Research on EEG emotion recognition based on CNN+BiLSTM+self-attention model*

LI Xueqing[1], LI Penghai[1]**, FANG Zhendong[2], CHENG Longlong[3], WANG Zhiyong[1], and WANG Weijie[4]

1. School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin 300384, China
2. China France BOHAI Geoservices Co., Ltd., Tianjin 300457, China
3. China Electronics Cloud Brain (Tianjin) Technology Co., Ltd., Tianjin, China
4. Institute of Motion Analysis and Research, University of Dundee, Scotland DD14HN, UK

(Received 5 December 2022; Revised 13 January 2023)

To address the problems of insufficient dimensionality of electroencephalogram (EEG) feature extraction, the tendency to ignore the importance of different sequential data segments, and the poor generalization ability of the model in EEG based emotion recognition, the model of convolutional neural network and bi-directional long short-term memory and self-attention (CNN+BiLSTM+self-attention) is proposed. This model uses convolutional neural network (CNN) to extract more distinctive features from both spatial and temporal dimensions. The bi-directional long short-term memory (BiLSTM) is used to further preserve the long-term dependencies between the temporal phases of sequential data. The self-attention mechanism can change the weights of different channels to extract and highlight important information and address the often-ignored importance of different channels and samples when extracting EEG features. The subject-dependent experiment and subject-independent experiment are performed on the database for emotion analysis using physiological signals (DEAP) and collected datasets to verify the recognition performance. The experimental results show that the model proposed in this paper has excellent recognition performance and generalization ability.

Document code: A **Article ID:** 1673-1905(2023)08-0506-7
DOI https://doi.org/10.1007/s11801-023-2207-x

Emotion is a general term to describe a series of subjective cognitive experiences. It is a psychological and physiological state produced by a combination of multiple feelings, thoughts and behaviors. The signals used in emotion recognition mainly include physiological and non-physiological signals. Non-physiological signals[1] include facial expression, posture and other image signals, which can be collected conveniently but these signals can be misleading because people may hide their real emotional expressions through mental training. Physiological signals[2] include electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), etc. Among them, EEG[3] has advantages in temporal resolution, reliability and accuracy, so it can truly reflect human psychological activities and cognitive behaviors.

EEG signal processing includes data preprocessing, feature extraction, classification and output, among which feature extraction is the key step in emotion recognition. None-deep learning methods for feature extraction heavily rely on manual feature extraction, such as extracting higher-order cross features and Hjorth features in the time domain, and using power spectral density (PSD), differential entropy (DE), and wavelet transform (WT) in the frequency domain, among which DE features have the best performance[4,5]. Features cannot be sufficiently extracted in the manual extraction case, while automatic feature extraction using deep learning algorithms not only eliminates the complex and time-consuming manual feature extraction process, but also enables the extraction of more comprehensive features, so automatic feature extraction is getting more and more popular.

In recent years, traditional deep learning methods such as convolutional neural network (CNN), long short-term memory (LSTM) have been used to improve EEG emotion recognition effect. CNN plays a great role in the feature extraction of EEG signals. WEN et al[6] proposed a stochastic CNN to classify EEG signals, which can extract features and classify them end-to-end. Many researchers currently are starting to use bi-directional long short-term memory (BiLSTM) networks more and more, due to their ability to learn past and future information in time series and get better results than LSTM. CUI et al[7]

* This work has been supported by the National Key Research and Development Program of China (No.2021YFF1200600), the National Natural Science Foundation of China (No.61806146), and the Natural Science Foundation of Tianjin City (Nos.18JCYBJC95400 and 19JCTPJC56000).
** E-mail: lph1973@tju.edu.cn

used BiLSTM to serialize feature-specific information processed by CNN and then learn the relationship between past and future in the information. The model proposed in this paper also utilizes the features of BiLSTM to achieve good results in EEG emotion recognition.

As stated above, the introduction of traditional deep learning algorithms makes the model focus more on important parts of signals rather than treat all data as equally important, thus avoiding the inefficiency and inadequate features extraction in manual feature extraction. However, the traditional deep learning algorithms also has some deficiency, as the importance of different channels and samples is often neglected in EEG features extraction, which leads to low accuracy of emotion recognition. The attention mechanism can effectively solve the poor recognition problem caused by the tendency to ignore the importance of different data segments. For example, CHEN et al[8] applied the attention mechanism to further capturing deep features from EEG signals in the BCI field. LI et al[9] used a combination of capsule network (CapsNet) and attention mechanism. CapsNet can yield fast learning speed and increase data efficiency, the attention mechanism can explore the information of the feature map by changing the weights of different channels, so as to extract more important information about the EEG signal channels. YI et al[10] proposed a transformer capsule network for EEG emotion recognition, and the transformer model based on the attention mechanism is used to capture global features among different windows and highlight important features. Nowadays, self-attention mechanisms are less frequently used in EEG emotion recognition to improve classification accuracy. In contrast to CNN, recurrent neural network (RNN) and other attention mechanisms mentioned above, the self-attention mechanism autonomously acquires a set of weight coefficients by learning to focus on regions of interest in a dynamically weighted manner while suppressing irrelevant regions. In the study of this paper, the incorporation of self-attention mechanism into the fusion model significantly improved the accuracy of emotion recognition.

Although the above recognition methods have made some progress in EEG emotion recognition, there are still some problems. None-deep learning feature extraction could only be conducted manually, and the extracted features are not sufficient and not holistic, resulting in low recognition accuracy. In contrast, deep learning algorithms can extract and enhance important features more comprehensively, but poor generalization performance due to inter-individual differences is rarely taken into account when mining features. The successful applications of CNNs, BiLSTM networks, and self-attention mechanisms have provided new ideas for EEG-based emotion recognition. Therefore, we propose a CNN+BiLSTM+self-attention model for emotion recognition, using CNN to extract more discriminative EEG features from both spatial and temporal dimensions, and adding a self-attention mechanism to analyze the importance of samples after using BiLSTM to further extract temporal features. The self-attention mechanism is an internal attention mechanism that associates different positions of a single sequence and encodes the sequence data according to the importance of different parts. Moreover, most of the existing literature uses subject-dependent experiments to train and test models, so those constructed models cannot effectively characterize significant differences between multiple subjects' data when used directly for cross-subject emotion recognition, resulting in poor generalization. In contrast, the model proposed in this paper can obtain similar features among different individuals, thus overcoming the poor generalization problem caused by inter-subject differences. In this research, subject-dependent experiments are conducted to verify the emotion recognition performance of the proposed model, and subject-independent experiments are conducted to verify the generalization ability of the model.

The process of EEG emotion recognition proposed in this paper mainly includes the following parts: EEG data preprocessing, feature extraction, emotion classification and output. The most important part is to use the CNN+BiLSTM+self-attention model to extract temporal and spatial features, which gives effective and in-depth information, so as to facilitate the emotion classification. The overall framework for emotion recognition is shown in Fig.1.

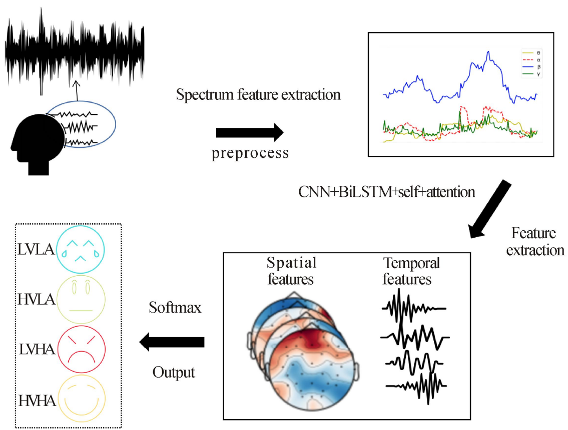

Fig.1 The overall framework for emotion recognition

The feature extraction based on CNN+BiLSTM+self-attention model includes CNN module, BiLSTM module, and self-attention module. The model's structure is shown in Fig.2. The CNN module encodes temporal-spatial features, the BiLSTM module is used to further learn the temporal correlation of EEG signals, and the self-attention mechanism module is designed to highlight important time periods of EEG signal.

The CNN module[11,12] encodes data by importing EEG feature vectors from various frequency bands. The processed data is delivered to the CNN module, which is made up of two convolutional layers that capture distinct features of the data. The first convolutional layer has 16 kernels with 1×32 filters. The convolutional layer chronologically filtered the EEG signals of each channel using multichannel convolution kernels to extract the variation characteristics of the EEG sequence. The second convolutional layer has 32 kernels with 32×1 filters, which can integrate the information between the EEG signal electrode channels. Since the EEG data distribution varies greatly from one subject to another, the network needs to adapt to different data distributions, resulting in a decrease in training speed and overfitting. To deal with this issue, a batch normalization layer[13] is added after using the Relu activation function in convolutional layer.

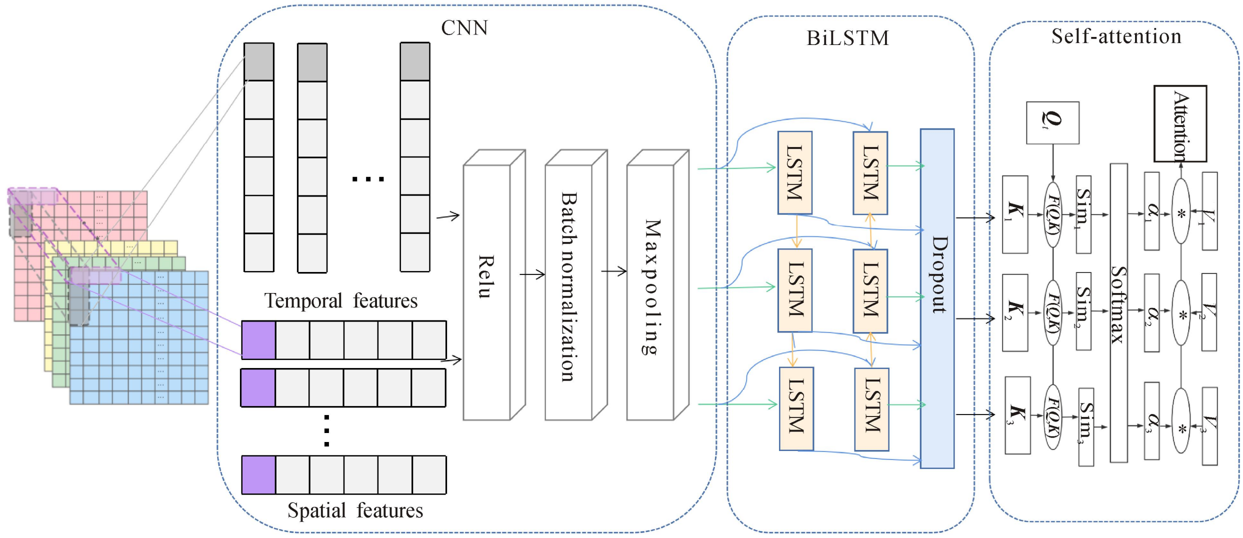

Fig.2 Structure of the proposed CNN+BiLSTM+self-attention model

The traditional LSTM[14,15] is a one-way RNN, which can only process information in chronological order, so it tends to ignore the information of future moments. The BiLSTM network expands the traditional LSTM network by introducing the second layer containing reversely inputted LSTM, which helps the model to simultaneously cover the information transmission from the past and the future moments, hence improving the model's predictive ability. The calculation process is shown in Eqs.(1) and (2), where h_t is the hidden state at current moment, $\boldsymbol{X}_t$ is the input vector, and $h_{t\text{-}1}$ is the hidden state at the preceding moment, $C_{t\text{-}1}$ is the unit state at the preceding moment, h_1 is the forward LSTM output state, h_2 is the backward LSTM output state, $\boldsymbol{W}_1$ is the forward output weight matrix, and $\boldsymbol{W}_2$ is the backward output matrix. $\boldsymbol{Y}_t$ represents the final output result. The LSTM module includes 3 bidirectional LSTM layers, the first layer has 128 cells, the second layer has 64 cells, and the third layer has 32 cells.

$$h_t = \mathrm{LSTM}(\boldsymbol{X}_t, h_{t\text{-}1}, C_{t\text{-}1}), \tag{1}$$

$$\boldsymbol{Y}_t = \sigma(\boldsymbol{W}_1 h_1 + \boldsymbol{W}_2 h_2). \tag{2}$$

To improve emotion recognition models' efficiency and accuracy, it's crucial to rapidly screen key emotional information out of multimodal signals containing huge redundant information. In this paper, the self-attention mechanism is adopted to highlight the important time phases of the characteristic EEG signals to locate the key emotional information, which facilitates emotion recognition. Self-attention mechanism has been widely used recently in various deep learning tasks, and it achieved satisfactory recognition results in tasks ranging from sequence modeling to image-based[16] tasks. First, the correlation Sim_t is obtained as a dot product of the feature vector $\boldsymbol{K}_t$ and the query vector $\boldsymbol{Q}_t$. Then the softmax function is used to normalize the Sim_t's weight to obtain the weight coefficient of the corresponding α_t. Finally, α_t and $\boldsymbol{K}_t$ are weighted to achieve the attention value. Attention is the final output representation used for classification. The calculation process is shown in Eqs.(3)—(5).

$$\mathrm{Sim}_t = \boldsymbol{Q}_t \cdot \boldsymbol{K}_t, \tag{3}$$

$$\alpha_t = \frac{\mathrm{e}^{\mathrm{Sim}_t}}{\sum_{j=1}^{T} \mathrm{e}^{\mathrm{Sim}_j}}, \tag{4}$$

$$attention = \sum_{t=1}^{T} \alpha_t * \boldsymbol{K}_t. \tag{5}$$

This research tested the effectiveness of the CNN+BiLSTM+self-attention model on the database for emotion analysis using physiological signals (DEAP), which collected data from 32 healthy participants. The subjects were asked to watch 40 one-minute videos. Emotion was divided into four categories based on the level of arousal and valence, low valence low arousal (LVLA), low valence high arousal (LVHA), high valence low arousal (HVLA), high valence high arousal

(HVHA), as shown in Tab.1. We designed both the subject-dependent and subject-independent experiments. In subject-dependent experiments, the data in the training and test sets are from the same subject but with different trials. The common features between different trials of the same subject can be extracted and used to verify the recognition performance of the model. Since emotions bear strong individual differences, each subject needs to train a specific classifier for emotion recognition, which is a big obstacle in cross-subject application, so this research designs a subject-independent experiment to train the classifier model for emotion recognition. In subject-independent experiments, the data in the training set and test set are obtained from different subjects, and the common features among subjects are extracted to overcome individual differences and improve the generalization ability of the model. Subject-independent experiments can be used to verify the ability of the model to be generalized to different individuals. In order to measure the ability of the model in the cross-subject emotion recognition task, the leave-one-subject-out (LOSO) cross-validation method was used, which means that all the data from one subject is selected for testing in each round, and the data from all remaining subjects are used for training, and the cycle is continued until all samples are tested. In this paper, the accuracy of the LOSO method of subjects-independent experiments is used as an evaluation index, and usually the higher the accuracy of experimental results, the better the generalization ability. The latter task is more demanding and difficult than the former one, but it fits more people and is better in generalization. The average accuracy was taken as the evaluation result for all experiments.

Tab.1 Four categories of valence and arousal

Valence / Arousal	Low (1-5)	High (5-9)
Low (1-5)	LVLA	HVLA
High (5-9)	LVHA	HVHA

To verify the effect of emotion recognition more straightforwardly, subject-dependent experiments were performed on the DEAP using six recognition models, support vector machine (SVM)[17], K-nearest neighbor (KNN)[18], multi-layer perceptron (MLP)[19], CNN, LSTM and CNN+BiLSTM+self-attention models. In the model training and testing process, 80% of the experimental samples of each subject were taken as training data, and the remaining 20% were taken as test data. When training the models with each subject's training data, 10-fold cross-validation was adopted to select appropriate hyper-parameters. The test results were recorded and the average accuracy was calculated to show the experimental results. As shown in Fig.3, the CNN+BiLSTM+self-attention model has achieved an average accuracy of 92.83% on DEAP, which greatly surpasses that of traditional machine learning. Compared with a single CNN or LSTM model, the accuracy is improved by 13.79% and 10.51%, respectively. Additionally, the precision, recall and F1-score of this model were calculated. As shown in Tab.2, the accuracy of the model is 90.85%, the recall is 90.68%, and the F1-score is 90.71%. These numbers further indicate that this novel emotion recognition mode can more accurately recognize the information of emotion feature in the EEG signal, thus attaining a better emotion recognition accuracy rate.

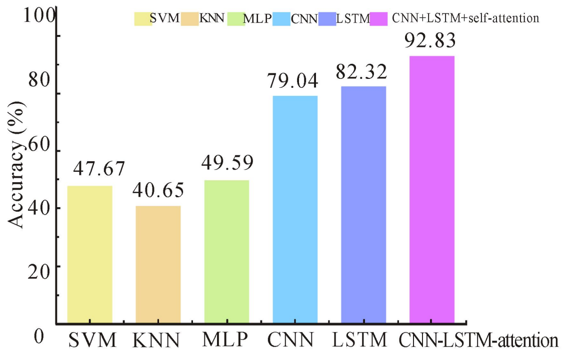

Fig.3 The average accuracy of different models (%)

Tab.2 Subject-independent experiments results on the DEAP

Accuracy	Precision	Recall	F1-score
92.83%	90.85%	90.68%	90.71%

Specifically, in LOSO cross-validation with 32 subjects, 31 subjects' EEG data was used as the training dataset, and the remaining one subject's EEG data was used as the test dataset. The average recognition accuracy based on temporal-spatial features in all frequency bands was used as the recognition result. The average recognition accuracy of all the 32 subjects in the four-category classification task is 87.76%, and the standard deviation is 1.02%. The test results of different subjects are shown in Tab.3.

In order to test the robustness and generalization ability of this method, an EEG acquisition system is constructed in this paper. 15 healthy subjects (aged 21—27, 10 males and 5 females) participated in the experiment. All participants read and signed the informed consent forms. The experimental flow is shown in Fig.4. During data collection, the subjects of each experiment needed to watch twelve 60 s movie clips containing the four emotions. Therefore, there are 12 trials in total in each experiment. Movie clips showing various emotions were shuffled for watching, and two adjacent clips showed different emotions. After watching the video, all the subjects were required to mark the valence, arousal and dominance of the watched video ranging from 1 to 9 according to the self-assessment manikin (SAM) scale. The data sampling frequency is 128 Hz and the size of the data collected from each person is 12×32×7 680, where 12 is the number of videos, 32 is the leads, and 7 680 is the sampling points of 60 s EEG data. The link to download the data is http://sicse.tjut.edu.cn/kxyj/xzzq.htm.

Tab.3 Subject-dependent experimental results on the DEAP

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Accuracy	87.07%	88.39%	88.04%	88.12%	88.59%	87.65%	86.71%	88.43%	87.42%	87.26%	87.37%
Subject	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
Accuracy	89.65%	88.19%	88.01%	90.66%	87.58%	86.2%	87.26%	88.59%	88.59%	89.03%	86.48%
Subject	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32	Mean
Accuracy	87.26%	88.59%	86.56%	86.32%	87.56%	87.51%	88.34%	88.09%	87.53%	85.48%	87.76%

In order to verify the generalization ability of the model, identification tests were performed on the collected data. In the subject-dependent experiment, the data of different subjects were identified, and the recognition accuracy was 92.61%. The results are shown in Tab.4. In the subject-independent experiment, the data of each subject was identified. As shown in Tab.5, the average accuracy of the 15 subjects is 88.44%, and the standard deviation was 1.26. This experiment demonstrated that the network model proposed in this paper gives stable emotion recognition results on different subjects' data. The experiments on the collected data further proved that the above model boasts good generalization ability and can effectively avoid individual differences among subjects.

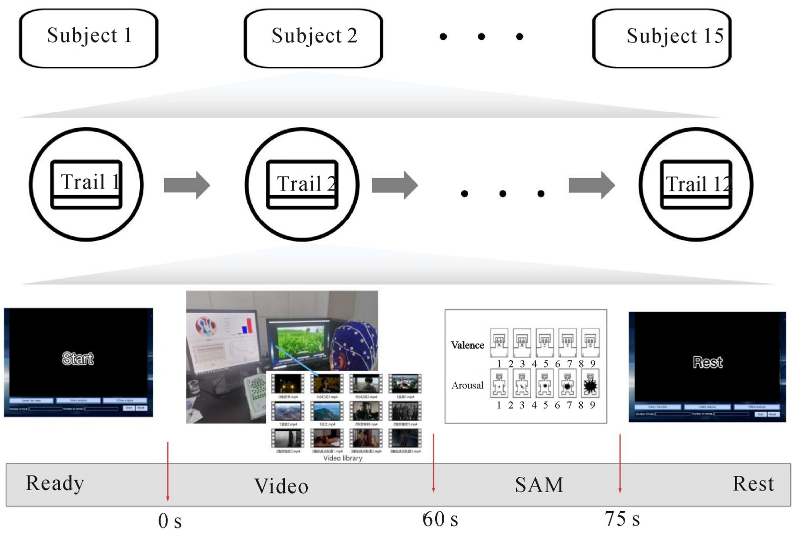

Fig.4 Experimental flow

Tab.4 Subject-dependent experimental results on the collected data

Accuracy	Precision	Recall	F1-score
92.61%	91.05	90.62	90.89

In order to capture the temporal-spatial features of different emotions during model learning, this paper proposes a hybrid CNN+BiLSTM+self-attention model, which can extract EEG signals' temporal-spatial features through CNN, then further extract temporal feature and deal with long-term dependences in the input sequences with BiLSTM, and the self-attention mechanism is able to determine critical time periods. In order to prove the advantage of this algorithm, this paper compares it with

Tab.5 Subject-independent experimental results on the collected data

Subject	S1	S2	S3	S4
Accuracy	89.24%	87.50%	90.17%	88.50%
Subject	S5	S6	S7	S8
Accuracy	89.10%	86.74%	85.50%	88.70%
Subject	S9	S10	S11	S12
Accuracy	89.16%	88.70%	87.69%	89.20%
Subject	S13	S14	S15	Mean
Accuracy	87.00%	89.55%	89.98%	88.44%

other advanced emotion recognition methods in existing literatures, and the comparison results are listed in Tab.6. Comparing our results with those of previous work, the accuracy of both the subject-dependent and subject-independent experiments was significantly improved, and the results showed that the model proposed in this paper outperformed the other representative methods for comparison. The CNN+BiLSTM+self-attention model used in this paper is able to extract temporal features and global dependence of emotion recognition on different EEG channels, fusing the spatiotemporal information of EEG signals, which enables the model to maintain the independence between EEG channels and obtain the temporal importance of EEG data, and achieve better feature extraction results. The model can acquire similar features among different individuals, thus alleviating the bottleneck caused by inter-subject differences in EEG data in subject-independent studies. The experimental results show that the recognition performance of the model is effectively improved in subject-dependent experiments, meanwhile the recognition accuracy is significantly improved and the generalization ability is enhanced in subject-independent experiments. Compared with current methods of emotion recognition, this method facilitates the screening out of disguised emotions and has better accuracy.

Tab.6 Comparison with current existing emotion recognition

Work	Model	Dataset	Accuracy	
GAO et al[4]	SPD+SVM	DEAP	Subject-dependent: 86.71%	/
RAHUL et al[5]	TOC+LSTM	DEAP	Subject-dependent: 82.01%	
NALINI et al[20]	DCERNet	DEAP	/	Subject-independent: 86.5%
HU et al[21]	CNN-BiLSTM-MASH	DEAP	Subject-dependent: 89.33%	/
LIU et al[22]	DCCA	SEED-IV	Subject-dependent: 87.5%	/
LI et al[23]	TANN	SEED-IV	Subject-dependent: 73.94%	Subject-independent: 68.00%
Proposed	CNN+BiLSTM+self-attention	DEAP	Subject-dependent: 92.83%	Subject-independent: 87.76%
		Collection	Subject-dependent: 92.61%	Subject-independent: 88.44%

In this paper, a CNN+BiLSTM+self-attention model for EEG-based emotion recognition is proposed. Compared with machine learning techniques, this emotion recognition model greatly improves the classification accuracy and does not require manually extracting features, which allows quick acquisition of classification results in practical applications. The results show that the accuracy of the model can reach 92.83% in subject-dependent experiments, and 87.76% in subject-independent experiment on the DEAP. On the collected data, the average accuracy of 92.61% and 88.44% was achieved in subject-dependent and subject-independent experiments, respectively. In summary, the hybrid CNN+BiLSTM+self-attention model advanced in this paper exhibits excellent recognition performance and generalization ability, and can effectively solve the problems of insufficient features extraction and poor recognition accuracy in existing emotion recognition methods. This recognition algorithm will be applied to the online analysis system of emotion recognition to facilitate real-time monitoring of emotional states. The EEG and other physiological response signals can effectively detect the real human emotions and states.

Ethics declarations

Conflicts of interest

The authors declare no conflict of interest.

References

[1] IBRAHIM A, ELIJAH O, OMOBOLAJI F O. Performance evaluation of different support vector machine kernels for face emotion recognition[C]//2015 SAI Intelligent Systems Conference (IntelliSys), November 10-11, 2015, London, UK. New York：IEEE, 2015：15670349.

[2] WEI C, CHEN L L, SONG Z Z, et al. EEG-based emotion recognition using simple recurrent units network and ensemble learning[J]. Biomedical signal processing and control, 2020, 58(3)：101756.

[3] SORAIA M A, MANUE J F. Emotions recognition using EEG signals：a survey[J]. IEEE transactions on affective computing, 2019, 10(3)：374-393.

[4] GAO Y Y, SUN X Y, MENG M, et al. EEG emotion recognition based on enhanced SPD matrix and manifold dimensionality reduction[J]. Computers in biology and medicine, 2022, 146(3)：105606.

[5] RAHUL S, RAM B P, PRADIP S. Automated emotion recognition based on higher order statistics and deep learning algorithm[J]. Biomedical signal processing and

control, 2020, 58：101867.

[6] CHENG W X, GAO R B, SUGANTHAN P N, et al. EEG-based emotion recognition using random convolutional neural networks[J]. Engineering applications of artificial intelligence, 2022, 116(5-6)：105349.

[7] CUI F, WANG R, DING W, et al. A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition[J]. Mathematics, 2022, 10(4)：582

[8] CHEN P Y, GAO Z K, YIN M M, et al. Multiattention adaptation network for motor imagery recognition[J]. IEEE transactions on systems, man, and cybernetics：system, 2022, 52(8)：5127-5139.

[9] LI C, WANG B, ZHANG S L, et al. Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism[J]. Computers in biology and medicine, 2022, 143(22)：105303.

[10] WEI Y, LIU Y, LI C, et al. TC-Net：a transformer capsule network for EEG-based emotion recognition[J]. Computers in biology and medicine, 2023, 152：106463.

[11] ZHONG M Y, YANG Q Y, LIU Y, et al. EEG emotion recognition based on TQWT-features and hybrid convolutional recurrent neural network[J]. Biomedical signal processing and control, 2023, 79(10)：104211.

[12] LIU S Q, WANG X, ZHAO L, et al. Subject-independent emotion recognition of EEG signals based on dynamic empirical convolutional neural network[J]. IEEE/ACM transactions on computational biology and bioinformatics, 2021, 18(5)：1710-1721.

[13] ORHAN A, ABDULKADIR Ș. Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition[J]. Applied acoustics, 2021, 182(1)：108260.

[14] KONG W C, DONG Z Y, JIA Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network[J]. IEEE transactions on smart grid, 2019,10(1)：841-851.

[15] ELESSAWY R H, ELDAWLATLY S, ABBAS H M. A long short-term memory auto encoder approach for EEG motor imagery classification[C]//International Conference on Computation, Automation and Knowledge Management (ICCAKM), January 9-10, 2020, Dubai, United Arab Emirates. ICCAKM, 2020：9051489.

[16] DADEBAYEV D, GOH W W, TAN E X. EEG-based emotion recognition：Review of commercial EEG devices and machine learning techniques[J]. Journal of King Saud University-computer and information sciences, 2022, 34(7)：4385-4401.

[17] BHATT A, ANANT R B. EEG based emotion recognition using SVM and LibSVM[J]. International journal of computer applications, 2019, 178(45)：1-3.

[18] PAN C, SHI C, MU H, et al. EEG-based emotion recognition using logistic regression with Gaussian kernel and Laplacian prior and investigation of critical frequency bands[J]. Applied sciences, 2020, 10(5)：1619.

[19] HUY N P, MAY D T, HUONG N T, et al. Combination of wavelet and MLP neural network for emotion recognition system[J]. International journal on future revolution in computer science & communication engineering, 2018, 4(11)：105-109.

[20] NALINI P, ANURAG S, SHRIVISHAL T. Learning DenseNet features from EEG based spectrograms for subject independent emotion recognition[J]. Biomedical signal processing and control, 2022, 74(2)：103485.

[21] HU Z, CHEN L, LUO Y, et al. EEG-based emotion recognition using convolutional recurrent neural network with multi-head self-attention[J]. Applied sciences, 2022, 12(21)：11255.

[22] LIU W, QIU J L, ZHENG W L, et al. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition[J]. IEEE transactions on cognitive and developmental systems, 2022, 14(2)：715-729.

[23] LI Y, FU B X, LI F, et al. A novel transferability attention neural network model for EEG emotion recognition[J]. Neurocomputing, 2021, 447(2)：92-101.