Multimedia Tools and Applications (2023) 82:15439–15456
https://doi.org/10.1007/s11042-022-14011-7

Valence-arousal classification of emotion evoked by Chinese ancient-style music using 1D-CNN-BiLSTM model on EEG signals for college students

Ruoyu Du[1,2] · Shujin Zhu[1,2] · Huangjing Ni[1,2] · Tianyi Mao[1,2] · Jiajia Li[1] · Ran Wei[3] ·

Received: 12 January 2022 / Revised: 22 June 2022 / Accepted: 23 September 2022 /
Published online: 4 October 2022

Abstract

During the COVID-19 pandemic, young people are using multimedia content more frequently to communicate with each other on Internet platforms. Among them, music, as psychological support for a lonely life in this special period, is a powerful tool for emotional self-regulation and getting rid of loneliness. More and more attention has been paid to the music recommender system based on emotion. In recent years, Chinese music has tended to be considered an independent genre. Chinese ancient-style music is one of the new folk music styles in Chinese music and is becoming more and more popular among young people. The complexity of Chinese-style music brings significant challenges to the quantitative calculation of music. To effectively solve the problem of emotion classification in music information search, emotion is often characterized by valence and arousal. This paper focuses on the valence and arousal classification of Chinese ancient-style music-evoked emotion. It proposes a hybrid one-dimensional convolutional neural network and bidirectional and unidirectional long short-term memory model (1D-CNN-BiLSTM). And a self-acquisition EEG dataset for Chinese college students was designed to classify music-induced emotion by valence-arousal based on EEG. In addition to that, the proposed 1D-CNN-BILSTM model verified the performance of public datasets DEAP and DREAMER, as well as the self-acquisition dataset DESC. The experimental results show that, compared with traditional LSTM and 1D-CNN-LSTM models, the proposed method has the highest accuracy in the valence classification task of music-induced emotion, reaching 94.85%, 98.41%, and 99.27%, respectively. The accuracy of the arousal classification task also gained 93.40%, 98.23%, and 99.20%, respectively. In addition, compared with the positive valence classification results of emotion, this method has obvious advantages in negative valence classification. This study provides a computational classification model for a music recommender system with emotion. It also provides some theoretical support for the brain-computer interactive (BCI) application products of Chinese ancient-style music which is popular among young people.

Keywords EEG · Chinese music · Emotion classification · BiLSTM

Extended author information available on the last page of the article

1 Introduction

The COVID-19 pandemic has challenged people's mental health [23]. Studies have confirmed the frequent use of various multimedia content on the Internet to convey information and emotions during the lockdown, with the consumption of music in particular. Music is often used as a means of self-regulation of negative emotions such as anxiety and pain, and many studies have shown that music is a good part of reducing stress [24]. In today's society, music has been used as the best therapeutic tool [19]. Further research shows that emotion determines what type of music we choose to listen to, and music can also be used to express the emotion we feel [11]. According to Juslin et al. [12], about 64% of musical experiences affect us emotionally, leading to happiness, joy, nostalgia, or longing. While the COVID-19 pandemic has brought rapid changes to travel, study environments, working conditions, and social support, it has also stressed many university students. A study of young people showed that listening to music is one of their most effective strategies for coping with stress [23]. Therefore, emotion-based music recommendation system has been paid more and more attention in the Internet era and multimedia application products [15]. The classification of music-induced emotion can provide data support for improving the music recommendation system.

In recent years, Chinese music has become an independent genre. It is different from other existing genres, such as Vienna classical music genre, Russian folk music genre, and Venetian music genre [26]. Chinese civilization has a long history and integrates the cultures of different nationalities. Chinese music has formed a diverse and complex system. Chinese ancient-style music is a kind of Chinese new folk style music, which can well integrate traditional cultural elements and modern music elements and is becoming more and more popular among young people [5]. The complexity of this kind of music brings significant challenges to the quantitative calculation of music.

With the development of science and technology and the ravages of COVID-19, people are increasingly aware of the vital role of music in emotional guidance [10, 28]. Therefore, the research of music and neuroscience, cognitive psychology, and signal processing has become a hot topic in the academic world. Relevant theoretical studies have shown that human brain activity plays an essential role in the generation and activity of emotion. EEG can be collected through brain-computer Interface (BCI) technology to detect and identify information related to changes in emotional states [1]. Galvo et al. [6] predicted the exact values of valence and arousal in a subject-independent scenario and identified four Emotional classes with an accuracy of 84.4% using the DEAP, AMIGOS, and DREAMER datasets. Zhou et al. [27] collected EEG data from 40 participants for regulating negative emotions, and a binary prediction of valence (high or low) of 78.75 ± 9.48% and 73.98 ± 5.54% for arousal was calculated through the machine learning method. Li et al. [16] review the recent representative works in the EEG-based emotion recognition research and provide a tutorial to guide the researchers to start from the beginning. In the above studies, valence and arousal are often used to describe emotional states. Therefore, the identification of music-induced emotional states can be based on the classification of valence-arousal.

In recent years, more and more researchers have applied deep learning models to emotion recognition. Many studies focused on extracting temporal and spatial features by combining CNN and LSTM models to prove the effectiveness and superiority of their schemes [4]. Anubhav et al. studied the classifier performance of the subject-independent model and subject-dependent model, respectively, for the problem of emotion recognition and classification based on EEG. they found that the accuracy of the LSTM model in terms of emotional potency and arousal was above 90% [3]. Grave et al. improved the LSTM model in phonemic classification and recognition and proposed the bidirectional LSTM model (BiLSTM) [9]. Sharma et al. [22] proposed an automated classification of population-labeled EEG signals using nonlinear higher-order statistics Deep learning algorithm. And the average classification accuracy is 82.01%, with a 10-fold cross-validation technique corresponding to four-labeled emotions classes.

Therefore, this paper first designed an emotional EEG experiment evoked by Chinese ancient-style music for college students, a young group, to provide some experimental data of specific groups to improve the music recommendation system. Secondly, a 1D-CNN-BILSTM hybrid model is proposed for emotional feature extraction and valence-arousal classification of emotion. This method makes use of the advantages of 1D-CNN in capturing local features and BiLSTM in comprehensively capturing temporal information to achieve the study of the valence-arousal classification of emotional states evoked by Chinese ancient-style music and verifies the performance of our proposed classification method through comparative analysis with various deep learning models. This paper provides design ideas and a research model for designing a music recommendation system based on emotion.

The rest of the paper is organized as follows: in the Section 2, we describe the two public available emotional EEG datasets, DEAP and DREAMER, which were used to compare the experimental data in this paper. And then, we describe DESC, an emotional EEG dataset evoked by Chinese ancient-style music based on pentatonic mode, which was collected through the self-designed experiment. Furthermore, the proposed method of the classification model, including data preprocessing, the structure description of the 1D-CNN-BiLSTM model, and the model validation results, is described in the Section 3. The Section 4 gives the experimental results and several published studies for comparison. The classification performance of the proposed method can be summarized in the Section 5. Finally, conclusions and future work are discussed in the Section 6.

The main contributions of this paper can be summarized as follows:

- Design a small sample EEG dataset based on pentatonic mode for emotion classification on specific topics (Section 2);
- Based on the LSTM model, combining the advantages of 1D-CNN in feature extraction and feature recognition capability of BiLSTM, a 1D-CNN-BILSTM hybrid model was proposed to classify the valence and arousal of emotion (Section 3);
- Study the classification accuracy for valence-arousal of emotion, and find that the proposed model has advantages in the valence-arousal classification of emotion in both public and self-acquisition EEG datasets, especially in the classification of negative valence (Sections 4.2 and 4.3);
- The repetition times of the feature extraction layer and classification optimization layer are modified to obtain more sensitive and accurate classification results with the model framework, which provides ideas for other deep learning models to study how to improve accuracy (Section 5).

2 Material and experiment description

Many researchers using EEG techniques for their work often lack adequate data support and validation. Many institutions or organizations, as well as researchers or research teams, make their datasets of conducted research available for open access. Therefore, this paper uses two publicly available emotional EEG datasets and one EEG dataset collected from a self-designed experiment to perform classification model validation.

2.1 DEAP dataset

The DEAP dataset [14] was collected experimentally by Koelstra et al. at Queen Mary University of London, UK, the University of Twente, Netherlands, the University of Geneva, Switzerland, and the Swiss Federal Institute of Technology to study multichannel physiological data on emotional states, and the data are publicly and freely available. The dataset contains 32 channels of EEG signals from 32 subjects and 8 channels of other physiological signals. In this paper, only the EEG signal from 14 of the 32 channels is used as the experimental data to standardize the data. The EEG signals were first sampled at 512 Hz and then resampled to 128 Hz; bandpass frequency filtering from 4 to 45 Hz was performed, and electrooculography (EOG) artifacts were removed. Each subject watched 40 emotional music videos that were 60 s in length. After viewing each video, the subjects rated VALENCE, AROUSAL, PREFERENCE, and DOMINANCE on a 9-point scale. In the experiment, using value 5 as the rating threshold, labels with ratings more significant than value 5 were labeled "positive valence", and those less than value 5 were tagged "negative valence".

2.2 DREAMER dataset

The DREAMER database [13], published by the University of the West of Scotland, provides the subjects' ratings of films regarding valence, arousal, and dominance, from which the corresponding emotional positivity or arousal and control are obtained. The movies consisted of 18 segments, ranging from 65 to 393 s [13, 20]. Using the Emotiv EPOC system with 14 channels, EEG and ECG data were collected from 23 subjects (14 males, 9 females; mean age 26.6; standard deviation 2.7) while watching the movie, with a sampling rate of 128 Hz. The last 60 s of each signal were intercepted and used as input data in this experiment. Since the dimensional scale of these data is 1–5, the threshold of 3.5 was used as the scale threshold; labels with a score greater than 3.5 were labeled "positive valence", and those with a score less than 3.5 were tagged "negative valence".

2.3 Experimental description

With the development of the Internet, Chinese ancient-style music in popular songs is a new folk music style in Chinese music that caters to the appreciation needs of Chinese young people. This kind of music has distinct Chinese national characteristics and the mark of The Times. This music has particular preferences in lyrics and tunes [5]. First, the lyrics generally consist of nostalgic poems and ancient Chinese as well as local dialects. Secondly, the tunes of most Chinese ancient-style music are created by using The national pentatonic mode. According to the two characteristics of Chinese ancient-style music, 30 music videos of Chinese ancient style were first selected before starting this experiment. Each music intercept takes 30 to 40 s. Volunteers who were not music majors were selected through a questionnaire to rate the valence and arousal of the chosen music perceptions. Eighteen of them with high valence differences were selected as material stimuli after statistical analysis. Among them, the shortest playing time was used as the standard, and each music video was intercepted for 34 s. The selected Chinese ancient-style music video Information for the experiment is shown in Fig. 1.

The experimental flowchart is shown in Fig. 2. The selected Chinese ancient-style music video clips were then randomly divided into three groups, each with six stimulus materials that were each approximately 6 min in length, sufficient to allow the subjects to reach the desired emotional state and maintain it to some extent while avoiding audiovisual fatigue. In this paper, 20 college students were selected as subjects for Chinese ancient-style music emotional experiment; there were 10 male and 10 female students between 19 and 22 years of age, all right-handed, with normal hearing and normal or corrected visual acuity, in good health, and having no physical diseases. They voluntarily participated in this experiment, and all signed an informed consent form before the experiment began. The experimental data acquisition equipment was the Emotiv Epoc + EEG device, and the recording channels were AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, with a total of 14 leads. The music video clip was played at the beginning of the experiment while recording the EEG signals of the subjects.

(1) Baseline recording: This procedure lasts 30 s, and the subject wears headphones while watching cross marks on the screen to facilitate concentration and calm while baseline recordings are taken.

(2) Music video stimulation and self-assessment: This procedure lasts 264 s (44 s × 6 songs). A 34-second music video is first played, and the subjects rate valence and arousal on a 9-point scale based on their true feelings promptly after the end of the music video stimulus; the rating time lasts 10 s. The next stimulus and rating steps are repeated until the 6 music videos are completed. Then, the subjects sit still for 30 s to wait for emotional recovery. Good EEG recordings are continuously collected during this process.

(3) The next set of experiments is started. Steps (1) and (2) are repeated until 3 sets of experiments are completed.

In the process of rating music video materials, the music video materials were presented randomly. The subjects rated the valence and arousal of the music video materials using a 9-

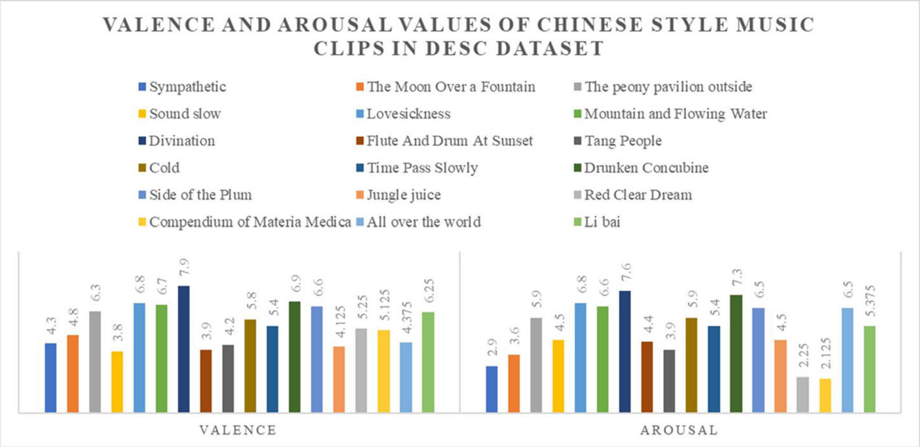


Fig. 1 Chinese Ancientry-style Music Information Selected by DESC dataset

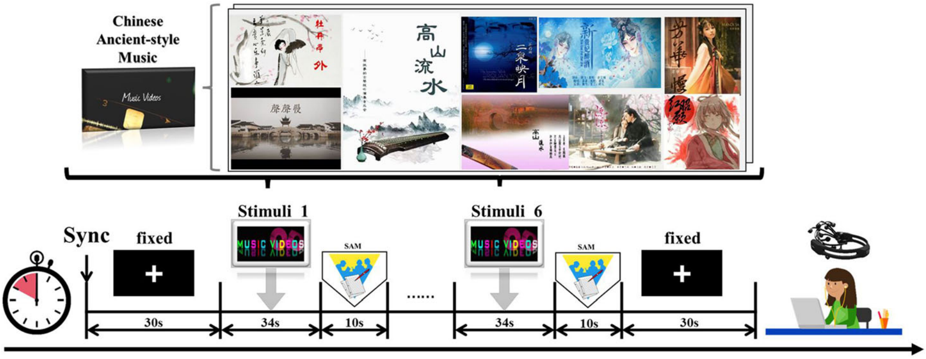


Fig. 2 Experimental Design Process

point scale after watching each music video. Valence refers to how pleasant or unpleasant the subject's feelings were after watching the music video, with the most pleasing being 9 and the least friendly being 1. Arousal refers to how excited or unmotivated the subject felt after watching the music video, with the more excited being closer to 9 and the less enthusiastic being closer to 1. This rating process was individual and was administered in a quiet and closed laboratory. Subject ratings were performed based on immediate feelings after viewing the music video without overthinking.

2.4 DESC dataset

In this paper, EEG signals were self-acquisition for the changes in emotional states induced by Chinese ancient-style music design. A Database of emotional EEG Stimulated by Chinese ancient-style music (DESC,) was constructed. The design idea of this dataset is similar to the process of the DEAP dataset, and the equipment used is the same as that used in the DREAMER dataset. This can eliminate unnecessary errors in the subsequent comparative evaluation of the models.

2.5 Experiment datasets selection

The selected datasets were split into a training set and a test set, respectively. We then trained the proposed model to classify the valence and arousal rating of emotion. A self-acquisition EEG dataset evoked by the Chinese ancient-style music designed in this paper and the two publicly available EEG datasets all contain emotion valence-arousal information, with DESC and DEAP containing 9-rank valence-arousal and DREAMER containing only 5-rank valence-arousal. The main characteristics of the three EEG datasets are shown in Table 1.

3 Proposed method

3.1 Data preprocessing

Before deep learning is performed, the data are preprocessed. The raw signals are segmented and filtered using traditional methods, and five EEG components are extracted from each

Table 1 Main characteristics of the DEAP, DREAMER, and DESC datasets

	DEAP	DREAMER	DESC
Stimuli	40	18	18
Type	Music videos	Film clips	Chinese ancient-style music videos
Duration	60s	65-393s	34s
Physiological Signals	EEG, GSR, BVP, RESP, SKT, EOG, EMG	EEG, ECG	EEG
Participants	32 (19 males,13 female)	23 (14 males, 9 females)	20 (10 males, 10females)

electrode based on different frequency bands: the theta wave (4–8 Hz), alpha wave (8–12 Hz), low beta wave (12–16 Hz), high beta wave (16–25 Hz), and gamma wave (25–45 Hz). Thus, there are 14*5 = 70 EEG signals for each type of sample. The selected EEG data (DEAP/DREAMER: 60 s, DESC: 34 s) are then computed with the standard Power spectral density (PSD) features using the Fast Fourier Transform (FFT) with a 2-s window with 50% overlap. Each sample size is 7680(60*128)*70 or 4352(34*128)*70. 7680/4352 is the time series (timesteps), and 70 is the spatial component.

3.2 LSTM

RNNs are a kind of neural network for sequential data, and LSTM [8] is a temporal recurrent neural network that can avoid the long-term dependency problem that exists in ordinary RNNs and has been successfully applied in the fields of speech recognition and sentiment analysis. The LSTM unit consists of forget gates, input gates, and output gates, as shown in Fig. 3, which control the proportions of discarded information and information passed to the next time step. At time t, the output f_t of the forget gate of the LSTM unit, the output i_t of the input gate, the output o_t of the output gate, the cell state c_t, and the hidden state h_t are updated. The specific calculation formulas (Eqs. 1–5) are as follows:

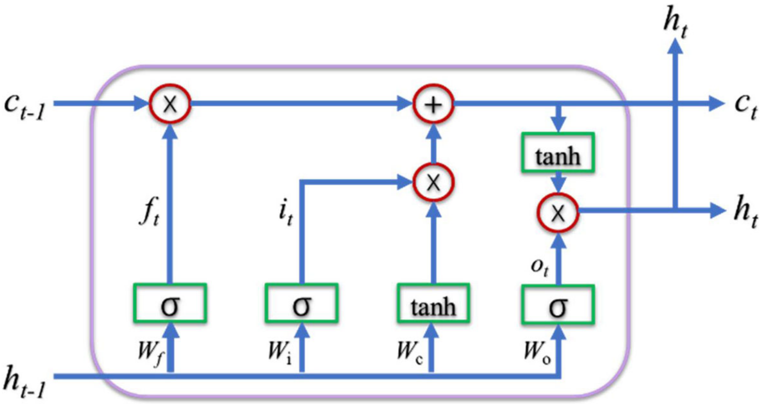


Fig. 3 LSTM Model Structure

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

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$h_t = O_t \times \tanh(c_t) \tag{5}$$

where x_t is the input at time t; W_f, W_i, $W_{o,}$ and W_c are the weights of the forget gate, input gate, output gate, and cell state, respectively; and b_f, b_i, $b_{o,}$ and b_c are the biases of the forget gate, input gate, output gate, and cell state, respectively.

3.3 1D-CNN-BiLSTM

For many temporal signal classification tasks, considering both past and future contextual information can effectively improve classification accuracy. In contrast, the hidden state h_t of the LSTM at the moment t considers only past information. The basic idea of BiLSTM [9] is to present each sequence forward and backward as two independent hidden states to capture the past information h_t and future information h'_t, respectively. Then, the two hidden states are connected to form the final output H_t; i.e., $H_t = h_t + h'_t$. The structure of the BiLSTM model was shown in Fig. 4.

CNN is good at identifying simple patterns in data and then using those simple patterns to generate more complex patterns in higher-level layers. 1D-CNN can obtain features of interest from shorter (fixed-length) segments of the overall dataset, and this property does not depend on the location information in the data segment. Considering the advantages of CNN and LSTM in feature extraction and processing dynamic temporal information, this

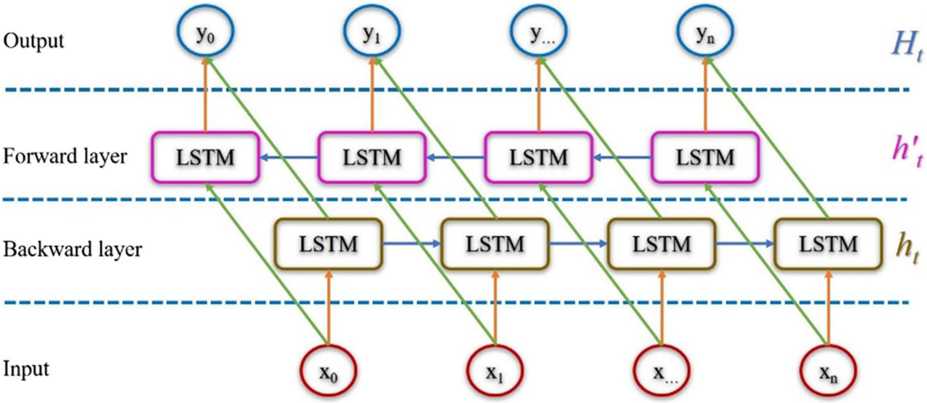


Fig. 4 BiLSTM Model Structure

paper proposed a valence and arousal classification of emotion method for EEG signals based on a 1D-CNN-BiLSTM model. The proposed method is shown in Fig. 5. First, the valence or arousal features of each channel of the EEG signal are automatically extracted using a 1D-CNN model. Then, the valence or arousal features with a high level of 14 channels are extracted by using the modeling ability of the BiLSTM framework on the sequences. Finally, the features of the multiple channels are classified using a softmax classifier. The details are given below.

- **The first part: The 1D-CNN layer**

First, the 1D time series of EEG data is directly used as the input to the model, and the shape of the input data is 70 × 1. Then, the input data are passed through the first convolutional layer to extract the abstract properties of the original data; the number of 1D convolutional kernels in the first Conv 1D sublayer is 32, the shape of each convolutional kernel is 15 × 1, and the step size of the convolutional kernel is 2. This convolutional layer is followed by a ReLU activation layer that can introduce nonlinearity to the proposed model. After convolutional activation, 32 feature maps of size 28 × 1 are output. After that, the output of the first Conv 1D layer is passed through a max-pooling layer. In the max-pooling layer, the size of the pooling window is 2, and the stride of the window is also 2. This method can significantly reduce the number of training parameters in the model and speed up the training process. After the first max-pooling operation, 32 feature mappings with a size of 14 × 1 are output. Then, high-level features are further extracted through the second Conv 1D sublayer to facilitate classification. The second

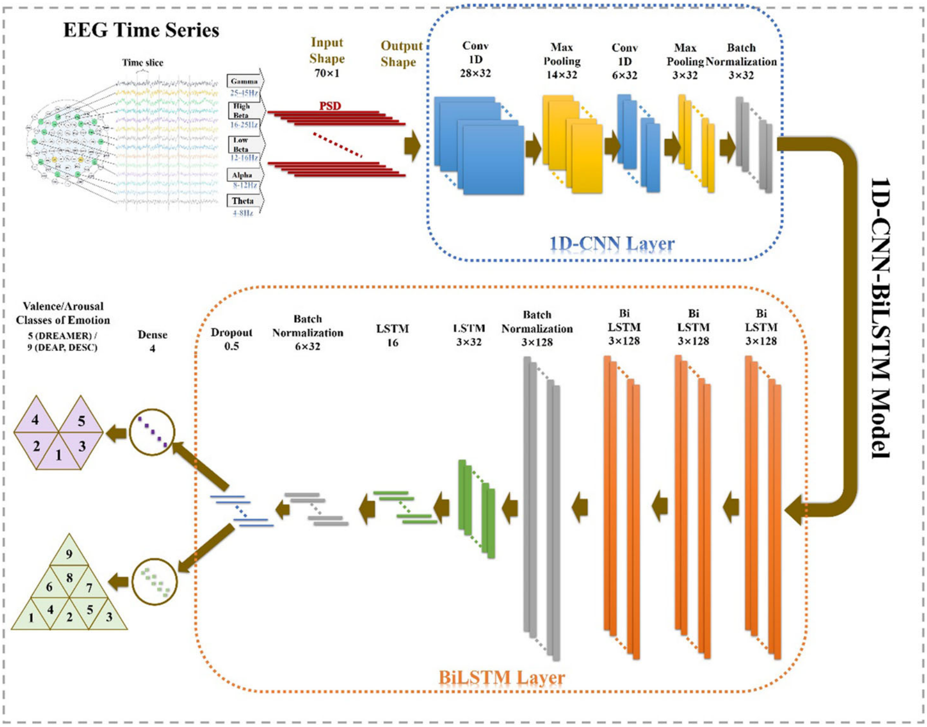


Fig. 5 Proposed Model Structure

Conv 1D sublayer has 32 kernels of size 6 × 1. The convolution operation is the same as that of the first Conv 1D sublayer, and ReLU is also used for nonlinear activation. Then, feature sorting is carried out through the second max-pooling sublayer, and finally, 32 feature maps with a size of 3 × 1 are output.

After all the 1D convolution layers, the 32 obtained feature maps of size 3 × 1 are fed into a batch normalization layer, which speeds up the convergence of the model during training, makes the model training process more stable, and plays a specific role in regularization, which can further prevent overfitting. The size of the output matrix is 3 × 32.

- **The second part: The BiLSTM layer**

After passing through the 1D-CNN layer, the output characteristics are fed into the BiLSTM layer, consisting of three BiLSTM sublayers and two LSTM layers. This operation avoids the long-term dependency problem in the standard RNN. There are four gates in the LSTM unit, including the cell state gate, forget gate, input data gate, and output gate. They can collaborate to preserve previous information, further improving the ability to learn valuable information from EEG time-series data. The three BiLSTM sublayers each have 128 neurons and are followed by a batch normalization sublayer. Then, the first LSTM sublayer, with 32 neurons, is used after data normalization, and the second LSTM sublayer, with 16 neurons, then returns to the hidden state of the last step. Finally, the third batch normalization sublayer is used to standardize the data.

Dropout is then applied to the output of the BiLSTM layer. The second LSTM sublayer can reduce the dimensionality of the feature mapping to fit the input of the first LSTM sublayer, and dropout can alleviate the overfitting concern to some extent. Through this operation, the model becomes less sensitive to small changes in the data. Thus, this method can further improve the accuracy of the processing of invisible data. Once the features have passed through dropout processing, the output features are fed into the dense layer. Finally, an output layer with a softmax function is added to the model for final classification.

4 Results

In this paper, the performance of the proposed method is evaluated through experiments conducted on public emotional EEG datasets and self-acquisition datasets, and the training and testing results of the proposed method are given. In addition, comparative experimental results with classical deep learning methods are presented to show the superiority of the proposed method.

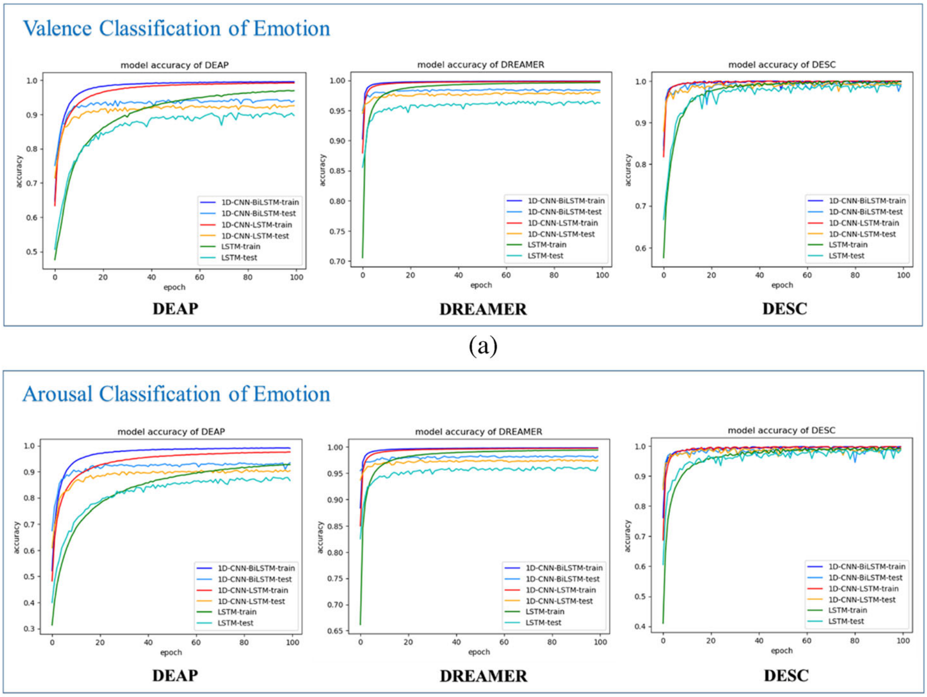
4.1 Experimental setup

To validate the effectiveness of the proposed 1D-CNN-BiLSTMs method for the valence-arousal classification of emotion from EEG signals, experiments are conducted on three datasets, DEAP, DREAMER, and DESC. The hardware devices used for the experiments are an Intel(R) Core(TM) i9-10885 H CPU and an NVIDIA GeForce GTX 1650 GPU. The software environment used is Python 3.6, while the Keras framework is used to build the neural network model. In the experiments of this paper, the dropout operation retention rate is set to 0.5, and the optimizer uses Adam, so the learning rate is 0.001. In this paper, we conducted multi-category experiments using all subjects' data labeled with valence and arousal classes of emotion labels in the DEAP, DREAMER, and DESC datasets and evaluated the classification performance of the model using a 5-fold cross-validation technique. The samples are divided equally into 5 subsets: 1 subset is taken as the test set, and the remaining 4 are the training set. The above operation is repeated 5 times until all subsets have been used as the test set for the experiments. In addition, to verify that the model proposed in this paper is better for the negative valence of emotion, the positive valence of emotion-labeled data from the three datasets are subjected to classification experiments in turn, and the classification results are obtained using the 5-fold cross-validation are used to evaluate the performance.

4.2 Experimental results

In this section, two traditional deep learning models are implemented for valence-arousal classification of emotion and compared with the proposed models, which are the standard LSTM and 1D-CNN-LSTM. To further evaluate the classification performance of these three models, we computed and compared the accuracy model, accuracy, precision, recall, and F1 scores and Cohen's kappa values from three EEG datasets, which were shown in Fig. 6; Tables 2 and 3.

To understand the accuracy advantage of the proposed 1D-CNN-BiLSTMs model over the standard LSTM model and the 1D-CNN-LSTM model in more detail, the accuracy model of the three models on the valence-arousal classification task from three EEG datasets was shown



(a)

(b)

Fig. 6 Model Accuracy Results with Valence-Arousal Classification of Emotion from Three EEG Datasets, (**a**) Valence; (**b**) Arousal

Table 2 The performance of LSTM, 1D-CNN-LSTM, and the proposed 1D-CNN-BiLSTM model on the valence classification task with the DEAP, DREAMER, and DESC datasets

Dataset	Method	Accuracy	Precision	Recall	F1 score	Cohen's kappa
DEAP	LSTM	91.71%	91.71%	91.59%	0.9159	0.8874
	1D-CNN-LSTM	92.67%	92.52%	92.67%	0.9259	0.9005
	1D-CNN-BiLSTM	94.85%	94.65%	94.58%	0.9461	0.9300
DREAMER	LSTM	96.86%	96.84%	96.87%	0.9685	0.9528
	1D-CNN-LSTM	97.69%	97.69%	97.68%	0.9769	0.9653
	1D-CNN-BiLSTM	98.41%	98.42%	98.40%	0.9841	0.9760
DESC	LSTM	97.77%	98.26%	98.74%	0.9850	0.9615
	1D-CNN-LSTM	99.19%	98.87%	99.57%	0.9922	0.9861
	1D-CNN-BiLSTM	99.27%	99.15%	99.60%	0.9937	0.9874

in Fig. 6. Figure 6a showed that the proposed model achieved the highest test accuracy of valence classification in most of the training and test processes. At the same time, Fig. 6b showed that the proposed model achieved the highest test accuracy of arousal classification in most training and test processes. Compared with the standard LSTM model and the 1D-CNN-LSTM model, There is a significant increase in the accuracy of the proposed model in the DEAP dataset. Combining Tables 2 and 3, the highest accuracy of valence-arousal classification was obtained on the DESC dataset, 99.27% for valence and 99.20% for arousal. In addition, the proposed model also achieved the best accuracy in both public EEG datasets, which is 98.41% for valence and 98.23% for arousal in DREAMER, 94.85% for valence, and 93.40% for arousal in DEAP.

From Table 2, it can be seen that the proposed model used in the valence classification has a precision of 99.15%, recall of 99.60%, F1 score of 0.9937, and Cohen kappa value of 0.9874 in the self-acquisition dataset DESC, and the model, with both the self-acquisition dataset and public datasets, has significantly better performance than the standard LSTM model and the 1D-CNN-LSTM model. In particular, on the DEAP dataset, compared with the standard LSTM model and the 1D-CNN-LSTM model, the proposed model improves precision by 2.94% and 2.13%, recall by 2.98% and 1.90%, F1 score by 0.03 and 0.04, and Cohen kappa value by 0.043 and 0.029, respectively.

From Table 3, it can be seen that the proposed model used in the arousal classification has a precision of 99.27%, recall of 99.17%, F1 score of 0.9922, and Cohen kappa value of 0.9907 in the self-acquisition dataset DESC, and the model, with both the self-acquisition dataset and public datasets, has significantly better performance than the standard LSTM model and the

Table 3 The performance of LSTM,1D-CNN-LSTM, and the proposed 1D-CNN-BiLSTM model on the arousal classification task with DEAP, DREAMER, and DESC datasets

Datasets	Methods	Accuracy	Precision	Recall	F1-score	Cohens kappa
DEAP	LSTM	86.61%	85.33%	85.70%	0.8548	0.8457
	1D-CNN-LSTM	90.38%	89.49%	89.64%	0.8956	0.8892
	1D-CNN-BiLSTM	93.40%	92.70%	92.96%	0.9283	0.9240
DREAMER	LSTM	96.16%	96.04%	96.15%	0.9609	0.9493
	1D-CNN-LSTM	97.38%	97.48%	97.12%	0.9730	0.9654
	1D-CNN-BiLSTM	98.23%	98.29%	98.05%	0.9817	0.9765
DESC	LSTM	98.08%	98.08%	97.99%	0.9803	0.9775
	1D-CNN-LSTM	98.70%	98.86%	98.57%	0.9871	0.9848
	1D-CNN-BiLSTM	99.20%	99.27%	99.17%	0.9922	0.9907

1D-CNN-LSTM model. In particular, on the DEAP dataset, compared with the standard LSTM model and the 1D-CNN-LSTM model, the proposed model improves precision by 7.37% and 3.21%, recall by 7.26% and 3.32%, F1 score by 0.07 and 0.03, and Cohen kappa value by 0.078 and 0.035, respectively.

In addition, the performance advantage of the proposed method is noticed in the negative valence classification of emotion. To verify this conclusion, the positive valence of emotion with EEG data was defined as the rating higher than 5 in the self-acquisition dataset DESC and the public dataset DEAP, and the positive valence of emotion with EEG data was defined as the rating higher than 3.5 in another public dataset DREAMER are used for classification task with the proposed method in this paper. The average accuracy of negative and positive valence classification of emotion is compared with the same dataset, as shown in Fig. 7. From Fig. 7, it can be seen that the proposed method has the most significant difference in the positive and negative valence classification of emotion for DESC, and the average accuracy of negative valence classification improves by 1.61% over the positive emotion. The average accuracy of negative valence classification in the DEAP dataset by this model is 1.19% higher than the positive valence, and the smallest difference in the DREAMER dataset is only an improvement of 0.24%. Overall, the model has a better negative valence classification of emotion than the positive valence.

4.3 Comparison with several published studies

Finally, we compare the proposed method with several published studies using the same dataset, i.e., the DEAP dataset and DREAMER dataset, and using the Self-acquisition dataset. Table 4 shows the details of several published studies on DEAP, DREAMER, and Self-acquisition datasets, respectively. From the results of EEG emotion recognition summarized in Table 4, we can see that our method improves the outcomes of valence and arousal classification on both DEAP and DREAMER. Specifically, on the DREAMER dataset, our method achieves the highest accuracy of 98.41%, and 98.23% for valence and arousal, respectively. The accuracy of our method is 3.82% for valence and 2.97% for arousal higher than the second-highest accuracy with DREAMER [18] listed in Table 2. On the DEAP dataset, our method achieves the best performance of 94.85% and 93.4% for valence and arousal, respectively, which also improves the accuracy of valence and arousal by 2.61% and 0.48% compared with the second-highest accuracy with DEAP [7] listed in Table 4. Moreover,

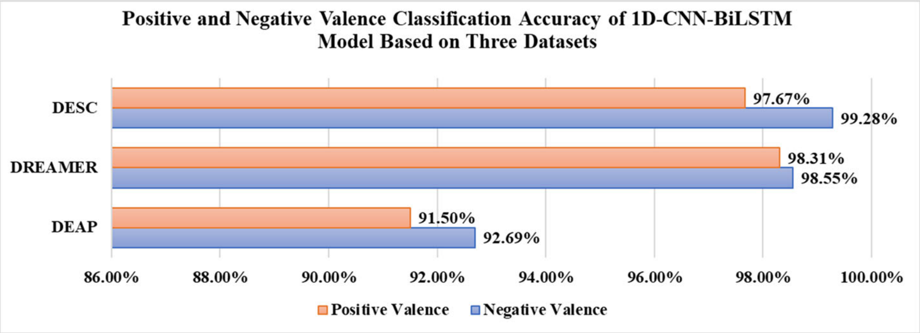


Fig. 7 Positive and Negative Valence Classification Accuracy of the Proposed Method based on DEAP, DREAMER, and DESC Datasets

Table 4 Details of several published studies on DEAP, DREAMER, and Self-acquisition datasets

Reference	Dataset	Device	Stimuli	Inputs	Classifier	Accuracy(%)	
						Valence	Arousal
Salma et al. [2]	DEAP	Biosemi	Audio-Visual (music and video clips)	Raw EEG signals	LSTM	85.45%	85.65%
Zhan et al. [25]	DEAP	Biosemi	Audio-Visual (music and video clips)	PSD	CNN	82.95%	84.07%
Sharma et al. [22]	DEAP	Biosemi	Audio-Visual (music and video clips)	Higher order statistics	Bi-LSTM	84.16%	85.21%
Gao et al. [7]	DEAP	Biosemi	Audio-Visual (music and video clips)	DE	Dense CNN	92.24%	92.92%
The proposed method	DEAP	Biosemi	Audio-Visual (music and video clips)	PSD	1D-CNN-BiLSTM	94.85%	93.40%
Katsigiannis and Ramzan [13]	DREAMER	Emotiv EPOC	Audio-Visual (film clips)	PSD	SVM-RBF	62.49%	62.17%
Y. Liu et al. [18]	DREAMER	Emotiv EPOC	Audio-Visual (film clips)	Raw EEG signals	MLF-CapsNet	94.59%	95.26%
Pandey et al. [21]	DREAMER	Emotiv EPOC	Audio-Visual (film clips)	Raw EEG signals	1D-CNN	75.93%	81.48%
The proposed method	DREAMER	Emotiv EPOC	Audio-Visual (film clips)	PSD	1D-CNN-BiLSTM	98.41%	98.23%
Liu et al. [17]	Self-acquisition dataset	Emotiv EPOC	Audio-Visual (film clips)	PSD, ASM	SVM-RBF	Positive: 86.43% Negative: 65.09%	None
Zhou et al. [27]	Self-acquisition dataset	Biosemi	Audio-Visual (film and music clips)	PSD, DE	Random Forest	78.75%	73.98%
The proposed method	Self-acquisition dataset (DESC)	Emotiv EPOC	Audio-Visual (music and video clips)	PSD	1D-CNN-BiLSTM	99.27%	99.20%

compared with the methods in references [17, 27], Our method used self-designed and collected EEG signals as input, which achieved higher accuracy than other references that used Self-acquisition data for research, demonstrating the superiority of our method.

5 Discussion

It can be seen from the experimental results, that the proposed method on the small sample data classification task is more obviously superior to the performance of several reported works, especially our method, compared with the traditional model framework, simply modify the feature extraction and classification optimization layer repetitions, successfully in reducing the number of samples at the same time improve the classification accuracy. It is necessary to discuss why the proposed method can achieve such excellent performance in the valence - arousal emotion classification task under musical stimulation. The superior classification performance of our method is most likely due to the following:

1. CNN is a particularly effective means of feature extraction. LSTM is good at processing time-series data, while BiLSTM trains two models on the input sequence instead of one LSTM. The first of the input sequences is in the original sample, and the second is the reverse sample of the input sequence. It provides additional context for the network and allows for faster and more comprehensive learning of the problem. BiLSTM is very suitable for modeling time series data and was first used in emotion classification tasks in natural language processing. Its advantage lies in the consideration of context information in the modeling process. EEG, as a nonlinear time series signal, is also suitable for BiLSTM framework modeling. Because each person's EEG signal is affected by individual factors, there are apparent unique characteristics in the signal. These unique characteristics will affect the classification effect of the classifier and the generalization ability of the model. In layman's terms, the model may not have learned meaningful target characteristics, but instead learned irrelevant information that made the model less able to migrate to new data. More attention should be paid to its accuracy in feature construction. The working principle of the 1D-CNN-BiLSTM model is to extract local features of EEG signal space from 1D-CNN. BiLSTM is then used to capture the relationship between two directional representations, and the global features of EEG signals are learned in time. According to the global feature, EEG data can be judged whether they are from the same label labeled by the same subject enhancing the feature representation ability. The negative emotion generated by music video stimulation with the EEG used in this paper has more complex and less sensitive internal representations than positive emotion. Therefore, this method uses three BiLSTM sublayers to improve classification sensitivity. This results in better performance of our method on self- acquisition data sets with small sample sizes and specific populations than on public datasets with sufficient sample sizes.
2. Since the feature dimension of data extracted from the convolution layer is very high, to solve this problem and reduce the training cost in the model, a pooling layer is generally added after convolution to reduce the number of features. The proposed method uses two convolution - pooling layers to reduce the number of channels in the feature map. Therefore, this method can significantly reduce the number of parameters without sacrificing the performance of the emotion classification task based on valence-arousal.

6 Conclusions and future work

In this paper, a valence-arousal classification method of emotion using EEG for Chinese ancient-style music is proposed. The proposed method can better classify the relationship between music-induced emotion using EEG signals based on the valence-arousal index. We first preprocessed raw EEG data to obtain PSD values and then input them into two convolution–pooling layers to extract the features. Finally, these features were transformed into three BiLSTM sublayers for optimal classification. The proposed framework reduces the number of parameters and improves the accuracy and sensitivity of classification. Validation experiments are carried out on the DEAP dataset, DREAMER dataset, and DESC dataset. The average accuracy of our method was 94.85% for valence classification and 93.40% for arousal classification on the DEAP dataset, and 98.41% for valence classification and 98.23% for arousal classification on the DREAMER dataset, respectively. The average accuracy of the DESC dataset was 99.27% for valence classification and 99.20% for arousal classification, respectively. The experimental results show that the accuracy of the 1D-CNN-BILSTMs method is higher than that of CNN, Dense CNN, LSTM, BI-LSTM, SVM-RBF, and MLF-CapsNet methods. In addition, compared with traditional LSTM and 1D-CNN-LSTM models, the accuracy of our method on the valence-arousal classification task on the DEAP dataset is increased by 3.14% and 2.18% for valence, 6.79%, and 3.02% for arousal, respectively, the accuracy of the valence-arousal classification task on DREAMER dataset is increased by 1.55% and 0.72% for valence, 2.07% and 0.85% for arousal, respectively. The accuracy of the valence-arousal classification task on DESC dataset is increased by 1.50% and 0.07% for valence, 1.12% and 0.50% for arousal, respectively, which verified the effectiveness of our method.

Since the training of this model belongs to supervised training, it needs to prepare a large number of labeled EEG data to build, and it is time-consuming and laborious to collect enough labeled EEG data. Therefore, based on these two limitations, future work will focus on two areas: First, the model was further modified and optimized to improve its performance in the emotion classification task on the data collected by different EEG devices, to improve its classification ability on other datasets. Secondly, attention mechanisms or transfer learning techniques can be introduced to the model to enhance recognition efficiency and reduce the reliance on labeled signal data.

Acknowledgements We thank American Journal Experts (AJE) for English language editing.

This work was supported by a grant from the National Natural Science Foundation of China (Grant No. 61977039), and the Natural Science Foundation of Jiangsu Province (Grant No. BK20190736, BK20200745), This work was also sponsored by the Natural Science Foundation for Colleges and Universities of Jiangsu Province, 20KJB510022, NUPTSF (Grant No. NY218138).

Data availability DEAP dataset generated and analysed during the current study is available in the Queen Mary University of London repository, (http://www.eecs.qmul.ac.uk/mmv/datasets/deap/) [14].

DREAMER dataset generated and analysed during the current study IS available from the corresponding author on reasonable request [13].

DESC dataset generated and analysed during the current study is available on GitHub (https://github.com/yakyou15/DECS-DATASET).

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Algarni M, Saeed F (2021) Review on emotion recognition using eeg signals based on brain-computer interface system. https://doi.org/10.1007/978-3-030-70713-2_42
2. Alhagry S, Aly A, Reda A (2017) Emotion recognition based on eeg using lstm recurrent neural network. Int J Adv Comput Sci Appl 8(10):355–358. https://doi.org/10.14569/IJACSA.2017.081046
3. Anubhav, Nath D, Singh M, Sethia D, Indu S (2020) An efficient approach to EEG-based emotion recognition using LSTM network. 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, pp 88–92. https://doi.org/10.1109/CSPA48992.2020.9068691
4. Bai Z, Sun G, Zang H, Zhang M, Shen P, Liu Y et al (2019) Identification technology of grid monitoring alarm event based on natural language processing and deep learning in china. Energies MDPI 12(17):1–19. https://doi.org/10.3390/EN12173258
5. Chen Y (2019) Understanding and thinking of ancient-chinese-style music in popular songs. Proceedings of the 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019). https://doi.org/10.2991/iccese-19.2019.71
6. Galvo F, Alarco SM, Fonseca MJ (2021) Predicting exact valence and arousal values from eeg. Sensors 21(10):3414. https://doi.org/10.3390/s21103414
7. Gao Z, Wang X, Yang Y, Li Y, Ma K, Chen G (2020) A channel-fused dense convolutional network for eeg-based emotion recognition. IEEE Trans Cogn Dev Syst PP(99):1. https://doi.org/10.1109/TCDS.2020.2976112
8. Graves A (2012) Long short-term memory[J]. Springer, Berlin Heidelberg
9. Graves A, Fernández S, Schmidhuber J (2005) Bidirectional LSTM networks for improved phoneme classification and recognition. Artificial neural networks: formal models & their applications-icann, International Conference, Warsaw, Poland, September. DBLP. 3697, pp 799–804. https://doi.org/10.5555/1986079.1986220
10. Hennessy S, Sachs M, Kaplan J, Habibi A (2021) Music and mood regulation during the early stages of the covid-19 pandemic. PLoS ONE 16(10):e0258027. https://doi.org/10.1371/journal.pone.0258027
11. Juslin PN, Sloboda JA (2001) Music and emotion: theory and research. Oxford University Press, Oxford
12. Juslin PN, Liljeström S, Västfjäll D, Barradas G, Silva A (2008) An experience sampling study of emotional reactions to music: listener, music, and situation. Emotion 8(5):668. https://doi.org/10.1037/a0013505
13. Katsigiannis S, Ramzan N (2017) Dreamer: a database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. IEEE J Biomedical Health Inf 22(1):98–107. https://doi.org/10.1109/JBHI.2017.2688239
14. Koelstra S (2012) Deap: a database for emotion analysis ;using physiological signals. IEEE Trans Affect Comput 3(1):18–31. https://doi.org/10.1109/T-AFFC.2011.15
15. Lampropoulos AS, Lampropoulou PS, Tsihrintzis GA (2012) A cascade-hybrid music recommender system for mobile services based on musical genre classification and personality diagnosis. Multimed Tools Appl 59(1):241–258. https://doi.org/10.1007/s11042-011-0742-0
16. Li X, Zhang Y, Tiwari P, Song D, Hu B, Yang M et al (2022) EEG based emotion recognition: a tutorial and review. https://doi.org/10.48550/arXiv.2203.11279
17. Liu YJ, Yu M, Zhao G, Song J, Shi Y (2017) Real-time movie-induced discrete emotion recognition from eeg signals. IEEE Trans Affect Comput PP(99):1. https://doi.org/10.1109/TAFFC.2017.2660485
18. Liu Y, Ding Y, Li C, Cheng J, Chen X (2020) Multi-channel eeg-based emotion recognition via a multi-level features guided capsule network. Comput Biol Med 123:103927. https://doi.org/10.1016/j.compbiomed.2020.103927
19. Martín JC, Ortega-Sánchez D, Miguel IN, GMG Martín (2021) Music as a factor associated with emotional self-regulation: a study on its relationship to age during covid-19 lockdown in spain. Heliyon 7(2):e06274. https://doi.org/10.1016/j.heliyon.2021.e06274
20. Song TF, Zheng WM, Song P, Cui Z (2018) EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks[J]. IEEE Transactions on Affective Computing, pp 532–541. https://doi.org/10.1109/TAFFC.2018.2817622
21. Pandey P, Seeja KR (2022) A one-dimensional CNN model for subject independent emotion recognition using EEG signals. In: Khanna A, Gupta D, Bhattacharyya S, Hassanien AE, Anand S, Jaiswal A (eds) International conference on innovative computing and communications. Advances in intelligent systems and computing, vol 1388. Springer, Singapore, pp 509–515. https://doi.org/10.1007/978-981-16-2597-8_43
22. Sharma R, Pachori RB, Sircar P (2020) Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomed Signal Process Control 58:101867. https://doi.org/10.1016/j.bspc.2020.101867
23. Strasser MA, Sumner PJ, Meyer D (2022) Covid-19 news consumption and distress in young people: a systematic review. J Affect Disord 300:481–491. https://doi.org/10.1016/j.jad.2022.01.007
24. Yehuda N (2011) Music and stress. J Adult Dev 18(2):85–94. https://doi.org/10.1007/s10804-010-9117-4

25. Zhan Y, Vai MI, Barma S, Pun SH, Li JW, Mak PU (2019) A computation resource friendly convolutional neural network engine for EEG-based emotion recognition. IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp 1–6. https://doi.org/10.1109/CIVEMSA45640.2019.9071594
26. Zhang Y, Zhou Z, Sun M (2022) Influence of musical elements on the perception of 'chinese style' in music. Cogn Comput Syst. https://doi.org/10.1049/ccs2.12036
27. Zhou W, Qiu C, Liu G (2021) Efficient regulation of emotion by positive music based on EEG valence-arousal model. In: 2021 3rd International Conference on Image, Video and Signal Processing (IVSP 2021). Association for Computing Machinery, New York, pp 81–86. https://doi.org/10.1145/3459212.3459225
28. Ziv N, Hollander-Shabtai R (2022) Music and covid-19: changes in uses and emotional reaction to music under stay-at-home restrictions. Psychol Music 50(2):475–491. https://doi.org/10.1177/03057356211003326

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

Affiliations

Ruoyu Du[1,2] • Shujin Zhu[1,2] • Huangjing Ni[1,2] • Tianyi Mao[1,2] • Jiajia Li[1] • Ran Wei[3]

✉ Ruoyu Du
dury@njupt.edu.cn

✉ Shujin Zhu
shujinzhu@njupt.edu.cn

[1] School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China

[2] Smart Health Big Data Analysis and Location Services Engineering Laboratory of Jiangsu Province, Nanjing, China

[3] School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, China