+SVM	83.33%	–
Song et al. [25]	DGCNN	–	79.95%
Li et al. [32]	VAE+LSTM	85.81%	–
Zhang et al. [42]	CNN-DDC	–	82.1%
Li et al. [43]	BiDANN	–	83.28%
Cimtay and Ekmekcioglu [44]	Pretrained InceptionResnetV2	86.5%	78.3%
Javier Fdez et al. [45]	Neural Networks With Stratified Normalization	–	79.6%
Proposed work	DCoT	88.37%	83.03%

accuracy of 83.03% with a standard deviation of 5.70% in three classification tasks. In addition, other matrixes are applied to display our model performance. In two classification tasks, the proposed method yields an F1-score of 88.81% and an AUC value of 98.72%. In three classification tasks, the model obtains an MF1-score of 83.18% and a macro-averaging AUC value of 95.37%. The confusion matrix of three emotions in subject-independent experiments is depicted in Fig. 5(b). The AUV values are 86.17%, 96.50%, and 93.25% for negative, positive, and natural emotions. Similar to the subject-dependent experiments, positive emotion yields the best performance.

Additionally, we compare the methods of other works to demonstrate the superiority of the DCoT model in emotion recognition. Table 3 lists the highly cited state-of-art works and the corresponding performance attained. For two classification tasks, the average accuracy of our work ranks first in the table. For three classification tasks, the average accuracy of our work is in the second 0.25% lower than the first one. We think it is due to the fact that the number of samples is limited, as Transformer structures perform better on large datasets [37]. For instance, most models with transformer structures perform better on datasets with about millions of samples in the computer vision field. We believe that our method will have improved performance in practical applications where a huge amount of samples can be obtained easily.

In addition, the proposed model has another advantage, a more interpretable feature extraction process, from which we can visualize the deep-level features and present the significant brain areas in emotional activities. The brain science experts can evaluate the reliability of the emotion recognition results by combining the shown active brain area and the

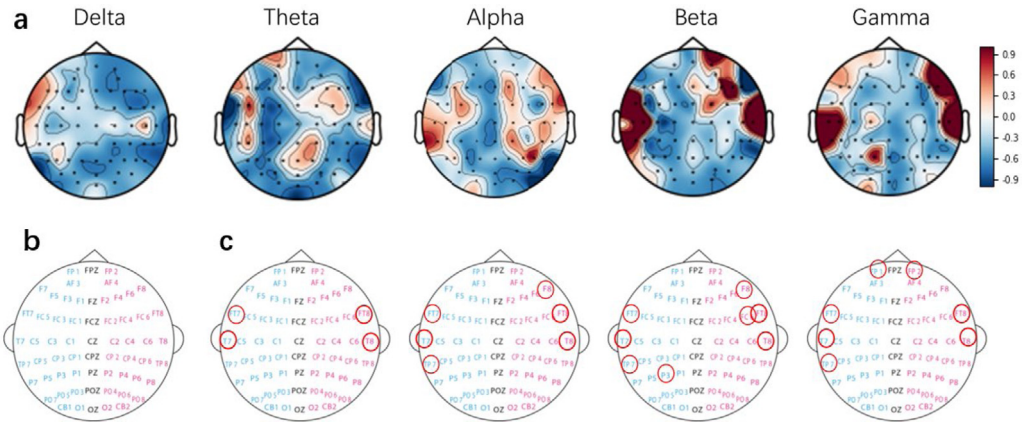

Fig. 6. The crucial EEG channels visualization. (a) The importance of each EEG channel captured by the DCoT model. (b) The positions of EEG electrodes. (c) The selected crucial EEG channels.

clinical brain science knowledge, and this is a unique application provided by the proposed method. The details will be depicted in the following section.

4.3. Crucial EEG channels identify and visualization

As we know, many EEG signal classifiers in state-of-the-art works of emotion recognition are black-box models, in which the learning processes are difficult to interpret. We can attain classification accuracies from these classifiers, but we do not know how to explain the meaning of deep features extracted by these models. For instance, many previous studies consider the multi-channel EEG signals as graphs and applied CNNs for classification. However, the in-depth local features extracted from the EEG signals via CNNs are hard to interpret its meaning in these works. In order to reveal the concerns in the process of model learning different emotions and make the classification process interpretable, the DCoT model is designed. The DCoT model focuses on exploring the hidden interrelations between EEG channels and the dependences of emotion recognition on each EEG channel. The model structures make it easy for us to visualize the captured deep-level features, and these features are the importance of each EEG channel in detecting emotional activities. Fig. 6(a) displays the visualization of critical EEG channels and brain areas for emotion recognition.

From the multi-head self-attention (MSA) layer, we can obtain the importance of each input sequence of this layer while classifying. The MSA layer inputs are the features of different EEG channels. Thus, from the MSA layer outputs, we attain the importance of each EEG channel while recognizing emotions. Fig. 6(a) illustrates the distribution of the critical EEG channels in five frequency bands, which displays the active brain areas in emotional activities. The redder the EEG channel (brain area), the more important it is in emotional activities. Obviously, the lateral temporal lobe and the prefrontal lobe are more active than other brain areas, especially in gamma and beta bands. Fig. 6(b) shows the corresponding names of each EEG channel. Consequently, the feature interpretability captured by the proposed model gives our model more practical application value.

4.4. Emotion recognition with crucial EEG channels

From the previous process, we can obtain the weight of each EEG channel from the MSA layers in emotional activities, which represents the importance of the EEG channel. Then we test whether the emotion recognition performance could be enhanced by using signals of selected significant electrodes in this section. Fig. 6(c) presents four different EEG channels combinations based on the channel significations obtained from the multi-head self-attention block: plan (A) four electrodes: FT7, T7, FT8, and T8; plan (B) six electrodes: FT7, T7, TP7, FT8, T8, and F8; plan (C) eight electrodes: FT7, T7, TP7, P3, FC6, FT8, T8, and F8; plan (D) seven electrodes: FT7, T7, TP7, FT8, T8, TP1, and TP2. We select DE features of all bands corresponding to these EEG channels as classifier inputs. The mean classification accuracies and stander deviations of positive and negative emotions in four electrodes plans are 93.08%/4.36%, 97.64%/4.43%, 99.33%/3.27%, and 96.74%/4.94%, respectively, in subject-dependent two classification tasks. The mean classification accuracies and stander deviations of positive and negative emotions in four electrodes plans are 87.27%/4.32%, 92.46%/4.21%, 92.62%/3.69%, and 90.88%/3.19%, respectively, in subject-dependent three classification tasks. Obviously, plan (C) is advantageous in three classification tasks.

Although the average accuracies of using selected EEG electrodes are slightly lower than using all 62 electrodes, the average classifier training time of using the selected eight electrodes in plan(C) is remarkably reduced than using 62 electrodes. Besides, most of these electrodes are located in the lateral temporal lobe, which is easy for EEG data collection. In brief, almost similar accuracies and shorter training are obtained by using few EEG channels. This result promotes the development of portable as well as wearable EEG devices for emotion recognition.

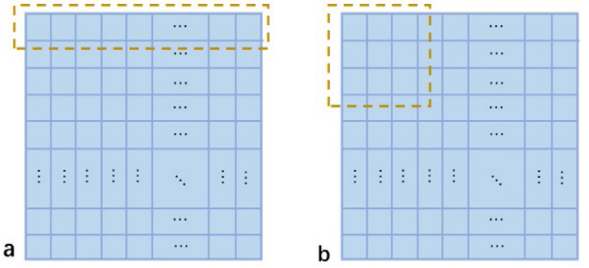

Fig. 7. Different segmentation methods of data in the classifier input layer.

4.5. Different designs of the input layer

This section analyzes the different performances of the proposed DCoT model with different segmentation methods for input data. We segment the input EEG features into a series of sequences according to EEG channels in the model input layer. The sketch map is shown in Fig. 6(a). The independence between EEG channels is considered in this way. However, the input EEG features are commonly segmented into square patches in other models of emotion recognition. The sketch map is presented in Fig. 7(b). We conduct experiments to explore the difference between the two kinds of segmentation approaches to input data.

First, we design a special input layer that splits the input DE features into a series of two-dimensional patches, as shown in Fig. 7(b), and the subsequent model structures are the same as the DCoT model we proposed. Both the sequence in Fig. 7(a) and the patch in Fig. 7(b) contain the same number of data points. Then we utilize the two methods for positive, neutral, and negative emotions recognition. On average, using sequences achieves a 1.48% and 3.21% higher classification accuracy than using patches in subject-dependent and subject-independent three classification experiments, respectively. According to the result, we speculate that there are independences between DE features of different channels. Some information will be ignored if the input DE features are treated as images and processed with patches in the model input layer. Hence, considering the independence of signals in different EEG channels while designing models may enhance the results. Moreover, it also makes more physiological sense to segment the input DE features by channel.

5. Discussion

The aforementioned experiments indicate that the proposed method has a satisfactory emotion recognition performance compared with the previous works. In addition, the model has advantages in terms of feature interpretability, as the crucial features captured by the DCoT model characterize the importance of each EEG channel in emotion recognition. We expect that the proposed model not only can learn significant brain areas in emotion recognition but also can be used to capture active brain areas in the diagnosis of mental disorders such as Alzheimer's and so on. This allows us to conduct in-depth studies of the human mind. However, we are also facing some challenges in emotion recognition. Despite the remarkable capability of deep learning model-based emotion recognition performed in recent years, lots of related works still rely on data preprocessing to strengthen the distinction of EEG characteristics for different emotions [24,25,43,46]. For instance, discrete wavelet transformation, entropy measures method, and so on have been wildly adopted in state-of-art works. So does our method. We believe that deep learning-based EEG classification will get rid of the dependence on preprocessing when a huge amount of EEG signal collection is available. We will explore whether preprocessing is necessary when more EEG data is available for deep learning models in future investigations. Besides, human emotions are associated with cultures. However, like most benchmark datasets, the SEED dataset used in this work only involved a limited number of participants. Therefore, the cultural backgrounds of these participants are not various. We would like to collect EEG signals from subjects with various cultural backgrounds and use them to test the emotion recognition capability of the proposed method in the future.

6. Conclusion

In this work, we have designed a novel model for EEG-based emotion recognition. We innovatively introduce the depthwise convolution and Transformer structure to extract discriminative EEG features, classify emotional EEG signals and visualize the importance of EEG channels in emotion recognition. Then we conduct subject-dependent experiments as well as subject-independent experiments on the SEED dataset to prove the validity of the DCoT model. For subject-dependent experiments, the average accuracies of two classification tasks and three classification tasks can be as high as 99.82% and 93.83%, respectively. For subject-independent experiments, the average accuracy of two classification tasks achieves 88.37%, and the average accuracy of three classification tasks can be as high as 83.03%, which indicates that the DCoT model performs better in emotion recognition than lots of methods. Additionally, we especially visualize

the extracted in-depth features from the DCoT. The visualization presents the importance of each EEG channel and the active brain areas in emotion recognition. Then, we select the critical EEG channels, design four kinds of channel combinations, and test the classification effect. We achieve the best subject-dependent classification performance, 99.33% (two classification tasks) and 92.62% (three classification tasks), while using the EEG channels: FT7, T7, TP7, P3, FC6, FT8, T8, and F8. The classifier training time is remarkably reduced as well.

CRediT authorship contribution statement

Jia-Yi Guo: Methodology, Software, Validation, Data curation, Writing – original draft. **Qing Cai:** Software, Data curation. **Jian-Peng An:** Visualization. **Pei-Yin Chen:** Supervision, Writing – reviewing. **Chao Ma:** Formal analysis, Writing – reviewing and editing. **Jun-He Wan:** Supervision, Writing – reviewing. **Zhong-Ke Gao:** Conceptualization, Investigation, Formal analysis, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding

This work was supported in part by the National Natural Science Foundation of China [grant numbers: 61922062, 61903270, 61873181] and the Natural Science Foundation of Tianjin, China [grant number: 21JCJQJC00130].

References

[1] W.L. Zheng, W. Liu, Y. Lu, B.L. Lu, A. Cichocki, EmotionMeter: A multi-modal framework for recognizing human emotions, IEEE Trans. Cybern. 49 (2019) 1110–1122, http://dx.doi.org/10.1109/TCYB.2018.2797176.
[2] K. Anderson, P.W. McOwan, A real-time automated system for the recognition of human facial expressions, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 36 (2006) 96–105, http://dx.doi.org/10.1109/TSMCB.2005.854502.
[3] X. Huang, S.J. Wang, X. Liu, G. Zhao, X. Feng, M. Pietikainen, Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition, IEEE Trans. Affect. Comput. 10 (2019) 32–47, http://dx.doi.org/10.1109/TAFFC.2017.2713359.
[4] Y.J. Liu, J.K. Zhang, W.J. Yan, S.J. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Trans. Affect. Comput. 7 (2016) 299–310, http://dx.doi.org/10.1109/TAFFC.2015.2485205.
[5] F. Deligianni, Y. Guo, G.Z. Yang, From emotions to mood disorders: A survey on Gait analysis methodology, IEEE J. Biomed. Health Inf. 23 (2019) 2302–2316, http://dx.doi.org/10.1109/JBHI.2019.2938111.
[6] A.K. Maddirala, R.A. Shaik, Separation of sources from single-channel EEG signals using independent component analysis, IEEE Trans. Instrum. Meas. 67 (2018) 382–393, http://dx.doi.org/10.1109/TIM.2017.2775358.
[7] R. Maskeliunas, R. Damaševičius, I. Martisius, M. Vasiljevas, Consumer-grade EEG devices: are they usable for control tasks, Peerj 4 (2016) http://dx.doi.org/10.7717/peerj.1746.
[8] Z.K. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, G. Chen, A channel-fused dense convolutional network for EEG-based emotion recognition, IEEE Trans. Cogn. Dev. Syst. 13 (2021) 945–954, http://dx.doi.org/10.1109/TCDS.2020.2976112.
[9] D. Sammler, M. Grigutsch, T. Fritz, S. Koelsch, Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music, Psychophysiol. 44 (2007) 293–304, http://dx.doi.org/10.1111/j.1469-8986.2007.00497.x.
[10] N. Ahmadi, Y.L. Pei, M. Pechenizkiy, Effect of linear mixing in EEG on synchronization and complex network measures studied using the Kuramoto model, Physica A 520 (2019) 289–308, http://dx.doi.org/10.1016/j.physa.2019.01.003.
[11] S.V. Bozhokin, I.B. Suslova, Wavelet-based analysis of spectral rearrangements of EEG patterns and of non-stationary correlations, Physica A 421 (2015) 151–160, http://dx.doi.org/10.1016/j.physa.2014.11.026.
[12] Z.K. Gao, X.M. Wang, Y.X. Yang, C.X. Mu, Q. Cai, W.D. Dang, S.Y. Zuo, EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation, IEEE Trans. Neural Netw. Learn. Syst. 30 (2019) 2755–2763, http://dx.doi.org/10.1109/TNNLS.2018.2886414.
[13] P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from EEG using higher order crossings, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 186–197, http://dx.doi.org/10.1109/TITB.2009.2034649.
[14] B. Hjorth, EEG analysis based on time domain properties, Electroencephalogr. Clin. Neurophysiol. 29 (1970) 306–310, http://dx.doi.org/10.1016/0013-4694(70)90143-4.
[15] C.A. Frantzidis, C. Bratsas, C.L. Papadelis, E. Konstantinidis, C. Pappas, P.D. Bamidis, Toward emotion aware computing: An integrated approach using multi-channel neurophysiological recordings and affective visual stimuli, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 589–597, http://dx.doi.org/10.1109/TITB.2010.2041553.
[16] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (2007) 1084–1093, http://dx.doi.org/10.1016/j.eswa.2006.02.005.
[17] O.A. Rosso, M.T. Martin, A. Plastino, Brain electrical activity analysis using wavelet-based informational tools, Physica A 313 (2002) 587–608, http://dx.doi.org/10.1016/S0378-4371(02)00958-5.
[18] U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, Expert Syst. Appl. 38 (2011) 13475–13481, http://dx.doi.org/10.1016/j.eswa.2011.04.149.
[19] H. Chen, Y. Song, X.L. Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network, Neurocomputing 356 (2019) 83–96, http://dx.doi.org/10.1016/j.neucom.2019.04.058.
[20] P.Y. Li, H. Liu, Y.J. Si, C.B. Li, F.L. Li, X.Y. Zhu, X.Y. Huang, Y. Zen, D.Z. Yao, Y.S. Zhang, P. Xu, EEG based emotion recognition by combining functional connectivity network and local activations, Ieee Trans. Biomed. Eng. 66 (2019) 2869–2881, http://dx.doi.org/10.1109/tbme.2019.289765.

[21] E. Aravind, S. Deepak, A. Sudheer, EEG-based emotion recognition using statistical measures and auto-regressive modeling, in: Proc. Int. Conf. Comput. Intell. Commun. Technol, 2015, pp. 587–591.
[22] J. Hu, J. Min, Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model, Cogn. Neurodyn. 12 (2018) 431–440, http://dx.doi.org/10.1007/s11571-018-9485-1.
[23] R.N. Duan, J.Y. Zhu, B.L. Lu, Differential entropy feature for EEG-based emotion classification, in: Proc. IEEE 6th Int. IEEE/EMBS Conf. Neural Eng. (NER), 2013, pp. 81–84.
[24] W.L. Zheng, B.L. Lu, Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, IEEE Trans. Autonom. Mental Dev. 7 (2015) 162–175, http://dx.doi.org/10.1109/TAMD.2015.2431497.
[25] T. Song, W. Zheng, P. Song, Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks, Ieee Trans. Affect. Comput. 11 (2020) 532–541, http://dx.doi.org/10.1109/TAFFC.2018.2817622.
[26] J. Atkinson, D. Campos, Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers, Expert Syst. Appl. 47 (2016) 35–41, http://dx.doi.org/10.1016/j.eswa.2015.10.049.
[27] J.X. Liu, H.Y. Meng, A. Nandi, M.Z. Li, Emotion detection from EEG recordings, in: Fuzzy Systems and Knowledge Discovery, ICNC-FSKD, in: 12th International Conference on Natural Computation, PEOPLES R CHINA, Changsha, 2016, pp. 1722–1727.
[28] T. Tuncer, S. Dogan, A. Subasi, A new fractal pattern feature generation function based emotion recognition method using EEG, Chaos Solitons Fractals 144 (2021) http://dx.doi.org/10.1016/j.chaos.2021.110671.
[29] T. Zhang, W. Zheng, Z. Cui, Y. Zong, Y. Li, Spatial-temporal recurrent neural network for emotion recognition, IEEE Trans. Cybern. 49 (2019) 839–847, http://dx.doi.org/10.1109/TCYB.2017.2788081.
[30] N.M. Krishna, K. Sekaran, A.V.N. Vamsi, G.S.P. Ghantasala, P. Chandana, S. Kadry, T. Blazauskas, R. Damaševičius, An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals, Ieee Access 7 (2019) 77905–77914, http://dx.doi.org/10.1109/ACCESS.2019.2922047.
[31] X. Li, Z.G. Zhao, D.W. Song, Y.Z. Zhang, J. Pan, L. Wu, J.D. Huo, C.Y. Mu, D. Wang, Latent factor decoding of multi-channel EEG for emotion recognition through autoencoder-like neural networks, Front. Neurosci. 14 (2020) http://dx.doi.org/10.3389/fnins.2020.00087.
[32] D. Komolovaite, R. Maskeliūnas, R. Damaševičius, Deep convolutional neural network-based visual stimuli classification using electroencephalography signals of healthy and Alzheimer's disease subjects, Life 12 (2022) 374, http://dx.doi.org/10.3390/life12030374.
[33] D. Maheshwari, S.K. Ghosh, R.K. Tripathy, M. Sharma, U.R. Acharya, Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel EEG signals, Comput. Biol. Med. 134 (2021) http://dx.doi.org/10.1016/j.compbiomed.2021.104428.
[34] T.J. Siddharth T.-P. Jung, Sejnowski, Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing, Ieee Trans. Affect. Comput. 13 (2022) 96–107, http://dx.doi.org/10.1109/TAFFC.2019.2916015.
[35] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, Comput. Sci. (2014) http://dx.doi.org/10.48550/arXiv.1409.0473.
[36] A. Vaswani, N. Shazeer, N. Parmar, J. Yszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: 31st Annual Conference on Neural Information Processing Systems, Vol. 30, NIPS, Long Beach, CA, 2017, https://dl.acm.org/doi/10.5555/3295222.3295349.
[37] X. Li, W. Zhang, Q. Ding, Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism, Signal Process. 161 (2019) 136–154, http://dx.doi.org/10.1016/j.sigpro.2019.03.019.
[38] P.Y. Chen, Z.K. Gao, M.M. Yin, J.L. Wu, K. Ma, C. Grebogi, Multiattention adaptation network for motor imagery recognition, IEEE Trans. Syst. Man Cybern. Syst. (2022) http://dx.doi.org/10.1109/TSMC.2021.3114145.
[39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: International conference on learning representations, 2021, pp. 1–7, http://dx.doi.org/10.48550/arXiv.2010.11929.
[40] R. Collobert, F. Sinz, J. Weston, L. Bottou, Large scale transductive svms, J. Mach. Learn. Res. 7 (2006) 1687–1712, http://dx.doi.org/10.1007/s10846-006-9063-3.
[41] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, B. Hu, Exploring EEG features in cross-subject emotion recognition, Front. Neurosci. 12 (2018) http://dx.doi.org/10.3389/fnins.2018.00162.
[42] W. Zhang, F. Wang, Y. Jiang, Z. Xu, S. Wu, Y. Zhang, Cross-subject EEG-based emotion recognition with deep domain confusion, in: 12th international conference on intelligent robotics and applications, ICIRA, PEOPLES R CHINA, Shenyang, 2019, p. 11740, 558-570.
[43] Y. Li, W.M. Zheng, Y. Zong, Z. Cui, T. Zhang, X.Y. Zhou, A Bi-hemisphere domain adversarial neural network model for EEG emotion recognition, Ieee Trans. Affect. Comput. 12 (2021) 494–504, http://dx.doi.org/10.1109/TAFFC.2018.2885474.
[44] Y. Cimtay, E. Ekmekcioglu, Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition, Sensors 20 (2020) http://dx.doi.org/10.3390/s20072034.
[45] J. Fdez, N. Guttenberg, O. Witkowski, A. Pasquali, Cross-subject EEG-based emotion recognition through neural networks with stratified normalization, Front. Neurosci. 15 (2021) http://dx.doi.org/10.3389/fnins.2021.626277.
[46] J.H. Zhang, Z. Yin, P. Chen, S. Nichele, Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Inf. Fusion 59 (2020) 103–126, http://dx.doi.org/10.1016/j.inffus.2020.01.011.